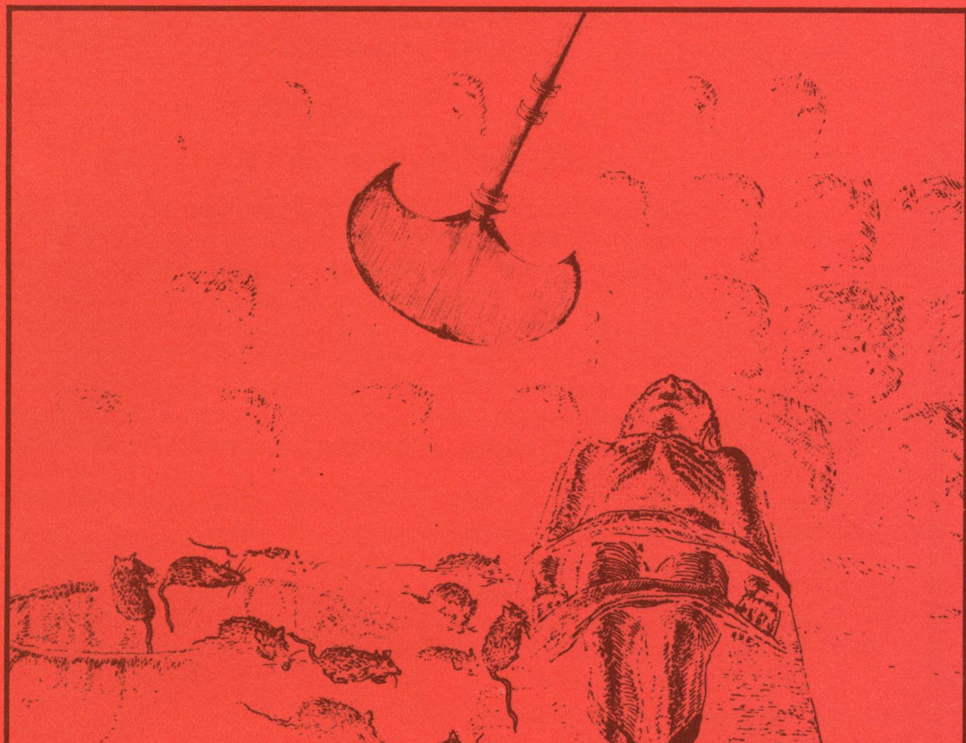
MATHEMATICS MAGAZINE



Vol. 58 No. 2
MARCH 1985

POE'S PENDULUM • VECTOR PROOFS IN GEOMETRY
COMPUTING TOPOLOGIES • FACTORING FACTOR RINGS

MATHEMATICS MAGAZINE

THE MATHEMATICAL ASSOCIATION OF AMERICA

Volume 58, Number 2, March 1985
(ISSN— 0025- 570X)

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

COVER: The terrifying descent of Poe's pendulum. See illustrations, p. 66, and pp. 78–83.

EDITORIAL POLICY

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

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

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

AUTHORS

Linda W. Brinn ("Computing Topologies") received her Ph.D. in complex analysis at Wayne State University under W. Seidel. After several years of teaching (part-time) while raising a family (full time) followed by several years as an industrial mathematician, she joined the faculty of the University of Michigan-Dearborn. Her interests include both classical mathematics and computer science and, especially, the relation between these subjects.

ILLUSTRATIONS

Our cover illustration is from the sound filmstrip "The Time, Life, & Works of Edgar Allan Poe," courtesy of Educational Audio Visual Inc., Pleasantville, N.Y. The illustrator is **Kenneth Marcus-Daly**.

Vic Norton pictures the topological wood on p. 72.

Gary Kraud penned the musical artwork on p. 110.

All other illustrations were provided by the authors.

The MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, five times a year: January, March, May, September, and November.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $11 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The non-member/library subscription price is $28 per year. Bulk subscriptions (5 or more copies) are available to colleges and universities for classroom distribution to undergraduate students at a 41% discount ($6.50 per copy—minimum order $32.50).
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Company, Middletown, CT 06457. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.
Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, The Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

Second class postage paid at Washington, D.C. and additional mailing offices.
Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.
PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Computing Topologies

Two roads diverged in a mathematical wood, and I—
I took the one less traveled by.

L. W. BRINN
University of Michigan-Dearborn
Dearborn, MI 48128

How far is it from computer science to topology? Surprisingly, it is not far at all—if you take the right path. For a mathematician, the traditional path to topology begins at calculus, goes through real analysis, and continues to the study of abstract spaces. Throughout the journey, the mathematician studies infinite sets of points, and these seem far indeed from the finite world of the computer scientist.

But there is another path to topology. This path begins at computer programming, goes through discrete mathematics, and continues to the study of abstract spaces. The point sets studied are finite, and this is to be expected since only a finite topology can have a computer representation. However, many of the properties of finite spaces are shared by a certain interesting class of infinite spaces. The second path leads eventually to the study of infinite topologies.

The milestones on the mathematician's road to topology are definitions and theorems. The milestones on the computer scientist's road are definitions, theorems, and the computational methods which result from those theorems. For instance, what is meant by the closure of a set? What properties does the closure have? And how (asks the computer scientist) can one write a computer program to find the closure of any given set? Finite topologies can be computed in a way that infinite topologies cannot, and it is reasonable to ask that theorems lead to programs.

As we set out on the computer scientist's path, our first step will be to find suitable data structures for finite topologies. The key to these data structures is the fact that any finite topology can be specified in terms of the dominance relation which generates it. Our aim is not only to *define* topologies in terms of these relations, but also to *represent* the topologies clearly and efficiently both for computer calculation and for our own visualization. We assume knowledge of the mathematics customarily taught in a first course in discrete mathematics for computer science students, and a level of programming skill which can reasonably be expected after a first course in some high-level language such as Pascal, PL/1, or FORTRAN.

As we continue our journey, we look at finite topologies in two ways. First, we ask standard topologists' questions and find computer scientists' answers. Suppose that you are given a finite space. How can you determine which subsets of the space are open? What separation properties does the space have? Is it connected? Which functions from it to other spaces are continuous? As answers we seek both theorems and computational methods.

Next, we take the computer scientist's answers and find the corresponding questions from topology. That is, we see what facts from discrete mathematics can tell us about the topological structure of finite spaces.

Data structures

We begin by defining a dominance relation on a set and giving two standard data structures for the relation: its table and its directed graph. We then show how a dominance relation generates a topology and how the two data structures for the relation function as data structures for the topology. We also show that any finite topology can be generated by a dominance relation. Thus, our data structures can be used to represent any finite topology.

A relation R on a set A is called a **dominance relation** (or a **preorder**) whenever R is both reflexive and transitive. If y is related to x, we write $y \to x$ and say that y **dominates** x. In particular, any equivalence relation or partial order on A is a dominance relation. For instance, let $A = \{1,2,3,4,5,6\}$ and say that y dominates x whenever y divides x. The relation is a dominance relation, in fact a partial order.

If A is finite, a relation on A may easily be represented in matrix form or graphical form. If $A = \{x_1, x_2, \ldots, x_n\}$, the **table** (or **adjacency matrix**) of the relation R is a square matrix (b_{ij}) in which $b_{ij} = 1$ if x_i is related to x_j, and $b_{ij} = 0$ otherwise. The table is a convenient computer representation for the relation and is easy to construct and manipulate in almost any programming language. A relation R may also be represented by a **directed graph** (or **digraph**) in which there is a directed edge from x_i to x_j if and only if x_i is related to x_j.

Since a dominance relation is reflexive, its digraph will have a loop at each vertex. Since the relation is transitive, if there are directed edges from x_i to x_j and from x_j to x_k there will be a directed edge from x_i to x_k. Even for small sets A, the digraph of a dominance relation can become visually confusing. In order to simplify the picture, it is customary to omit the loop at each vertex as well as any arrows implied by the transitivity of the relation. The matrix and the digraph of our example $A = \{1,2,3,4,5,6\}$ with R the division relation are in FIGURE 1. For more about the matrices and digraphs of relations, see **[1]**, **[6]**, **[10]**, **[11]**, or **[12]**.

Some high-level computer languages (for instance, Pascal and PL/1) allow digraphs to be represented as complex linked lists and to be investigated with relative ease. For details about list handling in Pascal and PL/1, see **[1]** and **[2]**, respectively. Since list handling is not usually taught in an introductory programming course, and since the process is inconvenient in certain common languages (for instance, FORTRAN) we will use the adjacency matrix for machine representation of a relation. The digraph remains a valuable intuitive aid in visualizing the structure of a dominance relation and the topology it generates.

An open set U in a topology is characterized by the property that each point of U is an interior point. Intuitively, each point is "insulated" from the complement of U. We can use the dominance relation to give an appropriate meaning to the term "insulated." We will say that a point x in a subset U of A is **insulated** from $A - U$ if and only if there is no point y in $A - U$ such that y dominates x. We define the set U to be **open** whenever all of its points are insulated from $A - U$. With this definition, both the empty set and A itself are open. The definition implies (as should be the case) that the union of open sets is open, and the intersection of open sets is open. (Here unions and intersections may be arbitrary.)

Note that the process of defining a dominance relation on a set A and using the relation to generate a topology does not require A to be finite. If A is infinite, there will, of course, be no

R	1	2	3	4	5	6
1	1	1	1	1	1	1
2	0	1	0	1	0	1
3	0	0	1	0	0	1
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1

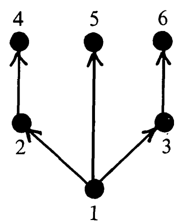


FIGURE 1. **The adjacency matrix and digraph of the dominance relation R on the set $A = \{1,2,3,4,5,6\}$ defined by: y dominates x whenever y divides x.**

computer representation for the relation or for the topology. The topology will have the property that the arbitrary intersection of open sets is open. Thus, not all infinite topologies can be generated by dominance relations. In particular, the ordinary topology on E^n cannot be generated by a dominance relation, nor can any infinite metric topology (except, of course, the discrete topology). The computer scientist's road to topology can lead directly to the study of infinite spaces, but these are not the spaces of classical analysis.

In fact, any topology in which the arbitrary intersection of open sets is open can be generated by a dominance relation. Given such a topology, define the relation by saying that y dominates x whenever y is an element of every open set about x. It is straightforward to verify that this relation is indeed a dominance relation and generates the original topology. In particular, every finite topology can be generated by a dominance relation. Thus, every finite topology has the simple tabular and graphical representations discussed below.

As an example in which a finite topology is used to define a relation, let $B=\{a,b,c,d,e\}$ and let $U=\{a,b,c,\}$, $V=\{d\}$, and $W=\{e\}$. Let the topology $\mathcal{T}$ on B be the collection of open sets $\mathcal{T}=\{\varnothing,U,V,W,U\cup V,U\cup W,V\cup W,B\}$. This is indeed a topology on B and is generated by a certain dominance relation, in fact by an equivalence relation. What is the relation?

We can use the table of a dominance relation to actually construct the open sets of the corresponding topology, and we can also visualize these open sets in the corresponding digraph. For each point x of A, we expect to find a unique smallest open set U_x about x; the set U_x is simply the intersection of all open sets containing x. But, if U is any open set containing x, and if $y\to x$, then $y\in U$ (otherwise, x would not be insulated from $A-U$). Thus, $U_x=\{y|y\to x\}$. This provides a simple criterion for deciding whether or not a given set U is open. The set U is open if and only if

$$U=\bigcup_{x\in U}U_x.$$

In the case of our example in FIGURE 1, $U_1=\{1\}$, $U_3=\{1,3\}$, and $U_4=\{1,2,4\}$. We see immediately that the set $U=\{1,3,4\}$ is not open since $U\neq U_1\cup U_3\cup U_4$.

Since every open set can be written as a union of sets U_x, the sets U_x form a basis for the topology. If A has n elements, there are n of these basis sets (not necessarily all distinct). That is, any topology on an n-element set A has a basis of at most n sets.

Moreover, the sets U_x form a computationally useful basis for the topology. If $A=\{x_1,x_2,\dots,x_n\}$ and if S is a subset of A, the **Boolean** or **bit string representation** of S is an array $(b_1,b_2,\dots,b_n)$ in which $b_i=1$ if x_i is an element of S, and $b_i=0$ otherwise. For instance, in the case of our example in FIGURE 1, the representation of $S=\{1,2,4\}$ is $(1,1,0,1,0,0)$ and the representation of $T=\{4,6\}$ is $(0,0,0,1,0,1)$. The representation of a union of sets is the bit by bit disjunction of the representations of those sets, and the representation of the intersection is the bit by bit conjunction. Thus, for example, the representation of $S\cup T$ is

$$(1\vee 0,1\vee 0,0\vee 0,1\vee 1,0\vee 0,0\vee 1)=(1,1,0,1,0,1)$$

and the representation of $S\cap T$ is

$$(1\wedge 0,1\wedge 0,0\wedge 0,1\wedge 1,0\wedge 0,0\wedge 1)=(0,0,0,1,0,0).$$

The representation of the complement of a set is the bit by bit complement of its representation. Thus, the representation of $A-S$ is

$$(\bar{1},\bar{1},\bar{0},\bar{1},\bar{0},\bar{0})=(0,0,1,0,1,1).$$

In the matrix of a dominance relation R the jth column provides the Boolean representation of the basis set U_{x_j}. Recall that $b_{ij}=1$ if and only if $x_i\to x_j$. That is, $b_{ij}=1$ if and only if x_i is an element of U_{x_j}. In the case of our example in FIGURE 1, we see immediately that the representation of U_3 is simply the third column of the matrix of R, or $(1,0,1,0,0,0)$. Thus, the matrix of R is a data structure not only for the relation itself, but also for the topology it generates. The matrix "stores" the topology in the sense that its columns are the Boolean representations of sets which form a basis.

Our simple criterion for determining whether or not a set U is open can be developed immediately into a computer program. In the matrix of the relation R, form the disjunction of the columns corresponding to the elements of U. If the result is the Boolean representation of U, then U is open; otherwise U is not open. In the case of our example in FIGURE 1, if $U = \{1,3,4\}$, then the disjunction of columns 1, 3, and 4 is $(1,1,1,1,0,0)$ which does not equal U, so that U is not open. (What does the bit string $(1,1,1,1,0,0)$ represent? It does *not* represent the interior of U.)

Not only do the basis sets U_x appear in a natural way in the table of the relation R, but they are also easy to see in the digraph of R. The basis set U_x consists of all those points y such that x is "reachable" from y, i.e., there is a path from y to x. Intuitively, the set U_x begins at x and "spreads" backwards along the directed edges of the graph. The digraph can help one to answer, by inspection, simple questions about the corresponding topology. In our example in FIGURE 1, is there an open set which does not contain the point 1? Since the digraph shows every point is reachable from 1, there is no such open set. Consequently, no two open sets are disjoint.

Before we leave the matrix and the digraph of the relation R, let us look at them from another point of view. We have seen that the columns of the matrix have a topological significance, but do the rows? In the digraph, the points y such that x is reachable from y have a significance. What about the points y which are reachable from x? We expect a certain duality, and we find it by means of the dual topology.

The Dual topology

Topologies generated by dominance relations can also be approached by means of their closed sets. If F is a subset of A, recall that F is closed whenever $A - F$ is open. If F is closed, x is an element of F, and $x \to y$, then y cannot be an element of $A - F$ (since $A - F$ is open). Thus, y is also an element of F. If we let $F_x = \{y | x \to y\}$, then F_x is the smallest closed set about x. That is, $F_x = \overline{\{x\}}$, where $\overline{\{x\}}$ denotes the closure of $\{x\}$. This is clear since F_x is closed (as the relation R is transitive), contains x (as the relation R is reflexive), and is a subset of every closed set about x. A set F is closed if and only if

$$F = \bigcup_{x \in F} F_x.$$

This is the dual of our representation of open sets. Note that, in a general topology, a closed set F has the representation given above. However, a union of closures of single point sets is not necessarily closed.

If the set A is finite, the Boolean representation of the set F_{x_i} is simply the ith row of the matrix of the relation R. A set F is closed if and only if its Boolean representation is the disjunction of the rows corresponding to its elements. Closed sets, like open sets, can be easily constructed using Boolean operations on the matrix of R. In the digraph, the points y contained in F_x are those points y which are reachable from x. Intuitively, F_x "spreads" forwards from x along the edges of the digraph. The key to the duality between open and closed sets is the dual of the relation R.

If R is a dominance relation on a set A, then its **dual** or **converse** R^d is defined by the requirement that $y \to^d x$ if and only if $x \to y$. The dual is also a dominance relation, and its table is the transpose of the original table. Its digraph is generated from the digraph of R by reversing the orientation of each edge. The topology it generates is called the dual of that generated by the original relation R.

If we let $U^d{}_x$ and $F^d{}_x$ be the smallest open and smallest closed sets about x in the dual topology, then $U^d{}_x = F_x$ and $F^d{}_x = U_x$. This is immediate from the definitions and is what we expect from the representations of these sets in the table of R and its transpose. Thus, the closed sets of the original topology are exactly the open sets of its dual, and conversely.

The ability of the topology to distinguish between two points x and y by open sets is equivalent to its ability to distinguish between x and y by closed sets, and both are equivalent to the ability of the relation R to distinguish between x and y. More precisely, the following are equivalent:

$$x \leftrightarrow y$$

$$U_x = U_y$$

$$F_x = F_y.$$

The second condition says that any open set contains both points x and y or else neither point. The third condition says the same thing about closed sets. An easy proof is to notice that the first condition is equivalent to the second and also to the condition that $x \leftrightarrow {}^d y$.

Christie, in [**3**], uses a slightly different approach to generating finite topologies from preorders. Christie uses the preorder to define an interior operator on the subsets of A and notices that, if A is finite, the topology is completely determined by the interiors of the complements of single point sets. A topology on A can also be defined in terms of a closure operator on the subsets of A or in terms of a boundary operator. For more about such operators, see [**4**], [**7**], and [**9**]. Wilansky, in [**13**], p. 50, shows how a topology on a finite set A generates a preorder on A. The fact that every finite topology is generated by a preorder has also been used to count the number of topologies on a finite set. The problem is reduced to one of counting the number of matrices of a certain form. For the details, see [**5**] and [**8**].

As an exercise, suppose that S is a subset of A. How would one write a computer program to find the interior of S? Recall that $\bigcup_{x \in S} U_x$ is *not* the interior of S. However, if $T = A - S$, then $\bigcup_{x \in T} F_x$ is the closure of T. Also, it is well known that the interior of S is simply $A - \overline{T}$. How would one write a program to find the boundary of S? How does one visualize the interior, the closure, or the boundary of S in the digraph of the relation R?

If R is any relation on the set A, we cannot expect that R will generate a topology. There is, however, a smallest preorder R' on A which contains R. R' is simply the intersection of all preorders on A which contain R (here considered as subsets of $A \times A$). The matrix of R' can be found efficiently by first placing 1's along the main diagonal of the matrix of R (extending R to a reflexive relation) and then finding the transitive closure of the reflexive extension. The transitive closure can be found efficiently by means of Warshall's algorithm (see [**1**] or [**2**]). Thus, for each point x of a finite set A, one can specify arbitrarily those points y which must lie in every open set about x (in a reasonable specification, x would be one such point). One can then construct a topology on A in which each basic open set U_x (and thus each open set) contains the required points y and as few others as possible.

Friends and relations

Suppose that the dominance relation is required to be antisymmetric and so is a *partial order*, denoted by $\leqslant$. In this case, for each point x of A, $U_x = \{y \mid y \to x\} = \{y \mid y \leqslant x\}$, and $F_x = \{y \mid x \to y\} = \{y \mid x \leqslant y\}$. (Technically, U_x is the principal ideal generated by the element x.) The single-element open (closed) sets are those consisting of a minimal (maximal) element alone. Two points can be separated by disjoint open (closed) sets if and only if they have no common lower (upper) bound. Thus, in particular, in a lattice no two points can be separated by either disjoint open sets or by disjoint closed sets. In our example in FIGURE 1, there is one single element open set, $\{1\}$, and three single element closed sets. No two points can be separated by disjoint open sets, but some points can be separated by disjoint closed sets (for instance, 6 and 4 can be so separated).

Since a partial order is antisymmetric, for each element x of A, $U_x \cap F_x = \{y \mid y \leftrightarrow x\} = \{x\}$. Conversely, if for each element x of A $U_x \cap F_x = \{x\}$, then the relation is a partial order. A **total** or **linear order** is characterized by the fact that, for each element x of A, $U_x \cup F_x = A$. Here note that, if x is not a minimal element, then U_x is the smallest open neighborhood about one of its points (namely x) but not about any of its other points (for if $y < x$, then $U_y \subsetneq U_x$).

Suppose that the dominance relation R is required to be symmetric and so is an **equivalence relation**. Then, for each point x of A, $U_x = F_x = C_x$, where C_x is the equivalence class containing

x. This follows directly from the definitions of U_x and F_x. If x and y are equivalent, these points lie in exactly the same open sets (since $U_x = U_y$) and exactly the same closed sets (since $F_x = F_y$). If x and y are not equivalent, they are separated by disjoint open sets and by disjoint closed sets (since $U_x \cap U_y = \varnothing$ and $F_x \cap F_y = \varnothing$). Since an equivalence relation is its own dual, the dual topology is simply the original. Any subset S of A is either both open and closed (in case it is the union of equivalence classes) or neither open nor closed. Here each basic neighborhood U_x is the smallest open neighborhood about not just x but about any of its points. The topology $\mathscr{T}$ of our earlier example space $B = \{a, b, c, d, e\}$ is generated by the equivalence relation for which the corresponding equivalence classes are $U = \{a, b, c\}$, $V = \{d\}$, $W = \{e\}$. The points a and b lie in exactly the same open sets and exactly the same closed sets. However, the points a and d can be separated by disjoint open sets and by disjoint closed sets. The topology was originally given in terms of its open sets, and these are precisely the sets which are unions of equivalence classes.

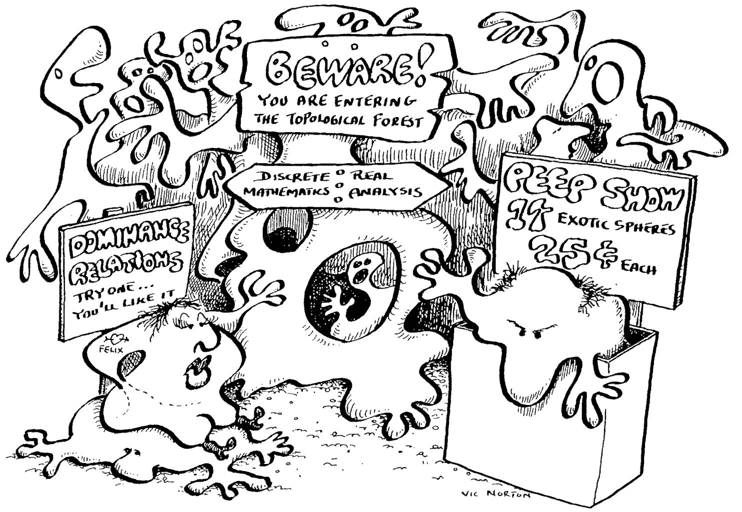
In what other ways do partial orders and equivalence relations generate different types of topologies?

Topological properties

What do relations reveal about the topologies that they generate? One can approach this question from the point of view of the topologist or from that of the computer scientist. That is, one can ask questions about basic topological properties, or one can explore the topological interpretations of familiar theorems from discrete mathematics.

Let us begin with the first point of view. What do preorders reveal about the separation properties of the topologies they induce? What do they reveal about connectivity, compactness, or the continuity of functions?

The familiar spaces of analysis are at the very least Hausdorff. Our dominance relation spaces are hardly so tidy. If the induced topology is Hausdorff, then the relation is the trivial relation in which each element is related only to itself. Recall that a **Hausdorff space** is one in which every two distinct points can be separated by disjoint open sets and that, as a direct consequence of the definition, single point sets are closed. If the space is Hausdorff, for each element x of A, $\{x\} = \overline{\{x\}} = F_x = \{y | x \to y\}$. But the trivial relation induces the discrete topology (since $U_x = \{y | y \to x\} = \{x\}$). Thus, if the space is Hausdorff, it is a discrete space. Since any finite Hausdorff space is a discrete space, the result is not unexpected.



TWO ROADS DIVERGED IN A MATHEMATICAL WOOD, AND I –
I TOOK THE ONE LESS TRAVELED BY.

The separation properties to explore are clearly those below Hausdorff or T_2. Recall that a $\boldsymbol{T_0}$ **space** is one in which, whenever x and y are distinct, at least one point has a neighborhood not containing the other. In our case, what is required is that whenever $x \neq y$, then either $x \notin U_y$ or $y \notin U_x$. That is, if $x \neq y$, then either $y \nrightarrow x$ or $x \nrightarrow y$. If the relation is a nontrivial equivalence relation, the resulting space is not even T_0. The T_0 spaces are precisely those induced by partial orders (see [**13**], p. 50). Recall that a $\boldsymbol{T_1}$ **space** is one in which, whenever x and y are distinct, each point has a neighborhood not containing the other. In our case, the requirement is that $x \notin U_y$ and $y \notin U_x$ whenever $x \neq y$. Again, the relation must be the trivial one and the topology discrete. In particular, our example in FIGURE 1 is a T_0 space, but no more, and our space B is not even T_0.

Using preorders, we may easily construct nonhomogeneous topologies in which the neighborhood structures at distinct points x and y are quite different. Separation properties might better be studied locally. If we define $h(x, y)$ to be the number of elements in $U_x \cap U_y$, then any neighborhoods of x and y will have in their intersection at least this many points. Thus, $h(x, y)$ is a measure of how separated x and y are. What properties must the function h have? For a finite space, the function is easily computed. In the matrix of the relation $R, h(x_i, x_j)$ is the number of 1's in the conjunction of columns i and j.

Connectivity

Whether or not the space A is connected is revealed in a straightforward way by the digraph of the corresponding relation. The space A is connected exactly when the digraph of the relation R is (weakly) connected, that is, whenever the underlying undirected graph is connected. For a discussion of connectivity in graphs, see [**6**] or [**11**]. What is required for connectivity is that, for any points x and y in A, there is a semipath in the digraph from x to y. A semipath differs from a path in that it is permissible to travel along an edge from its terminal point to its initial point. If A is finite, then the connectivity of the digraph can be checked using Warshall's algorithm. (The algorithm is to be applied, not to the original digraph, but to the corresponding symmetric digraph in which each edge is given both orientations. Again, see [**1**], or [**2**].) Thus, our example space of FIGURE 1 is connected, but our space B is not.

In order to see that the connectivity of the space is equivalent to the connectivity of the digraph, note first that if K is any component of the digraph, then K is both open and closed in the space. If x is an element of K and $y \to x$ or $x \to y$, then y is also an element of K since there is an edge in the undirected graph between x and y. Thus, K is open and closed. Conversely, if a set K is both open and closed in the space, there can be no edge in the undirected graph between K and $A - K$, and K is a component of the digraph. The space A is connected if and only if A itself is the only nonempty set both open and closed in A. The graph is connected if and only if its only component is A.

Suppose that we construct "at random" a relation R on an n-element set A. That is, we write a computer program to assign each entry of an $n \times n$ matrix the value 0 or 1 with specified probabilities $p(0) = p$ and $p(1) = 1 - p$. If we then construct the smallest preorder R' containing R, what is the probability that the space A is connected under the topology generated by R'? We can find an approximate answer, simply and by elementary methods. The process of constructing the relation R, generating R', and checking whether the associated digraph is connected can be programmed efficiently, as outlined above, using Warshall's algorithm. The process can be considered a Bernoulli trial. Our program can make m trials, and we can (by the methods of elementary undergraduate statistics) find confidence intervals for the probability that the space is connected. For an easier project, what is the probability that the space is T_0? It is trivial to check that a matrix is antisymmetric.

If the set A has at least two elements, then an equivalence relation on A produces either the indiscrete topology or a topology in which A is not connected. Here the components of the graph are exactly the equivalence classes. Partial orders may clearly produce disconnected spaces, but lattice structures never do.

However, any preorder produces a locally connected space. The sets U_x form a basis of connected open sets. If we had $x \in V \subsetneq U_x$, where V was both open and closed in U_x, then V would be open in A, and U_x would not be the smallest open set about x. In particular, then, all finite spaces are locally connected. There are, of course, infinite locally connected spaces (such as E^n) whose topology is not generated by a preorder.

By considering the dual topology we know that, for each element x of A, $F_x = \overline{\{x\}}$ is connected. This is not peculiar to topologies generated by dominance relations. Closures of single point sets are always connected (see [4], p. 109). Sometimes the dual topology leads us to new information, and sometimes it leads us back to well known facts.

Compactness

What can be said about the compactness of our dominance relation spaces? If A is finite, it is always a compact space. If A is infinite, a preorder may generate either a compact or a noncompact topology for A. For example, both the discrete and indiscrete topology for A are generated by preorders (in fact, by equivalence relations).

If A is a finite space, let us call A ***m*-compact** if and only if every open cover of A has a subcover of at most m sets. For a given space A, what is the least value of m, and how may the sets of a subcover be generated? One approach to this question is in terms of maximal elements. An element x of A is called a **maximal element** if and only if, whenever y is an element of A and $x \to y$, then $y \to x$. In our terminology, x is maximal if and only if F_x is a subset of U_x, and this condition can be checked easily using Boolean operations on the matrix of the relation R. If the relation R is restricted to maximal elements, then it is an equivalence relation. If the number of equivalence classes is k, then it is straightforward to verify that every open cover of A contains a subcover of at most k sets. That is, A is k-compact.

Given an open cover of A, a subcover can be found by choosing, for each equivalence class C_i $(1 \leqslant i \leqslant k)$ a maximal element x_i of C_i and a set U_i from the open cover with x_i an element of U_i. Note that if x_i is an element of U_i, then C_i is a subset of U_i since U_i is open. If x is an element of A, then either x is maximal and $x \in C_i \subseteq U_i$ for some i, or else there is a maximal element x_i with $x \to x_i$. In this case, x is an element of U_i since U_i is open. The number k is best possible since the cover $\mathscr{U} = \{U_x | x \in A\}$ of basic open sets has no subcover of fewer than k sets. In particular, our space in FIGURE 1 is 3-compact, and our example space B is also 3-compact. (What does the dual topology reveal about closed covers of A?)

The process for generating the subcover is straightforward. We inspect the elements of A in order, and we produce a collection of nonequivalent maximal elements as follows: given an element x, if $x \to y$, where y is already part of our collection, we discard x. Otherwise, we add x to our collection. If the open sets of the cover $\mathscr{U}$ are given as a list of Boolean representations, we search the list until we have found enough open sets to cover our collection of maximal elements. This will require us to find, at most, one open set per maximal element. The number of sets we find may, of course, depend on the order of the list of open sets. For instance, try the above process with our example space B and the cover $\mathscr{U} = \{U, V, W, B\}$. If the cover sets are listed in the order given, the subcover will consist of U, V, and W. If the cover sets are listed in the opposite order, the subcover will consist of B alone.

Continuity

In order to investigate a function f from some finite set A to a finite set B, we need a representation of the function itself as well as representations of both topologies. If the function has a simple closed form, for instance $f(x) = x^2$, this presents no problem. For a more arbitrary function f there is a standard tabular representation of f as a relation from A to B which is convenient in that both the image of each point x of A and the preimage of each point y of B are easily accessible. If $A = \{x_1, x_2, \ldots, x_n\}$ and $B = \{y_1, y_2, \ldots, y_m\}$, the **table** of f is a matrix (f_{ij}) in which $f_{ij} = 1$ if $f(x_i) = y_j$, and otherwise $f_{ij} = 0$. Thus, the ith row is the Boolean representation of the (one element) image set of the point x_i, and the jth column is the Boolean representation of the preimage of the point y_j. The disadvantage of this representation is the storage requirement. A more compact representation of f is simply a one-dimensional array (a_i) in which $a_i = j$ when $f(x_i) = y_j$. Here the preimage of the point y_j is not so easily accessible.

Since the sets A and B are finite, their topologies are given by dominance relations. This fact helps us to determine the continuity of the function f without explicitly recovering preimages. Let $\rightarrow$ be used to denote each dominance relation. Then the function f is continuous if and only if $f(y) \rightarrow f(x)$ whenever $y \rightarrow x$. That is, f is continuous if and only if f preserves dominance. The proof follows directly from the fact that open sets V are those for which z is an element of V whenever w is an element of V and $z \rightarrow w$ (see [**13**], p. 60).

The continuity of a function f can be checked easily with the array representation. Suppose that a function f from our space A in FIGURE 1 to our space B is given by the array $F = (1,1,2,4,4,3)$. That is, $f(1) = a$, $f(2) = a$, $f(3) = b$, $f(4) = d$, $f(5) = d$, $f(6) = c$. Let the matrices of the dominance relations on A and B be denoted by T^A and T^B, respectively. We check each entry of T^A. If for some i and j $(1 \leqslant i \leqslant 6, 1 \leqslant j \leqslant 6)$ we have $T^A{}_{ij} = 1$ and $T^B{}_{F_iF_j} = 0$, then the function f is not continuous. The check can be programmed as a simple double do-loop.

In particular, a homeomorphism f from A to B is a one-to-one correspondence for which $y \rightarrow x$ if and only if $f(y) \rightarrow f(x)$. In other words, f is an isomorphism between the digraphs of the corresponding dominance relations. This characterization allows the use of graph-theoretical methods in determining whether two spaces are homeomorphic. Although a number of isomorphism invariants are known for graphs, there is as yet no simple complete set of such invariants.

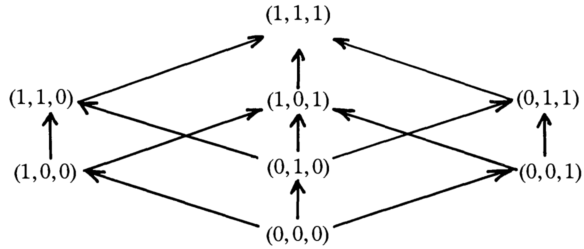
If the dominance relations on both A and B are equivalence relations, then any continuous function f from A to B must preserve equivalence classes. If x is an element of A, there is some element w of B such that $f(C_x)$ is a subset of C_w, where C_x and C_w denote the equivalence classes containing x and w, respectively. This follows directly from the fact that $f(x) \equiv f(y)$ whenever $x \equiv y$. Since any partitions of A and B are induced by corresponding equivalence relations, it is easy to find topologies on finite sets of real numbers (or even on the reals themselves) under which such functions as $f(x) = 2x$ or $f(x) = x^2$ are not continuous. The derivation in terms of equivalence relations even makes these topologies seem (in some sense) natural.

If the dominance relations on A and B are partial orders, the function f is continuous if and only if it is order-preserving. That is, f is continuous if and only if $f(x) \leqslant f(y)$ whenever $x \leqslant y$, where $\leqslant$ is the generic symbol for a partial order.

If the relation on A is an equivalence relation while that on B is a partial order, then under any continuous function f each equivalence class must map to a single point in B. If $x \equiv y$, then $f(x) = f(y)$ since a partial order is antisymmetric. What if the relation on A is a partial order while that on B is an equivalence relation? What if A is a lattice? Why must A then map entirely into a single equivalence class?

The above discussion illustrates how the properties of dominance relations illuminate the topological structures of all finite (and some infinite) spaces, and how the properties of continuous functions illuminate the interplay between two different structures on different spaces or on the same space. A point set topology text (such as [3], [4], [7], [9], or [**13**]) will suggest a number of other topological properties that can be explored using dominance relations.

As an exercise, how would we apply what we have learned to a type of finite set which is a major focus of discussion in elementary discrete mathematics, namely, the finite Boolean algebra? Consider the algebra as a subset algebra. If A is any finite set, then the power set of A with the subset relation $\subseteq$ is a partially ordered set, in fact a distributive complemented lattice (for details, see [**10**] or [**11**]). As a space with the topology induced by the relation $\subseteq$, what properties does it have? (See FIGURE 2 for the case of A a three-element set.) If S is any subset of A, U_S is simply the collection of all subsets of S, and F_S is the collection of all supersets of S. If S and T are any subsets of A, then S and T cannot be separated by disjoint open sets. In this case, $h(S,T)$ is the number of subsets of $S \cap T$. Since the subset relation $\subseteq$ is a partial order, the space is T_0. Since the set A itself is a universal upper bound, the space is 1-compact. The space is connected since its digraph is connected.



R	(0,0,0)	(0,0,1)	(0,1,0)	(0,1,1)	(1,0,0)	(1,0,1)	(1,1,0)	(1,1,1)
(0,0,0)	1	1	1	1	1	1	1	1
(0,0,1)	0	1	0	1	0	1	0	1
(0,1,0)	0	0	1	1	0	0	1	1
(0,1,1)	0	0	0	1	0	0	0	1
(1,0,0)	0	0	0	0	1	1	1	1
(1,0,1)	0	0	0	0	0	1	0	1
(1,1,0)	0	0	0	0	0	0	1	1
(1,1,1)	0	0	0	0	0	0	0	1

FIGURE 2. **The digraph and adjacency matrix of the set inclusion relation R on the power set of A, where A is any three element set. The Boolean representations of the subsets are used, and these representations are the key to decomposing the relation as a product relation and the corresponding topological space as a product space. Notice that only the part of the matrix above the main diagonal need be stored.**

If A is large, the digraph is extremely complex, and this suggests that the resulting topology is also complex. To the contrary, the space can be naturally decomposed as the product of n copies of a simple two element space. The key is to represent the subset order as a product order. This is the approach often taken in elementary discrete mathematics (for instance, see **[11]** and **[12]**). As a project, verify that the product of dominance relations is also a dominance relation, and the topology it generates is the product topology, if the product is finite (for information about product topologies, see [4] or [7]). Use the product representation to investigate the topology of the power set of A.

From discrete mathematics to topology

In the previous discussion, we have asked topological questions about spaces whose topologies are given by preorders. We may also study these spaces by considering the topological interpretations of known facts from discrete mathematics. We will simply illustrate this second approach by an example.

If A is a partially ordered set with order relation denoted by $\leqslant$, then a **chain** in A is a subset of A in which any two elements are related, and an **antichain** is a subset in which no two elements are related. A familiar theorem from discrete mathematics is a dual of Dilworth's theorem: *If A is a partially ordered set in which the length of a longest chain is n, then A can be partitioned into n (but no fewer) disjoint antichains* (see **[11]**). In topological terms, what does this say? The elements $x_1 < x_2 < \cdots < x_n$ form a chain if and only if the corresponding basic open sets $U_{x_1} \subsetneq U_{x_2} \cdots \subsetneq U_{x_n}$ form a chain. The subset S of A forms an antichain if and only if S has, as a subspace, the discrete topology. This follows immediately from the fact that, for each element y of S, $U_y \cap S = \{y\}$. Thus, in topological terms, if the length of a longest chain of basic open sets is n, then A can be partitioned into n (but no fewer) subspaces, each of which has the discrete topology. Can we omit the requirement that the open sets be basic open sets U_x? We expect that our space A in FIGURE 1 can be partitioned into three (but no fewer) discrete subspaces. What are the subspaces? How would you write a computer program to generate them? (The method for finding the antichains described in **[11]** leads immediately to a program.) What are the topological interpretations of other theorems from discrete mathematics?

Postscript

We have seen how any topology on a finite set A (and some topologies on an infinite set A) can be generated by dominance relations or preorders on A. If A is finite, the properties of the topology are reflected in both the matrix and the digraph of the relation. In particular, the matrix is an efficient, easily maintained data structure for the topology. Basis sets appear as columns, and open sets are disjunctions of columns. Closures of single point sets appear as rows, and closed sets are disjunctions of rows.

In asking standard topological questions, we found as answers both theorems and computational methods. We have also briefly illustrated how theorems from discrete mathematics can have topological interpretations.

How far is it from computer science to topology? Not far at all.

I would like to thank the editor and the referees for their helpful comments and suggestions. I would also like to thank Professor R. Fakler for introducing me to Christie's book and to the connection between dominance relations and topologies.

References

[1] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, Data Structures and Algorithms, Addison-Wesley, Reading, Mass., 1983.
[2] M. J. Augenstein and A. M. Tenenbaum, Data Structures and PL/1 Programming, Prentice-Hall, Englewood Cliffs, N.J., 1979.
[3] D. E. Christie, Basic Topology, Macmillan, New York, 1976.
[4] J. Dugundji, Topology, Allyn and Bacon, Boston, Mass., 1966.
[5] J. W. Evans, F. Harary, and M. S. Lynn, On the computer enumeration of finite topologies, Comm. of the ACM, 10, #5 (1967) 295–297.
[6] F. Harary, Graph Theory, Addison-Wesley, Reading, Mass., 1969.
[7] J. L. Kelly, General Topology, Van Nostrand, Princeton, N.J., 1955.
[8] V. Krishnamurthy, On the number of topologies on a finite set, Amer. Math. Monthly, 73 (1966) 154–157.
[9] K. Kuratowski, Topology, vol. 1, Academic Press, New York, 1966.
[10] L. S. Levy, Discrete Structures of Computer Science, Wiley, New York, 1980.
[11] C. L. Liu, Elements of Discrete Mathematics, McGraw-Hill, New York, 1977.
[12] R. E. Prather, Discrete Mathematical Structures for Computer Science, Houghton Mifflin, Boston, Mass., 1976.
[13] A. Wilansky, Topology for Analysis, Ginn, Waltham, Mass., 1970.

Author's Note

In preparing this manuscript
we were not assisted
by any grant;
in fact,
we were seriously distracted
by teaching duties,
inadequate facilities,
requirements that we attend
to details of administrative fervor
and making do
with inferior pay
and insecure position.

Editor's note: The author of the poem, while real, shall remain anonymous. Any resemblance to authors published in this journal is purely intentional.

Postscript

We have seen how any topology on a finite set A (and some topologies on an infinite set A) can be generated by dominance relations or preorders on A. If A is finite, the properties of the topology are reflected in both the matrix and the digraph of the relation. In particular, the matrix is an efficient, easily maintained data structure for the topology. Basis sets appear as columns, and open sets are disjunctions of columns. Closures of single point sets appear as rows, and closed sets are disjunctions of rows.

In asking standard topological questions, we found as answers both theorems and computational methods. We have also briefly illustrated how theorems from discrete mathematics can have topological interpretations.

How far is it from computer science to topology? Not far at all.

I would like to thank the editor and the referees for their helpful comments and suggestions. I would also like to thank Professor R. Fakler for introducing me to Christie's book and to the connection between dominance relations and topologies.

References

[1] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, Data Structures and Algorithms, Addison-Wesley, Reading, Mass., 1983.

[2] M. J. Augenstein and A. M. Tenenbaum, Data Structures and PL/1 Programming, Prentice-Hall, Englewood Cliffs, N.J., 1979.

[3] D. E. Christie, Basic Topology, Macmillan, New York, 1976.

[4] J. Dugundji, Topology, Allyn and Bacon, Boston, Mass., 1966.

[5] J. W. Evans, F. Harary, and M. S. Lynn, On the computer enumeration of finite topologies, Comm. of the ACM, 10, #5 (1967) 295–297.

[6] F. Harary, Graph Theory, Addison-Wesley, Reading, Mass., 1969.

[7] J. L. Kelly, General Topology, Van Nostrand, Princeton, N.J., 1955.

[8] V. Krishnamurthy, On the number of topologies on a finite set, Amer. Math. Monthly, 73 (1966) 154–157.

[9] K. Kuratowski, Topology, vol. 1, Academic Press, New York, 1966.

[10] L. S. Levy, Discrete Structures of Computer Science, Wiley, New York, 1980.

[11] C. L. Liu, Elements of Discrete Mathematics, McGraw-Hill, New York, 1977.

[12] R. E. Prather, Discrete Mathematical Structures for Computer Science, Houghton Mifflin, Boston, Mass., 1976.

[13] A. Wilansky, Topology for Analysis, Ginn, Waltham, Mass., 1970.

Author's Note

In preparing this manuscript
we were not assisted
by any grant;
in fact,
we were seriously distracted
by teaching duties,
inadequate facilities,
requirements that we attend
to details of administrative fervor
and making do
with inferior pay
and insecure position.

Editor's note: The author of the poem, while real, shall remain anonymous. Any resemblance to authors published in this journal is purely intentional.

NOTES

Poe's Pendulum

ROBERT L. BORRELLI
COURTNEY S. COLEMAN
DANA D. HOBSON
Harvey Mudd College
Claremont, CA 91711

The study of the motion of a steadily descending pendulum has practical applications (e.g., to the oscillations of a load being lowered by a crane). Our interest, however, was stimulated not by practicality but by a striking literary work in which a descending pendulum plays the role of an infernal machine. The story is Edgar Allan Poe's classic tale "The Pit and the Pendulum," written in 1842 and familiar to generations of *aficionados* of the macabre. In his customary florid style Poe tells of a prisoner bound flat on the floor of a chamber of horrors:

> ... Looking upward, I surveyed the ceiling of my prison. It was some thirty or forty feet overhead, and constructed much as the side walls. In one of its panels a very singular figure riveted my whole attention. It was the painted figure of Time as he is commonly represented, save that, in lieu of a scythe, he held what, at a casual glance, I supposed to be the pictured image of a huge pendulum, such as we see on antique clocks. There was something, however, in the appearance of this machine which caused me to regard it more attentively. While I gazed directly upward at it ... I fancied that I saw it in motion. In an instant afterward the fancy was confirmed. Its sweep was brief, and of course slow ... It might have been half an hour, perhaps even an hour ... before I again cast my eyes upward. What I then saw confounded and amazed me. The sweep of the pendulum had increased in extent by nearly a yard. As a natural consequence its velocity was also much greater. But what mainly disturbed me was the idea that it had perceptibly descended. I now observed—with what horror it is needless to say—that its nether extremity was formed of a crescent of glittering steel, about a foot in length from horn to horn; the horns upward, and the under edge as keen as that of a razor ... and the whole hissed as it swung through the air ... long, long hours of horror more than mortal during which I counted the rushing oscillations of the steel! Inch by inch—line by line—with a descent only appreciable at intervals that seemed ages—down and still down it came! ... The vibration of the pendulum was at right angles to my length. I saw that the crescent was designed to cross the region of my heart. ... its terrifically wide sweep (some thirty feet or more) ... Down—steadily down it crept. I took a frenzied pleasure in contrasting its downward with its lateral velocity. To the right to the left—far and wide—with the shriek of a damned spirit! ... Down—certainly, relentlessly down! ...

Poe's narrator-prisoner was hardly in any condition to make a mathematical analysis of the motion of the pendulum, but that is what we shall do. Can Poe be right that the *sweep was brief and of course slow*, but that, as the pendulum descended, *the sweep ... had increased by nearly a yard ... its velocity also was much greater*, until at last it had a *terrifically wide sweep* (*some thirty feet or more*)? Try an experiment with a weight at the end of a cord held taut over the edge of a table. Start the weight swinging and slowly pay out the cord. Some puzzling inconsistencies arise between the actual motion of the weight and Poe's descriptions. We shall see what a mathematical model implies about all this.

The equations of motion

Suppose a pendulum is supported by a wire whose length at time t is $L(t)$. Let $\theta(t)$ be the angle the wire makes with the downward vertical from the point of support. The equations may be derived as follows [5]. Suppose $\hat{r}$ and $\hat{\theta}$ are unit vectors at the point of support and oriented as indicated in FIGURE 1. The forces acting on the pendulum bob are the vertical force of gravity, the tension along the support wire, and a frictional force due to the air and tangent to the direction of motion. We shall ignore the frictional force, since its damping effect would appear to be small. The tensile force may also be ignored, not because its magnitude is small (it isn't), but because we shall only use the components of forces in the $\hat{\theta}$ direction orthogonal to the wire.

The equation of motion in the $\hat{\theta}$ direction is obtained from Newton's Second Law after some preliminary work with derivatives (all taken with respect to time). The position vector for the pendulum bob is $\vec{R} = L\hat{r}$. Since $\hat{r}' = \theta'\hat{\theta}$ and $\hat{\theta}' = -\theta'\hat{r}$, the velocity and acceleration vectors are given respectively by

$$\vec{R}' = (L\hat{r})' = L'\hat{r} + L\hat{r}' = L'\hat{r} + L\theta'\hat{\theta},$$

$$\vec{R}'' = L''\hat{r} + L'\hat{r}' + L'\theta'\hat{\theta} + L\theta''\hat{\theta} + L\theta'\hat{\theta}' = (L'' - L\theta'\theta')\hat{r} + (2L'\theta' + L\theta'')\hat{\theta}. \tag{1}$$

If the pendulum bob has mass m, then the component in the $\hat{\theta}$ direction of the gravitational force on the bob is $-mg\sin\theta$, where g is the gravitational constant at the surface of the earth. Since from (1) the component of the acceleration vector along $\hat{\theta}$ is $2L'\theta' + L\theta''$, Newton's Second Law yields the equation of angular motion,

$$m(2L'\theta' + L\theta'') = -mg\sin\theta. \tag{2}$$

After canceling the mass factor and replacing $\sin\theta$ by θ (a reasonable approximation for small $|\theta|$), the linear equation of motion is

$$L\theta'' + 2L'\theta' + g\theta = 0, \tag{3}$$

where some consistent set of units is assumed.

Poe's descriptions of the sweep and the velocity of the descending pendulum in the prison cell can be compared with the sweep, $L\theta$, and the curvilinear velocity, $(L\theta)'$, of the mathematical pendulum modeled by equation (3), once $L(t)$ has been given functional form.

The steadily descending pendulum: Bessel functions

Poe does not explicitly state that the descent is steady, but that is a plausible assumption:

$$L(t) = a + bt, \quad a \text{ and } b \text{ positive constants}, \quad t \geq t_0 \geq 0. \tag{4}$$

Poe uses feet as a measure of length, and so shall we. With length in feet, time in seconds, angles in radians, g is 32 ft/sec^2, $a + bt_0$ is the initial length of the pendulum, and b is the constant rate of extension of the pendulum support wire. The equation of angular motion is

$$(a + bt)\theta'' + 2b\theta' + g\theta = 0, \qquad t \geq 0. \tag{5}$$

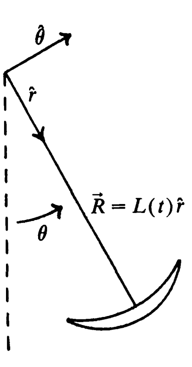


FIGURE 1. **Poe's pendulum.**

This equation may not be familiar, but changes in the time and angle variables transform it to a Bessel equation [4], [5]. In particular, let new variables x and y be defined by

$$x = \frac{2}{b}\sqrt{(a+bt)g}, \qquad y = \theta\sqrt{a+bt}.$$

Application of the chain rule then converts equation (5) to

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2-1)y = 0, \tag{6}$$

the Bessel equation of order one. Any solution $y(x)$ of (6) determines a solution of (5),

$$\theta(t) = \frac{1}{\sqrt{a+bt}}\,y\left(\frac{2}{b}\sqrt{(a+bt)g}\right), \tag{7}$$

and properties of solutions of Bessel's equation (6) translate to properties of the angular motion of a steadily descending pendulum.

Three of these properties are of significance in the current context. They are most easily expressed in terms of the general solution,

$$y_n(x) = \alpha J_n(x) + \beta Y_n(x), \qquad n = 0,1,\ldots, \qquad x > 0,$$

of Bessel's equation of order n,

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2-n^2)y = 0$$

where J_n and Y_n are the Bessel functions of order n of the first and second kind, and α and β are arbitrary constants independent of n. The functions $y_n(x)$ have the following properties [1], [2], [3], [5], [6]:

- The extreme values of $|y_n(x)|$ decay monotonically to zero as x increases to infinity.
- For any $c > 0$ the function $y_n(x)$ oscillates like a decaying sinusoid about $y = 0$,

  $$y_n(x) = \frac{A_n}{x^{1/2}}\cos(x-\phi_n) + \frac{r_n(x)}{x^{3/2}}, \qquad x \geq c,$$

  where A_n and ϕ_n are constants and $r_n(x)$ is bounded.
- The functions y_{n+1} and y_n satisfy the recursion relation,

  $$y'_{n+1}(x) = -\frac{n+1}{x}y_{n+1}(x) + y_n(x). \tag{8}$$

The function $y(t)$ in (7) has the form of $y_1(t)$ for constants α and β determined by initial conditions.

In light of (7) and the properties above we see that angle, sweep, and curvilinear velocity behave with the advance of time in the following ways.

- The angle $\theta(t)$ decays sinusoidally to zero, the magnitudes of its extrema monotonically decreasing like $(a+bt)^{-3/4}$. (See FIGURE 2.)
- The sweep, $L\theta \equiv (a+bt)\theta(t)$, grows sinusoidally in time, the magnitudes of its extrema monotonically increasing like $(a+bt)^{1/4}$. (See FIGURE 3.)
- The curvilinear velocity, $(L\theta)'$, decays sinusoidally to zero, the magnitudes of its extrema decreasing monotonically like $(a+bt)^{-1/4}$. (See FIGURE 4.)

The first and second properties follow directly from (7) and the properties given above of solutions of a Bessel equation. The third property is a consequence of (7), the chain rule, the relation $L = a + bt$, and the recursion relation. For, we have that

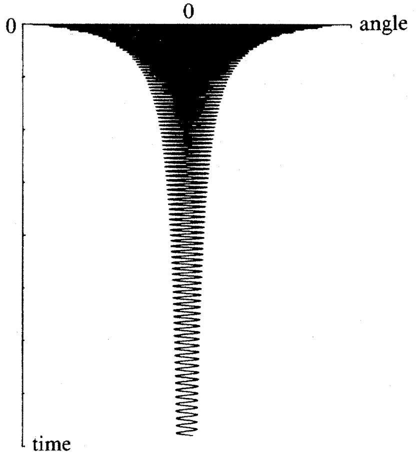


FIGURE 2. **The decaying angle of a steadily descending pendulum.**

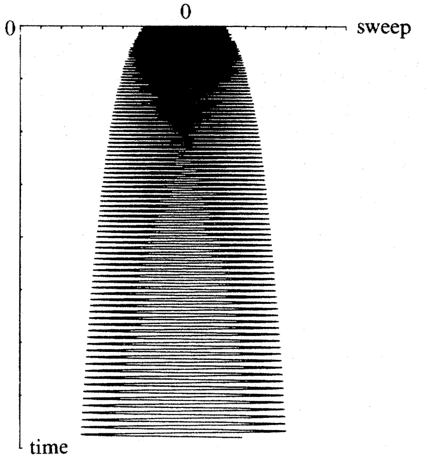


FIGURE 3. **The slowly growing sweep.**

$$\begin{aligned}\frac{d}{dt}(L\theta) &= \frac{d}{dL}(L\theta)\frac{dL}{dt} = \frac{d}{dL}\left[\sqrt{L}\,y_1\left(\frac{2}{b}\sqrt{gL}\right)\right]b\\ &= \frac{b}{2\sqrt{L}}y_1\left(\frac{2}{b}\sqrt{gL}\right) + \sqrt{g}\,y_1'\left(\frac{2}{b}\sqrt{gL}\right)\\ &= \sqrt{g}\,y_0\left(\frac{2}{b}\sqrt{gL}\right)\end{aligned}$$

where recursion relation (8) with $n = 0$ is used to obtain the last equality. The decay characteristics of the curvilinear velocity follow.

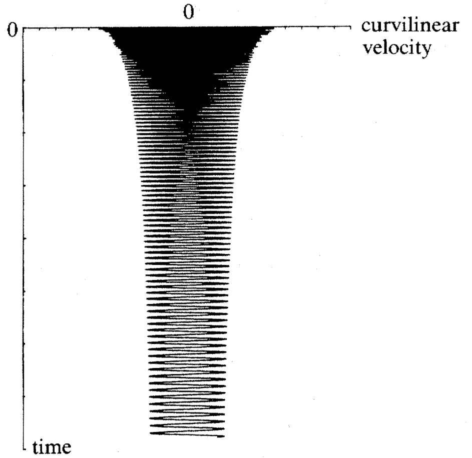


FIGURE 4. **The slow decay of the curvilinear velocity.**

How do these mathematical properties of the mathematical pendulum compare to the imagery of Poe's tale? Oscillations? Yes. Decaying angle? Poe's narrator has nothing to say about the angle. But the narrator does claim that there is a *terrifically wide sweep* (*thirty feet or more*). This seems mathematically implausible in light of the slow growth of the sweep proportional to $(a+bt)^{1/4}$. Specifically, let the pendulum support wire be initially one foot long, while the Bessel function y in (7) is initially at one of its extreme values, say M. Since the sweep is given by $L\theta = \sqrt{L}y$ and $L(t_0)=1$, then we certainly have that $|M| \le \pi/2$. From the monotone decay of the extrema of $|y|$, we conclude that for all $t \ge t_0$

$$|\text{sweep}| = |\sqrt{L}\,y| \le \sqrt{L}\,|M| \le \frac{\sqrt{L}\,\pi}{2}.$$

By the time the swinging pendulum has fallen forty feet to the floor of the cell (and bisected the prisoner), the sweep is no larger than $\sqrt{40}\,\pi/2 \approx 10$ feet, hardly the *terrifically wide sweep* (*some thirty feet or more*) of poetic imagination.

How about the *rushing oscillations of steel*, and the change from *its sweep was brief and of course slow* to *its velocity also was much greater*? The phrases suggest increasing extremes of curvilinear velocity, but, as pointed out above, $(L\theta)'$ has extrema whose magnitudes actually decay in time. Poe's pendulum is clearly not the pendulum of our mathematical model, but some other machine altogether.

The accelerating pendulum: Euler's equation

The length of the support wire of a plunging pendulum may be modeled by $L = bt^2$, where b is a positive constant, and to avoid a singularity, the initial time $t_0 = 1$. The equation of angular motion becomes the Euler equation,

$$bt^2\theta'' + 4bt\theta' + g\theta = 0,$$

which reduces to the constant coefficient equation

$$\frac{d^2\theta}{ds^2} + 3\frac{d\theta}{ds} + \frac{g}{b}\theta = 0, \tag{9}$$

where $s = \ln t$. There are three cases to consider, depending upon whether $4g > 9b$, $4g = 9b$, or $4g < 9b$. The reader may show that in the last two cases the pendulum is falling so fast that there are no oscillations at all. There are oscillations if $4g > 9b$, but they have little resemblance to those of Poe's pendulum.

In the oscillatory case corresponding to $4g > 9b$, the solutions of (9) have the form

$$e^{-3s/2}[C_1\cos cs + C_2\sin cs],$$

where the constants C_1 and C_2 are determined by the initial data and $c = \frac{1}{2}\sqrt{-9+4g/b}$. Replacing s by $\ln t$ and setting

$$C = \sqrt{C_1^2 + C_2^2}, \qquad \cos\phi = C_1/C, \qquad \sin\phi = C_2/C,$$

we have that

$$\theta(t) = Ct^{-3/2}\cos(c\ln t - \phi),$$
$$L(t)\theta(t) = bt^2\theta = bCt^{1/2}\cos(c\ln t - \phi),$$
$$(L\theta)' = bKt^{-1/2}\cos(c\ln t - \phi + \psi),$$

where $K = \sqrt{1/4 + c^2}$, $\cos\psi = 1/2K$, $\sin\psi = c/K$.

The rapidly falling oscillating pendulum behaves qualitatively like the steadily descending pendulum. The magnitudes of the extreme values of angle and of curvilinear velocity decay towards zero like $t^{-3/2}$ and $t^{-1/2}$, respectively, while the extrema of the sweep increase in magnitude but only like $t^{1/2}$. Once again, this is not Poe's pendulum.

The descending pendulum: the Sonin-Pólya theorem

Neither the model of the steadily descending pendulum nor that of the rapidly descending pendulum fits the descriptions given in the tale of the sweep and the velocity. Suppose now that we assume nothing about the nature of $L(t)$ except that $L(t) \geq A > 0$ and $L'(t) \geq B > 0$ for all $t \geq t_0$. Here is an open problem. *Is there a smooth function $L(t)$ for which the solutions of the equation of motion* (3) *produce the sweep and the velocity of Poe's pendulum*? Probably not, but we do not know how to prove it.

One can show, however, even in this general setting that the extreme values of the angle decay in magnitude. The result follows from the Sonin-Pólya theorem [**2**], [**3**]:

Suppose $p(t)q(t)$ is strictly increasing on an open interval where the smooth functions, p and q, satisfy $p > 0$, $q \neq 0$. Then the magnitudes of the extrema of any solution of the self-adjoint equation, $(p\theta')' + q\theta = 0$, form a strictly decreasing sequence.

The equation of motion (3) is not self-adjoint, but takes on that form upon multiplication by $L(t)$:

$$(L^2\theta')' + gL\theta = 0.$$

Since $L^2 \cdot gL$ is strictly increasing (recall that $L' \geq B > 0$), the magnitudes of the angular extrema of this generalized descending pendulum always decrease.

In fact this is true for the nonlinear simple pendulum of equation (2), which, after multiplication by L, has the "nonlinear self-adjoint" form,

$$(L^2\theta')' + gL\sin\theta = 0, \qquad |\theta| < \pi/2.$$

The proof is a straightforward modification of the usual proof [**2**], [**3**] of the Sonin-Pólya theorem and is omitted. The function $\sin\theta$, $|\theta| < \pi/2$, may be replaced by any smooth, strictly increasing odd function on an interval centered at 0 and the conclusion of the Sonin-Pólya theorem still holds. The unanswered question in this general setting is: *how do $L\theta$ and $(L\theta)'$ behave*?

Other approaches

Poe's narrative conveys the vivid impression of a pendulum of steadily increasing total energy. Perhaps an approach along the lines of energy balance would work, but the authors found it curiously difficult to construct a satisfactory model.

Computer routines for solving differential equations and plotting the solutions may also be used. In fact, given a function $L(t)$ for the length of the pendulum, a computer analysis may be the most effective way to determine the behavior of sweep and velocity, even if the differential equations reduce to classical equations such as those of Bessel or Euler. This analysis may lead not only to a deeper understanding of the motions of the descending pendulum, but to an appreciation of the foibles of numerical routines and their computer implementation.

References

[1] M. Abramowitz and I. A. Stegun, eds., Handbook of Mathematical Functions, Dover Publications, Inc., New York, 1965.

[2] G. Birkhoff and G.-C. Rota, Ordinary Differential Equations, 3rd ed., John Wiley & Sons, New York, 1978.

[3] R. L. Finney and D. R. Ostberg, Elementary Differential Equations with Linear Algebra, Addison-Wesley, Reading, Mass., 1981.

[4] E. Kamke, Differentialgleichungen: Lösungsmethoden und Lösungen, Chelsea Publ. Co., New York, 1959.

[5] F. E. Relton, Applied Bessel Functions, Dover Publications, Inc., New York, 1956.

[6] C. R. Wylie and L. C. Barrett, Advanced Engineering Mathematics, 5th ed., McGraw-Hill, New York, 1982.

A Method for Vector Proofs in Geometry

N. J. LORD
Tonbridge School
Kent TN9 1JP, England

In his book on problem-solving, Loren C. Larson observes that, for proofs in plane geometry, it is often convenient to have a notation to represent the rotation of a vector through 90° (see [**6**], Example 8.3.5, pp. 305–6). Rather than introduce such a notation and face the necessity of developing both its properties and facility in its use, we suggest the following natural alternative which relies on the exploitation of standard properties of the vector cross-product, and which can lead to elegant vector proofs. Our approach also serves as an alternative to introducing complex numbers, where multiplication by i corresponds to rotation through 90° (as utilised, for example, in one of the standard proofs of Napoleon's Theorem—see [**6**], Example 8.4 pp. 314–5 and compare with our proof below), and can achieve a succinctness close to that attained by using two-dimensional transformation geometry in the spirit of [**2**], [**3**], and [**5**].

We start by reworking Larson's Example 8.3.5 (a problem originally due to M. Slater in [**10**]) using our method, and then give several further examples to illustrate its message: that it may sometimes be helpful to 'three-dimensionalise' the proofs of certain results in plane geometry by introducing a fixed vector perpendicular to the plane and that such a vector, far from complicating the solution, may actually serve as a catalyst in the proof.

In Slater's Problem, we are asked to prove: *if similar isosceles triangles OAB', OBA', ABO' are erected on the sides of a triangle OAB (respectively externally, externally, internally), as in* FIGURE 1, *then $OA'O'B'$ is a parallelogram.*

Introduce $\mathbf{k}$, a unit vector 'up' out of the plane of $\overrightarrow{OA} = \mathbf{a}$ and $\overrightarrow{OB} = \mathbf{b}$. Then $\mathbf{k}\times\mathbf{a}, \mathbf{k}\times\mathbf{b}$ are normal to $\mathbf{a}, \mathbf{b}$, respectively, and lie in the plane of $\mathbf{a}$ and $\mathbf{b}$.

Similarity of the constructed isosceles triangles implies the existence of a positive scalar m such that:

$$\overrightarrow{OB'} = \tfrac{1}{2}\mathbf{a} + m(\mathbf{k}\times\mathbf{a})$$

$$\overrightarrow{OA'} = \tfrac{1}{2}\mathbf{b} - m(\mathbf{k}\times\mathbf{b})$$

$$\overrightarrow{OO'} = \tfrac{1}{2}(\mathbf{a}+\mathbf{b}) + m\mathbf{k}\times(\mathbf{a}-\mathbf{b}).$$

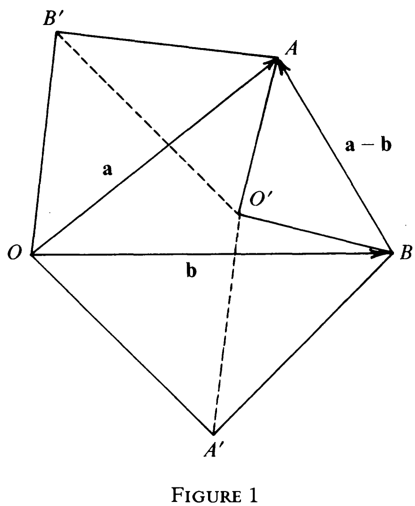


FIGURE 1

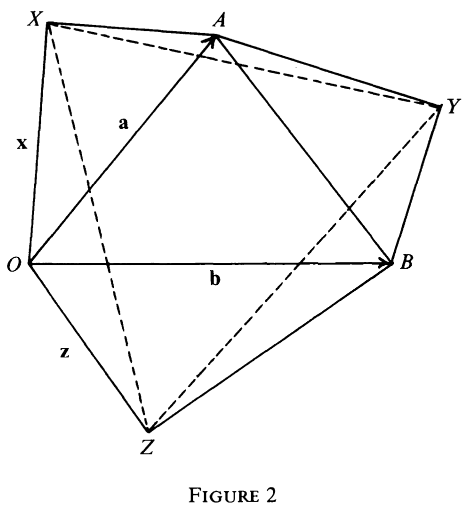


FIGURE 2

It is thus clear that $\overrightarrow{OO'} = \overrightarrow{OA'} + \overrightarrow{OB'}$ and hence that $OA'O'B'$ is a parallelogram.

Notice how smoothly a scaling factor can be incorporated into $\mathbf{k}$ and the cross product terms; this will be a recurring feature of the proofs that follow, too.

The same approach works as well on Exercises 8.3.12–8.3.16 in Larson's book and, as a second example, we prove an extension of Exercise 8.3.15 attributed to W. L. Ferrar by E. A. Maxwell in [**8**]. *Given any triangle OAB, let X, Y, Z denote the new vertices of externally erected similar triangles with bases OA, AB, OB.* (See FIGURE 2.) *Then triangles XYZ and OAB have the same centroid.*

Here, we introduce the unit vector $\mathbf{k}$ as before and observe that similarity of the constructed (not necessarily isosceles!) triangles this time implies the existence of two positive scalars m, n such that:

$$\mathbf{x} = \overrightarrow{OX} = m\mathbf{a} + n(\mathbf{k} \times \mathbf{a})$$

$$\mathbf{y} = \overrightarrow{OY} = \mathbf{a} + m(\mathbf{b} - \mathbf{a}) + n\mathbf{k} \times (\mathbf{b} - \mathbf{a})$$

$$\mathbf{z} = \overrightarrow{OZ} = (1 - m)\mathbf{b} - n\mathbf{k} \times \mathbf{b}.$$

Adding these equations gives $\frac{1}{3}(\mathbf{x} + \mathbf{y} + \mathbf{z}) = \frac{1}{3}(\mathbf{a} + \mathbf{b})$ or, equivalently, by a standard vector characterization of the centroid, that XYZ and OAB have the same centroid.

Our remaining examples all involve use of the triple scalar product. We write $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ to denote the triple scalar product, $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$, of three vectors and to emphasize two properties that we shall use repeatedly in what follows:

(i) The triple scalar product is linear in each term (meaning, for example, that $[\mathbf{a} + m\mathbf{a}', \mathbf{b}, \mathbf{c}] = [\mathbf{a}, \mathbf{b}, \mathbf{c}] + m[\mathbf{a}', \mathbf{b}, \mathbf{c}]$), and

(ii) The triple scalar product changes sign if two vectors are interchanged (in particular, it vanishes if any two constituent vectors are equal).

We also employ, without further comment, the trivial observation that

(iii) $(\mathbf{k} \times \mathbf{a}) \cdot (\mathbf{k} \times \mathbf{b}) = |\mathbf{k}|^2(\mathbf{a} \cdot \mathbf{b})$ if $\mathbf{k}$ is perpendicular to $\mathbf{a}$ and $\mathbf{b}$.

With these remarks in mind, we give a proof of the following result from [**2**]: *if, on adjacent sides AC and CB of the parallelogram $ACBO$, external equilateral triangles ACX and CBY are erected, then triangle OXY is also equilateral.* (See FIGURE 3.)

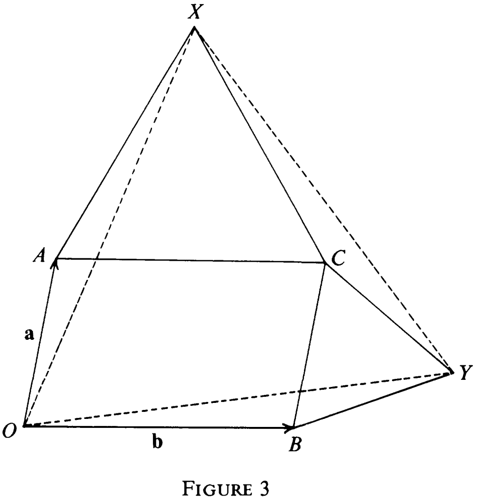


FIGURE 3

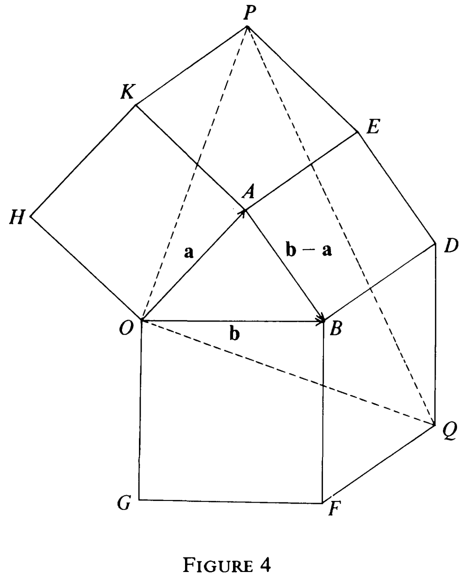


FIGURE 4

This time, take $\mathbf{k}$ to be 'up' out of the plane of $\mathbf{a}$ and $\mathbf{b}$ and of magnitude $\frac{1}{2}\sqrt{3}$. The position vectors of X and Y are then easily seen to be:

$$\mathbf{x} = \mathbf{a} + \tfrac{1}{2}\mathbf{b} + (\mathbf{k} \times \mathbf{b})$$
$$\mathbf{y} = \tfrac{1}{2}\mathbf{a} + \mathbf{b} - (\mathbf{k} \times \mathbf{a}),$$

whence

$$\mathbf{x} - \mathbf{y} = \tfrac{1}{2}\mathbf{a} - \tfrac{1}{2}\mathbf{b} + \mathbf{k} \times (\mathbf{a} + \mathbf{b}).$$

Thus

$$\begin{aligned} |\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x} &= \left(\mathbf{a} + \tfrac{1}{2}\mathbf{b} + \mathbf{k} \times \mathbf{b}\right) \cdot \left(\mathbf{a} + \tfrac{1}{2}\mathbf{b} + \mathbf{k} \times \mathbf{b}\right) \\ &= |\mathbf{a}|^2 + \tfrac{1}{4}|\mathbf{b}|^2 + |\mathbf{k} \times \mathbf{b}|^2 + \mathbf{a} \cdot \mathbf{b} + 2[\mathbf{a}, \mathbf{k}, \mathbf{b}] \\ &= |\mathbf{a}|^2 + |\mathbf{b}|^2 + \mathbf{a} \cdot \mathbf{b} + 2[\mathbf{a}, \mathbf{k}, \mathbf{b}]. \end{aligned}$$

But the latter expression is unchanged by the substitutions

$$\begin{cases} \mathbf{a} \to & \mathbf{a} + \mathbf{b} \\ \mathbf{b} \to & -\mathbf{a} \end{cases} \quad \text{and} \quad \begin{cases} \mathbf{a} \to & -\mathbf{b} \\ \mathbf{b} \to & \mathbf{a} + \mathbf{b} \end{cases}$$

corresponding to the evaluation of $|\mathbf{y}|^2$ and $|\mathbf{x} - \mathbf{y}|^2$, respectively, which establishes the result.

In a similar vein, we prove the somewhat surprising theorem stated in Ogilvy ([**9**], p. 120). *Given any triangle OAB, construct the exterior squares $ABDE$, $OBFG$, $AOHK$ and complete the parallelograms $FBDQ$, $EAKP$.* (See FIGURE 4.) *Then POQ is always a right-angled isosceles triangle.*

In this case, let $\mathbf{k}$ be a unit vector 'up' out of the plane of $\mathbf{a}$ and $\mathbf{b}$. The position vectors $\mathbf{p}, \mathbf{q}$ of P and Q are then given by:

$$\mathbf{p} = \mathbf{a} + (\mathbf{k} \times \mathbf{a}) + \mathbf{k} \times (\mathbf{b} - \mathbf{a}) = \mathbf{a} + \mathbf{k} \times \mathbf{b},$$
$$\mathbf{q} = \mathbf{b} + (\mathbf{b} \times \mathbf{k}) + \mathbf{k} \times (\mathbf{b} - \mathbf{a}) = \mathbf{b} - \mathbf{k} \times \mathbf{a}.$$

Then, as above, $|\mathbf{p}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 + 2[\mathbf{a}, \mathbf{k}, \mathbf{b}]$, which, being invariant under the substitution

$$\begin{cases} \mathbf{a} \to & \mathbf{b} \\ \mathbf{b} \to & -\mathbf{a}, \end{cases}$$

corresponding to the evaluation of $|\mathbf{q}|^2$, shows that $|\mathbf{p}| = |\mathbf{q}|$. Also, $\mathbf{p} \cdot \mathbf{q} = \mathbf{b} \cdot \mathbf{a} - (\mathbf{k} \times \mathbf{b}) \cdot (\mathbf{k} \times \mathbf{a}) = \mathbf{b} \cdot \mathbf{a} - \mathbf{b} \cdot \mathbf{a} = 0$, which concludes the proof. (Here, the vector proof is noticeably shorter and more elegant than a 'Euclidean' proof: compare [**9**], p. 169.)

We next fulfill our promise in the introduction by providing a proof of 'Napoleon's Theorem': *the centroids of equilateral triangles erected (externally) on the sides of any triangle again form an equilateral triangle.*

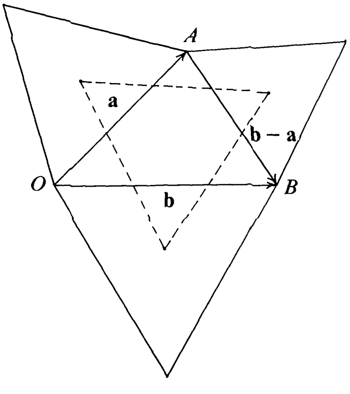


FIGURE 5

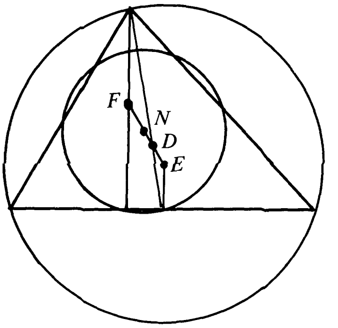


FIGURE 6

With notation as in FIGURE 5, let $\mathbf{k}$ be a vector 'up' out of the plane of $\mathbf{a}$ and $\mathbf{b}$, of magnitude $\sqrt{3}/6$. Then the position vectors of the three centroids involved are readily seen to be $\frac{1}{2}\mathbf{a} + (\mathbf{k}\times\mathbf{a})$, $\frac{1}{2}\mathbf{b} - (\mathbf{k}\times\mathbf{b})$, $\frac{1}{2}(\mathbf{a}+\mathbf{b}) + \mathbf{k}\times(\mathbf{b}-\mathbf{a})$, so that the sides of the triangle formed by these centroids have representing vectors $\frac{1}{2}\mathbf{a} + \mathbf{k}\times(2\mathbf{b}-\mathbf{a})$, $\frac{1}{2}\mathbf{b} + \mathbf{k}\times(\mathbf{b}-2\mathbf{a})$ and $\frac{1}{2}(\mathbf{a}-\mathbf{b}) + \mathbf{k}\times(\mathbf{a}+\mathbf{b})$. A calculation along now familiar lines quickly shows that:

$$\begin{aligned} 3\left|\tfrac{1}{2}\mathbf{a} + \mathbf{k}\times(2\mathbf{b}-\mathbf{a})\right|^2 &= 3\left(\tfrac{1}{4}|\mathbf{a}|^2 + \tfrac{1}{12}(2\mathbf{b}-\mathbf{a})\cdot(2\mathbf{b}-\mathbf{a}) + 2[\mathbf{a},\mathbf{k},\mathbf{b}]\right) \\ &= |\mathbf{a}|^2 + |\mathbf{b}|^2 - \mathbf{a}\cdot\mathbf{b} + 6[\mathbf{a},\mathbf{k},\mathbf{b}] \end{aligned}$$

which, being invariant under the respective substitutions

$$\begin{cases} \mathbf{a}\to\mathbf{b} \\ \mathbf{b}\to\mathbf{b}-\mathbf{a} \end{cases} \quad\text{and}\quad \begin{cases} \mathbf{a}\to\mathbf{a}-\mathbf{b} \\ \mathbf{b}\to\mathbf{a}, \end{cases}$$

establishes that the sides of the triangle concerned have equal length. (The proof is readily modified to obtain the analogous theorem for internally erected equilateral triangles.)

For a more classical application, we next give a short proof of the ratio properties of the Euler Line. We take for granted the existence of the centroid D, circumcentre E, orthocentre F and nine-point centre N of a triangle (recall that the latter may be defined as the centre of that circle which, rather impressively, can always be drawn through the midpoints of sides and feet of altitudes of a triangle—see [**9**], pp. 117–20 for further details), and aim to prove the theorem that *D, E, F, N lie on a line (the Euler Line of the triangle) with $ED:DN:NF = 2:1:3$.* (See FIGURE 6.)

We label the vertices of the triangle O, A, B and use the obvious letters to denote the position vectors of points relative to O. Take $\mathbf{k}$ to be a fixed (non-zero) vector normal to the plane of $\mathbf{a}$ and $\mathbf{b}$.

The position vector of the centroid of OAB is given by $\mathbf{d} = \frac{1}{3}(\mathbf{a}+\mathbf{b})$. The position of the circumcentre is determined by the point of intersection of the perpendicular bisectors of OA and OB, that is, from the existence of scalars p, q such that $\mathbf{e} = \frac{1}{2}\mathbf{b} + p(\mathbf{b}\times\mathbf{k}) = \frac{1}{2}\mathbf{a} + q(\mathbf{a}\times\mathbf{k})$; from which $\frac{1}{2}\mathbf{b}\cdot\mathbf{b} + p[\mathbf{b},\mathbf{k},\mathbf{b}] = \frac{1}{2}\mathbf{a}\cdot\mathbf{b} + q[\mathbf{a},\mathbf{k},\mathbf{b}]$ and hence

$$q = \frac{(\mathbf{b}-\mathbf{a})\cdot\mathbf{b}}{2[\mathbf{a},\mathbf{k},\mathbf{b}]}.$$

The position of the orthocentre is determined by the point of intersection of the altitudes through A and B, that is, from the existence of scalars r, s such that $\mathbf{f} = \mathbf{a} + r(\mathbf{b}\times\mathbf{k}) = \mathbf{b} + s(\mathbf{a}\times\mathbf{k})$, from which

$$s = \frac{(\mathbf{a}-\mathbf{b})\cdot\mathbf{b}}{[\mathbf{a},\mathbf{k},\mathbf{b}]} = -2q.$$

Similarly, using the definition of the nine-point centre given above, we obtain $\mathbf{n} = \frac{1}{4}\mathbf{a} + \frac{1}{2}\mathbf{b} - \frac{1}{2}q(\mathbf{a}\times\mathbf{k})$. Thus:

$$\begin{aligned} \mathbf{d} &= \tfrac{1}{3}\mathbf{a} + \tfrac{1}{3}\mathbf{b} \\ \mathbf{e} &= \tfrac{1}{2}\mathbf{a} \qquad\quad + q(\mathbf{a}\times\mathbf{k}) \\ \mathbf{f} &= \qquad\quad \mathbf{b} - 2q(\mathbf{a}\times\mathbf{k}) \\ \mathbf{n} &= \tfrac{1}{4}\mathbf{a} + \tfrac{1}{2}\mathbf{b} - \tfrac{1}{2}q(\mathbf{a}\times\mathbf{k}) \end{aligned}$$

and the required property:

$$\tfrac{1}{3}(\mathbf{f}-\mathbf{n}) = \mathbf{n}-\mathbf{d} = \tfrac{1}{2}(\mathbf{d}-\mathbf{e})$$

follows immediately.

It is interesting to compare this proof with the more standard vector approaches to the theorem in [**7**] and [**11**].

As our penultimate example we tackle the problem in [**1**]: *if, in the isosceles triangle OAB of* FIGURE 7 *with circumcentre E, OA = OB, D is the midpoint of OA and C is the centroid of triangle OBD, then EC and BD are perpendicular.*

Retaining the notation of the previous example, we have:

$$\mathbf{e}=\frac{1}{2}\mathbf{a}+\frac{(\mathbf{b}-\mathbf{a})\cdot\mathbf{b}}{2[\mathbf{a},\mathbf{k},\mathbf{b}]}(\mathbf{a}\times\mathbf{k}) \quad \text{and} \quad \mathbf{c}=\frac{1}{3}\left(\mathbf{b}+\frac{1}{2}\mathbf{a}\right),$$

whence

$$\overrightarrow{EC}=\frac{1}{3}\mathbf{a}-\frac{1}{3}\mathbf{b}+\frac{(\mathbf{b}-\mathbf{a})\cdot\mathbf{b}}{2[\mathbf{a},\mathbf{k},\mathbf{b}]}(\mathbf{a}\times\mathbf{k}).$$

Since $\overrightarrow{BD}=\frac{1}{2}\mathbf{a}-\mathbf{b}$ we deduce:

$$\begin{aligned}6\overrightarrow{EC}\cdot\overrightarrow{BD}&=3\overrightarrow{EC}\cdot 2\overrightarrow{BD}\\&=|\mathbf{a}|^2-\mathbf{a}\cdot\mathbf{b}-2\mathbf{a}\cdot\mathbf{b}+2|\mathbf{b}|^2-3(\mathbf{b}-\mathbf{a})\cdot\mathbf{b}\\&=|\mathbf{a}|^2-|\mathbf{b}|^2\\&=0\end{aligned}$$

because, and indeed only because, $|\mathbf{a}|=|\mathbf{b}|$.

Finally, we rework problem 1966/6 in [**4**], p. 8 (see also pp. 94–7): *if, in the interior of sides OA, AB, BO of triangle OAB points P, Q, R respectively are selected, then the area of at least one of the triangles BRQ, OPR, APQ is less than or equal to one-quarter of the area of OAB.* (See FIGURE 8.)

As usual, we take **k** to be a unit vector 'up' out of the plane of **a** and **b**. By hypothesis, there are scalars x, y, z with $0<x, y, z<1$, and $\mathbf{p}=x\mathbf{a}$, $\mathbf{r}=y\mathbf{b}$, $\mathbf{q}=\mathbf{a}+z(\mathbf{b}-\mathbf{a})$.

Then, writing ΔOAB to denote the area of triangle OAB, it is easily checked that:

$$\Delta OAB=\tfrac{1}{2}[\mathbf{b},\mathbf{a},\mathbf{k}]=c, \text{ say},$$

$$\Delta OPR=\tfrac{1}{2}xy[\mathbf{b},\mathbf{a},\mathbf{k}]=xyc,$$

$$\Delta APQ=\tfrac{1}{2}[-(1-x)\mathbf{a},z(\mathbf{b}-\mathbf{a}),\mathbf{k}]=\tfrac{1}{2}(1-x)z[\mathbf{b},\mathbf{a},\mathbf{k}]=(1-x)zc,$$

$$\Delta BRQ=\tfrac{1}{2}[(1-z)(\mathbf{a}-\mathbf{b}),-(1-y)\mathbf{b},\mathbf{k}]=\tfrac{1}{2}(1-z)(1-y)[\mathbf{b},\mathbf{a},\mathbf{k}]=(1-z)(1-y)c.$$

But the product of the coefficients of c in the last three equations is

$$xy(1-x)z(1-z)(1-y)=x(1-x)y(1-y)z(1-z),$$

and this product must be less than or equal to $\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}$, by the inequality of arithmetic and geometric means. Thus at least one of xy, $(1-x)z$, $(1-z)(1-y)$ is less than or equal to a quarter, as required.

Further geometric theorems which are amenable to proof by our methods may be found in [**6**] (pp. 310–11) and [**5**].

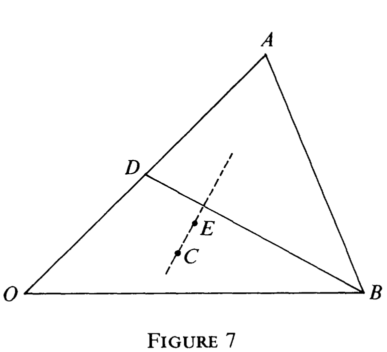


FIGURE 7

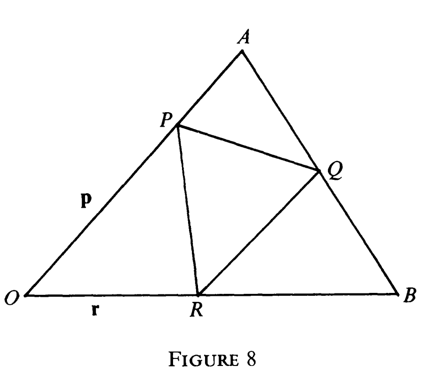


FIGURE 8

References

[1] British Mathematical Olympiad 1983, Question 1.
[2] Bulletin of the Institute of Mathematics and its Applications, 17 (1981), p. 93 Problem EFG 27 and p. 125 for a solution.
[3] Ross Finney, Dynamic proofs of Euclidean theorems, this MAGAZINE, 43 (1970) 177–185.
[4] S. L. Greitzer, International Mathematical Olympiads 1959–1977, The Mathematical Association of America, 1978.
[5] Oscar F. Kaputski, To the vectors go the spoils, California Math., 5 (1980) 26–34.
[6] Loren C. Larson, Problem-Solving Through Problems, Springer-Verlag, 1983.
[7] A. M. Macbeath, Elementary Vector Algebra, 2nd ed., Oxford, 1967, pp. 23–24, 31, 35.
[8] E. A. Maxwell, Geometry for Advanced Pupils, Oxford, 1960, pp. 45–46.
[9] C. S. Ogilvy, Excursions in Geometry, Oxford, 1969.
[10] M. Slater, Amer. Math. Monthly, 88 (1981) 66–67.
[11] D. E. Wallis, Vectors and the nine-point circle, Math. Gazette, 66 (1982) 224–225.

Incenters and Excenters Viewed from the Euler Line

ANDREW P. GUINAND
Trent University
Peterborough, Ontario, K9J 7B8, Canada

In the geometry of the triangle the relative positions of the circumcenter O, the centroid G, the nine-point center N, and the orthocenter H have been known for over a century and a half. They form the Euler line and they are so spaced that $OG:GN:NH = 2:1:3$. Euler also found other formulas for the distances between these points and for their distances from the incenter I_0 and from the excenters I_1, I_2, I_3. For a quick refresher course on terminology, readers should see [**5**]. Only recently have the full implications of such results on the possible positions of these so-called 'tritangent centers' relative to the Euler line been clarified [**2**], [**3**]. This clarification came through the investigation of a problem posed by Wernick [**6**] in this MAGAZINE—to reconstruct a triangle given its incenter and any two of the centers forming the Euler line. Since the Euler line of an equilateral triangle is indeterminate (the points O, G, N, H coincide) we exclude the equilateral case from consideration.

There are two main results presented in [**3**]. First, that *the incenter I_0 always lies inside the circle on diameter GH, and the excenters I_1, I_2, I_3 all lie outside it*. Second, that *relative to the Euler line there is a curiously shaped region, bounded by a closed quartic curve, inside which no tritangent center can lie*. Once it is envisaged, the first result is not difficult to prove, but the original proof of the second result involved truly horrendous algebraic calculations in finding and factorizing the discriminant of a cubic for the cosines of the angles of the triangle.

The object of this note is to give an easier and more geometric (or trigonometric) approach to the second result, together with a simple ruler-and-compass method of constructing points on the boundary of the forbidden zone for tritangent centers.

Let the triangle be ABC, and R the circumradius, r the inradius, r_1 the exradius corresponding to the excenter I_1, and let K and L be the points on the Euler line which are respectively the reflections of H and G in O; that is to say $KL:LO:OG:GN:NH = 4:2:2:1:3$. We use Σ to denote cyclic sums over the angles A, B, C of the triangle, i.e., $\Sigma \cos A$ means $\cos A + \cos B + \cos C$. Then standard trigonometric results are [**4**]:

References

[1] British Mathematical Olympiad 1983, Question 1.
[2] Bulletin of the Institute of Mathematics and its Applications, 17 (1981), p. 93 Problem EFG 27 and p. 125 for a solution.
[3] Ross Finney, Dynamic proofs of Euclidean theorems, this MAGAZINE, 43 (1970) 177–185.
[4] S. L. Greitzer, International Mathematical Olympiads 1959–1977, The Mathematical Association of America, 1978.
[5] Oscar F. Kaputski, To the vectors go the spoils, California Math., 5 (1980) 26–34.
[6] Loren C. Larson, Problem-Solving Through Problems, Springer-Verlag, 1983.
[7] A. M. Macbeath, Elementary Vector Algebra, 2nd ed., Oxford, 1967, pp. 23–24, 31, 35.
[8] E. A. Maxwell, Geometry for Advanced Pupils, Oxford, 1960, pp. 45–46.
[9] C. S. Ogilvy, Excursions in Geometry, Oxford, 1969.
[10] M. Slater, Amer. Math. Monthly, 88 (1981) 66–67.
[11] D. E. Wallis, Vectors and the nine-point circle, Math. Gazette, 66 (1982) 224–225.

Incenters and Excenters Viewed from the Euler Line

ANDREW P. GUINAND
Trent University
Peterborough, Ontario, K9J 7B8, Canada

In the geometry of the triangle the relative positions of the circumcenter O, the centroid G, the nine-point center N, and the orthocenter H have been known for over a century and a half. They form the Euler line and they are so spaced that $OG:GN:NH=2:1:3$. Euler also found other formulas for the distances between these points and for their distances from the incenter I_0 and from the excenters I_1, I_2, I_3. For a quick refresher course on terminology, readers should see [**5**]. Only recently have the full implications of such results on the possible positions of these so-called 'tritangent centers' relative to the Euler line been clarified [**2**], [**3**]. This clarification came through the investigation of a problem posed by Wernick [**6**] in this MAGAZINE—to reconstruct a triangle given its incenter and any two of the centers forming the Euler line. Since the Euler line of an equilateral triangle is indeterminate (the points O, G, N, H coincide) we exclude the equilateral case from consideration.

There are two main results presented in [**3**]. First, that *the incenter I_0 always lies inside the circle on diameter GH, and the excenters I_1, I_2, I_3 all lie outside it*. Second, that *relative to the Euler line there is a curiously shaped region, bounded by a closed quartic curve, inside which no tritangent center can lie*. Once it is envisaged, the first result is not difficult to prove, but the original proof of the second result involved truly horrendous algebraic calculations in finding and factorizing the discriminant of a cubic for the cosines of the angles of the triangle.

The object of this note is to give an easier and more geometric (or trigonometric) approach to the second result, together with a simple ruler-and-compass method of constructing points on the boundary of the forbidden zone for tritangent centers.

Let the triangle be ABC, and R the circumradius, r the inradius, r_1 the exradius corresponding to the excenter I_1, and let K and L be the points on the Euler line which are respectively the reflections of H and G in O; that is to say $KL:LO:OG:GN:NH=4:2:2:1:3$. We use Σ to denote cyclic sums over the angles A, B, C of the triangle, i.e., $\Sigma \cos A$ means $\cos A+\cos B+\cos C$. Then standard trigonometric results are [**4**]:

$$\Sigma \cos^2 A = 1 - 2\cos A \cos B \cos C, \tag{1}$$

$$r = 4R \sin\tfrac{1}{2}A \sin\tfrac{1}{2}B \sin\tfrac{1}{2}C = R(\Sigma \cos A - 1), \tag{2}$$

$$\begin{aligned} r_1 &= 4R \sin\tfrac{1}{2}A \cos\tfrac{1}{2}B \cos\tfrac{1}{2}C \\ &= R(\cos B + \cos C - \cos A + 1). \end{aligned} \tag{3}$$

Also

$$OI_0^2 = R^2 - 2Rr, \tag{4}$$

$$OI_1^2 = R^2 + 2Rr_1 \tag{5}$$

$$OH^2 = R^2(1 - 8\cos A \cos B \cos C), \tag{6}$$

$$I_0N = \tfrac{1}{2}R - r, \tag{7}$$

$$I_1N = \tfrac{1}{2}R + r_1. \tag{8}$$

Formulas (4), (5), and (6) stem from Euler, and (7) and (8) are equivalent to Feuerbach's theorem on the contact of tritangent circles with the nine-point circle. From these results we can deduce that certain regions are forbidden to tritangent centers.

LEMMA 1. *$I_0K \leqslant 4R + r$, and $I_1K \leqslant 4R - r_1$, but equality only occurs for degenerate triangles with at least one zero angle.*

Proof. Since O divides KN in the ratio $2:1$,

$$I_0K^2 + 2I_0N^2 = 3OI_0^2 + \tfrac{2}{3}KN^2 = 3OI_0^2 + \tfrac{3}{2}OH^2$$

by Stewart's theorem [**1**]. Hence by (4), (6), (7),

$$\begin{aligned} I_0K^2 &= 3(R^2 - 2Rr) + \tfrac{3}{2}R^2(1 - 8\cos A \cos B \cos C) - 2\left(\tfrac{1}{2}R - r\right)^2 \\ &= 4R^2 - 4Rr - 2r^2 - 12R^2 \cos A \cos B \cos C, \end{aligned}$$

whence

$$\begin{aligned} (4R + r)^2 - I_0K^2 &= 3\left[(2R + r)^2 + 4R^2 \cos A \cos B \cos C\right] \\ &= 3R^2\left[(\Sigma \cos A + 1)^2 + 4\cos A \cos B \cos C\right], \quad \text{by (2)} \\ &= 3R^2\left[\Sigma \cos^2 A + 2\Sigma \cos B \cos C + 2\Sigma \cos A + 1 + 4\cos A \cos B \cos C\right] \\ &= 6R^2[1 + \Sigma \cos A + \Sigma \cos B \cos C + \cos A \cos B \cos C], \quad \text{by (1)} \\ &= 6R^2(1 + \cos A)(1 + \cos B)(1 + \cos C) \\ &\geqslant 0, \end{aligned}$$

with equality only if one of the cosines is -1. That is, equality occurs only if one of the angles is π, and consequently the other two angles are both zero. Hence for all non-degenerate triangles, $I_0K < 4R + r$.

If, in the formulas (2), (4), (7), we replace r by $-r_1$ and the angles B and C by their supplements (i.e. change the signs of $\cos B$ and $\cos C$) we obtain the corresponding formulas (3), (5), (8) which pertain to the excenter I_1. Consequently an argument similar to the preceding gives

$$(4R - r_1)^2 - I_1K^2 = 6R^2(1 + \cos A)(1 - \cos B)(1 - \cos C) \geqslant 0$$

with equality only if one of B or C is zero, and the other is the supplement of A. By (3) $4R \geqslant r_1$ with equality only if $A = \pi$, $B = C = 0$. Hence for all non-degenerate triangles, $I_1K < 4R - r_1$.

LEMMA 2. *If I is any tritangent center then*

$$9OI^2 - 4IN^2 \geqslant 4IN \cdot IK. \tag{9}$$

Proof. If I is the incenter I_0, then by (4) and (7)

$$9OI_0^2 - 4I_0N^2 - 4I_0N \cdot I_0K = 9(R^2 - 2Rr) - 4(\tfrac{1}{2}R - r)^2 - 4(\tfrac{1}{2}R - r)I_0K$$
$$= 2(R - 2r)(4R + r - I_0K).$$

But $R \geqslant 2r$ by (4), so (9) follows by Lemma 1. Similarly if I is an excenter I_1 then

$$9OI_1^2 - 4I_1N^2 - 4I_1N \cdot I_1K = 2(R + 2r_1)(4R - r_1 - I_1K),$$

and (9) again follows by Lemma 1.

LEMMA 3. *If the point K and the line KOH are taken as origin and axis of polar coordinates (ρ, θ), and $KO = OH = \kappa$, then*

$$3\rho^2 - 4\kappa\rho\cos\theta - 12\kappa^2\sin^2\theta = 0 \tag{10}$$

is the equation of a closed curve enclosing all points I for which

$$9OI^2 - 4IN^2 < 4IN \cdot IK. \tag{11}$$

Proof. FIGURE 1 illustrates the situation described. The boundary between the sets of points I satisfying the inequalities (9) and (11) consists of those I for which

$$9OI^2 - 4IN^2 = 4IN \cdot IK. \tag{12}$$

Now $IK = \rho$, $KN = \frac{3}{2}\kappa$, so by the cosine rule

$$OI^2 = \rho^2 + \kappa^2 - 2\kappa\rho\cos\theta, \qquad IN^2 = \rho^2 + \tfrac{9}{4}\kappa^2 - 3\kappa\rho\cos\theta. \tag{13}$$

Hence by (12), $5\rho^2 - 6\kappa\rho\cos\theta = 4\rho IN$. So either $\rho = 0$ and I coincides with K, or $IN = \frac{1}{4}(5\rho - 6\kappa\cos\theta)$. Substituting this in (13) and rearranging gives the required polar equation (10). Solving (10) as a quadratic in ρ gives

$$\rho = \frac{2\kappa}{3}\cos\theta \pm 2\kappa\sqrt{\left(1 - \tfrac{8}{9}\cos^2\theta\right)}. \tag{14}$$

This meets the axis $\theta = 0$ where $\rho = 0$ and $\rho = 4\kappa/3$, so the possibility of I coinciding with K is still allowed for. Also the values of ρ are bounded and continuous in θ, and each line through K (except the axis itself) meets the curve in just two other points, as in FIGURE 2. Hence the curve is closed and the point $O(\kappa, 0)$ lies inside it, and the point $H(2\kappa, 0)$ outside it. But if I were to coincide with O the left-hand side of (9) would be negative, so the inequality would fail. By continuity it also fails for all other I inside the curve. Similarly if I is at H then (9) is satisfied, and remains so for all other I outside the curve (10).

By the lemmas we have:

THEOREM 1. *For non-degenerate triangles with a given Euler line, no tritangent center can lie inside the closed curve* (10) *of Lemma* 3, *and tritangent centers on that curve correspond only to degenerate triangles.*

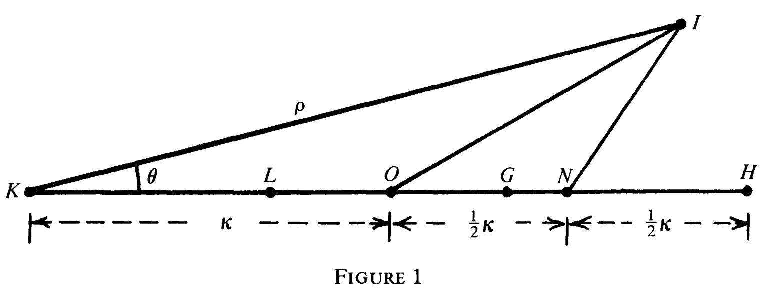


FIGURE 1

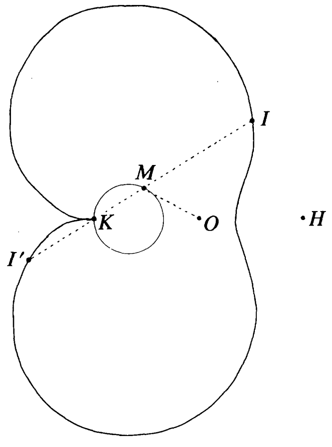


FIGURE 2

The polar equation (10) leads to a simple geometric characterization of the curve, thus:

THEOREM 2. *The locus of the midpoints M of chords through K of the curve of Lemma 3 is the circle on diameter KL, and the length of each chord equals $4OM$.*

Proof. Let I, I' be the endpoints of a chord through K at an angle θ to the axis. Then the radial coordinates of I, I' are given by (14), so their midpoint M satisfies $\rho = \frac{2}{3}\kappa\cos\theta$. This is the polar equation of the circle on diameter KL. The length of the chord II' is the difference of the values of ρ given by (14); that is

$$II' = 4\kappa\sqrt{\left(1 - \tfrac{8}{9}\cos^2\theta\right)}\,. \tag{15}$$

By the cosine rule

$$\begin{aligned} OM^2 &= KO^2 + KM^2 - 2KO \cdot KM\cos\theta \\ &= \kappa^2 + \left(\frac{2\kappa}{3}\cos\theta\right)^2 - 2\kappa\left(\frac{2\kappa}{3}\cos\theta\right)\cos\theta \\ &= \kappa^2\left(1 - \tfrac{8}{9}\cos^2\theta\right). \end{aligned} \tag{16}$$

By (15) and (16), $II' = 4OM$, as required.

Remarks. If points O, K and the circle of Theorem 2 are drawn, then points I, I' on the curve are easily constructed by ruler and compass, as in FIGURE 2.

It can also be shown that all points outside the curve can be tritangent centers of some sort of real triangle, but I know no proof appreciably simpler than that in reference [**3**].

References

[1] N. Altshiller-Court, College Geometry, Barnes & Noble, 1952, pp. 152–3

[2] A. P. Guinand, Triangles from centres out, James Cook Math. Notes, University of North Queensland, 30 (1982) 3127–32.

[3] ______, Euler lines, tritangent centers, and their triangles, Amer. Math. Monthly, 91 (1984) 290–300.

[4] E. W. Hobson, A Treatise on Plane Trigonometry, Cambridge, 1939, chapters IV and XII.

[5] D. Paine, Triangle Rhyme, this MAGAZINE, 56 (1983) cover and 235–238.

[6] W. Wernick, Triangle constructions with three located points, this MAGAZINE, 55 (1982) 227–230.

Factoring Finite Factor Rings

JUDY L. SMITH
JOSEPH A. GALLIAN
University of Minnesota, Duluth
Duluth, MN 55812

The most descriptive way to represent a finite Abelian group is as a direct product of cyclic groups. In this form, many important characteristics such as order, exponent, and rank are immediately evident. For instance, consider the group G of units of the ring of integers modulo 180. With this description of G, very little about its structure is readily apparent. On the other hand, if we are told that G is isomorphic to $Z_{12} \times Z_2 \times Z_2$ (see [**9**] or [**11**, p. 46]), then we may observe that G has order 48, exponent 12, and rank 3.

More generally, one could ask for the structure of the group of units of any commutative ring R. When R is finite, we know from the Fundamental Theorem of Finite Abelian Groups the beautiful fact that the group of units of R, $U(R)$, is isomorphic to a direct product of cyclic groups. So, it is natural to look for a method to express the group of units of any finite commutative ring as a direct product of cyclic groups. This problem has not been solved in general, but the solution to the special case in which R is a finite field is well known. In this case, $U(R)$ is cyclic [**8**, p. 405]. Other familiar finite rings are those of the form $R = F[x]/\langle h(x)\rangle$, where F is a finite field and $h(x)$ is a nonconstant polynomial in $F[x]$. For example, consider $R = Z_7[x]/\langle (x+4)^8(x^2+4)(x^3+3)^2\rangle$. What is $U(R)$ in this case? It turns out that a nice blend of elementary ring theory and group theory gives an algorithm which yields the answer:

$$U(R) \cong Z_6 \times Z_7 \times Z_7 \times Z_7 \times Z_7 \times Z_7 \times Z_7 \times Z_7 \times Z_7 \times Z_{49} \times Z_{48} \times Z_{342}.$$

In addition, this algorithm can easily be implemented on a machine, provided that $h(x)$ is given as a product of irreducible polynomials over F. In this note we derive this algorithm.

To this end, let F be a finite field with q elements and $h(x)$ an element of $F[x]$, the ring of polynomials over F. When $h(x)$ is irreducible over F and has degree n, then the factor ring $F[x]/\langle h(x)\rangle$ is a finite field of order q^n, so every nonzero element is a unit and the group of units is isomorphic to the cyclic group Z_{q^n-1}. But, what about the case in which $h(x)$ is not irreducible over F? A routine argument yields the following reduction which is a special case of the Chinese Remainder Theorem for rings.

LEMMA. *If $h(x) = h_1(x)^{m_1}h_2(x)^{m_2}\cdots h_k(x)^{m_k}$, where all $h_i(x)$ are distinct irreducibles in $F[x]$, then*

$$\frac{F[x]}{\langle h(x)\rangle} \cong \frac{F[x]}{\langle h_1(x)^{m_1}\rangle} \times \cdots \times \frac{F[x]}{\langle h_k(x)^{m_k}\rangle}.$$

Since the group of units of a direct sum of rings is the direct product of the groups of units of the factors, we have reduced our problem to that of finding the structure of $U(F[x]/\langle g(x)^m\rangle)$, where $g(x)$ is irreducible over F.

The next theorem further simplifies the problem.

THEOREM. *Let F be a finite field and $g(x)$ an irreducible polynomial in $F[x]$. If a is a root of $g(x)$ and $K = F(a)$, then $F[x]/\langle g(x)^m\rangle \cong K[x]/\langle x^m\rangle$.*

To prove this theorem, let $g(x)$ be irreducible over F, let a be a root of $g(x)$ in some extension of F, and let $K = F(a)$, the smallest extension of F containing a. We then claim that the natural ring homomorphism ϕ from $F[x]$ into $K[x]/\langle (x-a)^m\rangle$ which takes $f(x)$ to $f(x) + \langle (x-a)^m\rangle$ has kernel $\langle g(x)^m\rangle$. To see this, suppose $t_1(x) \in \operatorname{Ker}\phi$. Then $(x-a)^m$ divides $t_1(x)$ in $K[x]$, and since $g(x)$ is the irreducible polynomial of a over F, it follows that $g(x)$ divides $t_1(x)$. Now, because the zeros of an irreducible polynomial over a finite field have multiplicity 1 [**8**, pp. 391–393], we know that $(x-a)^{m-1}$ divides $t_2(x)=t_1(x)/g(x)$. Then, as before, $g(x)$ divides $t_2(x)$ and, therefore, $g(x)^2$ divides $t_1(x)$. Continuing in this fashion, we see that $g(x)^m$ divides $t_1(x)$. So, $\operatorname{Ker}\phi \subseteq \langle g(x)^m\rangle$. The other inclusion is obvious. This proves that $F[x]/\langle g(x)^m\rangle$ is isomorphic to a subring of $K[x]/\langle (x-a)^m\rangle$. However, a simple calculation shows that these two rings have the same number of elements, so they are isomorphic.

Finally, we note that the mapping from $K[x]/\langle (x-a)^m\rangle$ onto $K[x]/\langle x^m\rangle$ given by $f(x)+\langle (x-a)^m\rangle \to f(x+a)+\langle x^m\rangle$ is a ring isomorphism.

All that remains to be done is to find an explicit factorization of the group $U(F[x]/\langle x^m\rangle)$ as a direct product of cyclic groups. We begin by observing that

$$\begin{aligned} G &= U(F[x]/\langle x^m\rangle) \\ &= \left\{ a_{m-1}x^{m-1}+\cdots+a_1x+a_0+\langle x^m\rangle \mid a_i\in F, \quad a_0\neq 0\right\}. \end{aligned}$$

So, if $|F|=p^n$, there are $p^{n(m-1)}(p^n-1)$ elements in G. Next, we define a sequence $\{G_i\}$ of subgroups of G as follows: Let

$$\begin{aligned} G_0 &= \left\{ a_{m-1}x^{m-1}+\cdots+a_1x+1+\langle x^m\rangle \mid a_i\in F\right\}, \\ G_i &= G_{i-1}{}^p = \left\{ g^p \mid g\in G_{i-1}\right\}, \qquad i\geqslant 1. \end{aligned}$$

Since $|G_0|=p^{n(m-1)}$, we see that G_0 is the Sylow p-subgroup of G. Now, because F is a field, we know that $\{a+\langle x^m\rangle \mid a\in F, a\neq 0\}$ is isomorphic to Z_{p^n-1}. So, it follows that $G\cong Z_{p^n-1}\times G_0$ and $G_0\cong Z_{p^{n_1}}\times Z_{p^{n_2}}\times\cdots\times Z_{p^{n_k}}$, where the n_i's are positive integers whose sum is $n(m-1)$. Our next step is to determine the n_i's exactly by examining the G_i's more closely. To facilitate the discussion, we let $rk(G_i)$, the rank of G_i, denote the minimum number of factors necessary to express G_i as a direct product of cyclic groups. (Since G_i is a p-group, we could also define $rk(G_i)$ as the dimension of G_i/G_{i+l} as a vector space over Z_p.) Observe that

$$G_1 = G_0^p \cong Z_{p^{n_1-1}}\times Z_{p^{n_2-1}}\times\cdots\times Z_{p^{n_k-1}}$$

so that in passing from G_0 to G_1, every Z_p summand in G_0 is lost. Thus, the multiplicity of Z_p in the cyclic factorization of G_0 is $rk(G_0)-rk(G_1)$. Similarly, $rk(G_1)-rk(G_2)$ is the multiplicity of Z_p in the cyclic factorization of G_1, which is, in turn, the multiplicity of Z_{p^2} in the cyclic factorization of G_0. Continuing in this manner, we see that the multiplicity of Z_{p^i} in G_0 is $rk(G_{i-1})-rk(G_i)$. For convenience, let $m_i=rk(G_{i-1})-rk(G_i)$. Then, letting s be the first integer where $G_s=\{1\}$, we have

$$G_0 \cong \mathop{\times}_{i=1}^{s} m_i * Z_{p^i}, \tag{1}$$

where $m_i * Z_{p^i}$ means Z_{p^i} occurs m_i times.

The only thing left to do is to calculate the m_i's. To this end, notice that

$$\begin{aligned} G_i &= \left\{\left(a_{m-1}x^{m-1}+\cdots+a_1x+1\right)^{p^i}+\langle x^m\rangle \mid a_i\in F\right\} \\ &= \left\{ a_{m-1}{}^{p^i}x^{(m-1)p^i}+\cdots+a_1{}^{p^i}x^{p^i}+1+\langle x^m\rangle \mid a_i\in F\right\}. \end{aligned}$$

But, in $F[x]/\langle x^m\rangle$, $x^r\equiv 0$, when $r\geqslant m$ and $\{a^{p^i}\mid a\in F\}=F$. So,

$$G_i = \left\{ b_{k_i}x^{k_ip^i}+b_{k_i-1}x^{(k_i-1)p^i}+\cdots+b_1x^{p^i}+1+\langle x^m\rangle \mid b_j\in F\right\},$$

where $k_i=\max\{h\in Z \mid hp^i<m\}$. (Notice that the first time $G_i=\{1\}$ is the first time $p^i\geqslant m$. So, s is $\min\{h\in Z \mid p^h\geqslant m\}$ and $s=0$ if and only if $m=1$.) The order of G_i is now readily seen to be p^{nk_i}. Thus, it follows that $rk(G_{i-1})=nk_{i-1}-nk_i$, and from this, we get $m_i=n(k_{i-1}-2k_i+k_{i+1})$. But, $U(F[x]/\langle x^m\rangle)\cong Z_{p^n-1}\times G_0$ and from (1) we obtain the finished product.

THEOREM. *Let F be a finite field with p^n elements, where p is a prime. Then, for any positive integer m, we have*

$$U(F[x]/\langle x^m\rangle) \cong Z_{p^n-1} \times \bigtimes_{i=1}^{s} n(k_{i-1} - 2k_i + k_{i+1}) * Z_{p^i},$$

where $s = \min\{h \in Z | p^h \geqslant m\}$, $k_i = \max\{h \in Z | hp^i < m\}$, *and* $t * Z_{p^i}$ *means* Z_{p^i} *occurs in the product t times.*

Let us return to the example mentioned before. Consider the ring

$$R = Z_7[x]/\langle (x+4)^8(x^2+4)(x^3+3)^2\rangle.$$

Since each of $x+4$, x^2+4, and x^3+3 are irreducible over Z_7, our reduction and replacement results tell us that

$$U(R) \cong U(Z_7[x]/\langle x^8\rangle) \times U(K_0[x]/\langle x\rangle) \times U(K_1[x]/\langle x^2\rangle),$$

where $|K_0| = 7^2$ and $|K_1| = 7^3$. For $U(Z_7[x]/\langle x^8\rangle)$, we have $k_i = \max\{h \in Z | h \cdot 7^i < 8\}$, which yields $k_0 = 7$, $k_1 = 1$, and $k_i = 0$ for $i \geqslant 2$. So,

$$U(Z_7[x]/\langle x^8\rangle) \cong Z_6 \times 1 \cdot (7 - 2 + 0) * Z_7 \times 1(1 - 0 + 0) * Z_{49}.$$

For $U(K_0[x]/\langle x\rangle)$, we have $k_i = \max\{h \in Z | h \cdot 7^i < 1\}$, which gives us $k_0 = 0$. So,

$$U(K_0[x]/\langle x\rangle) \cong Z_{7^2-1}.$$

Finally, for $U(K_1[x]/\langle x^2\rangle)$, we have $k_i = \max\{h \in Z | h \cdot 7^i < 2\}$, so $k_0 = 1$ and $k_1 = 0$. Thus,

$$U(K_1[x]/\langle x^2\rangle) \cong Z_{7^3-1} \times 3(1 - 0 + 0) * Z_7.$$

Putting it all together, we see the group we seek is isomorphic to $Z_6 \times 8 * Z_7 \times Z_{49} \times Z_{48} \times Z_{342}$.

In addition to our result, numerous other special cases of the general problem of determining the group of units of a commutative ring have been solved. Gilmer [**10**] discusses the problem of decomposing a finite commutative ring into a direct sum of primary rings and the associated decomposition of the group of units. Eldridge and Fischer [7] extend Gilmer's results to Artinian rings. Eggert [**5**] partially characterizes the structure of the group of units of a finite commutative ring. Ayoub [**2**] determines the structure of the group of units of a primary Noetherian ring which is homogeneous. Ditor [**4**] characterizes the groups of odd order which can appear as groups of units. Rings which have cyclic groups of units have been determined by Gilmer [**10**], Ayoub [**1**], and Pearson and Schneider [**12**]. Cross [**3**] determines the structure of the group of units of factor rings of the Gaussian integers.

The authors wish to thank David Witte for his suggestions in writing this paper and the referees for their comments.

References

[1] C. Ayoub, On finite primary rings and their groups of units, Composito Math., 21 (1969) 247–252.
[2] ______, On the group of units of certain rings, J. Number Theory, 4 (1972) 383–403.
[3] J. T. Cross, The Euler ϕ-function in the Gaussian integers, Amer. Math. Monthly, 90 (1983) 518–528.
[4] S. Z. Ditor, On the group of units of a ring, Amer. Math. Monthly, 78 (1971) 522–524.
[5] N. H. Eggert, Jr., Some results on finite rings whose groups of units are abelian, Ph.D. dissertation, University of Colorado, 1966.
[6] K. E. Eldridge, On ring structures determined by groups, Proc. Amer. Math. Soc., 23 (1969) 472–477.
[7] K. E. Eldridge and I. Fischer, D. C. C. rings with a cyclic group of units, Duke Math. J., 34 (1967) 243–248.
[8] J. B. Fraleigh, A First Course in Abstract Algebra, 3rd edition, Addison-Wesley, Reading, Mass., 1982.
[9] J. A. Gallian and D. Rusin, Factoring groups of integers modulo n, this MAGAZINE, 53 (1980) 33–36.
[10] R. W. Gilmer, Finite rings having a cyclic multiplicative group of units, Amer. J. Math., 85 (1963) 447–452.
[11] K. Ireland and M. Rosen, Elements of Number Theory: Including an Introduction to Equations over Finite Fields, Bogden and Quigley, Tarrytown, N.Y., 1972.
[12] K. Pearson and J. Schneider, Rings with a cyclic group of units, J. Algebra, 16 (1970) 243–251.

On Fermat's Last Theorem

Michael H. Brill
Science Applications International Corporation
1710 Goodridge Drive
McLean, VA 22102

A previously unresolved conjecture in mathematics is Fermat's Last Theorem, which states that $x^n + y^n = z^n$ has positive integer solutions x, y, z, and n only for $n \leqslant 2$. No counterexample has been found in searches through large values of n, but there is at least one counterexample at $n = \infty$, in the neighborhood of which insufficient investigation has taken place.

Since, for positive x and y,

$$\lim_{n \to \infty} (x^n + y^n)^{1/n} = \max(x, y),$$

we may take x and y as positive integers, $z = \max(x, y)$ and $n = \infty$ for a solution.

There are purists who would deny $n = \infty$ the right to claim the title of counterexample to Fermat's Last Theorem. Notwithstanding the generous inclusion of the point at infinity in such formalisms as projective geometry, the purists might even doubt that $n = \infty$ is decidably an integer. In answer to this objection, we advance the plea that searchers for counterexamples to the Theorem ought to look in the neighborhood of $n = \infty$ for a more canonical counterexample, instead of starting with $n = 3$, where the conjecture seems safe enough. We are presently recruiting Pearys (*not* Amundsens) of the Riemann sphere to perform the exploration.

Sampling Bias and the Inspection Paradox

William E. Stein
Ronald Dattero
College of Business Administration
Texas A&M University
College Station, TX 77843

Elementary statistics textbooks rarely spend an adequate amount of time on the important topic of sampling. When faced with the problem of estimating the average size of a family, students frequently suggest polling their classmates or standing on a corner and asking people who walk past. In using these methods, bias may be introduced since we are sampling people and not families. Or ask the students how to measure the average speed of all the cars on a freeway. Measuring the speed of cars passing a fixed point will give biased results. It is interesting to note that both these examples illustrate the same kind of sampling bias!

These examples of sampling bias are a special case of the well-known **inspection paradox** in probability. The purpose of this note is to clarify the results in [4] by placing them in a more general context. This also enables us to spot situations where we might expect such bias but none is actually present.

As a model, consider a hypothetical process: a machine repeatedly places a random number of items, N, in boxes. We assume that the numbers of items in successive boxes are independent and identically distributed random variables with known probability mass function: $f(k)$, $k \geqslant 1$. Let μ and σ^2 be the mean and variance, respectively, of N:

On Fermat's Last Theorem

MICHAEL H. BRILL
Science Applications International Corporation
1710 Goodridge Drive
McLean, VA 22102

A previously unresolved conjecture in mathematics is Fermat's Last Theorem, which states that $x^n + y^n = z^n$ has positive integer solutions x, y, z, and n only for $n \leqslant 2$. No counterexample has been found in searches through large values of n, but there is at least one counterexample at $n = \infty$, in the neighborhood of which insufficient investigation has taken place.

Since, for positive x and y,

$$\lim_{n \to \infty} (x^n + y^n)^{1/n} = \max(x, y),$$

we may take x and y as positive integers, $z = \max(x, y)$ and $n = \infty$ for a solution.

There are purists who would deny $n = \infty$ the right to claim the title of counterexample to Fermat's Last Theorem. Notwithstanding the generous inclusion of the point at infinity in such formalisms as projective geometry, the purists might even doubt that $n = \infty$ is decidably an integer. In answer to this objection, we advance the plea that searchers for counterexamples to the Theorem ought to look in the neighborhood of $n = \infty$ for a more canonical counterexample, instead of starting with $n = 3$, where the conjecture seems safe enough. We are presently recruiting Pearys (*not* Amundsens) of the Riemann sphere to perform the exploration.

Sampling Bias and the Inspection Paradox

WILLIAM E. STEIN
RONALD DATTERO
College of Business Administration
Texas A&M University
College Station, TX 77843

Elementary statistics textbooks rarely spend an adequate amount of time on the important topic of sampling. When faced with the problem of estimating the average size of a family, students frequently suggest polling their classmates or standing on a corner and asking people who walk past. In using these methods, bias may be introduced since we are sampling people and not families. Or ask the students how to measure the average speed of all the cars on a freeway. Measuring the speed of cars passing a fixed point will give biased results. It is interesting to note that both these examples illustrate the same kind of sampling bias!

These examples of sampling bias are a special case of the well-known **inspection paradox** in probability. The purpose of this note is to clarify the results in [4] by placing them in a more general context. This also enables us to spot situations where we might expect such bias but none is actually present.

As a model, consider a hypothetical process: a machine repeatedly places a random number of items, N, in boxes. We assume that the numbers of items in successive boxes are independent and identically distributed random variables with known probability mass function: $f(k)$, $k \geqslant 1$. Let μ and σ^2 be the mean and variance, respectively, of N:

$$\mu = E(N) = \sum_{k=1}^{\infty} kf(k) \quad \text{and} \quad \sigma^2 = \sum_{k=1}^{\infty} k^2 f(x) - \mu^2. \tag{1}$$

The above model includes the examples in [**4**] as special cases. The model can represent passengers in cars, children in families, or cookies in a box.

There are two ways to estimate the mean number of items per box.

METHOD I: Select a *box* at random. The formulas (1) give the mean and variance of the quantity, N, contained within.

METHOD II: Select an *item* at random. Let S be the quantity in the box containing this item (the count includes the selected item).

In method II, selecting an item is an indirect way of selecting a box; given the numbers of items in the boxes, boxes are being selected with probability proportional to the (random) numbers of items they contain. (If the boxes are the 50 states and the items are residents, then Texas is much more likely to be selected than Idaho.)

It is fairly easy to demonstrate that when using method II sampling, the mean of S, $E(S)$, need not equal μ. For example, suppose that only two boxes are involved and that

$$f(1) = \frac{1}{2} \quad \text{and} \quad f(2) = \frac{1}{2}.$$

Let B_i be the number of items in box i. Then

$$\begin{aligned}
P[S=1] = {} & P[\text{box with one item selected}|B_1 = 1 \text{ and } B_2 = 1]\, P[B_1 = 1 \text{ and } B_2 = 1] \\
& + P[\text{box with one item selected}|B_1 = 1 \text{ and } B_2 = 2]\, P[B_1 = 1 \text{ and } B_2 = 2] \\
& + P[\text{box with one item selected}|B_1 = 2 \text{ and } B_2 = 1]\, P[B_1 = 2 \text{ and } B_2 = 1] \\
& + P[\text{box with one item selected}|B_1 = 2 \text{ and } B_2 = 2]\, P[B_1 = 2 \text{ and } B_2 = 2] \\
= {} & 1 \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = \frac{5}{12}.
\end{aligned}$$

So

$$E(S) = \frac{5}{12} + 2 \cdot \frac{7}{12} = \frac{19}{12} > \frac{3}{2} = \mu.$$

The difference $E(S) - \mu = 1/12$ can be considered to be the (positive) bias in the estimation of μ due to the use of S with method II sampling.

If we consider an idealized situation where the number of boxes is infinite it is possible to give simple expressions for the probability mass function of S and $E(S)$ in terms of $f(k)$, μ and σ^2. These are

$$h(k) = P(S = k) = \frac{kf(k)}{\mu}, \qquad k \geqslant 1 \tag{2}$$

and

$$E(S) = \sum_{k=1}^{\infty} kh(k) = \sum_{k=1}^{\infty} \frac{k^2 f(k)}{\mu} = \frac{\sigma^2 + \mu^2}{\mu} = \mu + \frac{\sigma^2}{\mu}. \tag{3}$$

Notice that (2) comes about by reasoning that "amongst an infinite number of boxes a fraction $f(k)$ will contain k items, so that a fraction $kf(k)/\Sigma_{j=1}^{\infty} jf(j)$ of the infinite number of items reside in boxes of content k." The positive quantity $\sigma^2/\mu = E(S) - \mu$ is the bias involved in estimating μ in this idealized "infinite number of boxes" situation.

Since

$$E\left(\frac{1}{S}\right) = \sum_{k=1}^{\infty} \frac{1}{k} h(k) = \sum_{k=1}^{\infty} \frac{f(k)}{\mu} = \frac{1}{\mu},$$

we have

$$\frac{1}{E\left(\frac{1}{S}\right)} = \mu. \tag{4}$$

The left-hand side of (4) is the **harmonic mean** of S. This suggests that if we have a sample of observations $s_1, \dots, s_n$ selected via method II, then in order to estimate μ we might well compute the harmonic mean HM of the data

$$HM = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{s_i}\right]^{-1}, \tag{5}$$

rather than the arithmetic mean.

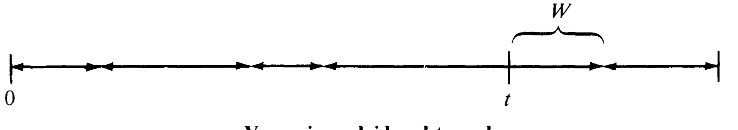
The case where the sampling bias equals 1 is of some interest. This can be given the following interpretation. People arrive at a station to take a bus to another city. A bus leaves every hour (even if empty). From the bus company's view, the random variable N describes the passenger load. For a typical rider, $S-1$ is the number of passengers on the bus, excluding the rider. If N and $S-1$ have the same expected value, then the sampling bias is 1, which occurs if N has a Poisson distribution. (We now are generalizing the original box model to permit empty boxes.) It is interesting to note that N and $S-1$ have the same probability distribution f if and only if f is Poisson. (The easy proof follows by induction.)

The continuous variable version of the above discussion is well known. Consider lengths of yarn having density $f(x)$. If one were to reach into a pile of such yarn cuttings and randomly select one, what is the probability distribution of its length? A reasonable model of the selection process is that a 6″ piece of yarn is three times as likely to be selected as one 2″ long. So the selected length has density

$$h(x) = \frac{xf(x)}{\mu}, \qquad x > 0, \tag{6}$$

which is to be expected in light of (2). Formulas (3) and (4) also hold here. This is an example of **length-biased sampling** ([**2**], p. 61).

There is an equivalent way in which to describe the yarn-selecting process. Imagine laying the pieces of yarn end-to-end in arbitrary order. Suppose we select a distance t from the start of this aggregation. If t is "large" then a "random" piece of yarn will be selected. More precisely, the above is assumed to be modeled by a **stationary renewal process**: the successive yarn lengths are independent and identically distributed.



Yarn pieces laid end-to-end.

Equation (3) then may be obtained from standard results in renewal theory. The length of the portion of the chosen piece of yarn that exceeds the selection point t is a random variable W. The density g of W is given by the equation ([**1**], p. 304):

$$g(W) = \frac{1-F(W)}{\mu}, \tag{7}$$

where $F(W) = \int_0^W f(x)\,dx$. Integration by parts yields

$$E(W) = \frac{\sigma^2 + \mu^2}{2\mu}. \tag{8}$$

Due to symmetry, the length of the selected yarn cutting is twice $E(W)$. This gives another explanation of equation (3). As a special case, if $f(x) = \lambda e^{-\lambda x}$ then the renewal process becomes a

Poisson process. By direct computation, $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$. Therefore, using (7), the mean length of the selected piece of yarn is $2/\lambda$, twice the actual mean length. In this context, the difference between sampling methods I and II is called the **inspection paradox** ([**6**], p. 202).

Here is another example of sampling bias of a similar nature. Suppose cars travel (one way) along a freeway, each with a constant speed. Assume these speeds follow a density f and are independent of one another. The density f represents the speed distribution if we could simultaneously measure all the speeds. On the other hand, if one measures (with a radar gun) the speed of cars passing a fixed point, the observed distribution is given by equation (6) ([**3**], p. 116). The reason is that the faster cars are sampled more frequently than the slower cars. (In a fixed interval of time, the distance traveled is proportional to speed. These distances play the role of the yarn cuttings in the previous example.) However, at typical freeway speeds the bias σ^2/μ is only about 1 mph. Even if we used a more refined model which permitted cars to change their speed, we would still obtain a bias in favor of the faster cars.

We now discuss a possible misconception concerning sampling bias. Suppose students in a college class are asked "How many brothers and sisters do you have?" Since this seems to be a method II sampling procedure to estimate the number of children per family, will we get biased results? Maybe! The bias of method II is due to overrepresentation of the larger families. However, it is likely that large families send a smaller percentage of their children to college than do small families. Such sociological considerations make this a difficult question to answer.

Our last example will suggest a further generalization of the box model. Suppose a school library wishes to measure information (time spent in library, favorite periodicals, etc.) about its visitors. By surveying arriving students at random, frequent visitors are more likely to be included in the sample. Instead, the library could select a sample of students from the student directory. To see the connection with the box model, imagine a guest register of all visitors to the library. Frequent visitors will have their names in the register many times. Group all such duplicate listings for a given person into a "box." Selecting a box corresponds to selecting a person without regard to frequency of visit (method I). Selecting a name from the guest register will result in a larger representation of frequent visitors (method II). If frequent visitors have library usage patterns different from other students, these two sampling methods will yield different results.

The difference between the library example and the previous ones is that the size of the box is no longer the variable of interest. Instead, we are interested in another variable which is probabilistically related to the box size. A common example of this is found in marketing surveys carried out in shopping centers. When surveyed in a shopping center, the frequent shoppers will be overrepresented. Shoppers who spend a long time in the center also will be overrepresented. ("... the probability of selection of a respondent at a random point within the shopping center is proportional to the length of time that the respondent has been in the center" [**7**], p. 428.) If, for example, membership in either of these two groups of shoppers is related to income level, a bias in the survey may result. Related sampling models can be found in Johnson and Kotz ([**5**], p. 284).

The examples in this article have shown that size-biased and length-biased sampling occur in many seemingly unrelated situations. An estimation bias may arise if method II sampling is used, since the data will not follow the same distribution as that generated by method I sampling. This problem may be attacked by matching an appropriate estimator to the chosen sampling technique (e.g., using the harmonic mean with method II sampling). The usefulness of survey data depends upon a careful consideration of the probabilistic implications of the sampling technique used.

References

[1] E. Cinlar, Introduction to Stochastic Processes, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[2] D. R. Cox and P. A. W. Lewis, The Statistical Analysis of Series of Events, Methuen, London, 1966.

[3] F. A. Haight, Mathematical Theories of Traffic Flow, Academic Press, New York, 1963.

[4] D. Hemenway, Why your classes are larger than 'average', this MAGAZINE, 55 (1982) 162–164.

[5] N. L. Johnson and S. Kotz, Urn Models and Their Application, John Wiley and Sons, New York, 1977.

[6] S. Ross, Introduction to Probability Models, Academic Press, New York, 1972.

[7] S. Sudman, Improving the quality of shopping center sampling, J. of Marketing Research, 17 (1980) 423–431.

Child's Play

WARNER CLEMENTS
Beverly Hills, CA 90210

Consider the problem of finding the total number of ways each possible outcome (spot total) can be rolled using three standard dice. For the modern mathematician this problem is a fairly easy one. Nevertheless, it is scarcely child's play; the number of ways to get certain outcomes runs as high as twenty-seven. Dice are the oldest gaming instruments known to man, yet no solution to the problem was generally known until the time of Galileo. It took that great man himself to demonstrate one.

So it is perhaps bold of me to assert, as I now do, that any reasonably intelligent child, working on his (or her) own, could find all the answers to the problem in ten or fifteen minutes! He could, that is, if briefly instructed in advance as to how to proceed. As you might suspect, he would have to employ a computational tool. But it wouldn't be what you'd think. Not a computer. Not even a calculator. Rather, an apparatus more primitive than an abacus: a supply of children's toy blocks! For the sake of a better explanation, it will be assumed in what follows that the blocks are available in a variety of colors. In addition, a cardboard box would be helpful (plus, of course, a pencil and paper to record results).

The child is to stack the blocks in the simple and repetitive manner to be described, while counting the number of blocks required to complete each separate stage of construction. The numbers thus obtained represent the required solutions. That's all there is to it.

Specifically, the child, having received instruction, starts by placing a single red block in an interior corner of the box. This constitutes the first stage of construction. The count of one block indicates the number of ways a "3" can be rolled with three dice and is so recorded. There will be three faces of this first block left visible. Next, using blocks of a second color, say, blue, the child places just enough blocks atop and beside the red one to cover these faces and to leave only the blue of the second course visible. The number of blocks (three) required for this second stage indicates the number of ways a "4" can be rolled. And, again, more blocks are laid, just sufficient to cover the blue, say, with a yellow this time. Now the number of blocks required indicates the number of ways a "5" can be rolled.

And so on, color by color. It will be noted that like colors accumulate in diagonally sloping layers. Part of the instruction given will stipulate that the number of blocks in any coordinate direction is not to exceed six, i.e., no block is to be placed outside the dimensions of a $6 \times 6 \times 6$ cube (which limits can be marked on the inside of the box). However, it is by no means necessary to set up the whole cube; half is sufficient. It will be obvious that beyond the midpoint of construction the block count will be only a duplication (in reverse order) of the figures already tabulated.

When the child is through, he will have set up an interesting and ornamental structure, and will probably have had fun in the process. He will also have produced an accurate and legitimate solution to the original problem. But he won't be entitled to think he's smarter than Galileo! It's the blocks that are "smart"; they deserve the credit.

To understand how they do the job, consider first the regularity of the structure as it grows. Just as the structure of a crystal or snowflake is said to be determined by a few molecules of "seed," so here it is the initial block which establishes the final structure. Already by the time blocks of the second or third color have been added it is established that the blocks of each color will lie with their centers in a plane exclusive to that color; that all such planes will have identical slope and spacing from each other. Each of these **color planes** is normal to those interior diagonals of the blocks which extend in the general direction of added construction. The spacing between planes will be one-third of a block's diagonal (not the one-half one might intuitively expect).

The most useful coordinates for the study of the resultant structure will be the usual rectangular ones, with axes chosen parallel to the edges of the blocks, and the origin chosen so that the coordinates of the center of the first (red) block are $(1,1,1)$. This choice produces integer coordinates for the centers of all the blocks. (Each block edge is measured as 1 unit.)

To find the place of any block in the stack, we may give it an address. For instance, a given block might be the third to the right from the starting corner, the second back and the fourth up. Such an address can be abbreviated as $(3,2,4)$. But, thanks to the position chosen for the origin, the integers representing the address of any block are also the xyz coordinates which locate its center.

A fact which may be initially surprising and which underlies the counting process is: *the xyz coordinates of the centers of blocks of the same color have the same sum*. How do we know the statement is true? From analytic geometry, the general equation for a plane in rectangular coordinates is

$$Ax + By + Cz + D = 0. \tag{1}$$

But our earlier observation about the slope of the color planes says that each color plane cuts each of the coordinate axes at an equal distance from the origin. This implies that $A = B = C = 1$, so the equation of each color plane is of the form

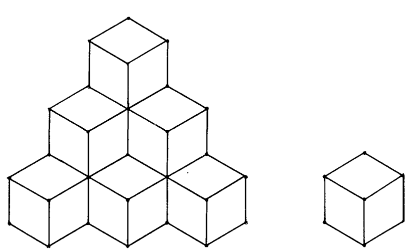
$$x + y + z = k. \tag{2}$$

It follows, then, that the addresses for all blocks of a given color must add up to a constant peculiar to its layer. Starting with the initial red block, k is 3. Since a block in the second layer adds one unit in just one of the three coordinate directions, the sum k of its center coordinates is 4. Similarly, k for the third layer is 5, and so on.

To apply these facts to our problem of rolling three dice, we need only assign one address direction (right, back, or up) to each of the dice, and regard the integer distance of a block from the origin in this direction as the outcome of a roll of that die. Then, for each block, the total of its address coordinates will represent the outcome of a roll of the three dice. Of course, most outcomes can be reached in a plurality of ways. Our system neatly counts permutations, not combinations. It places in the stack one and only one block for each possible way a given outcome can be rolled with the three dice. Finally, the system gathers the blocks representing the ways to make a given outcome together in a layer for convenient counting. Rather a slick performance, particularly when compared with the laborious methods even a mathematical pro would have to use if proceeding with pencil and paper!

The toy-block approach is designed for the three-die problem. But we need not limit its application to dice which have the usual six faces. The number of faces on the dice enters into our solution only by limiting the expansion of the stack. This limitation is readily adjustable in three dimensions, each corresponding to the number of faces on a respective die. So even the problem of three dice with different numbers of faces—say, one with four, one with five, and one with six—could be handled with only a simple adjustment of construction boundaries.

One can't claim any practical applications for the block-counting procedure; not in this day of the computer. Nevertheless it does suggest exploration of the implications of assigning various numbers of faces to the respective three dice. Someone might even find a way to enjoy our "child's play" with tesseracts or hypercubes, thus extending the fun into one more dimension.



Hand Computation of Generalized Inverses

D. R. BARR
Air Force Institute of Technology
Wright-Patterson Air Force Base, OH 45433

A growing interest in the theory of generalized inverses ([2], [13]) and in applications of generalized inverses in such fields as engineering ([4], [5], [9], [17]) and statistics ([1], [6], [7], [12], [15]) has stimulated an interest in teaching about generalized inverses in linear algebra courses ([3], [11], [16]).

When A is a nonsingular matrix, the equation $Ax = b$ has a unique solution, namely, $x = A^{-1}b$. The generalized inverse A^+ of a singular matrix A can be motivated as a generalization of this situation. In fact, $x = A^+b$ will be the unique solution to $Ax = b$ when a unique solution exists, will be a solution when there is more than one solution, and will be a least squares solution (*i.e.*, will minimize $\|Ax - b\|^2$) when no solution exists. Moreover, when either the solution is not unique, or the least squares solution is not unique, the generalized inverse will be the shortest solution, or the shortest least squares solution (*i.e.*, will minimize $\|x\|^2$, subject to the restriction that x is a solution, or that x is a least squares solution, as appropriate) [11], [16]. A geometric interpretation of the generalized inverse also helps to motivate the theory; $x = A^+b$ lies in the row space of A, while $Ax = AA^+b$ is the projection of b onto the column space of A [16].

Some instructors prefer to relegate all computations of inverses to computers. However, many instructors feel that students gain an understanding from computing an inverse by hand that cannot be gained from observation of computer output, and teach an elementary algorithm for computing inverses of matrices. This note is of interest only to the latter set of instructors. It contains an algorithm for the calculation of generalized inverses, suitable for use by students with hand-held calculators, which uses the information available when the usual (Gauss-Jordan) classroom algorithm for computing inverses has established that the inverse fails to exist. There are two equivalent procedures, which enable a student to calculate the generalized inverse using one procedure, and to check the work using the other procedure.

This algorithm is meant for classroom use only; there are more accurate, computationally stable, and efficient algorithms available in various computer packages (for example, the GINV function in the MATRIX procedure of SAS [14]) for use in real world problems.

A classroom discussion of the use of the results of a singular value decomposition in finding a generalized inverse is enlightening ([10], [11], [16]), but finding a singular value decomposition using only a hand-held calculator restricts attention to trivial examples when this approach is used.

An attempt to invert a matrix A using the usual classroom algorithm will either yield an inverse (when A is nonsingular), or else will at least provide us with the row-echelon form of A. The columns of A corresponding to columns of the identity matrix appearing in the row-echelon form are a basis for the column space of A; we form a matrix B from these columns. As we shall show later, the generalized inverse A^+ of A can then be expressed using B in the expression

$$A^+ = E(E^TE)^{-1}B^T \tag{1}$$

where $E = A^TB$.

EXAMPLE 1. Given the matrix

$$A = \begin{bmatrix} 0 & 1 & 3 & 4 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 1 & 2 \\ -1 & 0 & 4 & 5 \end{bmatrix}$$

find the inverse if it is nonsingular, or find the generalized inverse if it is singular.

We start by trying to find the inverse; we row-reduce $[A \quad I]$ to

$$\begin{bmatrix} 1 & 0 & -4 & -5 & -2 & 1 & 0 & 0 \\ 0 & 1 & 3 & 4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & -2 & 1 & 0 & 1 \end{bmatrix} \tag{2}$$

and see that A is singular. We can extract from (2) all that we need in order to be able to compute the generalized inverse of A.

Since the first 4 columns of the matrix in (2) are the row-echelon form of A, we see that the first two columns of A are a basis for the column space of A. Thus, we obtain

$$B = \begin{bmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 3 \\ -1 & 0 \end{bmatrix}.$$

Using equation (1), we compute

$$E = A^T B = \begin{bmatrix} 6 & 8 \\ 8 & 14 \\ 0 & 10 \\ 2 & 16 \end{bmatrix}$$

from which we obtain

$$E^T E = \begin{bmatrix} 104 & 192 \\ 192 & 616 \end{bmatrix}.$$

The inverse of this 2×2 matrix is

$$(E^T E)^{-1} = 3400^{-1} \begin{bmatrix} 77 & -24 \\ -24 & 13 \end{bmatrix}. \tag{3}$$

Substitution of (3) into (1) yields

$$A^+ = \begin{bmatrix} 6 & 8 \\ 8 & 14 \\ 0 & 10 \\ 2 & 16 \end{bmatrix} 3400^{-1} \begin{bmatrix} 77 & -24 \\ -24 & 13 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 & -1 \\ 1 & 2 & 3 & 3 \end{bmatrix}$$

$$= 340^{-1} \begin{bmatrix} -4 & 19 & 42 & -27 \\ -1 & 26 & 53 & -28 \\ 13 & 2 & -9 & 24 \\ 16 & 9 & 2 & 23 \end{bmatrix}.$$

An expression of an $m \times n$ matrix A of rank k as the product $A = BC$ of two matrices B and C, where B is $m \times k$ of rank k and C is $k \times n$ of rank k, is called a **full rank factorization of A**, since each of the factors, B and C, is of full rank (i.e., its rank is equal to the smaller of its dimensions). It is well-known [**11**, p. 339; **16**, p. 137] that if $A = BC$ is a full rank factorization of A, then $B^T B$ and CC^T are nonsingular, and the generalized inverse A^+ of A can be expressed as

$$A^+ = C^T (CC^T)^{-1} (B^T B)^{-1} B^T. \tag{4}$$

Using the nonsingularity of $B^T B$, we can solve $A = BC$ for C: we have

$$B^T A = B^T BC$$

from which it follows that

$$(B^T B)^{-1} B^T A = C.$$

This enables us to eliminate C from equation (4), and leads to the following:

$$
\begin{aligned}
A^+ &= \left[(B^TB)^{-1}B^TA\right]^T\left[\{(B^TB)^{-1}B^TA\}\{(B^TB)^{-1}B^TA\}^T\right]^{-1}(B^TB)^{-1}B^T \\
&= A^TB(B^TB)^{-1}\left[(B^TB)^{-1}B^TAA^TB(B^TB)^{-1}\right]^{-1}(B^TB)^{-1}B^T \\
&= A^TB\left[B^TB(B^TB)^{-1}B^TAA^TB(B^TB)^{-1}B^TB\right]^{-1}B^T \\
&= A^TB[B^TAA^TB]^{-1}B^T \\
&= (A^TB)\left[(A^TB)^TA^TB\right]^{-1}B^T,
\end{aligned}
$$

which, for $E = A^TB$, equals $E(E^TE)^{-1}B^T$. This is equation (1). A similar argument yields

$$
\begin{aligned}
A^+ &= C^T\left[CA^T(CA^T)^T\right]^{-1}(CA^T) \\
&= C^T(FF^T)^{-1}F \qquad (5)
\end{aligned}
$$

where $F = CA^T$.

We only need to find one of the two factors of some full rank factorization of A in order to compute A^+ using either equation (1) or equation (5); if we choose to use equation (5) as a check of the results obtained by use of equation (1), we are not restricted to using the other factor of the full rank factorization whose first factor has already been used.

The main advantage that these algorithms have over equation (4) is that only one matrix inverse needs to be calculated.

Although there are infinitely many possible full rank factorizations, it seems most appropriate to use the effort already put into obtaining the row-echelon form through application of the Gauss-Jordan algorithm. The selection of an appropriate B or C for use in these formulas is guided by the following Theorem.

THEOREM. *Let A be an $m \times n$ matrix of rank k.*

(a) *If column vectors $B^1,\dots,B^k$ form a basis for the column space of A, then there exists a full-rank factorization of A in which the matrix B, whose jth column is B^j, is the first factor.*

(b) *If row vectors $C_1,\dots,C_k$ form a basis for the row space of A, then there exists a full-rank factorization of A in which the matrix C, whose ith row is C_i, is the second factor.*

Proof. (a) Since $B^1,\dots,B^k$ form a basis for the column space of A, we can express the jth column of A as a linear combination of the columns of B:

$$
A^j = \sum_{i=1}^{k} B^i c_{ij} = [B^1,\dots,B^k]\begin{bmatrix} c_{1j} \\ \vdots \\ c_{kj} \end{bmatrix} = BC^j. \qquad (6)
$$

This serves to define a matrix C whose entries are the coefficients of the expression of the corresponding columns of A as linear combinations of columns of B [**8**]. It follows from (6) that for this matrix C, we have $A = BC$. Since the c_{ij} are unique, it follows that the choice of C is uniquely determined by the matrix B. Since the k columns of B are linearly independent, $\rho(B) = k$. Moreover, $k = \rho(A) = \rho(BC) \leqslant \rho(C)$. Since C is $k \times n$, with $k \leqslant n$, it also follows that $\rho(C) \leqslant k$, hence $\rho(C) = k$, and the factorization is of full rank. Assertion (b) follows from (a) via use of transposes.

EXAMPLE 1 (*continued*). Since the first two rows of the row-echelon form are a basis for the row space of A, we obtain

$$
C = \begin{bmatrix} 1 & 0 & -4 & -5 \\ 0 & 1 & 3 & 4 \end{bmatrix}.
$$

Using equation (5), we compute

$$F = CA^T = \begin{bmatrix} -32 & -22 & -12 & -42 \\ 26 & 20 & 14 & 32 \end{bmatrix}$$

from which we obtain

$$FF^T = \begin{bmatrix} 3416 & -2784 \\ -2784 & 2296 \end{bmatrix}.$$

The inverse of this 2×2 matrix is

$$(FF^T)^{-1} = 11560^{-1} \begin{bmatrix} 287 & 348 \\ 348 & 427 \end{bmatrix}. \tag{7}$$

Substitution of (7) into equation (5) yields

$$A^+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -4 & 3 \\ -5 & 4 \end{bmatrix} 11560^{-1} \begin{bmatrix} 287 & 348 \\ 348 & 427 \end{bmatrix} \begin{bmatrix} -32 & -22 & -12 & -42 \\ 26 & 20 & 14 & 32 \end{bmatrix}$$

$$= 340^{-1} \begin{bmatrix} -4 & 19 & 42 & -27 \\ -1 & 26 & 53 & -28 \\ 13 & 2 & -9 & 24 \\ 16 & 9 & 2 & 23 \end{bmatrix}.$$

Since we get the same answer both ways, we are confident that we have obtained the correct solution.

This algorithm can be used even when there is no chance that the matrix might be nonsingular; we start with row reduction to row-echelon form.

EXAMPLE 2. Find the generalized inverse of the matrix

$$A = [5 \ 0 \ 0].$$

The row-echelon form of this matrix is [1 0 0], so that the first column of A is a basis for the column space of A, and we have $B = [5]$. Using equation (1), we compute

$$E = A^T B = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix} [5] = \begin{bmatrix} 25 \\ 0 \\ 0 \end{bmatrix}$$

from which we obtain

$$E^T E = [25 \ 0 \ 0] \begin{bmatrix} 25 \\ 0 \\ 0 \end{bmatrix} = [625].$$

Substitution in equation (1) yields

$$A^+ = \begin{bmatrix} 25 \\ 0 \\ 0 \end{bmatrix} [625]^{-1} [5] = \begin{bmatrix} 0.2 \\ 0 \\ 0 \end{bmatrix}.$$

The row matrix [1 0 0] is a basis for the row space of A, so that we have $C = [1 \ 0 \ 0]$. Using equation (5), we compute

$$F = CA^T = [1 \ 0 \ 0] \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix} = [5]$$

from which

$$FF^T = [5][5] = [25],$$

and substituting in equation (5) yields

$$A^{+}=C^{T}(FF^{T})^{-1}F=\begin{bmatrix}1\\0\\0\end{bmatrix}[25]^{-1}[5]=\begin{bmatrix}0.2\\0\\0\end{bmatrix}.$$

Again, we get the same answer both ways.

As a practical matter, we do not need to row-reduce the matrix A all the way to row-echelon form in order to be able to construct either a suitable matrix B or a suitable matrix C, since all that is required is a basis for either the row space of A or the column space of A. In Example 2, we could have used A itself as a suitable matrix C, without doing any row reduction at all.

Now that we have calculated two generalized inverses, it is time to use them in an applied problem. First, we need to cite some more theory. If A is $m\times n$, X is $n\times p$, B is $p\times q$, and C is $m\times q$, and if A, B, and C are known, then $X=A^{+}CB^{+}$ is the **minimum norm least squares solution** of $AXB=C$ [**13**, pp. 60–61], and if $AA^{+}CBB^{+}=C$, then $X=A^{+}CB^{+}$ is the **minimum norm solution** of $AXB=C$ [**13**, pp. 24–25].

EXAMPLE 3. Find the minimum norm least squares solution of $AXB=C$ in which A is the 1×3 matrix of the second example, B is the 4×4 matrix of the first example, and C is the 1×4 matrix [51 85 51 85]; test to see if it is indeed a solution.

The minimum norm least squares solution is given by

$$X=\begin{bmatrix}0.2\\0\\0\end{bmatrix}[51\quad 85\quad 51\quad 85]340^{-1}\begin{bmatrix}-4&19&42&-27\\-1&26&53&-28\\13&2&-9&24\\16&9&2&23\end{bmatrix}$$

$$=\begin{bmatrix}0.102&0.238&0.374&-0.034\\0&0&0&0\\0&0&0&0\end{bmatrix}.$$

Since

$$[5\quad 0\quad 0]\begin{bmatrix}0.102&0.238&0.374&-0.034\\0&0&0&0\\0&0&0&0\end{bmatrix}340^{-1}\begin{bmatrix}-4&19&42&-27\\-1&26&53&-28\\13&2&-9&24\\16&9&2&23\end{bmatrix}=[51\quad 85\quad 51\quad 85],$$

it follows that X is, in fact, the minimum norm solution.

In classroom use, there is no need for the rank of the matrix A to exceed two, since, for rank three or greater, so much effort is spent in inverting either $E^{T}E$ or FF^{T} that the point of it all is forgotten.

The author would like to thank the referees for comments and suggestions which have improved the clarity of the note and which have led to the removal of minor errors.

References

[1] A. E. Albert, Regression and the Moore-Penrose Pseudoinverse, Academic Press, New York, 1972.

[2] A. Ben-Israel, and T. N. E. Greville, Generalized Inverses: Theory and Applications, John Wiley and Sons, New York, 1974.

[3] R. E. Cline, Elements of the Theory of Generalized Inverses for Matrices, Education Development Center, Newton, 1979.

[4] R. Deutsch, Estimation Theory, Prentice-Hall, Englewood Cliffs, New Jersey, 1965.

[5] C. N. Dorny, A Vector Space Approach to Models and Optimization, Wiley-Interscience, New York, 1975.

[6] F. A. Graybill, Introduction to Matrices with Applications to Statistics, Wadsworth Publishing Company, Belmont, California, 1969.

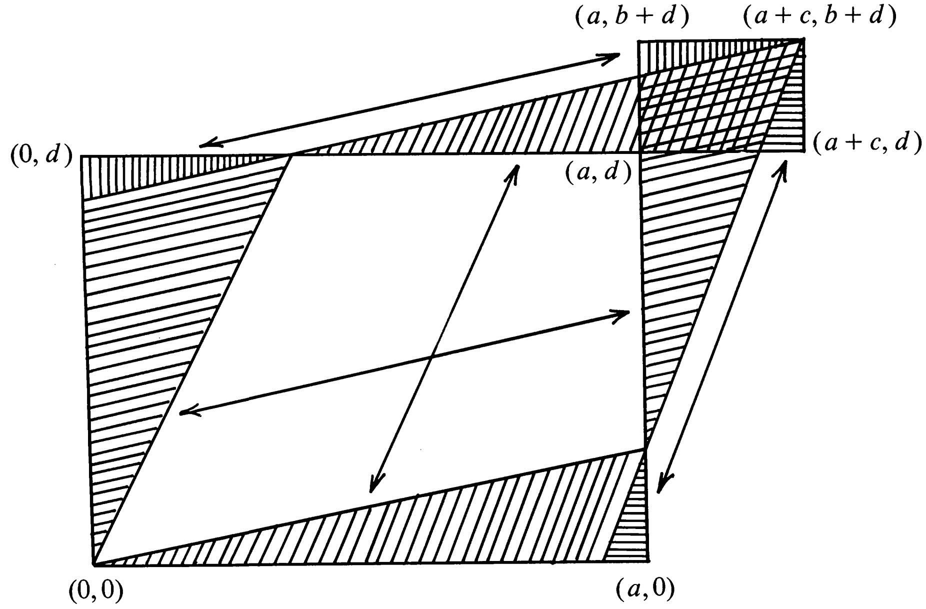
[7] ______, Theory and Application of the Linear Model, Duxbury Press, North Scituate, Massachusetts, 1976.

[8] I. I. Kolodner, A note on matrix notation, Amer. Math. Monthly, 71 (1964) 1031–1032.

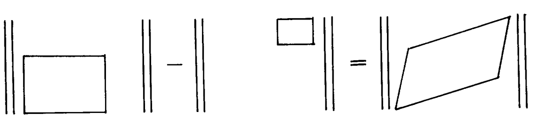
[9] T. O. Lewis, and P. L. Odell, Estimation in Linear Models, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

[10] C. Long, Visualization of matrix singular value decomposition, this MAGAZINE, 56 (1983) 161–167.
[11] B. Nobel and J. W. Daniel, Applied Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1977.
[12] C. R. Rao, Linear Statistical Inference and its Applications, 2nd ed., John Wiley and Sons, New York, 1973.
[13] C. R. Rao and S. K. Mitra, Generalized Inverse of Matrices and its Applications, John Wiley and Sons, New York, 1971.
[14] SAS User's Guide, SAS Institute, Cary, North Carolina, 1979.
[15] S. R. Searle, Matrix Algebra for the Biological Sciences, John Wiley and Sons, New York, 1966.
[16] G. Strang, Linear Algebra and its Applications, Academic Press, New York, 1976.
[17] D. M. Wiberg, Schaum's Outline of State Space and Linear Systems, McGraw-Hill, New York, 1971.

Proof without words: A 2 × 2 determinant is the area of a parallelogram



$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc =$

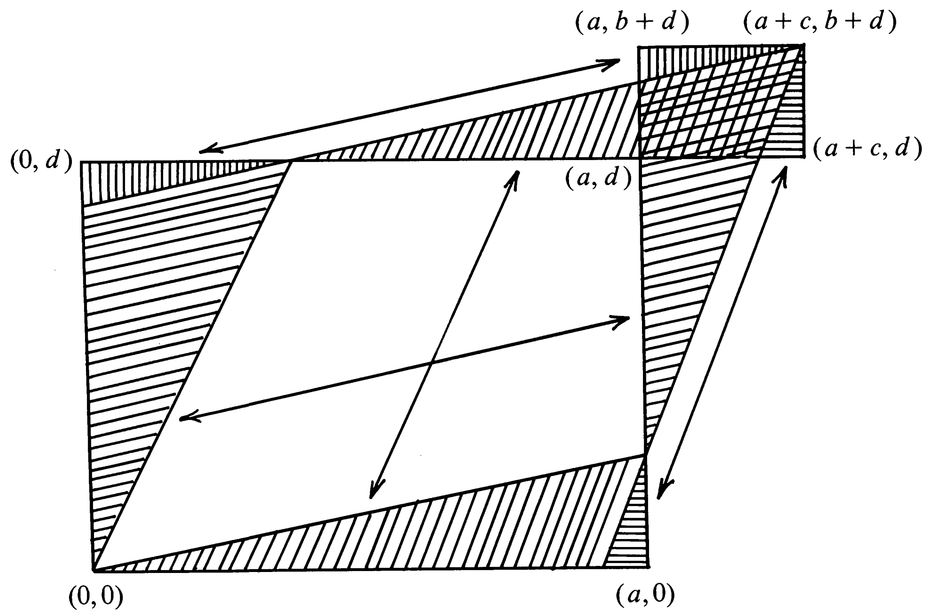


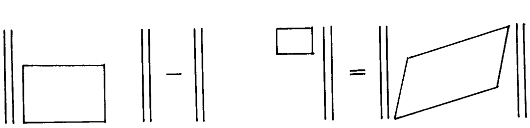
—SOLOMON W. GOLOMB
University of Southern California

Editor's note: This proof is for the case $0 < b < d$, $0 < c < a$. Professor Golomb has found dissections for the other cases as well, which the reader may seek to rediscover.

[10] C. Long, Visualization of matrix singular value decomposition, this MAGAZINE, 56 (1983) 161–167.
[11] B. Nobel and J. W. Daniel, Applied Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1977.
[12] C. R. Rao, Linear Statistical Inference and its Applications, 2nd ed., John Wiley and Sons, New York, 1973.
[13] C. R. Rao and S. K. Mitra, Generalized Inverse of Matrices and its Applications, John Wiley and Sons, New York, 1971.
[14] SAS User's Guide, SAS Institute, Cary, North Carolina, 1979.
[15] S. R. Searle, Matrix Algebra for the Biological Sciences, John Wiley and Sons, New York, 1966.
[16] G. Strang, Linear Algebra and its Applications, Academic Press, New York, 1976.
[17] D. M. Wiberg, Schaum's Outline of State Space and Linear Systems, McGraw-Hill, New York, 1971.

Proof without words:
A 2×2 determinant is the area of a parallelogram



$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc =$ 

—SOLOMON W. GOLOMB
University of Southern California

Editor's note: This proof is for the case $0 < b < d$, $0 < c < a$. *Professor Golomb has found dissections for the other cases as well, which the reader may seek to rediscover.*

Trapezoidal Numbers

CARLTON GAMER
DAVID W. ROEDER
JOHN J. WATKINS
The Colorado College
Colorado Springs, CO 80903

A strikingly simple but little-known result in elementary number theory has recently been discovered as a consequence of the musings of a composer on the Theme of Schoenberg's Variations for Orchestra, Op. 31. He began with the observation that its 12-tone set is partitioned into subsets containing three, four, and five tones. This partitioning into consecutive positive integers determines such thematic details as the number of pitches in a motive, the number of notes in a chord, and the number of measures in a phrase. For Schoenberg's own comments on this piece see his essay [**5**].

From there the composer turned his attention to tone sets containing some number of tones other than twelve and to the feasibility of partitioning each of these into subsets in a similar manner. The question he asked was this: *which numbers can be written as a sum of consecutive positive integers*? The composer (Gamer) then proceeded to check all numbers up to 100 by hand and a remarkable pattern emerged: the only numbers that cannot be written as a consecutive sum are the powers of 2. Incidentally, part of this conclusion appears as problem 78 on page 202 of [**2**].

One approach to answering the question is to realize that we have a natural generalization of **triangular numbers**: the nth triangular number is $T_n = 1 + 2 + \cdots + n$, a consecutive sum that begins with 1. A well-known formula is $T_n = \frac{1}{2}n(n+1)$. A number will be called **trapezoidal** if it is the sum of at least two consecutive positive integers. FIGURE 1 illustrates not only that $12 = 3 + 4 + 5$ is trapezoidal but also that it is the difference of two non-consecutive triangular numbers.

PROPOSITION 1. *All positive integers except the powers of* 2 *are trapezoidal.*

Proof. Let n be a trapezoidal number that is a sum of l consecutive numbers beginning with $k+1$. Then

$$\begin{aligned} n &= (k+1) + (k+2) + \cdots + (k+l) \\ &= T_{k+l} - T_k \\ &= \frac{(k+l)(k+l+1)}{2} - \frac{k(k+1)}{2} \\ &= \frac{l(2k+l+1)}{2}. \end{aligned}$$

Now, one of l or $2k + l + 1$ is odd (and the other is even). Therefore, if n is trapezoidal, then it is not a power of 2.

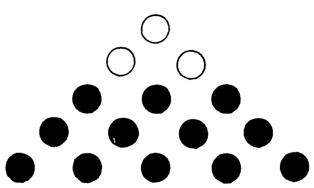


FIGURE 1. $\mathbf{12 = 3 + 4 + 5 = 15 - 3 = T_5 - T_2}$.

Conversely, let n be a positive integer with an odd factor; we show that it is trapezoidal. Since n has an odd factor, so does $2n$, and we can write

$$2n = f_1 f_2$$

where one of f_1 or f_2 is odd and $1 < f_1 < f_2$. To express n as a sum of l consecutive integers beginning with $k+1$, we simply let

$$l = f_1 \quad \text{and} \quad k = \frac{f_2 - f_1 - 1}{2}.$$

Then, $f_2 = 2k + l + 1$ so that

$$n = \frac{f_1 f_2}{2} = \frac{l(2k+l+1)}{2} = T_{k+l} - T_k$$
$$= (k+1) + (k+2) + \cdots + (k+l).$$

Since we get different values of l and k for different choices of f_1 and f_2, we can count the number of ways an integer may be expressed as a consecutive sum by counting its non-trivial odd factors. This result also appears in [**3**] and [**4**].

PROPOSITION 2. *Let $n = 2^r p_1^{r_1} p_2^{r_2} \cdots p_s^{r_s}$ be the prime decomposition of an integer n where the p_i are distinct odd primes. Then the number of ways that n can be written as a sum of at least two consecutive positive integers is*

$$\tau\left(\frac{n}{2^r}\right) - 1 = (r_1+1)(r_2+1)\cdots(r_s+1) - 1$$

where $\tau(m)$ is the number of positive divisors of m.

As an illustration, we consider $n = 225$. Since $225 = 3^2 \cdot 5^2$, there are $(2+1)(2+1) - 1 = 8$ ways to write 225 as a consecutive sum. TABLE 1 shows these sums, using the notation in the proof of Proposition 1.

Odd Factor	$f_1(=l)$	f_2	$k = \frac{1}{2}(f_2 - f_1 - 1)$	Trapezoidal sum
3	3	150	73	$74+75+76$
5	5	90	42	$43+44+45+46+47$
3^2	9	50	20	$21+\cdots+29$
$3 \cdot 5$	15	30	7	$8+\cdots+22$
5^2	18	25	3	$4+\cdots+21$
$3^2 \cdot 5$	10	45	17	$18+\cdots+27$
$3 \cdot 5^2$	6	75	34	$35+\cdots+40$
$3^2 \cdot 5^2$	2	225	111	$112+113$

TABLE 1

Another approach to the question of which numbers are sums of consecutive positive integers is to focus on the *length* of the sum. For example, for $l = 2$ we get $1+2=3$, $2+3=5$, $3+4=7, \ldots$ and so the odd numbers are consecutive sums. For $l=3$ we get $1+2+3=6$, $2+3+4=9$, $3+4+5=12, \ldots$ and so multiples of 3 are also such sums. In this way, we can easily list the numbers that are sums of consecutive positive integers:

$$\begin{array}{ll} 3,\ 5,\ 7,\ 9, 11, \ldots & (l=2) \\ 6,\ 9, 12, 15, 18, \ldots & (l=3) \\ 10, 14, 18, 22, 26, \ldots & (l=4) \\ 15, 20, 25, 30, 35, \ldots & (l=5) \\ \vdots & \end{array}$$

Since each row begins with a triangular number, we see that a consecutive sum of length l is a number of the form

$$T_l + kl = \frac{l(l+1)}{2} + kl = \frac{l(2k+l+1)}{2},$$

where $l \geqq 2$ and $k \geqq 0$.

The development above shows that a positive integer n is trapezoidal if and only if $2n$ factors non-trivially as a product of an even and an odd integer. Following this same approach we can answer a more general question: *which numbers are sums of finite arithmetic progressions of positive integers*? If we let d stand for the constant difference in such a progression, then such a sum beginning with b and of length l is given by

$$b+(b+d)+\cdots+(b+(l-1)d) = \frac{l(2b+d(l-1))}{2}.$$

We invite the reader to prove the following result.

PROPOSITION 3. *A positive integer n is the sum of a non-trivial arithmetic progression with constant difference d if and only if there is a factorization* $2n = f_1 f_2$ *with* $f_1 > 1$, $f_2 > d(f_1 - 1)$, *and either* f_2 *is even in the case d is even, or* f_1 *and* f_2 *have opposite parity when d is odd. The factor* f_1 *is the number of terms in the progression.*

We began with a musical observation. We conclude with another, this having to do with the overtone series. The trapezoidal numbers denote those partials within the overtone series that are not octave equivalents of the fundamental, that is, those partials whose frequencies are not multiples by a power of 2 of the frequency of the fundamental. As a corollary to this, and in accordance with the definition of trapezoidal numbers, we know that by adding the frequencies of any two or more consecutive partials of the overtone series we can generate another partial that will not be an octave equivalent of the fundamental. (We are thinking here of an extension of the concept of sine-wave frequency modulation; for an explanation see [**1**].) This principle may prove to be of interest in the physics of sound, particularly with regard to the design of new electronic timbres.

References

[1] John Chowning, The synthesis of complex audio spectra by means of frequency modulation, Audio Engineering Soc., XXI/7 (1973) 526–534.

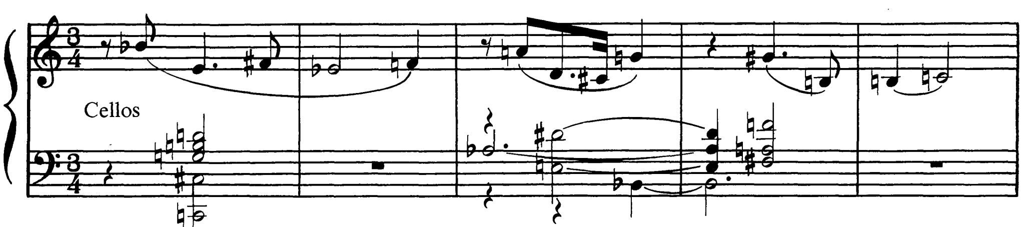
[2] Underwood Dudley, Elementary Number Theory, 2nd ed., W. H. Freeman and Company, San Francisco, 1978.

[3] Robert Guy, Sums of consecutive integers, Fib. Quart., 20 (1982) 36–38.

[4] W. J. Leveque, On representations as a sum of consecutive integers, Canad. J. Math., 4 (1950) 399–405.

[5] Arnold Schoenberg, Composition with twelve tones, Style and Idea: Selected Writings of Arnold Schoenberg, Leonard Stein, ed., St. Martins Press, New York, 1975, pp. 214–245.

Variations, Op. 31, Theme



PROBLEMS

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

Proposals

To be considered for publication, solutions should be mailed before August 1, 1985.

1211. Find the locus of points under which an ellipse is seen under a constant angle. [*Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*]

1212. Prove that if $x > 1$ and $0 < u < 1 < v$, then

$$\frac{v(x-1)(x^{v-1}-1)}{(v-1)(x^v-1)} < \log x < \frac{u(x-1)(1-x^{u-1})}{(1-u)(x^u-1)}.$$

[*L. Bass & R. Výborný, The University of Queensland, Australia; and V. Thomée, Chalmers Institute of Technology, Sweden.*]

1213. For each positive integer n let $S_n = \Sigma a_1 a_2 \dots a_n$, where the sum is taken over all n-tuples $(a_1, a_2, \dots, a_n)$ of positive integers such that $a_1 = 1$ and $a_{i+1} \leqq a_i + 1$ for $1 \leqq i < n$.

(a) How many terms are there in the sum?

(b) What is the value of S_n?

[*Nicholas K. Krier, Colorado State University, and Frank R. Bernhart, Rochester Institute of Technology.*]

1214. Let $A_0 = 1$ and $A_{n+1} = A_n + \sqrt{1 + A_n^2}$ for $n \geqq 0$. Show that $\lim_{n\to\infty} A_n/2^n$ exists, and find its value. [*Paul G. Nevai and G. A. Edgar, The Ohio State University.*]

1215. Prove that

$$\tan x < \frac{\pi x}{\pi - 2x} \quad \text{for } 0 < x < \frac{\pi}{2}.$$

[*Andrea Laforgia, Università di Torino, Italy.*]

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

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

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

Send all communications to the problems department to Leroy F. Meyers, Mathematics Department, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.

Quickie

Answer to Quickie is on p. 117.

Q696. If $f'(x) < 0 < f''(x)$ for all $x < x_0$ and $f'(x) > 0 > f''(x)$ for all $x > x_0$, then f is not differentiable at x_0. [*Edilio A. Escalona Fernandez, Maracay, Venezuela.*]

Solutions

Permutations by Derangement — March 1984

1186. (a) Show how to arrange the 24 permutations of the set $\{1,2,3,4\}$ in a sequence with the property that adjacent members of the sequence differ in each coordinate. (Two permutations (a_1, a_2, a_3, a_4) and (b_1, b_2, b_3, b_4) differ in each coordinate if $a_i \neq b_i$ for $i = 1,2,3,4$.)

*(b) For which n can the $n!$ permutations of the integers from 1 through n be arranged in a similar manner? [*Stanley Rabinowitz, Merrimack, New Hampshire.*]

Solution I (adapted by the editor): For all n except 3 we construct by induction an $n \times (n!)$ array S_n such that:

(a) the columns of S_n are the $n!$ permutations of $\{1,2,\ldots,n\}$, the first being $(1,2,\ldots,n)$; and

(b) for $n > 1$, each column of S_n is a derangement of the next column, the last column being a derangement of the first column.

Thus the columns of S_n, read in order, form a desired arrangement. For $n \leqq 4$ we may use the arrangements:

S_1 is 1; S_2 is $\begin{matrix} 1 & 2 \\ 2 & 1 \end{matrix}$; S_3 does not exist; and S_4 is

$$\begin{matrix} 1&2&1&2&3&4&3&4&1&2&4&3&2&1&2&1&3&4&3&4&2&1&4&3 \\ 2&1&3&4&1&2&1&2&3&4&3&4&1&2&3&4&2&1&2&1&3&4&3&4 \\ 3&4&2&1&2&1&4&3&4&3&1&2&3&4&1&2&1&2&4&3&4&3&2&1 \\ 4&3&4&3&4&3&2&1&2&1&2&1&4&3&4&3&4&3&1&2&1&2&1&2 \end{matrix}$$

Suppose that S_n has been constructed, where $n \geqq 2$. We show how to construct S_{n+2}. If i and j are distinct integers in the set $\{1,2,\ldots,n+2\}$, let $S_n(i,j)$ be the $(n+2) \times (n!)$ array whose ith row is

$$n+1 \quad n+2 \quad n+1 \quad n+2 \quad \ldots \quad n+1 \quad n+2$$

and whose jth row is

$$n+2 \quad n+1 \quad n+2 \quad n+1 \quad \ldots \quad n+2 \quad n+1,$$

the other rows being the rows of S_n in the order in which they appear in S_n. Let $S_n'(i,j)$ be the array obtained from $S_n(j,i)$ by reversing the order of the columns. Thus,

$$S_2(4,2) \text{ is } \begin{matrix} 1 & 2 \\ 4 & 3 \\ 2 & 1 \\ 3 & 4 \end{matrix} \quad \text{and} \quad S_2'(2,4) \text{ is } \begin{matrix} 2 & 1 \\ 3 & 4 \\ 1 & 2 \\ 4 & 3 \end{matrix}.$$

We now make a left-justified triangular array of the $\binom{n+2}{2}$ unordered pairs chosen from $\{1,2,\ldots, n+1, n+2\}$, arranged so that the pairs $\{i, j\}$ with $i<j$ are listed in row i in increasing order of j, thus:

$$\begin{aligned}
&\{1,2\}, \{1,3\},\ldots, \{1,n+1\}, \{1,n+2\},\\
&\{2,3\}, \{2,4\},\ldots, \{2,n+2\},\\
&\ldots\\
&\{n,n+1\}, \{n,n+2\},\\
&\{n+1,n+2\}.
\end{aligned}$$

We form a list of these ordered pairs by running through the triangular array in the reverse of the Cantor diagonal method: $\{n+1,n+2\},\{n,n+2\},\ldots, \{1,n+2\},\{1,n+1\},\ldots, \{n,n+1\},\{n-1,n\},\ldots, \{1,n\},\{1,n-1\},\ldots, \{1,2\}$. Note that each two consecutive pairs agree in one element, and the other elements differ from each other or the common element by 1. Beginning with $(n+1,n+2)$, we specify an order for each pair so that the common elements of two consecutive pairs are in the same position: $(n+1,n+2)$, $(n,n+2),\ldots,$ $(1,n+2)$, $(1,n+1),\ldots,$ $(n,n+1)$, $(n,n-1),\ldots.$

We construct S_{n+2} by writing down arrays $S_n(i,j)$ one to the right of the other, beginning with $S_n(n+1,n+2)$, according to the list of ordered pairs (i,j) above, and following this with the arrays $S_n'(i,j)$ in the same order. There will then be $2\binom{n+2}{2}(n!)=(n+2)!$ columns. Furthermore, the columns of $S_n(i,j)$ together with those of $S_n'(i,j)$ comprise all permutations of $(1,2,\ldots, n+2)$ having $n+1$ and $n+2$ in positions i and j in either order. Hence all permutations of $(1,2,\ldots, n+2)$ occur as columns of S_{n+2}. Since two successive $S_n(i,j)$'s differ only in that one of the "old" rows (from S_n) has been interchanged with one of the "new" rows (containing $n+1$ and $n+2$), each two successive columns in the left half of S_{n+2} are derangements of each other, and similarly in the right half. Finally, the first column of $S_n'(n+1,n+2)$ is obtained by shifting up by two rows the "old" numbers in the last column of $S_n(1,2)$ or $S_n(2,1)$ (according to n) and moving $n+1$ and $n+2$ to the bottom. Hence the first column in the right half of S_{n+2} is a derangement of the last column in the left half. Similarly, the first column of S_{n+2} is a derangement of the last column. This completes the induction step.

To begin the induction, note that S_4 was constructed from S_2 according to the above procedure. Since there is no S_3, we construct S_5 by using the procedure with S_3 replaced by

$$\begin{matrix} 1 & 2 & 3 & 1 & 2 & 3 \\ 2 & 3 & 1 & 2 & 3 & 1 \\ 3 & 1 & 2 & 3 & 1 & 2 \end{matrix} \quad \text{and} \quad \begin{matrix} 1 & 3 & 2 & 1 & 3 & 2 \\ 3 & 2 & 1 & 3 & 2 & 1 \\ 2 & 1 & 3 & 2 & 1 & 3 \end{matrix}$$

in the first and second halves of the construction.

R. B. EGGLETON
W. D. WALLIS
The University of Newcastle, Australia

Solution II: An arrangement is possible for all n except 3. Let S_n be the set of all permutations of the integers from 1 through n. Then S_n is made into a graph satisfying the conditions of the problem by joining with an edge any two permutations which differ in each coordinate. Such permutations are called *derangements* of each other. As is well known (see Brualdi, *Introductory Combinatorics*, 1st ed., §5.3), each permutation in S_n has exactly

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

derangements. Thus in the graph S_n we find exactly D_n edges meeting each vertex, i.e., the graph S_n is *D_n-regular.*

The problem is to produce a *hamilton path* in S_n: a sequence of the $n!$ vertices of S_n with consecutive vertices joined by an edge and with no vertex repeated. Such a path is clearly possible when n is 1 or 2. There is no hamilton path in S_3, since S_3 is not even connected (surely a necessary condition for such a path to exist). Assume now that $n \geqq 4$. Then S_n will be shown to have a hamilton path by use of a theorem of Jackson ("Hamilton cycles in regular 2-connected graphs," *J. Comb. Th.*, ser. B, vol. 29 (1980), pp. 27–46) which says that a 2-connected k-regular graph with not more than $3k$ vertices has a hamilton path.

Now the series $\Sigma_{j=0}^{\infty} (-1)^j/j!$ has alternating terms which decrease in absolute value, and its sum is $1/e$. Hence, for $n \geqq 4$ it is surely the case that $\Sigma_{j=0}^{n}(-1)^j/j! \geqq 1/e - 1/(5!)$, and so

$$3D_n = 3(n!) \sum_{j=0}^{n} \frac{(-1)^j}{j!} \geqq 3(n!)\left(\frac{1}{e} - \frac{1}{5!}\right) \geqq n!,$$

which shows that the D_n-regular graph S_n has no more than $3D_n$ vertices.

To complete the solution, we need only show that S_n is 2-connected, i.e., that removal of any vertex of S_n leaves a connected graph. The following lemma shows that more is true.

LEMMA. *If $n \geqq 4$, then for each two permutations α, β in S_n, there are at least two permutations in S_n which are derangements of both α and β.*

Proof. Evidently it is sufficient to find two permutations in S_n which are derangements of both $\alpha = (1,2,\ldots,n)$ and $\beta = (b_1, b_2, \ldots, b_n)$. We use induction on n. For $n = 4$, the conclusion is verified by a direct check of the 24 possibilities for β. Suppose now that the lemma holds for all n with $4 \leqq n < m$, and let $\beta = (b_1, b_2, \ldots, b_m)$. There are two cases. If $b_m = m$, find two permutations φ_1 and φ_2 in S_{m-1} which are derangements of $(1,2,\ldots,m-1)$ and $(b_1, b_2, \ldots, b_{m-1})$. Then $\varphi_1' = (m, \varphi_1(2), \ldots, \varphi_1(m-1), \varphi_1(1))$ and $\varphi_2' = (m, \varphi_2(2), \ldots, \varphi_2(m-1), \varphi_2(1))$ are distinct derangements of α and β. If $b_m = k \neq m$, first form $\gamma = (c_1, c_2, \ldots, c_m)$ from β by exchanging m and k. Find φ in S_{m-1} which is a derangement of both $(1,2,\ldots,m-1)$ and $(c_1, c_2, \ldots, c_{m-1})$. Now select j such that $\varphi(j) \neq b_m$ and $b_j \neq m$. There will be at least two such j's since there are $m - 1 \geqq 4$ candidates for j and at most two are ruled out. For each such j, the permutation $(\varphi(1), \ldots, \varphi(j-1), m, \varphi(j+1), \ldots, \varphi(m-1), \varphi(j))$ will be a derangement of both α and β.

JERRY METZGER
University of North Dakota

Also solved using group theory by Dorothy Wolfe; *and*, *citing Jackson's theorem*, *by Dopey*, *Anthony Gardiner* (*Australia*), *and Heiko Harborth & Arnfried Kemnitz* (*West Germany*); *solved partially* (*part* (*a*) *only*) *by Richard Gibson* (*student*), *M. R. Gopal*, *Gymnasium Bern-Kirchenfeld Problem Solving Group* (*Switzerland*), *Mark Kantrowitz* (*student*), *J. C. Linders* (*The Netherlands*), *Hubert J. Ludwig*, *Michael Vowe* (*Switzerland*), *Harry Zaremba*, *and the proposer.* There were two incorrect solutions to part (a) and two incorrect, seriously incomplete, or unreadable solutions to part (b). Hint to solvers: Check your solutions for correctness and clarity by reading them a few days after you have written them, or get a friend to read them.

Wolfe used group theory to find a suitable sequence of permutations. For $j = 2,3,\ldots,n$, let q_j be a j-cycle on $\{1,2,\ldots,n\}$, with the restriction that if $k > j$, then q_k moves every element that q_j moves. Let p_0 be the identity permutation, and for $1 \leqq k \leqq n!$, let $p_k = p_{k-1} r_k$ (operating from left to right), where $r_k = q_n q_{n-1} \cdots q_m$ if $k \equiv 0$ (mod $n!/m!$) and $k \not\equiv 0$ (mod $n!/(m-1)!$). If

$$q_n = (1,2,\ldots,n-1,n) \quad \text{if } n \text{ is even}, \qquad q_n = (1,3,\ldots,n,2,4,\ldots,n-1) \quad \text{if } n \text{ is odd},$$

$$q_j = (n-j+1, n-j+2, \ldots, n) \qquad \text{if } 2 < j < n,$$

and

$$q_2 = (n-1, n) \quad \text{if } n \text{ is even}, \qquad q_2 = (n-2, n-1) \quad \text{if } n \text{ is odd},$$

then each r_k fixes no element of $\{1,2,\ldots,n\}$, and the $n!$ permutations p_k (with $p_{n!} = p_0$) are distinct if $n \neq 3$.

1187. Let the chord AB of circle O be trisected at C and D. Let P be any point on the circle other than A and B. Extend the lines PD and PC to intersect the circle in E and F, respectively. Extend the lines EC and FD to intersect the circle in G and H, respectively. Let GF and HE intersect AB in L and M, respectively. Prove that $AL = BM$. [*R. S. Luthar, University of Wisconsin Center, Janesville.*]

Solution I: In *The Two Year College Mathematics Journal*, vol. 14 (1983), p. 3, Ross Honsberger states a lemma shown to him by his friend Professor Haruki.

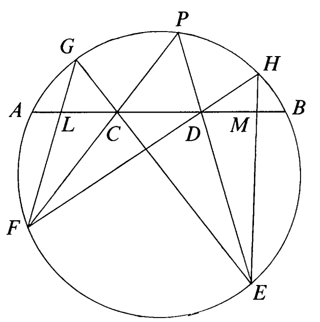
LEMMA. *Suppose AB and FE are nonintersecting chords in a circle, and that Q is a variable point on the arc AB remote from F and E. Then for each position of Q, the lines QF and QE cut AB into three segments of lengths x, y, and z (in order) such that xz/y is independent of Q.*

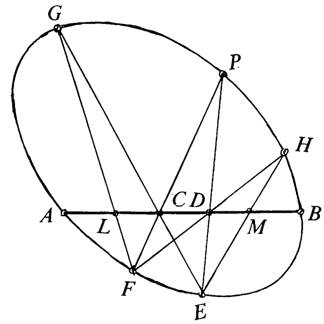
Applying this lemma twice (with Q taken as G and as H), we find

$$\frac{AL \cdot CB}{LC} = \frac{AD \cdot MB}{DM}.$$

Since $AC = DB$, it follows that $CB = AD$. Reducing and inverting the proportion gives $LC/AL = DM/MB$. Adding 1 and reducing yields $AL = MB$.

RONALD S. TIBERIO
Natick, Massachusetts





Solution II: More generally, the circle may be replaced by any conic, and it is sufficient to assume that $\overline{AC} = \overline{DB}$. The quadruples of lines $G(AFEB)$ and $H(AFEB)$ have the same cross ratio, as do their sections $(ALCB)$ and $(ADMB)$. But the cross ratios $(ALCB)$ and $(BCLA)$ are equal. Hence $(BCLA) = (ADMB)$. Since BA, CD, and AB are symmetric with respect to the midpoint of AB, so is LM.

JORDI DOU
Barcelona, Spain

Also solved by Anon (Erewhon), Don Chakerian, Benny N. Cheng, Tony Costa (student), Jan van de Craats (The Netherlands), Howard Eves, Mark Kantrowitz (student), L. M. Kelly, L. Kuipers (Switzerland), J. C. Linders (The Netherlands, two solutions), Gary Ling, Graham Lord, Richard Parris, Jan Söderkvist (student, Sweden), J. M. Stark, Robert L. Young, Harry Zaremba, and the proposer.

Several solvers pointed out that since the point P plays no role in the solution, the lines ED and FC need not meet on the circle (or conic). Desargues's involution theorem (see H. S. M. Coxeter, *Projective Geometry*, 2nd ed., p. 87) was used by van de Craats to prove a generalization, similar to one found by Eves. *Let m be a line and let O_1 and O_2 be conics intersecting in exactly four points nonincident with m. Let m intersect O_i in A_i and B_i. If each of the pairs (A_1, B_1) and (A_2, B_2) is symmetric about the (same) point N, then for every conic O passing through the intersections of O_1 and O_2, the intersection points of O with m are symmetric about N.* In the proposed problem, O_1 is the circle, O_2 is the degenerate conic consisting of the lines EG and FH, and O is the degenerate conic consisting of FG and EH. "The celebrated butterfly problem", by Léo Sauvé, *Crux Mathematicorum* (formerly *EUREKA*), vol. 2 (1977), pp. 2–5, contains an extensive bibliography for the well-known case in which $C = D$.

1188. Prove that for all real x and all integers $n > 1$,

$$|\cos(2x)|^{n/2} \leqslant |\cos^{2n}x - \sin^{2n}x|.$$

When does equality hold? [*Vania D. Mascioni, student, ETH Zürich, Switzerland.*]

Solution I: When $n = 2$, the inequality reduces to the identity $\cos 2x = \cos^2 x - \sin^2 x$. As we shall see, equality holds for $n > 2$ exactly when x is an integral multiple of $\pi/4$.

The graphs of $y = |\cos 2x|^{n/2}$ and $y = |\cos^{2n}x - \sin^{2n}x|$ are both symmetric with respect to every vertical line $x = k\pi/4$, as may be seen by replacing x by $k\pi/2 - x$ in each formula. Hence we may limit our attention to $0 \leqq x \leqq \pi/4$, where the absolute value signs are redundant.

Let $t = \sqrt{\cos 2x}$, so that $\cos^2 x = \frac{1}{2}(1 + t^2)$ and $\sin^2 x = \frac{1}{2}(1 - t^2)$. The inequality to be proved now becomes

$$t^n \leqq \left[(1 + t^2)^n - (1 - t^2)^n\right]/2^n,$$

or (if the solution $t = 0$, i.e., $x = \pi/4$, is ignored)

$$2^n \leqq \left(\frac{1}{t} + t\right)^n - \left(\frac{1}{t} - t\right)^n = f(t),$$

where $0 < t \leqq 1$ and $n > 2$. To finish the job, it is sufficient to calculate the derivative:

$$f'(t) = n(1 + t^2)(1 - t^2)\left[(1 - t^2)^{n-2} - (1 + t^2)^{n-2}\right]/t^{n+1}.$$

Observe that $f'(t) < 0$ for $0 < t < 1$, and that $f(1) = 2^n$. This establishes the desired inequality, and shows that equality occurs exactly when $t = 1$ (x is an even multiple of $\pi/4$) or $t = 0$ (x is an odd multiple of $\pi/4$).

Richard Parris
Phillips Exeter Academy

Solution II: Let $a = \cos^2 x$ and $b = \sin^2 x$. Then $a \geqq 0$, $b \geqq 0$, and $a + b = 1$, and the inequality becomes

$$|a - b|^{n/2} \leqq |a^n - b^n|,$$

by use of the double-angle formula. Since $|a + b|^{n/2} = 1$, we need merely prove

$$|a^2 - b^2|^{n/2} \leqq |a^n - b^n|.$$

If $a = b$, equality holds. If $a > b$, then let $c = \sqrt{a^2 - b^2}$; the inequality becomes

$$c^n \leqq (c^2 + b^2)^{n/2} - b^n.$$

But if $n \geqq 2$, then

$$(c^2 + b^2)^{n/2} = (c^2 + b^2)(c^2 + b^2)^{(n-2)/2} = c^2(c^2 + b^2)^{(n-2)/2} + b^2(c^2 + b^2)^{(n-2)/2}$$
$$\geqq c^2 c^{n-2} + b^2 b^{n-2} = c^n + b^n.$$

A similar argument works if $a < b$.

Equality holds if and only if (i) $(n - 2)/2 = 0$, i.e., $n = 2$; or, when $n > 2$, (ii) $a = b$, i.e., x is an odd multiple of $\pi/4$, or (iii) $a = 0$ or $b = 0$, i.e., x is an even multiple of $\pi/4$.

Peter Shor, student
MIT, jointly with
Mark Naigles, student
Tufts University

Also solved by Paul F. Byrd, Tony Costa (student), Richard Gibson (student), Chico Problem Group, L. Kuipers Switzerland), Weixuan Li & Edward T. H. Wang (Canada), Peter W. Lindstrom, Beatriz Margolis (France), Michael Powers, Robert E. Shafer, Jan Söderkvist (student, Sweden), J. M. Stark, Charles H. Toll, Michael Vowe (Switzerland), and the proposer; and partially by David Boduch (student). There was one incorrect solution.

Group, Lindstrom, Margolis, and Shafer noted that the inequality holds for all real $n \geqq 2$, provided that $\cos^{2n}x$ and $\sin^{2n}x$ are defined as $(\cos^2 x)^n$ and $(\sin^2 x)^n$. Shafer also noted that the inequality is reversed if $0 < n < 2$. Both remarks follow easily from each of the solutions above.

A Reversed Digit Problem **March 1984**

1190. (a) If abc is a three-digit number in base ten, where $a > c$, and if $def = abc - cba$ is always considered a three-digit number (even when $d = 0$), then it is known that $def + fed = 1089$. Generalize this result to any base $k > 1$.

(b) For which base(s) is $def + fed$ a perfect square? [*Robert L. Bernhardt, East Carolina University.*]

Solution: (a) The subtraction $abc - cba$ in base k is easily seen to yield:

$$\begin{array}{rccc} & a & b & c \\ - & c & b & a \\ \hline & a-c-1 & k-1 & c-a+k \end{array}$$

where borrowing is needed in the last two places, since $c < a$. Then the addition $def + fed$ yields:

$$\begin{array}{rcccc} & & a-c-1 & k-1 & c-a+k \\ + & & c-a+k & k-1 & a-c-1 \\ \hline & 1 & 0 & k-2 & k-1 \end{array}$$

where carrying is needed twice.

(b) From (a) we see that $def + fed = k^3 + (k-2)k + (k-1) = k^3 + k^2 - k - 1 = (k-1)(k+1)^2$. This is a perfect square if and only if $k-1$ is a square, for example, if $k = 10$.

J. C. LINDERS
Eindhoven, The Netherlands

Also solved by S. F. Barger, Robert E. Bernstein, Judith Biasotti, David J. Boduch (student), Duane M. Broline, Stephen D. Bronn, Stephen E. Eldridge (England), Richard Gibson (student), Westmont College Problem Solving Group, G. A. Heuer, Mark Kantrowitz (student), Edwin M. Klein, L. Kuipers (Switzerland), Peter W. Lindstrom, Gary Ling, Graham Lord, Vania D. Mascioni (student, Switzerland), Mark Naigles (student), Roger B. Nelsen, C. C. Oursler, P. J. Pedler (Australia), Daniel A. Rawsthorne, Daniel M. Rosenblum, Vincent P. Schielack, Jr., Jan Söderkvist (student, Sweden), Lawrence Somer, Dennis Spellman, Ronald S. Tiberio, Douglas H. Underwood, Michael Vowe (Switzerland), and the propser.

Schielack noted that if $k = n^2 + 1$, then the square root of the magic number is expressed in base k as nn.

Answer

Solution to the Quickie on p. 112.

Q696. Suppose that f is differentiable at x_0. Then $f'(x_0) = 0$ and, given $h > 0$, by the mean value theorem applied to f', there is a c in $(x_0, x_0 + h)$ such that

$$0 > f''(c) = \frac{f'(x_0+h) - f'(x_0)}{h} = \frac{f'(x_0+h)}{h} > 0,$$

contrary to hypothesis.

REVIEWS

Paul J. Campbell, Editor
Beloit College

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

Ognibene, Peter J., *Artificial intelligence: secret ciphers solved,* Omni 7:2 (November 1984) 38.

While mathematicians have been creating and breaking knapsack codes and 70-digit RSA codes, banks continue to use only short spoken passwords in their daily electronic fund transfers of billions of dollars daily. "There's good reason to believe that there's larceny ... happening on that insufficient protection, and the banks' response to it appears to be to deny its existence rather than to put in systems that will cure it." Note, though, that the Treasury Dept. has ordered all banks to use the Data Encryption Standard in transfers with the federal government (see "Speaking in codes," Datamation (1 December 1984) 40, 45).

Peterson, Ivars, *Shadows from a higher dimension: the journey into higher dimensions begins in Flatland,* Science News (3 November 1984) 285.

Recalls the 1884 publication of Edwin A. Abbott's Flatland: A Romance of Many Dimensions, an event whose centenary was celebrated by a symposium at Brown University. Several subsequent writers have invented analogous worlds, pursued the technology of 2-D, or continued the social satire at the center of Abbott's motivation.

Peterson, Ivars, *The unpacking of a knapsack,* Science News 126 (24 November 1984) 330-331.

Ernest Brickell (Sandia National Labs) has pulled the rug out from under knapsack codes which rely solely on modular multiplication. There may still exist a secure knapsack code--based on finite-field arithmetic, perhaps--but opinion is running against that possibility. The only other public-key cryptosystem, known as RSA, depends for its security on the complexity of factoring, which is still an open question.

Dembart, Lee, *Purdue professor proves touchy over proof of conjecture: mathematician claims colleagues stole idea,* Los Angeles Times (18 November 1984) 3, 30.

"The controversy that has grown out of de Branges' proof [of the Bieberbach conjecture] reveals that mathematics, one of mankind's highest intellectual achievements, is driven by the same human motivations, frailties and foibles that are found in more mundane work. It also reveals that individualists and creative people can have the same interpersonal problems in the supposedly sterile world of science that they have elsewhere in life."

Schniederjans, Marc J., Linear Goal Programming, Petrocelli, 1984; xvi + 229 pp, $24.95.

Linear goal programming is an extension of linear programming to solving multiple-objective and conflicting-goal linear programming problems. This book, designed for management-science students, includes a brief treatment of linear programming, followed by chapters on formulation and solution of linear goal programs. One-third of the book is devoted to 11 case studies based in part on real-world problem situations. Mathematical science students studying linear programming would benefit from seeing also the broader and more realistic framework of linear goal programming.

Taylor, John L. (ed), Teacher Shortage in Science and Mathematics: Myths, Realities, and Research; Summary and Proceedings, National Institute of Education, 1983 and 1984; vi + 54 pp (P), iv + 281 pp, (P).

Accounts of a 1983 conference. Like other reports that have built public concern over the issues of quantity and quality of teachers, this one offers no quick cures. It does suggest topics and general problems which cannot be adequately resolved by individual districts or states: curriculum reform, research on learning and instruction, and increasing public support of classroom teaching.

Watts, Lisa, and Wharton, Mike, Usborne Introduction to Machine Code for Beginners, EDC Publishing (8141 E. 44th St., Tulsa, OK 74145), 1983; 48 pp, (P).

Brilliant watercolor cartoons of robots performing machine-language instructions inside a computer make this booklet entertaining to read. It features both hex codes and mnemonics for 6502 and Z-80 machines, with adaptations noted for particular microcomputers.

Babbage, Henry Prevost, Babbage's Calculating Engines, Tomash Publishers, 1982; xxix + 359 pp, $45.

Second volume in the Charles Babbage Institute Series for the History of Computing, this volume is a facsimile of the 1889 edition, with some added plates and a new introduction by A. G. Bromley. Included among the papers reprinted is the famous annotated translation by Lovelace of the paper by Menabrae about the Analytical Engine.

Meirovitz, Marso, and Jacobs, Paul I., Brain Muscle Builders: Games to Increase Your Natural Intelligence, Prentice-Hall, 1983; ix + 244 pp, (P).

From the inventor of Mastermind comes a "gymnasium of the mind," a book of games designed to foster the skills of deductive logic, inductive logic, and strategic planning. More so than other game books, this one makes the reader aware of the skills each game is meant to encourage and also makes the connection back to using the same skills in the reader's daily life.

Ziemba, William T., and Hausch, Donald B., Beat the Racetrack, Harcourt Brace Jovanovich, 1984; xx + 392 pp, $22.95.

Written by a professor of operations research and portfolio management and a graduate student, this book explains the mathematically-based "Dr. Z" system for parimutuel betting: Make place and show bets in inefficient markets (the win market tends to be very efficient). Appropriately, the foreword is by E. Thorp, professor and author of a blackjack system (Beat the Dealer) based on card-counting.

Breaking Bieberbach, Scientific American 251:5 (November 1984) 75.

Short exposition of what the Bieberbach conjecture says, occasioned by the recent proof of de Branges.

Prof finds a new twist to solve an old riddle, Dayton (Ohio) Journal Herald (22 November 1984).

Announces counterexample by Henry Wente (Toledo) to a conjecture of Hopf in the 1940's that a sphere is the only surface with constant mean curvature.

Angier, Natalie, *Folding the perfect corner: a young Bell scientist makes a major math breakthrough*, Time (3 December 1984) 63.

Popular account of economic significance of discovery by 28-year-old Narendra Karmarkar (Bell Labs) of a new polynomial-time algorithm for linear programming, faster than the simplex method.

Dembart, Lee, *Experts chip away at ancient math puzzle: Fermat's Last Theorem*, Los Angeles Times (10 October 1984) 3, 22.

Researchers, including L. M. Adleman (U.S.C.), have proven that Fermat's Last Theorem holds for infinitely many (but not necessarily all) prime exponents in the case where the exponent divides none of the three bases. Previously this case had been verified for all primes less than 5.7×10^{10}.

Townend, M. Stewart, Mathematics in Sport, Wiley, 1984; 202 pp, $19.95 (P).

Based on a course given to British "sports science" students, this book shows how mathematics (through calculus) can explain known results and suggest optimal strategies in a wide variety of sports, from Olympic events to darts.

Jean, Roger V., Mathematical Approach to Pattern and Form in Plant Growth, Wiley, 1984; ix + 222 pp.

"Phyllotaxis" refers to the positioning of the leaves of a plant around the stem. This state-of-the-art treatment of phyllotaxis brings together descriptive, explanatory, and empirical mathematical models of the last decade. Its first chapter is a reprint of *UMAP* Module 571, "The use of continued fractions in botany." Remaining chapters include exercises and research projects, and the mathematics used ranges from the Chinese remainder theorem to partial differential equations and computer simulation.

Wussing, Hans, The Genesis of the Abstract Group Concept: A Contribution to the History of the Origin of Abstract Group Theory, MIT Pr, 1984; 331 pp.

"We are not concerned with the historical manifestation of some logical development but with the logical manifestation of the historical development." So saying, Wussing goes on to emphasize that abstract group theory has three equally important historical roots: theory of algebraic equations, number theory, and geometry. The book is a well-researched, outstanding achievement in history of mathematics; and this English translation is welcome even 15 years after the original German edition.

Edwards, Harold M., Galois Theory, Springer-Verlag, 1984; xiii + 152 pp.

Masterful and thorough exposition of Galois theory, based on Galois' memoir on solvability of equations by radicals (a translation is included). Exercises abound but "are not essential": "The only proofs that are relegated to the exercises are those that I believe to be too easy, or too much like other proofs already covered, to spend time on in the text."

NEWS & LETTERS

MATHEMATICS IN THE NEWS

Discoveries in space, breakthroughs in biological and chemical research, clues to subatomic particles, oceanic explorations, feats of computer technology -- all are reported without apology to the public in newspapers and magazines. But mathematical discoveries? That's another matter.

These are exciting times for mathematicians, yet little of that excitement is transmitted to the public. Lee Dembart, of the Los Angeles Times, is a refreshing exception to the stereotypical reporter -- he is fascinated by mathematicians and mathematics, and has written news stories and editorials to convey his message. On November 28, 1984, he wrote:

"For a science reporter there is nothing harder to sell a reader -- or for that matter, an editor -- than a story about an important development in mathematics. For some reason, not altogether clear, many otherwise educated people delight in rolling their eyes and proclaiming proudly, 'I don't know anything about mathematics.' Would they be so proud of their ignorance in literature, for example, or art?.."

On January 15, 1985, Dembart was again on the editorial page of the L.A. Times:

"The nation's mathematicians gathered in Anaheim last week for their annual meeting, which, like most academic confabulations, was a mixture of business and pleasure, an opportunity to hear new work and see old friends.

These are heady times for mathematics. In the last five years the pace of discovery has been phenomenal.. No one knows why things have proceeded so quickly. Some people guess that the large increase in mathematics funding in the 1960's and the consequent increase in the number of mathematicians, had a lot to do with it. They worry that cutbacks now will have the opposite effect on the future.. For weeks before the meeting, rumors had swept the mathematics community that Louis deBranges.. would announce a proof of the Riemann Hypothesis, widely regarded as the richest and most important outstanding problem in mathematics today.. There was also animated discussion about an assertion that.. Hideya Matsumoto has announced [its] proof.. in Paris.

As a group, the mathematicians are among the most fascinating people we know. They have a quirky turn of mind that makes them interested in mathematics and at the same time makes them interested in unusual things that you'd otherwise never hear of.. Besides being delightful, mathematics is also crucial to progress in science and technology. It is one of the most productive ways in which public money can be spent."

Mathematics needs more than one dedicated reporter to inform the public; perhaps this will change soon. The Joint Policy Board for Mathematics, an action committee of the MAA, AMS, and SIAM is seriously considering the creation of a public information office which would serve to receive and disseminate information to the media about all matters of importance to the mathematical community.

1984 W.L. PUTNAM PROBLEMS

A-1. Let A be a solid $a \times b \times c$ rectangular brick in three dimensions, where $a, b, c > 0$. Let B be the set of all points which are at a distance of at most one from some point of A (in particular, B contains A). Express the volume of B as a polynomial in a, b, and c.

A-2. Express

$$\sum_{k=1}^{\infty} \frac{6^k}{(3^{k+1} - 2^{k+1})(3^k - 2^k)}$$

as a rational number.

A-3. Let n be a positive integer. Let a, b, x be real numbers, with $a \neq b$, and let M_n denote the $2n \times 2n$

matrix whose (i, j) entry m_{ij} is given by

$$m_{ij} = \begin{cases} x \text{ if } i = j, \\ a \text{ if } i \neq j \text{ and } i + j \text{ is even}, \\ b \text{ if } i \neq j \text{ and } i + j \text{ is odd}. \end{cases}$$

Thus, for example,

$$M_2 = \begin{pmatrix} x & b & a & b \\ b & x & b & a \\ a & b & x & b \\ b & a & b & x \end{pmatrix}.$$

Express $\lim_{x \to a} \dfrac{\det M_n}{(x - a)^{2n-2}}$ as a polynomial in a, b, and n, where $\det M_n$ denotes the determinant of M_n.

A-4. A convex pentagon $P = ABCDE$, with vertices labeled consecutively, is inscribed in a circle of radius 1. Find the maximum area of P subject to the condition that the chords AC and BD be perpendicular.

A-5. Let R be the region consisting of all triples (x, y, z) of nonnegative real numbers satisfying $x + y + z \leq 1$. Let $w = 1 - x - y - z$. Express the value of the triple integral

$$\iiint_R x^1 y^9 z^8 w^4 \, dx \, dy \, dz$$

in the form $a!b!c!d!/n!$, where a, b, c, d, and n are positive integers.

A-6. Let n be a positive integer, and let $f(n)$ denote the last nonzero digit in the decimal expansion of $n!$. For instance, $f(5) = 2$.

(a) Show that if $a_1, a_2, \ldots, a_k$ are *distinct* nonnegative integers, then $f(5^{a_1} + 5^{a_2} + \ldots + 5^{a_k})$ depends only on the sum $a_1 + a_2 + \ldots + a_k$.

(b) Assuming part (a), we can define $g(s) = f(5^{a_1} + 5^{a_2} + \ldots + 5^{a_k})$, where $s = a_1 + a_2 + \ldots + a_k$. Find the least positive integer p for which $g(s) = g(s + p)$, for all $s \geq 1$, or else show that no such p exists.

B-1. Let n be a positive integer, and define $f(n) = 1! + 2! + \ldots + n!$. Find polynomials $P(x)$ and $Q(x)$ such that $f(n+2) = P(n)f(n+1) + Q(n)f(n)$, for all $n \geq 1$.

B-2. Find the minimum value of

$$(u - v)^2 + \left(\sqrt{2 - u^2} - \frac{9}{v}\right)^2$$

for $0 < u < \sqrt{2}$ and $v > 0$.

B-3. Prove or disprove the following statement. If F is a finite set with two or more elements, then there exists a binary operation * on F such that for all x, y, z in F,

(i) $x*z = y*z$ implies $x = y$ (right cancellation holds), and

(ii) $x*(y*z) \neq (x*y)*z$ (no case of associativity holds).

B-4. Find, with proof, all real-valued functions $y = g(x)$ defined and continuous on $[0, \infty)$, positive on $(0, \infty)$, such that for all $x > 0$ the y-coordinate of the centroid of the region

$$R_x = \{(s, t) \mid 0 \leq s \leq x, \ 0 \leq t \leq g(s)\}$$

is the same as the average value of g on $[0, x]$.

B-5. For each nonnegative integer k, let $d(k)$ denote the number of 1's in the binary expansion of k (for example, $d(0) = 0$ and $d(5) = 2$). Let m be a positive integer. Express

$$\sum_{k=0}^{2^m - 1} (-1)^{d(k)} k^m$$

in the form $(-1)^m a^{f(m)} (g(m))!$, where a is an integer and f and g are polynomials.

B-6. A sequence of convex polygons $\{P_n\}$, $n \geq 0$, is defined inductively as follows. P_0 is an equilateral triangle with sides of length 1. Once P_n has been determined, its sides are trisected; the vertices of P_{n+1} are the interior trisection points of the sides of P_n. Thus, P_{n+1} is obtained by cutting corners off P_n, and P_n has $3 \cdot 2n$ sides. (P_1 is a regular hexagon with sides of length 1/3.)

Express $\lim_{n \to \infty} \text{Area}(P_n)$ in the form $\sqrt{a}/b$, where a and b are positive integers.

Publications by the

Mathematical Association of America

1529 Eighteenth St., NW
Washington, DC 20036
202 387-5200

New Mathematical Library #30 . . .

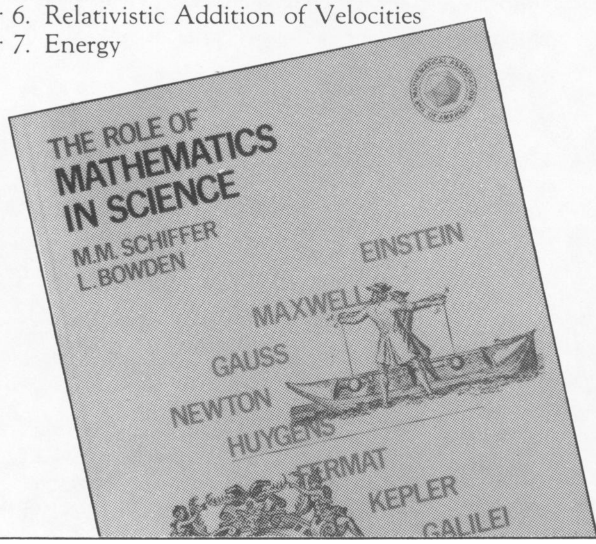
The Role of Mathematics in Science,
by Leon Bowden and M. M. Schiffer
220 pp. Paperbound.
List: $14.00   MAA Member: $11.00

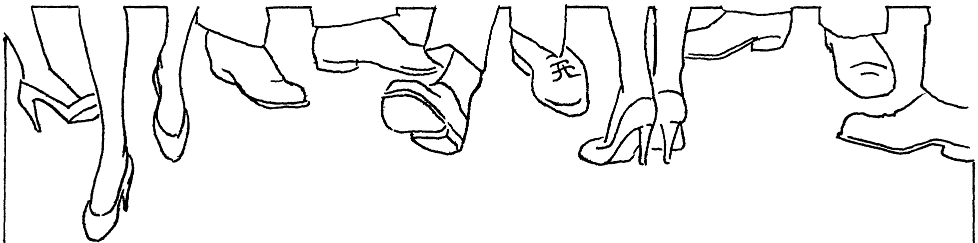
This book is based on a series of lectures given over a period of years to high school mathematics teachers. The lectures focused on topics in mechanics, questions of population growth, probability and other uses of exponential functions, optics and application of matrices to relativity theory.

The aim of the book is to illustrate the power and elegance of mathematical reasoning in science with some examples ranging from the work of Archimedes to that of Einstein. The book starts with problems of the lever, the mirror and the growth of populations and ends up with problems of space travel and atomic energy.

This excellent book will appeal to the advanced high school student, to the undergraduate and to their teachers. A sampling of the Table of Contents should inspire you to order your copy now.

Table of Contents





Random Walks and Electric Networks

by J. Laurie Snell and Peter Doyle

xiii + 159 pages. Hardbound

MAA Member: $16.00 List: $21.00

In this newest addition to the Carus Mathematical Monographs, the authors examine the relationship between elementary electric network theory and random walks, at a level which can be appreciated by the able college student. We are indebted to them for presenting this interplay between probability theory and physics in so readable and concise a fashion.

Central to the book is Polya's beautiful theorem that a random walker on an infinite street network in d-dimensional space is bound to return to the starting point when d = 2, but has a positive probability of escaping to infinity without returning to the starting point when d = 3. The authors interpret this theorem as a statement about electric networks, and then prove the theorem using techniques from classical electrical theory. The techniques referred to go back to Lord Rayleigh who introduced them in connection with an investigation of musical instruments.

In Part I the authors restrict themselves to the study of random walks on finite networks, establishing the connection between the electrical concepts of current and voltage and corresponding descriptive quantities of random walks regarded as finite state Markov chains. Part II deals with the idea of random walks on infinite networks.

Table of Contents

Part I: **Random Walks on Finite Networks**
- Random Walks in One Dimension
- Random Walks in Two Dimensions
- Random Walks on More General Networks
- Rayleigh's Monotonicity Law

Part II: **Random Walks on Infinite Networks**
- Poyla's Recurrence Problem
- Rayleigh's Short-Cut Method
- The Classical Proof of Polya's Theorem
- Random Walks on More General Infinite Networks

ORDER FROM

The Mathematical Association of America

1529 Eighteenth St., N.W.

Washington, D.C. 20036

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

MATHEMATICS MAGAZINE VOL. 58, NO. 2, MARCH 1985