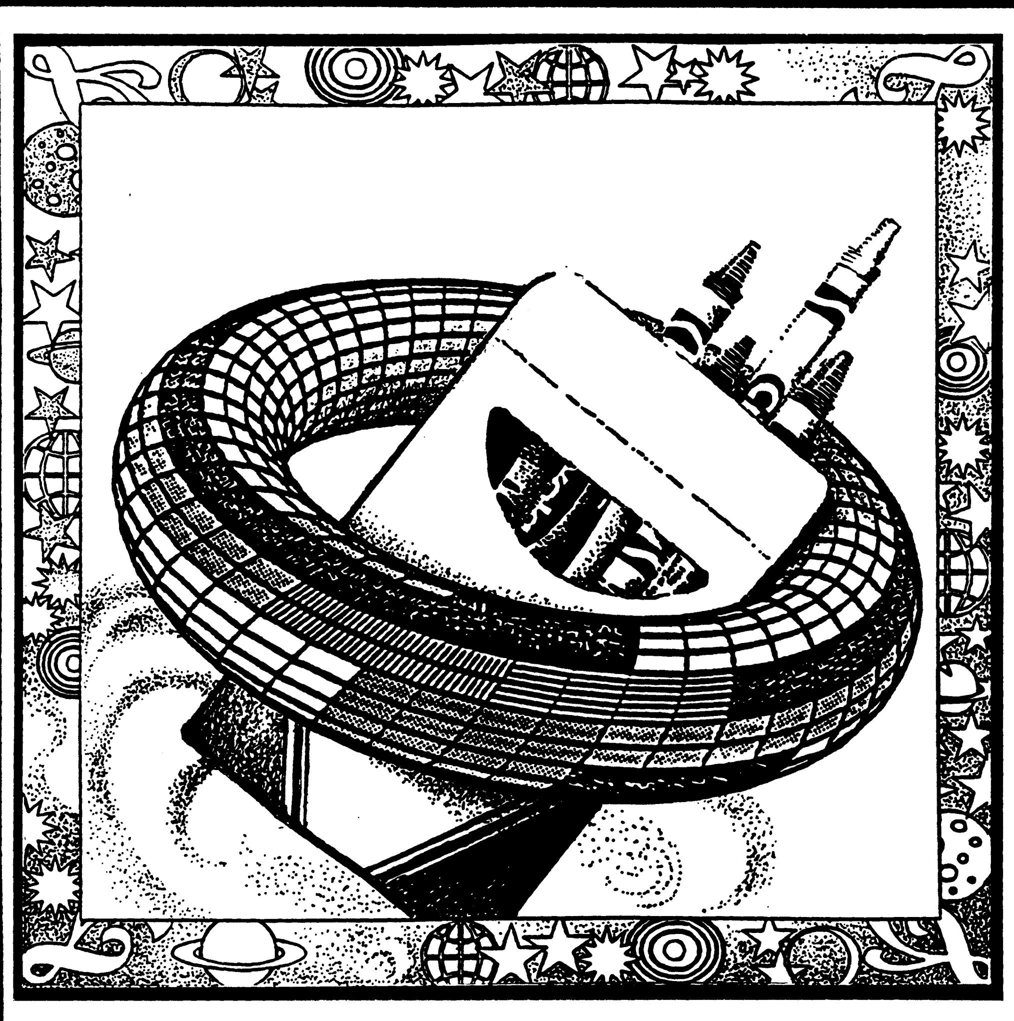
Vol. 66 No. 4 October 1993

MATHEMATICS MAGAZINE



- Coloring Ordinary Maps, Maps of Empires, and Maps of the Moon
- On Tangents
- Coordinate Systems and Primitive Maps

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

Send new manuscripts to: Martha Siegel, Editor, Mathematics Magazine, Towson State University, Towson, MD 21204. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and two copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added. Do not use staples.

Cover art by Carolyn Westbrook.

AUTHORS

Joan P. Hutchinson began seriously coloring maps in 1973, after receiving her Ph.D. from The University of Pennsylvania. She has taught primarily at Smith College and at Macalester College. Her many enjoyable sabbaticals at and visits to other colleges and universities, along with her interest in mountain climbing and backcountry skiing, have made her appreciate the importance of well-colored maps.

Hugh Thurston is old enough to have used Euclid as a textbook. His supervisor at Cambridge University, A. S. Besicovitch, showed him the delight of counter examples for plausible but wrong beliefs. He has always been a bit disconcerted by the superficial way geometry texts treat tangents and the nongeometrical way that calculus texts treat them. Put these factors together, and the result is the article in this issue.

Vol. 66 No. 4 October 1993

MATHEMATICS MAGAZINE

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

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

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

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

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Coloring Ordinary Maps, Maps of Empires, and Maps of the Moon[1]

JOAN P. HUTCHINSON
Macalester College
St. Paul, MN 55105

1. Introduction to Map Coloring

The following statement is *not* true:
"Every map can have one of four colors assigned to each country so that every pair of countries with a border arc in common receives different colors."

But isn't this the statement of the famous Four Color Theorem that was proved about 15 years ago using lots of computer checking? Has a flaw been found in that massive piece of work?

FIGURE 1 shows a small example of a map that needs five colors if every pair of adjacent countries is to receive different colors. The important feature is that one country (#5) is a disconnected country of two regions.

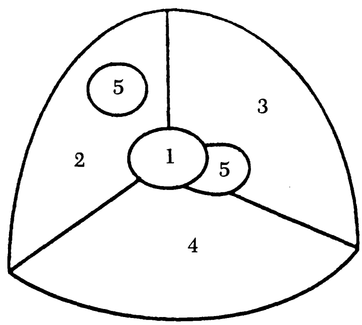


FIGURE 1
A planar map needing five colors.

Maps having disconnected countries are certainly possible; an example is the map of North America. This raises the following cartography question: Can today's map of the world be colored with four colors so that all pieces of each country receive the same color and so that no two different countries with a border arc in common receive the same color?

In any case, FIGURE 1 is not a counterexample to the Four Color Theorem; here's what that famous theorem really says [1].

FOUR COLOR THEOREM. *Every map drawn in the plane (or on the surface of the sphere) can have one of four colors assigned to each* ***connected region*** *so that every pair of regions with a border arc in common receives different colors.*

[1]This article is based on talks given at the Pi Mu Epsilon Conference at St. John's University, Collegeville, MN in March, 1990. This research was supported in part by an NSF Visiting Professorships for Women grant, #RII-8901458, and by the University of Washington, Seattle, WA.

This result was first conjectured in 1852 by Francis Guthrie; a proof was published in 1879 by A. B. Kempe, but, 11 years later, P. J. Heawood found a fatal flaw in that proof. Finally in 1976 the theorem was proved by K. Appel and W. Haken, although in an unusual manner. (A thorough history of the problem can be found in [**5**].) Briefly, Appel and Haken's proof consists of showing that every planar map contains one of a list of at most 2,000 configurations, and that each configuration admits a reduction, allowing a proof by induction. Although the numbers involved are unusually large, the most unusual (and controversial) part of the proof is that about 1,200 hours of computer time were used to generate the list of 2,000 configurations and to check that colorings on these admit the necessary reduction. Thus the proof depends upon electricity—that is, upon a physical experiment—and that dependence is most unusual in the history of mathematics. Both the theoretical and the computational aspects of the proof have been carefully checked, but mathematicians would very much like to see a simpler or more "natural" proof. Several papers ([**2**, **12**, **28**]) contain detailed and substantive analyses of the proof.

The map in FIGURE 1 might also be a representation of political boundaries in which area #5 represents an empire, two countries joined together. In the past, when countries were annexed to form empires, it was the custom of mapmakers to color all parts of an empire with the same color; thus maps of empires might require more than four colors. The mathematics of empire coloring was first studied by Heawood over a hundred years ago. In this paper we survey some of his work and more recent map-coloring variations by G. Ringel. Then one of their results will be applied to a modern problem in the testing of printed circuit boards.

2. Coloring Maps of Empires

When Heawood both found a flaw in Kempe's argument and discovered that he could not repair it, he invented some generalizations about map colorings that, to a certain extent, he could solve. His research began the field of topological graph theory as studied today. First he investigated the empire-coloring problem: If maps are made up of countries united into empires, then how many colors are needed to color such maps provided that all countries in an empire receive the same color and that empires with a border in common receive different colors? Heawood proved that if every empire consists of at most M connected regions (called an *M-pire* by B. Jackson and G. Ringel [**16**]), then the map can be colored with at most $6M$ colors.

We shall present a proof of Heawood's upper bound, but first, we ask whether $6M$ is the best possible bound. For $M = 1$, it is not: If empires consist only of single connected regions, then four colors are enough by the Four Color Theorem. But for $M = 2$, Heawood found an example of 12 empires, each consisting of two regions and so forming a 2-pire, such that every pair of empires has a border in common. Thus 12 colors are needed; however, Heawood's example was sufficiently irregular that he was unable to generalize it for $M > 2$.

After an introduction to some facts and techniques of graph theory, we'll prove Heawood's theorem and then summarize recent work of Jackson and Ringel [**16**] that settles the question of the best possible coloring bound for M-pires with $M > 1$.

We make a change now from maps to graphs as is customary with practitioners in this field. Given a map C drawn in the plane, we form a *planar graph* by creating a vertex for each connected region of C, and by joining two vertices by an edge if the corresponding regions have a border arc in common. This graph is called planar because it can be drawn in the plane without edge crossings. FIGURE 2 shows the planar graph that results from the map in FIGURE 1; notice that a vertex is created also for the outer, infinite region of the map. And given a planar graph, we could construct a map with countries having the prescribed borders in common.

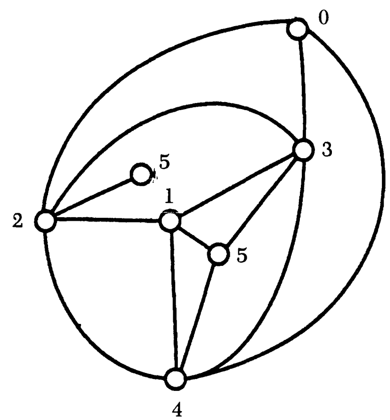


FIGURE 2

The planar graph derived from the map of FIGURE 1.

Previously we wanted to color regions of a map; now we color vertices of the graph. For k a positive integer, a graph is said to be *k-colored* (or *k-colorable*) if each vertex is (or can be) assigned one of k colors so that every pair of vertices that are joined by an edge receives different colors. Why bother with this change from maps to graphs and then coloring vertices rather than regions? This change certainly allows us to use less paint and, more to the point, this convention is better suited to current algorithmic approaches to graph theory problems, allowing for efficient storage and usage of graphs by computer programs.

Here then in the language of graph theory is Heawood's M-pire problem: Prove that every graph that is derived from an M-pire map can be $6M$-colored so that (in addition) all vertices corresponding to the same empire receive the same color.

Here is the first, crucially important, tool of the trade. A *face* in a planar graph G, drawn in the plane, is a connected region in the plane minus the edges and vertices of G.

EULER'S FORMULA. *If G is a connected planar graph, drawn in the plane with n vertices, e edges, and f faces, then*

$$n - e + f = 2.$$

For a proof, see [3, 7]. An immediate consequence of this formula is that the number of edges in a planar graph is limited. Suppose we count the number of edges bordering each face: Let e_i denote the number of edges on the ith face (counting an edge twice if both sides border on the same face). Then we get

$$2e = e_1 + e_2 + \dots + e_f \geq 3f,$$

since every edge is counted exactly twice in the sum of the e_i, and since each face has at least three bordering edges (provided the graph contains neither loops nor multiple edges). Thus

$$2e/3 \geq f.$$

Applying Euler's Formula,

$$n - e + 2e/3 \geq 2,$$

or

$$e \leq 3n - 6. \tag{1}$$

Similarly, if the ith vertex has $\deg_i$ incident edges, ($\deg_i$ is called the *degree* of the vertex), and we sum the degrees of all vertices, we get

$$2e = \deg_1 + \deg_2 + \cdots + \deg_n$$

since each edge is counted twice, once at each end-vertex. Then, by (1)

$$\deg_1 + \deg_2 + \cdots + \deg_n \leq 6n - 12. \tag{2}$$

We are almost ready to $6M$-color M-pires. Notice that our goal is to color certain graphs with an extra constraint imposed because the graph comes from an M-pire: Certain vertices are supposed to receive the same color. That's a bit unnatural to a graph colorer. Most graph-coloring theorems and algorithms assume that a graph just should be colored so that no two adjacent vertices receive the same color; this extra constraint is a bother. So we'll get rid of it appropriately.

The idea is a simple one. Recall that we began with a planar M-pire map, and from this formed a planar graph—call it G—with a vertex for each connected region. From G we form a new graph G^* by identifying the set of vertices of G corresponding to an empire into one vertex of G^* and removing any multiple edges. We call G^* the *M-pire graph* associated with an M-pire map. It is the graph G^* that we want to color (with as few colors as possible) since all countries of the same M-pire will necessarily receive the same color and all map adjacencies have been recorded in G^*. But G^* most likely will no longer be a planar graph.

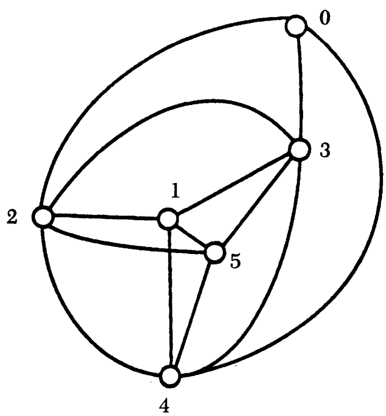


FIGURE 3
G^* formed from G of FIGURE 2.

We shall prove that G^* can always be $6M$-colored. Suppose G^* has n^* vertices and e^* edges. Since each vertex of G^* comes from at most M vertices of G, a planar graph with, say, n vertices and e edges, $Mn^* \geq n$, and $e^* \leq e$. Thus in G^* we have from (2)

$$\begin{aligned} \deg_1 + \deg_2 + \cdots + \deg_{n^*} = 2e^* \leq 2e &\leq 6n - 12 \\ &\leq 6Mn^* - 12. \end{aligned}$$

Hence the average degree of vertices in G^* is bounded:

$$\begin{aligned}(\deg_1 + \deg_2 + \cdots + \deg_{n^*})/n^* &\leq 6M - 12/n^* \\ &< 6M,\end{aligned}$$

and so G^* contains a vertex of degree at most $6M - 1$.

THEOREM 1 (Heawood) [15]. *Every M-pire graph G^* can be 6M-colored.*

Proof. By induction on n^*. If $n^* \leq 6M$, then G^* can easily be $6M$-colored by placing a different color on each vertex. Assume the theorem is true for every M-pire graph with fewer then n^* vertices, and let G^* be an M-pire graph with $n^* > 6M$ vertices.

Find a vertex v of G^* of degree at most $6M - 1$; delete v and all its incident edges. The resulting graph is still an M-pire graph (why exactly is this true?) and so, by induction, can be $6M$-colored. Since v in G^* is adjacent to at most $6M - 1$ different colors, there is a color available to place on v. Thus G^* is $6M$-colored.

COROLLARY 1. *Every (ordinary) planar graph (or map) can be 6-colored.*

But are $6M$ colors sometimes necessary (when $M > 1$)? Since 1980, examples have been known of 3-pires and 4-pires that need 18 and 24 colors respectively; colorful pictures of these are presented in [8]. In 1983 B. Jackson and G. Ringel were able to settle the whole question as follows.

M-PIRE THEOREM [16]. *For every $M > 1$ there is an M-pire graph that requires 6M colors. In fact, the graph consisting of 6M mutually adjacent vertices is an M-pire graph.*

The graph of $6M$ vertices with every pair of vertices joined by an edge is called the *complete graph* on $6M$ vertices and is denoted by K_{6M}. K_{6M} is $6M$-colorable, but cannot be colored with fewer colors. The M-pire theorem can be restated as follows: There is a planar map (or graph) that consists of $6M$ empires, each with at most M regions, such that every empire has a border in common with every other. Jackson and Ringel proved this theorem by constructing symmetrical maps using the theory of "current graphs" as developed by Ringel and J. W. T. Youngs for a solution of another problem invented by Heawood (this problem will be discussed in Section 5). Thus the problem of coloring M-pire maps has been completely and beautifully settled when $M > 1$. A thorough exposition of this subject is given in [14].

3. Coloring Maps of the Moon

Ringel suggested the following variation on the empire coloring problem. Suppose the Moon were colonized and we want to color a map of the Earth and the Moon so that

1. adjacent regions on Earth or on the Moon receive different colors, and
2. a country on Earth and its lunar colony receive the same color.

Suppose now that there are *no empires* on the Earth or Moon; each country and colony is a connected region.

But this is just a 2-pire problem. Let G_e and G_m be the corresponding planar graphs of the Earth and Moon maps; side-by-side they form one planar graph that comes from a 2-pire map (see FIGURE 4). When each vertex of G_m is identified with its country-vertex in G_e, the resulting graph G^* is a 2-pire graph and so can be 12-colored by Theorem 1.

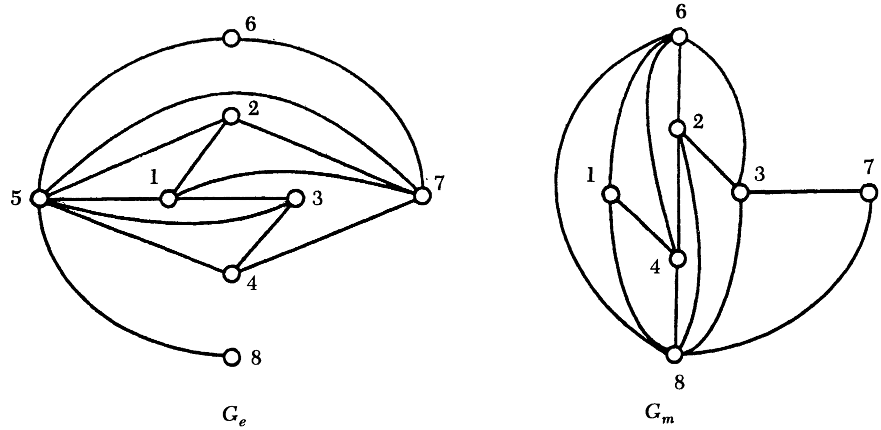


FIGURE 4
K_8 as an Earth/Moon graph.

A graph G^* is said to be an *Earth/Moon graph* (or to have *thickness* 2) if it can be divided into two planar graphs by making two copies of the vertex set of G^*, and then assigning each edge of G^* to one of the two copies so that two planar graphs result. Think of one planar graph as representing the terrestrial countries and the other the lunar colonies. More generally, a graph is said to have *thickness t* if its edges can be partitioned into t, but not fewer, planar graphs. This thickness parameter is studied by graph theorists both for its relevance to general graph theory and for applications in computer science. In the next section, we'll consider an application of thickness-2 graphs.

So what's new in Ringel's question about coloring Earth/Moon graphs if we already know they can be 12-colored? Again the important point is whether or not 12 colors are needed. Notice that the definitions of an Earth/Moon graph and of a 2-pire graph are not the same, and consequently a 2-pire graph might not be an Earth/Moon graph. If G^* is a 2-pire graph, some vertices can be split into two so that a planar graph, say, G results. But it may not be possible to divide G into two separate planar graphs, each containing one copy of each vertex, as is required for an Earth/Moon graph. For example, the complete graphs K_9, K_{10}, K_{11}, and K_{12} are all 2-pire graphs, but it is known that none of them is an Earth/Moon graph. They all have thickness 3 because of the following result. (For a summary of the several proofs involved, see [4].)

THEOREM (Beineke, Harary, Alekseev, Gonchakov, Vasak, Mayer). *The thickness of the complete graph K_n is given by*

$$t(K_n) = \begin{cases} \lfloor (n+7)/6 \rfloor, & \text{if } n \neq 9, 10, \\ 3, & \text{if } n = 9, 10. \end{cases}$$

FIGURE 4 shows how K_8 arises from two planar subgraphs. K_9 cannot be so created, but if the Earth had two moons, then it would be possible to find maps on the three spheres that would unite to form a K_9. Thus Earth/Moon graphs will need at least 8 colors, but since the graph K_{12} has thickness 3, we can't conclude that 12 colors are needed for Earth/Moon graphs.

The least number of colors that we need to have on hand so that we can color every Earth/Moon graph is called the *chromatic number* of Earth/Moon graphs (and of thickness-2 graphs). As reported in [8], T. Sulanke found a map of the Earth and Moon that needs 9 colors, although it does not come from K_9; FIGURE 5 shows a schematic version of the corresponding Earth/Moon graph. It is a worthwhile exercise to check that this graph can be 9-colored, but not 8-colored, and that its edges can be separated into two planar subgraphs.

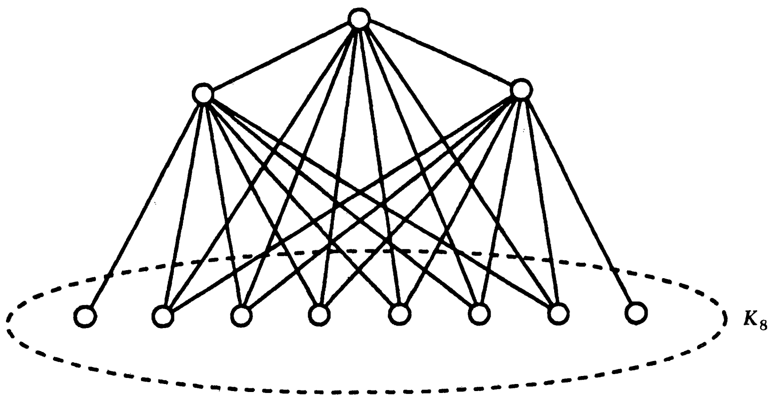


FIGURE 5
A 9-chromatic Earth/Moon graph.

Open Problem. What is the chromatic number of Earth/Moon graphs? 9? 10? 11? 12?

Progress would be made if any of these possibilities could be eliminated. In general the chromatic number of thickness-t graphs is not known exactly. By the preceding theorem and Theorem 1, the chromatic number is one of $6t-2$, $6t-1$, or $6t$ for $t>2$.

4. An Application of Earth/Moon Coloring to the Testing of Printed Circuit Boards

We turn to an application that, perhaps surprisingly, uses the coloring results of the previous section to develop an efficient procedure for the testing for certain errors in printed circuit boards. This example comes from three researchers at AT&T Bell Laboratories in Murray Hill and Whippany, NJ [9].

It is not hard to think of potential applications of the thickness parameter to electronics. Suppose you have a system of electrical units, certain pairs of which should be connected electrically. Such a design has a graph naturally associated with it. A goal could be to develop a layout for that graph with, say, electrical wires or solder joining the units. To avoid crossings and unwanted electrical connections, the joining should be done on multiple planar layers. Or, as in the design and fabrication of VLSI chips, connections should be etched onto separate layers of silicon.

For this application we focus on printed circuit boards (PCBs), small electrical circuits that are widely used in motorized, electrical, and electronic equipment. (For examples and illustrations, see [24].) We postulate a simple, but realistic, mathematical model for these circuit boards. We assume that a PCB consists of electrical units placed on a rectangular array with some of the units joined electrically along horizontal and vertical lines. Typically an array consists of a 100×100 grid with about 500 electrically connected components (called *nets*). And each net has a simple structure: Every pair of units in one net is joined by a unique conductor path; this structure is known to graph theorists as a *tree*. An example is shown in FIGURE 6.

We focus on one specific problem, that of finding certain extraneous and unwanted connections that may have mistakenly occurred during the manufacturing process. In practice a design is set for a PCB; thousands or even millions of these inexpensive PCBs are made; and as they roll off the production line, a quick and accurate test is needed to see whether there are errors on the board. If so, the board is discarded; no attempt is made to find and correct the error. (As an example of what happens when quality control is not effective, a front-page article in the *New York Times* **[20]** reported recently on errors in a \$12 circuit board that caused a delay in manufacturing 2,500 Air Force rockets, each worth about a million dollars.)

Here is a mathematical statement of the problem we consider.

Problem. Determine whether any extraneous horizontal or vertical conductor paths have been introduced in the manufacturing process, connecting two nets that should not be electrically connected (such a connection is known as a *short circuit*).

In practice, erroneous horizontal and vertical paths are the most common short circuit, and so this problem focuses on such mistakes.

There is a simple answer to the problem: Check all pairs of nets to see if any is incorrectly connected. If there are n nets, such a check requires $n(n-1)/2$ tests and is by far too slow a procedure. For example, with 500 nets, about 125,000 tests would be needed. The better solution presented in **[9]** necessitates only 11 checks per board, regardless of the number of nets. We also show how to reduce the checking to a mere four per board.

Not surprisingly, we make a graph from the intended PCB. The *PCB graph* is defined to have

1. a vertex for each net of the (correct) PCB, and
2. two vertices joined by an edge if there is a horizontal or vertical line going through the corresponding two nets and passing through no intermediate net.

FIGURE 7 shows the PCB graph for the example of FIGURE 6—for the moment, ignore the labels on the vertices. In this figure the vertices are positioned to correspond to the position of the nets in the PCB of FIGURE 6.

Perhaps a better name for this graph would be a "graph of possible mistakes." Imagine that in the fabrication process two passes are made over the PCB, one pass creating all horizontal connections and the second creating all vertical. The problem of concern is that too much connecting might be done; the fabrication process might not shut off correctly and so might connect more than is required. Thus in the PCB graph two vertices are joined if it is possible for the fabrication machine to mistakenly join the corresponding nets by a direct horizontal or vertical connection. But what about the condition of passing through no intermediate net? Look at FIGURE 6: Suppose nets x and y were mistakenly connected by a horizontal line. Then the connection would necessarily connect x with z and z with y. In other words, if we check that x and z are not connected and that z and y are also not connected, then we know for sure that x and y are not mistakenly connected. Hence there is no need for an x-to-y edge in the PCB graph.

The key observation now is that the PCB graph has thickness 2. Its edges can be divided into two planar graphs, one with all edges corresponding to possible vertical connections and one with edges from horizontal connections. That observation and the fact that a thickness-2 graph can be 12-colored leads to the following checking algorithm.

PCB-checking Algorithm. Given a plan for a PCB and the corresponding PCB graph G:

1. 12-color G, and transfer this coloring to the nets of the plan.
2. From the plan for the PCB, construct 12 external conductor tree-structures, called "supernets," so that when each supernet is attached to the PCB, all nets in the same color class become electrically connected.
3. Check all pairs of these supernets to see if any two of these are (mistakenly) electrically connected.

Here are a few more words of explanation. Consider all nets that receive one color, say, color #1. Since no two of the corresponding vertices in G are joined by an edge, no two of these nets need to be tested for a mistaken connection. A "supernet," such as those shown in FIGURE 8, is some simple, 3-dimensional electrical connection (shaped perhaps like an octopus or an n-pus) that attaches to the circuit board and electrically connects a set of nets.

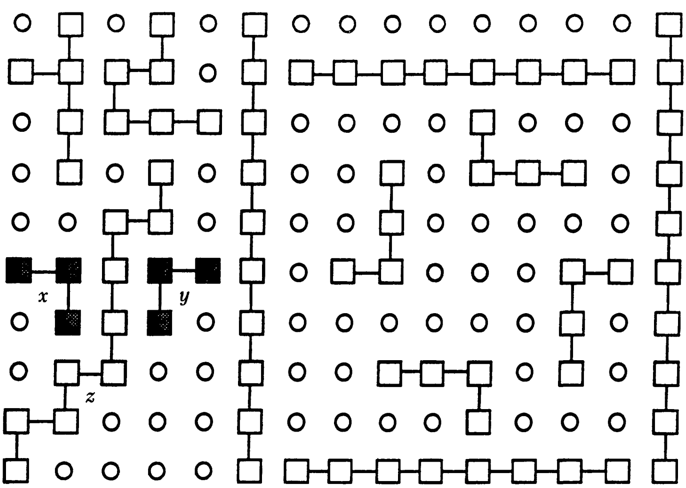


FIGURE 6
A printed circuit board model.

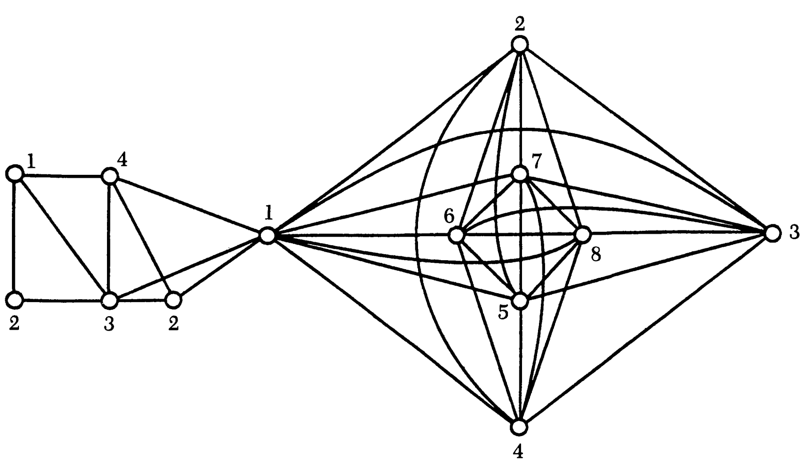


FIGURE 7
The PCB graph from FIGURE 6 with its vertices 8-colored.

In step (3) an electrical probe can check each pair of supernets. If, say, supernets #1 and #2 are electrically connected, then some net colored 1 and some net colored 2 are mistakenly connected (and we throw the board out). But if not, then no net of color 1 is connected to a net of color 2, and we proceed to test other pairs of supernets. With 12 supernets, there are $12 \cdot 11/2 = 66$ pairs that might need to be tested. Granted, the creating of these supernets takes some time and money, but recall that we are imagining that a machine (or several) is creating thousands of PCBs following one master plan so that the checking supernets can be used repeatedly as the PCBs roll off the line.

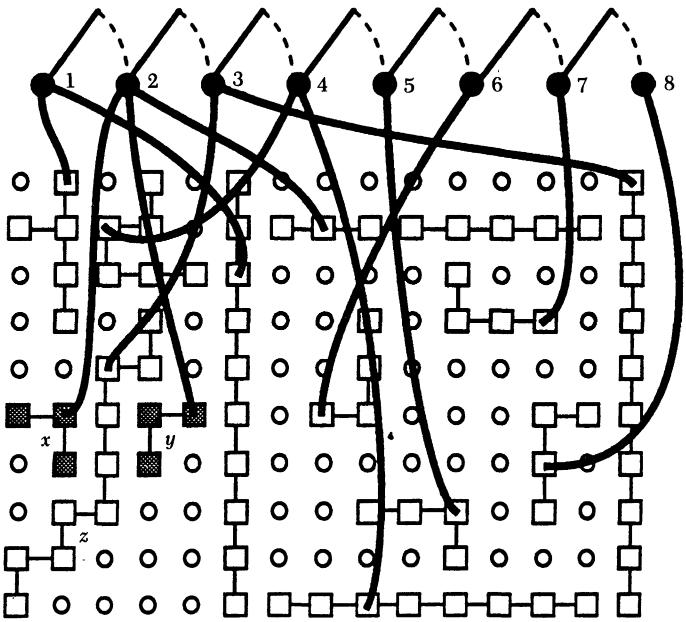


FIGURE 8
The printed circuit board of FIGURE 6 with supernets attached.

With another gadget we can reduce the number of tests even more. Suppose we test supernet #1 versus #2. If they are electrically connected, we throw the board out, but if not, we add a "gate", some simple electrical connection to make these two supernets electrically connected. Next we test this united #1–2 supernet versus supernet #3. If there is an electrical connection, then a net in color class 3 is mistakenly connected to one of color class 1 or 2. We don't care which; we just throw the board out. But if not, we connect supernet #3 up with #1 and #2. Continuing with this checking and connecting, we see that, with a correct board, after only 11 connections of supernets and 11 tests, we detect that we have a board free of the kinds of mistakes for which we were checking. FIGURE 7 shows the PCB graph from FIGURE 6 together with an 8-coloring of its vertices. Then FIGURE 8 shows the corresponding supernets and the connecting devices to reduce the checking to 11 steps.

Allen Schwenk (personal communication) has recently pointed out how to further reduce the number of checks to four, with additional gadgetry. Take the existing supernets, numbered $1, 2, \ldots, 12$, and think of these numbers, expressed in binary, each with exactly four binary digits (called bits). Make a supersupernet that connects the supernets labeled with numbers beginning with a 0 bit, and make a similar supersupernet that connects the supernets labeled with numbers beginning with a 1 bit. For the first test, check if these two supersupernets are electrically connected. If not, create two supersupernets, one connecting the supernets with a 0 in the second bit and one connecting the supernet with a 1 there. For the second test, check if these two are connected. Do the same creation and test for the third and fourth bits. If there is an erroneous connection, we will detect it: If there is a connection between a net colored i and a net colored j, the binary representation of i and j differ in some bit, and thus electricity would flow when the two supersupernets for that bit were tested. Notice that in these tests we have not detected all possible errors, such as those when too few connections are made within a net or when some extra zigzag-like connections are made between nets, but this approach does solve the problem of concern to the AT&T researchers.

5. Coloring Ordinary Maps and Empire Maps on Surfaces

Now we return to the realm of map coloring. Recall the first example of a map that needed more than four colors in FIGURE 1. There's another way in which maps might need more colors, and Heawood thought of this one too. Suppose a map were drawn on the surface of a torus (a donut) or a 4-holed torus (like a pretzel). Then the map might need as many as 7 or 10 colors, respectively. Here's some motivation for such topological map drawing and coloring.

Thinking now in terms of graphs, notice that a graph can be drawn in the plane (without edge crossings) if and only if it can be so drawn on the surface of a sphere. But what about graphs that can be drawn in neither place, such as K_5 or any K_n with $n \geq 5$? One trick would be to add bridges or handles to the plane or sphere so that edges can traverse these and so avoid edge crossings. The sphere plus a handle or plus g handles ($g > 0$) is essentially the same as the torus or the g-holed torus, respectively; topologists call this essential similarity a homeomorphism. For our purposes, we use the fact that a graph can be *embedded* (i.e., drawn without edge crossings) in the plane plus g bridges if, and only if, it can be embedded on the g-holed torus, also known as the surface of genus g. The *genus* of a graph is defined to be the least g such that the graph can be embedded on the sphere plus g handles.

For example, it is not hard to see that K_5 embeds on the torus; but with a little more effort one can also embed K_6 and K_7 there. So in particular, graphs that embed on the torus may need as many as 7 colors. FIGURE 9 shows the corresponding map situation, a map of seven mutually adjacent regions. Heawood discovered such a map and proved that every graph on the torus can be 7-colored.

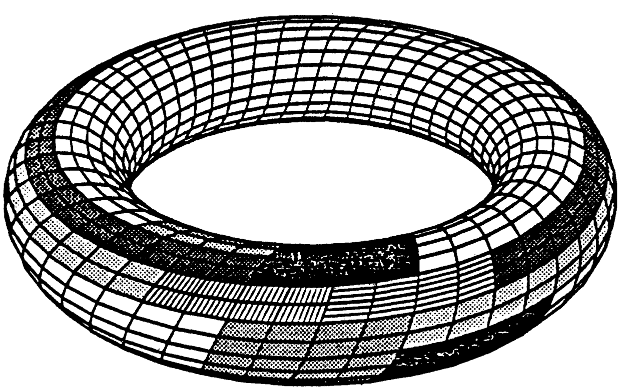


FIGURE 9
Seven mutually adjacent regions on the torus.

He also proved a related result for each surface of positive genus. Let $\chi(g)$ equal the minimum number of colors needed to color every graph that embeds on the sphere plus g handles, $g > 0$.

THEOREM 2 (Heawood) [**15**]. *For* $g > 0$, $\chi(g) \leq \lfloor (1/2)(7 + \sqrt{48g+1}) \rfloor$.

So, for example $\chi(1) \leq 7$, $\chi(2) \leq 8$, $\chi(6) \leq \chi(7) \leq 12$, and $\chi(17) \leq 17$. (Note that substituting $g = 0$ yields $\chi(0) \leq 4$, but Heawood's proof does not cover this case.) A good introduction to the mathematics and history of this problem is found in [**25**].

Why is Heawood's theorem true? Soon (in Theorem 3) we'll give a proof of this and a more general result about empires on surfaces that depends upon a generalization of Euler's Formula to surfaces of positive genus, but first, let's see where Heawood's bound comes from.

EULER-POINCARÉ FORMULA. *If G is a graph, embedded on the g-holed torus with n vertices, e edges, and f faces, then*

$$n - e + f \geq 2 - 2g.$$

If every face is a contractible region, then

$$n - e + f = 2 - 2g.$$

For a proof, see [**11, 29**]. FIGURE 10 shows three embeddings of K_4 on the torus; only the last has the nice property that every face is contractible (or homeomorphic to a planar disk). Another way to think of this property is that equality occurs in the Euler-Poincaré Formula when the embedding uses all the handles and is not really embedded on a sphere with fewer handles.

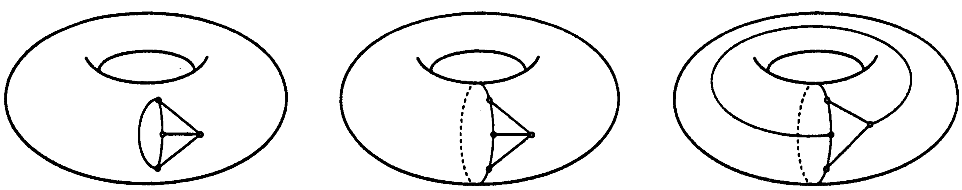


FIGURE 10
Embeddings of K_4 on the torus.

Using the Euler-Poincaré Formula and arguing just as we did for inequality (1) in section 2, we have in all cases that

$$e \leq 3n + 6(g-1).$$

How large a complete graph K_n could we embed on the sphere plus g handles? Such a graph has $n(n-1)/2$ edges and so it must be the case that

$$n(n-1)/2 \leq 3n + 6(g-1). \tag{3}$$

Thus

$$n^2 - 7n + 12(1-g) \leq 0,$$

and by the quadratic formula

$$n \leq (1/2)\left(7 + \sqrt{49 + 48(g-1)}\right)$$

so that

$$n \leq \lfloor (1/2)(7 + \sqrt{48g+1}) \rfloor.$$

Note that from (3) it also follows that

$$g \geq \lceil (1/12)(n-3)(n-4) \rceil. \tag{4}$$

But what is the genus of K_n? Is it the lower bound of line (4)? Equivalently, when $n = \lfloor (1/2)(7+\sqrt{48g+1}) \rfloor$, can K_n be embedded on the sphere plus g handles? If so, then $\lfloor (1/2)(7+\sqrt{48g+1}) \rfloor$ colors are needed for graphs that embed there. Heawood blithely assumed that since the answers to the preceding two questions were YES for the torus, $n=7$ and $g=1$, they must be YES for all larger n and g. His intuition was correct, but there was no proof of these facts until 1968 when the following deep and difficult result was proved principally by Ringel and Youngs, with help from others on a few cases.

MAP-COLOR THEOREM (Ringel, Youngs, Gustin, Guy, Mayer, Terry, Welch) **[23]**. *The genus of* K_n *is* $\lceil (1/12)(n-3)(n-4) \rceil$, *and consequently for* $g>0$, $\chi(g) = \lfloor (1/2)(7+\sqrt{48g+1}) \rfloor$.

Thus the equivalent of the Four Color Theorem was solved first for all surfaces except for the sphere (and without computer help).

But Heawood imagined even more, combining the ideas of maps on surfaces and empires. Suppose an M-pire map is drawn on the sphere plus g handles, and suppose G is the corresponding graph embedded on the same surface. Let the minimum number of colors needed for all such graphs be denoted by $\chi(g, M)$ where we require that all vertices coming from the same empire receive the same color.

THEOREM 3 **[15]**. *For all* $g \geq 0$ *and* $M \geq 1$, *except for the case* $g=0$ *and* $M=1$,

$$\chi(g, M) \leq \left\lfloor \frac{6M+1+\sqrt{48g+(6M+1)^2-48}}{2} \right\rfloor.$$

As in the case of $M=1$ and $g>0$, this formula can be motivated as above by supposing that K_n is an M-pire graph that arises from vertex-identifications of a graph embedded on the sphere plus g handles.

Notice first that for $g=0$ and $M \geq 2$,

$$6M \leq \frac{6M+1+\sqrt{48g+(6M+1)^2-48}}{2} < 6M+1,$$

and so the upper bound in these cases of Theorem 3 is $6M$. We have seen a proof that $\chi(0, M) \leq 6M$ in Theorem 1, and by the M-pire Theorem of Section 2, $\chi(0, M) = 6M$ for $M \geq 2$. Notice also that for $g>0$ and $M=1$, Theorem 3 coincides with that of Theorem 2, whose proof we haven't yet seen, and by the Map-Color Theorem $\chi(g, 1)$ actually equals this upper bound when $g>0$. Now we'll prove Theorem 3 in general, and then summarize what is known when $\chi(g, M)$ equals this upper bound. (Notice the parallels with the proof of Theorem 1 although, as we'll point out, this argument fails for $g=0$.)

Proof of Theorem 3. The case of $g=0$ was proved in Theorem 1, and so we assume that $g>0$.

If G is an n-vertex, e-edged graph embedded on the sphere plus $g>0$ handles that arises from an M-pire map, we may identify vertices from each empire to obtain a graph G^* that should be (normally) colored and that may no longer embed on the same surface. If G^* has n^* vertices and e^* edges, then $n \leq Mn^*$, $e^* \leq e$, and by the Euler-Poincaré Formula

$$e^* \leq e \leq 3n+6(g-1) \leq 3Mn^*+6(g-1).$$

The proof now proceeds by induction on n^*. If

$$n^* < \frac{6M + 1 + \sqrt{48g + (6M+1)^2 - 48}}{2},$$

the result is clearly true; so suppose

$$n^* \geq \frac{6M + 1 + \sqrt{48g + (6M+1)^2 - 48}}{2}.$$

Then $\dfrac{2e^*}{n^*} \leq 6M + \dfrac{12(g-1)}{n^*} \leq 6M + \dfrac{24(g-1)}{6M + 1 + \sqrt{48g + (6M+1)^2 - 48}}.$

(Notice that the second inequality holds only for $g \geq 1$.) Clearing the radical in the denominator of the previous fraction, one obtains:

$$\begin{aligned}\frac{2e^*}{n^*} &\leq 6M + \frac{24(g-1)\left\{6M + 1 - \sqrt{48g + (6M+1)^2 - 48}\right\}}{-48(g-1)} \\ &= 6M + \frac{-(6M+1) + \sqrt{48g + (6M+1)^2 - 48}}{2} \\ &= \frac{6M - 1 + \sqrt{48g + (6M+1)^2 - 48}}{2}.\end{aligned}$$

Thus there is a vertex v of degree at most

$$\left\lfloor \frac{6M - 1 + \sqrt{48g + (6M+1)^2 - 48}}{2} \right\rfloor = \left\lfloor \frac{6M + 1 + \sqrt{48g + (6M+1)^2 - 48}}{2} \right\rfloor - 1.$$

Remove v, use induction to color the remaining vertices with

$$\left\lfloor \frac{6M + 1 + \sqrt{48g + (6M+1)^2 - 48}}{2} \right\rfloor$$

colors, and extend this coloring to v since it has one fewer adjacent vertices than the number of colors being used.

This argument with $M = 1$ (the nonimperialist case) gives a complete proof of Heawood's bound (in Theorem 2) for coloring ordinary maps on surfaces. Good expositions of this and related aspects of topological graph theory can be found in [**3**, **14**, **25**], but, surprisingly, almost no introductory text on graph theory except for the recent book [**14**] contains a proof of this ($M = 1$) bound (although one text contains an incorrect proof). Proofs of this nonimperialist case can be found in more specialized texts [**6**, **11**, **22**, **29**], while the fully general case of Theorem 3 appears in [**21**].

Is the upper bound of Theorem 3 always achieved by some M-pire graph? Notice that for the torus, where $g = 1$, the upper bound is simply $6M + 1$. H. Taylor [**26**] has announced that for each M there are M-pire graphs on the torus, achieving the bound of $6M + 1$; his work is based on results of S. W. Golomb [**10**] on "graceful" labelings of graphs. And for the fully general case, Jackson and Ringel [**18**] have studied the situation intensively and have proved that the upper bound is achieved for at least 12.5% of the remaining cases.

THEOREM [18]. *For* $g > 0$,

$$\chi(g, M) = \left\lfloor \frac{6M + 1 + \sqrt{48g + (6M+1)^2 - 48}}{2} \right\rfloor \qquad (5)$$

(1) *when M is even and the right-hand side of* (5) *is congruent to* 1 *modulo* 12, *and*
(2) *when M is odd and the right-hand side of* (5) *is congruent to* 4 *or* 7 *modulo* 12.

It is also possible to consider empire maps on nonorientable surfaces such as the projective plane, the Klein bottle, or the sphere plus k cross-caps, $k > 0$. Then a formula analogous to that of Theorem 3 can be proved similarly, but again the hardest work is showing that equality can be achieved. Specific results are known, and in general the upper bound has been shown to be achieved in about 25% of the cases; a summary of these results is contained in [18].

Heawood thought up a variety of problems and questions that have intrigued researchers for years. Others since Heawood have embellished upon his ideas, some in equally fanciful map terms (e.g., [17]), others with more abstraction [13], and several with applications in theoretical computer science. For example, the concept of thickness has considerable applicability, beyond that of PCBs, in the area of complexity of algorithmic problems and in the theory of NP-completeness [19]. General graph-coloring questions and related algorithmic problems, not just in the context of maps, are some of the most widely studied today, because of both their difficulty and their applicability (see [27]).

Acknowledgements. I would like to thank Nora Hartsfield, Brad Jackson, Allen Schwenk, and Arthur White for helpful comments on this paper, and Stan Wagon for producing the diagrams.

REFERENCES

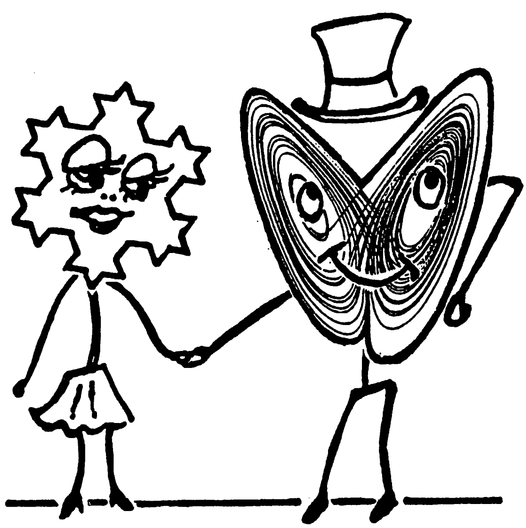
1. K. Appel and W. Haken, Every planar map is four colorable, *Bulletin Amer. Math. Soc.* 82 (1976), 711–712.
2. K. Appel and W. Haken, A proof of the four color theorem, *Discrete Math.* 16 (1976), 179–180.
3. L. W. Beineke, Are graphs finally surfacing?, *College Math. J.* 20 (1989), 206–225.
4. L. W. Beineke and A. T. White, Topological graph theory, *Selected Topics in Graph Theory*, L. W. Beineke and R. J. Wilson (eds.), Academic Press, London, 1978, 15–50.
5. N. L. Biggs, E. K. Lloyd, and R. J. Wilson, *Graph Theory, 1736–1936*, Clarendon Press, Oxford, 1976.
6. B. Bollobás, *Extremal Graph Theory*, Academic Press, London, 1978.
7. G. Chartrand, and L. Lesniak-Foster, *Graphs and Digraphs* (2nd edition), Wadsworth and Brooks/Cole, Publishing Co., Belmont, CA, 1986.
8. M. Gardner, Mathematical Games, *Scientific American* 242 (Febuary 1980), 14–19.
9. M. R. Garey, D. S. Johnson, and H. C. So, An application of graph coloring to printed circuit testing, *IEEE Trans. Circuits and Systems* CAS-23 (1976), 591–599.
10. S. W. Golomb, How to number a graph, *Graph Theory and Computing*, R. C. Read (ed.), Academic Press, New York, 1972.
11. J. L. Gross and T. W. Tucker, *Topological Graph Theory*, John Wiley & Sons, Inc., New York, 1987.
12. W. Haken, An attempt to understand the four color problem, *J. Graph Theory* 1 (1977), 193–206.
13. N. Hartsfield, B. Jackson, and G. Ringel, The splitting number of the complete graph, *Graphs and Combinatorics* 1 (1985), 311–329.
14. N. Hartsfield and G. Ringel, *Pearls of Graph Theory*, Academic Press, New York, 1990.
15. P. J. Heawood, Map colour theorem, *Quart. J. Pure and Applied Math.* 24 (1890), 332–333.
16. B. Jackson and G. Ringel, Solution of Heawood's empire problem in the plane, *J. Reine Angew. Math.* 347 (1983), 146–153.
17. ______, Coloring island maps, *Bull. Austral. Math. Soc.* 29 (1984), 151–165.
18. ______, Heawood's empire problem, *J. Combin. Theory Ser. B* 38 (1985), 168–178.
19. A. Mansfield, Determining the thickness of graphs is NP-hard, *Math. Proc. Camb. Phil. Soc.* 93 (1983), 9–23.

20. *The New York Times*, Sept. 15, 1989, p. 1.
21. G. Ringel, *Färbungsprobleme auf Flächen and Graphen*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1959.
22. G. Ringel, *Map Color Theorem*, Springer-Verlag, New York, 1974.
23. G. Ringel and J. W. T. Youngs, Solution of the Heawood map-color problem, *Proc. Nat. Acad. Sci. USA* 60 (1968), 438–445.
24. *Scientific American* 237 (Sept. 1977).
25. S. Stahl, The other map coloring theorem, this MAGAZINE, 58 (1985), 131–145.
26. H. Taylor, Synchronization patterns and related problems in combinatorial analysis and graph theory, *University of Southern California Electronic Science Laboratory Tech.* Report #509, June, 1981.
27. B. Toft, 75 Graph Colouring Problems, *Graph Colourings*, R. Nelson and R. J. Wilson, eds., Longman Scientific & Technical, Harlow, UK, 1990.
28. T. Tymoczko, The four-color problem and its philosophical significance, *J. Philosophy* 76 (1979), 57–83.
29. A. T. White, *Graphs, Groups and Surfaces* (revised edition), North-Holland, Amsterdam, 1984.

Fractal Basin Street Blues

MAURICE MACHOVER
St. John's University
Jamaica, NY 11439

There was a young fractal named Fracta,
Who certainly knew how to factor;
She tried out ceramics but switched to dynamics,
And married a strange attractor.

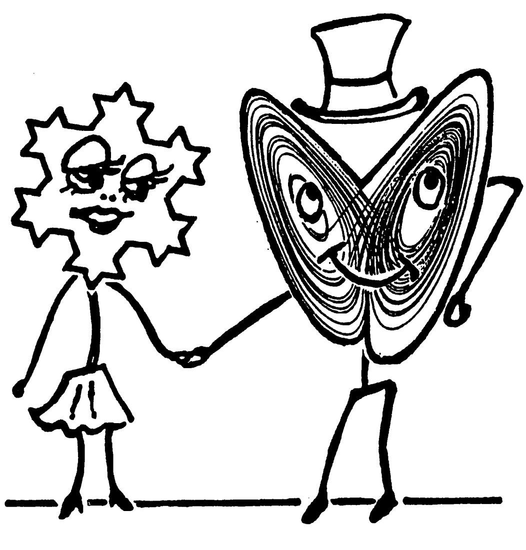


20. *The New York Times*, Sept. 15, 1989, p. 1.
21. G. Ringel, *Färbungsprobleme auf Flächen and Graphen*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1959.
22. G. Ringel, *Map Color Theorem*, Springer-Verlag, New York, 1974.
23. G. Ringel and J. W. T. Youngs, Solution of the Heawood map-color problem, *Proc. Nat. Acad. Sci. USA* 60 (1968), 438–445.
24. *Scientific American* 237 (Sept. 1977).
25. S. Stahl, The other map coloring theorem, this MAGAZINE, 58 (1985), 131–145.
26. H. Taylor, Synchronization patterns and related problems in combinatorial analysis and graph theory, *University of Southern California Electronic Science Laboratory Tech.* Report #509, June, 1981.
27. B. Toft, 75 Graph Colouring Problems, *Graph Colourings*, R. Nelson and R. J. Wilson, eds., Longman Scientific & Technical, Harlow, UK, 1990.
28. T. Tymoczko, The four-color problem and its philosophical significance, *J. Philosophy* 76 (1979), 57–83.
29. A. T. White, *Graphs, Groups and Surfaces* (revised edition), North-Holland, Amsterdam, 1984.

Fractal Basin Street Blues

MAURICE MACHOVER
St. John's University
Jamaica, NY 11439

There was a young fractal named Fracta,
Who certainly knew how to factor;
She tried out ceramics but switched to dynamics,
And married a strange attractor.



On Tangents

HUGH THURSTON
University of British Columbia
Vancouver, B.C., V6T 1Y4, Canada

Introduction

Tangents have interested mathematicians since antiquity. Classical Greek geometers investigated tangents to circles, conic sections, and spirals. Chinese mathematics, too, involved tangents: A typical Chinese problem was to find out how far west one has to walk from the north gate of a circular walled town in order to see a tree at a given location to the south of the town.

Because of the connection between tangents and derivatives, tangents became important in the calculus. It is this aspect of tangents that I propose to discuss. In elementary calculus, the graph of a function F has a tangent at the point with coordinates $(c, F(c))$ if the derivative $F'(c)$ exists (though the converse is not true [1]). The other way in which tangents are connected with derivatives is via tangent vectors. If a simple arc has a nonzero tangent vector at a point P then it has a tangent line at P.

Is the converse true? If P is a cusp, as for example at the origin on the graph of $y^2 = x^3$ (FIGURE 1), there cannot be a nonzero tangent vector at P, but if P is not a cusp does the existence of a tangent line at P imply that there is a nonzero tangent vector at P under some suitable parameterization? What if we insist that the tangent line is to be continuously turning? Again, does the existence of a continuously turning tangent line everywhere imply nice properties like rectifiability and regularity? If not, what if we also prohibit cusps?

Before getting down to details we must consider precisely what a tangent line is.

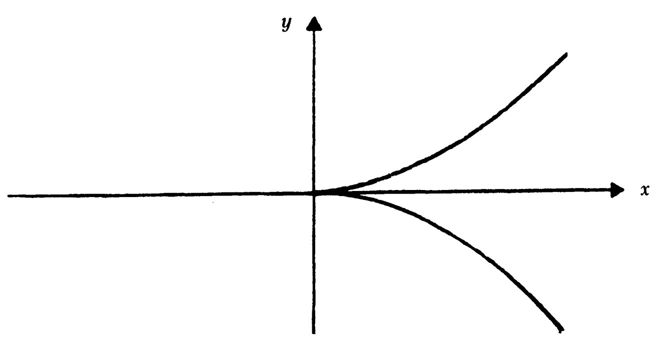


FIGURE 1
Graph of $y^2 = x^3$.

The Definition of Tangent Line

The definition of tangent line implicit in Chinese mathematics applies only to convex bodies. The definition implicit in Greek geometry is adequate for conic sections and spirals, but it does not enable us to deal with points where a curve crosses itself, as does the limaçon in FIGURE 2, nor with S-shaped curves like the graph of $y = x^3$, nor with cusps, let alone more complicated curves such as the graph of $y = x^2 \sin x^{-1}$ completed by including the origin (FIGURE 3).

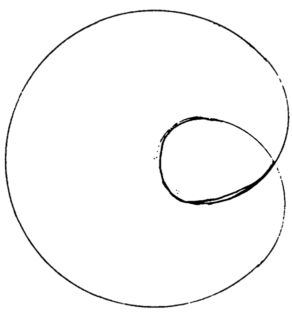


FIGURE 2

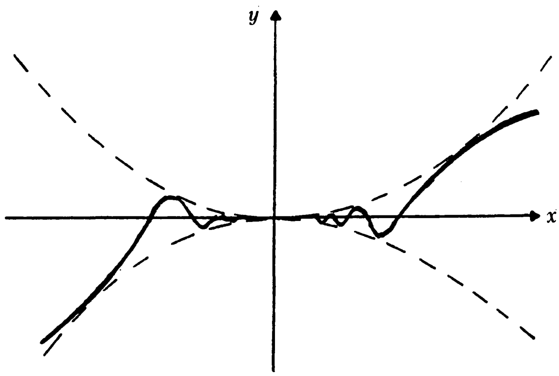


FIGURE 3
Graph of $y = x^2 \sin(x^{-1})$.

Eventually a good definition of tangent was devised. Succintly put,

the tangent is the limit of the secant.

Let us formalize this definition. I shall give a formal definition, not of a tangent to a curve, but of a tangent to a set of points. There are two reasons for this. The first is that there is no universally accepted definition of curve. A definition that was in favour for a long time is that a curve is the path of a continuously moving point. After Peano showed that such a curve can completely fill a square, attempts were made to modify the definition, none winning 100% acceptance. Luckily the most interesting curves for the study of tangents are simple arcs, about whose definition there is no controversy.

The second reason for not confining our attention to curves is to include tangent lines to graphs of functions. Curves and graphs are by no means the same thing, even in a plane. A circle is a curve, but is not the graph of a function; the graph of Dirichlet's function D (defined by setting $D(x) = 0$ if x is rational, $D(x) = 1$ if x is irrational) is not a curve.

Here is a formal definition. Just as a secant to a curve is a line through two points of the curve, so a *secant* to a set of points is a line through two points of the set.

Let $\mathscr{A}$ be a set of points, let P belong to $\mathscr{A}$, and let $\mathscr{T}$ be a line through P. Then $\mathscr{T}$ is a *tangent-line* to $\mathscr{A}$ at P if the following condition is satisfied. If ε is any positive number, there is a positive δ such that every secant PX for which the distance of X from P is less than δ makes an angle smaller than ε with $\mathscr{T}$.

It turns out that cusps of simple arcs play an important rôle in what follows.

A point P (not an endpoint) of a simple arc $\mathscr{A}$ is a *cusp* of $\mathscr{A}$ if there is a ray $\mathscr{R}$ with one end at P satisfying the following condition. If ε is any positive number there is a positive δ such that every ray that has one end at P and passes through a point of $\mathscr{A}$ within δ of P makes an angle smaller than ε with $\mathscr{R}$,

The definition of tangent line implies that $\mathscr{A}$ cannot have more than one tangent line at the point P unless P is an isolated point of $\mathscr{A}$, that if $\mathscr{A}$ lies in a plane all its tangent lines lie in that plane, that there is no tangent line at the crossing-point of a limaçon, that the x-axis is the tangent line at the origin to the S-shaped graph of $y = x^3$, and that the x-axis is the tangent line at the origin in FIGURE 3. Moreover, the

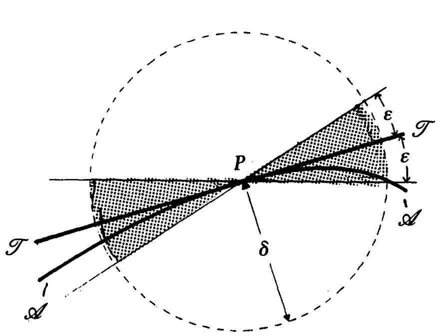


FIGURE 4

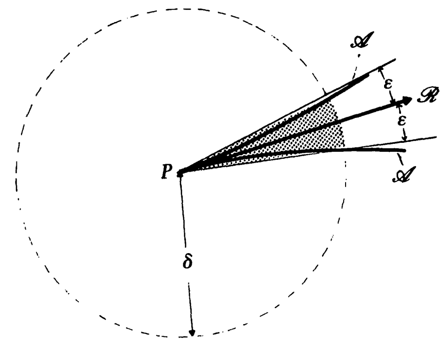


FIGURE 5

x-axis is the tangent line at the origin to the graph of $y^2 = x^3$ (FIGURE 1); the definitions allow tangent lines at cusps.

The Direction of Motion

There is an obvious tie-in between the direction of motion of a moving particle and tangents to its path. Set up coordinates and, for each t, let the coordinates of the particle at time t be

$$(X(t), Y(t), Z(t)).$$

The *direction of motion* at time t is the direction of the velocity-vector

$$X'(t)\vec{i} + Y'(t)\vec{j} + Z'(t)\vec{k},$$

if this vector exists (i.e. if X, Y, and Z are differentiable at t) and is nonzero. (If the path is in a plane we can set up coordinates in the plane and do without Z.)

If the moving particle passes through the point P exactly once, we might expect that the line through P in the direction of a velocity vector at P would be a tangent line, but this is not necessarily so. If

$$X(t) = (1 - 2\cos t)\cos t \quad \text{and} \quad Y(t) = (1 - 2\cos t)\sin t \quad \text{for } 0 < t < 5\pi/3,$$

the path is the portion of a limaçon shown in FIGURE 6.

The velocity-vector at P, where $t = \pi/3$, is $\frac{1}{2}(\sqrt{3}\,\vec{i} + 3\vec{j})$ but the path has no tangent-line there.

To avoid complications of this sort, let us confine our attention to simple arcs.

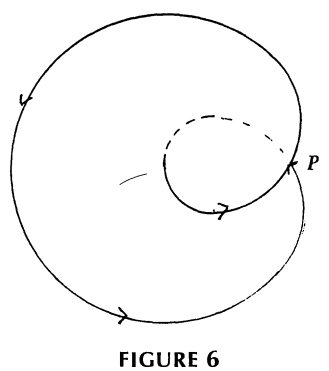


FIGURE 6

Simple Arcs

Let us, for conciseness, consider plane arcs. Details for three-dimensional arcs will be similar, with an extra coordinate.

A set of points is a (plane) *simple arc* if it has a parametrization of the form

$$(x,y)=(X(t),Y(t)); \qquad a\leqq t\leqq b$$

where (X,Y) is continuous, one-to-one, and has domain $[a,b]$.

Because the domain of (X,Y) is compact its inverse is continuous. It follows that if T and P are the points with coordinates $(X(t),Y(t))$ and $(X(p),Y(p))$ respectively then

$$t\to p \quad \text{if and only if } T\to P.$$

If the tangent vector

$$X'(p)\vec{i}+Y'(p)\vec{j} \tag{1}$$

is nonzero, the line $\mathscr{L}$ through P in the direction of this vector is a tangent line to the arc at P. To see this, suppose first that p is neither a nor b. For all t near enough to p, $X(t)-X(p)$ and $Y(t)-Y(p)$ are not both zero. Let

$$c=\left(\sqrt{[X(t)-X(p)]^2+[Y(t)-Y(p)]^2}\right).$$

The vector $\overrightarrow{PT}$ is the same direction as the vector

$$\frac{X(t)-X(p)}{c}\vec{i}+\frac{Y(t)-Y(p)}{c}\vec{j}.$$

If $t>p$,

$$\frac{X(t)-X(p)}{c}=\frac{X(t)-X(p)}{t-p}\div\frac{c}{\sqrt{(t-p)^2}}$$

and

$$\frac{c}{\sqrt{(t-p)^2}}=\sqrt{\left[\left(\frac{X(t)-X(p)}{t-p}\right)^2+\left(\frac{Y(t)-Y(p)}{t-p}\right)^2\right]}.$$

Therefore as t approaches p from above

$$\frac{X(t)-X(p)}{c}\to\frac{X'(p)}{\sqrt{\left[X'(p)^2+Y'(p)^2\right]}}.$$

Similarly for Y, and so the direction of the vector $\overrightarrow{PT}$ approaches that of the vector (1). We can show similarly that as t approaches p from below the direction of the vector $\overrightarrow{PT}$ approaches the direction precisely opposite to that of (1). Therefore as t approaches p the angle between the secant PT and $\mathscr{L}$ approaches zero. Then this angle approaches zero as T approaches P (it is this fact that precludes the situation shown in FIGURE 6), which shows that $\mathscr{L}$ is a tangent line. This is still true if p is a or b, in which case only one of the one-sided limits is needed.

Thus if a simple arc has a nonzero tangent vector at P, it has a tangent line there. The converse is not true, as is shown by an arc with a cusp. We shall see later that even if we disallow cusps it is still not true.

Continuously Turning Tangent Lines

The angle between the tangent line at a point X of a simple arc and the tangent line at the point P need not approach zero as X approaches P. This is shown by a well-known example, namely the part of the graph pictured in FIGURE 3 from, say, $x = -1$ to $x = 1$.

If the angle between the tangent line at X and the tangent line at P does approach zero as X approaches P, we say that the arc has a *continuously turning tangent line* at P.

If a simple arc has a nonzero tangent vector at a point P, it has a tangent line there. The converse is not true, even if the tangent line is continuously turning. The following example shows this.

On the graph of $y = x^{1/2}$ take the points where $x = 1$, $1/2$, $1/4$, etc. and with these points as opposite corners construct rectangles $\mathscr{R}_1, \mathscr{R}_2$ etc. as in FIGURE 7.

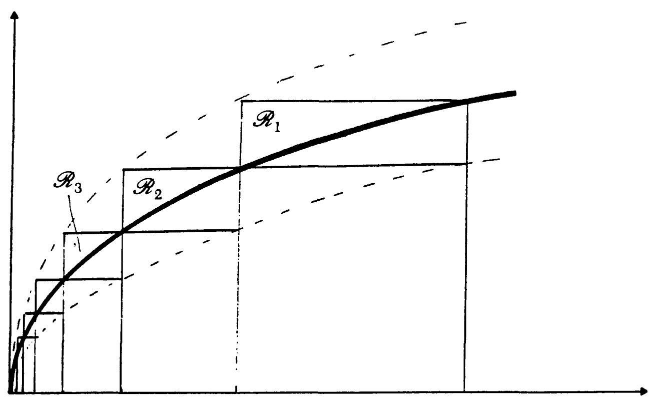


FIGURE 7

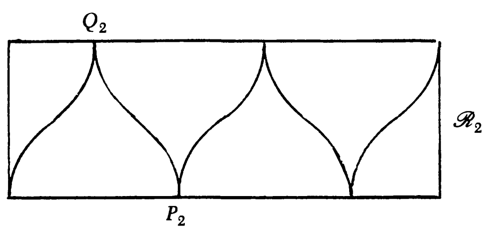


FIGURE 8

These rectangles are trapped between the parabolas $x = 2y^2$ and $x = \frac{1}{2}y^2$ shown as dashed lines. Divide $\mathscr{R}_n$ vertically into $2^n + 1$ parts and in each part insert a half-cycle of a sine curve with vertical tangents at the top and bottom as in FIGURE 8. The points on all these arcs, together with the origin and the reflections of these arcs in the origin form a simple arc. Clearly the y-axis is a tangent line at the origin and the tangent line is continuously turning. Let

$$x = X(t), \qquad y = Y(t); \qquad a \leqq t \leqq b$$

be any one-to-one parametrization of $\mathscr{A}$ running from $(-1, -1)$ to $(1, 1)$. Let c be the value of the parameter at the origin.

In each rectangle $\mathscr{R}_n$ let Q_n be the left-most point of $\mathscr{A}$ on the top edge and P_n the left-most point of $\mathscr{A}$, other than the corner, on the bottom edge, and let the values of the parameter at Q_n and P_n be q_n and p_n respectively. Then

$$Y(q_n) = 1/\sqrt{2^{n-1}}, \qquad Y(p_n) = 1/\sqrt{2^n} \quad \text{and} \quad Y(c) = 0.$$

Consequently

$$Y(q_n) - Y(c) = \sqrt{2}\,[Y(p_n) - Y(c)].$$

Because Q_n is between the origin and P_n, $c < q_n < p_n$ and so

$$0 < q_n - c < p_n - c.$$

Therefore

$$\frac{Y(q_n) - Y(c)}{q_n - c} > \frac{Y(q_n) - Y(c)}{p_n - c} = \sqrt{2}\,\frac{Y(p_n) - Y(c)}{p_n - c} > 0.$$

Consequently if $Y'(c)$ exists

$$Y'(c) \geqq \sqrt{2}\,Y'(c) \geqq 0$$

and so $Y'(c) = 0$. Then $X'(c)$ cannot be nonzero because if it were, $\vec{i}$ would be a tangent vector at the origin. Consequently, if there is a tangent vector at the origin, it can only be the zero vector. Even if we change to another (rectangular Cartesian) coordinate system the result still holds; if (X, Y) is a parametrization in the new system, with axes at an angle α to the old axes and origin at (a, b), then $(a + X\cos\alpha + Y\sin\alpha,\ b + X\sin\alpha - Y\cos\alpha)$ is a parametrization in the old system.

The same example shows that a simple arc can have a continuously turning tangent line everywhere without being rectifiable: The height of the rectangle $\mathscr{R}_n$ is $(\sqrt{2} - 1)/\sqrt{2}^{\,n}$ and so the total length of the portion of $\mathscr{A}$ inside $\mathscr{R}_n$ is greater than $(\sqrt{2} - 1)\sqrt{2^n}$, and the lengths of these portions form a divergent series.

A further property that a rectifiable arc might have is regularity: An arc is *regular* at P if arc PX/chord PX approaches 1 as X approaches P. If, in the example shown in FIGURE 9, we insert three (instead of $2^n + 1$) half-cycles in each $\mathscr{R}_n$, we obtain a rectifiable arc that is not regular at the origin but has a continuously turning tangent line everywhere.

These counter-intuitive results seem to be due to the fact that an arc with a continuously turning tangent line may have cusps. Let us therefore investigate cusp-free arcs. A cusp-free arc whose tangent line is not continuously turning may have the counter-intuitive properties just mentioned: To obtain appropriate examples, replace the vertical half-cycles of sine curves in the examples above by horizontal half-cycles, thus for instance turning FIGURE 8 into FIGURE 9. Therefore the arcs we investigate must be both free of cusps and continuously turning.

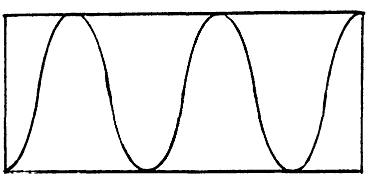


FIGURE 9

Arcs with Continuously Turning Tangent Lines and No Cusps

If neither the absence of cusps nor the existence everywhere of a continuously turning tangent line is enough to ensure that an arc is rectifiable, how about the two together? They are enough, and are enough to ensure that the arc has a nonzero tangent vector everywhere, as Georges Valiron showed in 1927—a result that deserves to be better known [2]. To be precise, he proved the following result, though he phrased it somewhat differently, using the equivalent concept of oriented tangent.

VALIRON'S THEOREM. *A simple arc with a continuously turning tangent line and no cusps is rectifiable, and if the arc-length parametrization is*

$$x = X(s),\ y = Y(s),\ z = Z(s)$$

then X, Y, and Z are continuously differentiable.

We are now on familiar ground. The arc has a continuously differentiable parametrization giving a tangent vector that is nowhere zero (because a tangent vector with respect to arc length is necessarily a unit vector). This is the type of curve usually treated in calculus texts and is well known to be rectifiable and regular (e.g. [3]). Thus the following properties are equivalent for a simple arc.

1. The arc has a continuously turning tangent line and no cusps.
2. The arc has a continuously differentiable parametrization in which the tangent vector is never zero.
3. The arc is rectifiable and the tangent vector with respect to arc length is continuously turning.
4. The arc has a parametrization giving a continuously turning tangent vector.

It is worth noticing that the parametrization in 4 need not itself be continuously differentiable. It might, for instance, be $x = y = F(t)$ where $F(t) = 2t + t^2 \sin(1/t)$ if $t \neq 0$ and $F(0) = 0$.

A property that is equivalent to the four listed but does not mention tangents is

5. At each point P of the arc there is a line $\mathscr{T}_P$ such that the angle between the secant XY and $\mathscr{T}_P$ approaches zero as X and Y approach P.

It is not hard to prove that if all tangent lines to a sufficiently small portion of an arc without cusps make arbitrarily small angles with $\mathscr{T}_P$, so do all secants in that portion; consequently 1 implies 5. Conversely, 5 clearly precludes cusps and implies that if P and Q are sufficiently close we can find a secant making arbitrarily small angles with both $\mathscr{T}_P$ and $\mathscr{T}_Q$, from which 1 follows.

Note on Oriented Tangents

Let $\mathscr{A}$ be a simple arc, let P be a point of $\mathscr{A}$, and let $\mathscr{T}$ be a line through P. Then $\mathscr{T}$ is an *oriented tangent* to $\mathscr{A}$ at P if the following condition is satisfied. If $\mathscr{C}$ is any cone with vertex P and axis $\mathscr{T}$ there is a sphere with centre P such that every point of $\mathscr{A}$ between P and one end-point that is in the sphere lies in one nappe of the cone and every point in the sphere between P and the other end-point lies in the other nappe.

Clearly if P is neither an end-point nor a cusp there is a tangent at P if, and only if, there is an oriented tangent at P. The formal proof of the "only if" clause takes a little ingenuity. Let the arc have a tangent at P and $\mathscr{C}$ be any cone with P for vertex and the tangent for axis. There is a sphere $\mathscr{S}_1$ such that all points of the arc that are in $\mathscr{S}_1$ are in $\mathscr{C}$, so no points of the arc lie in the shaded region in FIGURE 10.

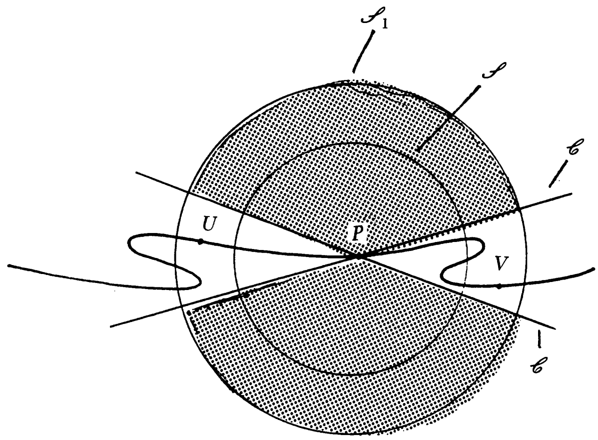


FIGURE 10

There is a segment UPV of the arc in $\mathscr{S}_1$ and therefore also in $\mathscr{C}$. It cannot cross the boundary of the unshaded region except at P, and so the segments UP and PV each lie in one nappe of $\mathscr{C}$. Because the inverse of the parametrization is continuous, there is a sphere $\mathscr{S}$ such that every point of the arc that is in $\mathscr{S}$ is a point of the segment UPV. Then every point of the arc that is in $\mathscr{S}$ and is on the same side of P as V lies in the nappe containing PV, and every point of the arc in $\mathscr{S}$ on the same side as U lies in the nappe containing U. U and V then cannot lie in the same nappe, because P is not a cusp. Therefore the tangent is oriented.

We can also show that if the tangent is continuously turning (in the absence of cusps) so is the oriented tangent. In other words, the oriented tangent cannot suddenly reverse direction. Although this is obvious enough, a formal proof is complicated. Such a proof is implicit in Theorem 6 of [4].

Possible Alternatives

The formal definition given earlier is not the only possible one. Instead of defining the tangent at P to be the limit of the secant PQ as Q approaches P we might have defined it to be the limit of the secant QR as Q and R approach P (i.e. by property 5 above). This ties in, not with the derivative, but with the *star-derivative* defined by

$$F^*(p) = \lim_{(q,r)\to(p,p)} \frac{F(q)-F(r)}{q-r}.$$

This kind of tangent cannot exist at a cusp, and if it exists everywhere it is continuously turning.

Instead of defining tangents to sets of points we could have adopted a definition of curve, such as the one cited earlier, and defined tangents to curves. In that case, we could not consider tangents to a graph unless the graph were a curve. The graph of F is a curve if and only if F is continuous and its domain is an interval. The reason is that if the graph has a parametrization

$$x = X(t),\ y = Y(t);\ t \in I$$

then $Y(t) = F(X(t))$ whenever $t \in I$, and the domain of F is $X(I)$, which is an interval. Then $Y = F \circ X$, so both X and $F \circ X$ are continuous. Consequently F is continuous [5].

Conclusion

A simple arc can have a tangent line at a point where there cannot be a nonzero tangent vector under any parametrization. This was demonstrated by two examples, each of which, incidentally, is also the graph of a function. In one case, the tangent line is not continuously turning; in the other, the arc has cusps. If we insist both (i) that the tangent be continuously turning and (ii) that the arc have no cusps, then the arc has a continuously differentiable nonsingular parametrization, with all that this implies. Together with the fact that a graph of a function can have a tangent line at a point where the function is not differentiable, these results suggest that the connection between tangents and derivatives is not as tight as is sometimes supposed. The tendency for the existence of a tangent line to imply the existence of a corresponding derivative is weaker than the tendency for the existence of a derivative to imply the existence of a tangent line.

REFERENCES

1. Hugh Thurston, Tangents to graphs, this MAGAZINE, 61 (1988) 292–294.
2. Georges Valiron, Sur les courbes qui admettent une tangente en chaque point, *Nouvelles Ann. de Math.*, 6 (1927) 46–51.
3. Philip Franklin, *Advanced Calculus*, John Wiley and Sons, Inc., New York, 1940 pp. 284–286.
4. T. M. Flett, The definition of a tangent to a curve, *Edinburgh Math. Notes*, 41 (1957) 1–19.
5. Answer to problem E3397, *Amer. Math. Monthly*, 99, no. 2 (1992) 164–165.

Math Bites

Contributions are solicited for a new feature called Math Bites. The printed equivalent of a sound bite, a Math Bite should be a striking quantum of mathematics. It can be a surprising or unexpected proof, a cogent example or counterexample, or a little-known mathematical connection. It should have the brevity and immediacy of a proof without words. A Math Bite can be an original contribution, part of the folklore, or quoted or paraphrased from another source (with attribution). A Math Bite should be expressible in half a page or less in the MAGAZINE.

NOTES

Coordinate Systems and Primitive Maps

JOHN A. BAKER
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada

Suppose U is an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^n$. Following [2], we say that F is *primitive* provided it fixes all but one coordinate; i.e. for some k between 1 and n and some $\varphi\colon U \to \mathbb{R}$ we have

$$F(x_1,\dots,x_n) = (x_1,\dots,x_{k-1},\varphi(x_1,\dots,x_n),x_{k+1},\dots,x_n) \quad \text{for all } (x_1,\dots,x_n) \in U.$$

We say F is a *flip* provided it interchanges two coordinates. For example, $(x, y, z) \to (z, y, x)$ is a flip of $\mathbb{R}^3$.

The inverse function theorem and the change of variable theorem for multiple integrals are relatively easy to prove for primitive maps and for flips. The general cases of these theorems can then be deduced from the following assertion (see [1] pp. 133–158 and [2] p. 249) and a partition of unity argument; if F is a C^1 map of an open subset U of $\mathbb{R}^n$ into $\mathbb{R}^n$ and the Jacobian determinant of F is nonzero at each point of U then F is locally (but, in general, not globally) a composition of at most n flips and at most n invertible C^1 primitive maps. The purpose of this note is to illustrate this assertion with the following examples.

Polar coordinates. Let H denote the half-plane $\{(x,y) \in \mathbb{R}^2\colon x > 0\}$. For $(x, y) \in H$ let $P(x,y) = (r,\theta)$ where $r = (x^2 + y^2)^{1/2}$ and $\theta = \arcsin(y/(x^2+y^2)^{1/2})$, polar coordinates of (x, y). Let $F_1(x,y) = ((x^2+y^2)^{1/2}, y)$ for $(x,y) \in H$. Then F_1 is a primitive bijection of H onto the quadrant $Q = \{(s,t) \in \mathbb{R}^2\colon 0 < s \text{ and } -s < t < s\}$ and $F_1^{-1}(s,t) = ((s^2-t^2)^{1/2}, t)$ for all $(s,t) \in Q$. Let $F_2(s,t) = (s, \arcsin(t/s))$ for $(s,t) \in Q$. Then F_2 is a primitive bijection of Q onto the band $B = \{(r,\theta)\colon r > 0, |\theta| < \pi/2\}$ and $F_2^{-1}(r,\theta) = (r, r\sin\theta)$ for $(r,\theta) \in B$. If $(x,y) \in H$ then $P(x,y) = ((x^2+y^2)^{1/2}, \arcsin(y/(x^2+y^2)^{1/2})) = F_2((x^2+y^2)^{1/2}, y) = F_2(F_1(x,y))$. That is, $P = F_2 \circ F_1$. Moreover $P^{-1} = F_1^{-1} \circ F_2^{-1}$; i.e., $P^{-1}(r,\theta) = (r\cos\theta, r\sin\theta)$ for $(r,\theta) \in B$.

Complex logarithms. Identify $\mathbb{R}^2$ with the complex plane and let L denote the principal logarithm on H, i.e. let $L(x,y) = (\ln(x^2+y^2)^{1/2}, \arcsin(y/(x^2+y^2)^{1/2}))$ for $(x,y) \in H$. If we let $F_3(r,\theta) = (\ln r, \theta)$ for $(r,\theta) \in B$ then F_3 is an invertible primitive bijection of B onto the strip $S = \{(u,v) \in \mathbb{R}^2\colon |v| < \pi/2\}$ and $L = F_3 \circ F_2 \circ F_1$. We may also express L as a composition of two primitive maps as follows. For $(x,y) \in H$ let $G_1(x,y) = (\ln(x^2+y^2)^{1/2}, y)$. Then G_1 is a primitive bijection of H onto $E = \{(s,t) \in \mathbb{R}^2\colon |t| < e^s\}$. If for $(s,t) \in E$ we let $G_2(s,t) = (s, \arcsin(te^{-s}))$ then G_2 is a primitive bijection of E onto S and $L = G_2 \circ G_1$.

Spherical coordinates. Let $U_1 = \{(x,y,z) \in \mathbb{R}^3\colon x > 0\}$. For $(x,y,z) \in U_1$ let $T(x,y,z) = (\rho,\theta,\phi)$ where

$$\rho = (x^2+y^2+z^2)^{1/2}, \qquad \theta = \arcsin\left[y/(x^2+y^2)^{1/2}\right] \quad \text{and}$$

$$\phi = \arcsin\left[(x^2+y^2)^{1/2}/\rho\right],$$

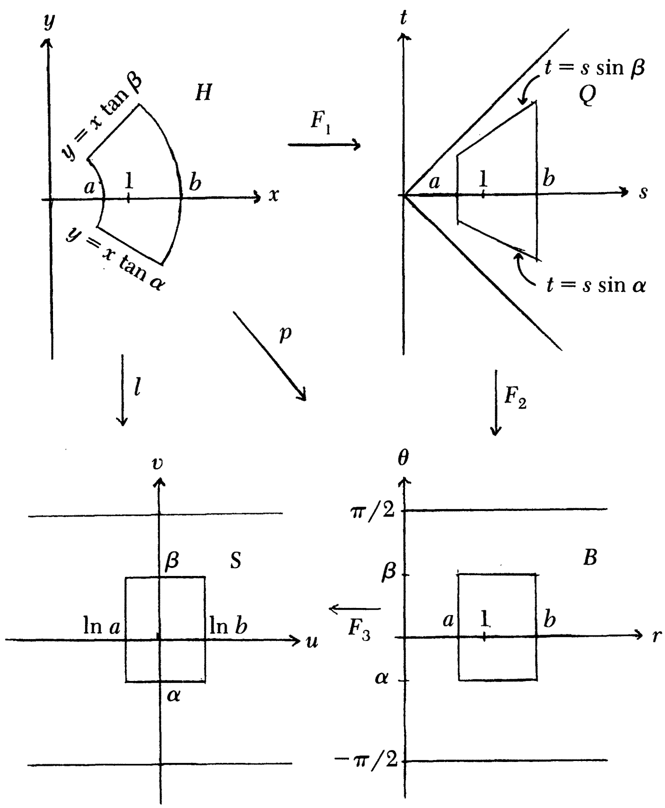
spherical coordinates of (x, y, z). Let

$$U_2 = \{(\rho, y, z) \in \mathbb{R}^3 \colon \rho > 0 \quad \text{and} \quad \rho^2 > y^2 + z^2\},$$
$$U_3 = \{(\rho, \theta, z) \in \mathbb{R}^3 \colon \rho > 0 \quad \text{and } |\theta| < \pi/2\} \quad \text{and}$$
$$U_4 = \{(\rho, \theta, \phi) \in \mathbb{R}^3 \colon \rho > 0, |\theta| < \pi/2 \text{ and } 0 < \phi < \pi\}.$$

For $1 \leq j \leq 3$ define $T_j \colon U_j \to U_{j+1}$ by

$$T_1(x, y, z) = \left((x^2 + y^2 + z^2)^{1/2}, y, z\right) \quad \text{for} \quad (x, y, z) \in U_1,$$
$$T_2(\rho, y, z) = \left(\rho, \arcsin\left[y/(\rho^2 - z^2)^{1/2}\right], z\right) \quad \text{for} \quad (\rho, y, z) \in U_2 \quad \text{and}$$
$$T_3(\rho, \theta, z) = \left(\rho, \theta, \arcsin\left[(\rho^2 - z^2)^{1/2}/\rho\right]\right) \quad \text{for} \quad (\rho, \theta, z) \in U_3.$$

Then T_j is an invertible primitive bijection of U_j onto U_{j+1} for $1 \leq j \leq 3$ and $T = T_3 \circ T_2 \circ T_1$.



REFERENCES

1. R. Courant, *Differential and Integral Calculus*, Vol. 2, John Wiley and Sons, Inc., Classics Library Edition, New York, 1988.
2. W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, McGraw-Hill, New York, 1976.

The Marriage Lemma for Polygamists

CLARK BENSON
University of Arizona
Tucson, AZ 85721

FRANKLIN LOWENTHAL
California State University
Hayward, CA 94542

It has long been the ambition of one of the authors (it is left as an exercise to the reader to determine which one) to appear as a guest on one of the "sensational" TV talk shows. But what could a mathematician talk about that would be of interest to Geraldo or Donahue? The first subject that comes to mind is the so-called marriage problem, alias the assignment problem; here the objective is to have a matching between a set of women and a set of men given that each woman likes some of the men and dislikes the rest. Let the integer m represent the number of women; the number of men must, of course, be at least m.

An obvious necessary condition that each woman gets herself a man is that for each $p < m$, every set of p women must together like at least p different men. Philip Hall's theorem, which has come to be known as the Marriage Lemma, asserts that this obvious necessary condition is sufficient. The proof of this result can most easily be accomplished by expressing the problem as that of determining the maximum flow in an appropriate network and then applying the maximum flow minimum cut theorem [**9**]. Another method of proof is based on the concept of maximum matchings in bipartite graphs that was developed by D. Konig and Philip Hall [**1**].

After further reflection, the authors suddenly had their moment of inspiration: Why not generalize the result to the polygamous case, i.e., where each man can be married to more than one woman? To be specific, for a fixed positive integer $d \geq 1$ we assume that up to d women can be married to one man. In the monogamous case we clearly needed at least m men, but now the only obvious condition is that the number n of men must satisfy $m \leq nd$. Note that we do not require each man to marry d women; a complete matching requires only that all women are married. Note also that we do not require the compatibility to be symmetric. A woman may be married to a man if she likes him, even if he does not like her.

THEOREM 1 (generalized Hall condition). *A complete matching or assignment of a set of m women to a set of n men where at most d women can be assigned to one man is possible if, and only if, for every positive integer $p \equiv 1 \bmod d$, $p \leq m$, every set of p women likes or is compatible with a set of at least $(p+d-1)/d$ men.*

Before considering the formal proof of this theorem, a number of comments are in order to clarify the meaning of the generalized Hall condition. First, note that if $m > nd$, then the condition cannot be satisfied since choosing $p = nd + 1 \leq m$ leads to the requirement that a set of $nd + 1$ women must like a set of $(nd + 1 + d - 1)/d = n + 1$ men, which is impossible. This is why it is unnecessary to separately include the obvious hypothesis $m \leq nd$. Second, the stipulation that $p \equiv 1 \bmod d$ can be safely eliminated, for if we interpret $(p + d - 1)/d$ to be the greatest integer less than or equal to that number when $p \not\equiv 1 \bmod d$, then increasing the set of women without increasing the size of the set of men that they must like poses no problem. Third, if we represent the compatibility or woman-likes-man relationship by an incidence or adjacency matrix with a one indicating "likes" and a zero indicating "dislikes," then the generalized Hall condition translates into the requirement that the matrix has no $p \times q$ block of zeros such that $p + dq > nd$ for $p \equiv 1 \bmod d$, $1 \leq p \leq m$. For if there were such a block, then the p women corresponding to the rows could like at most the remaining $n - q$ men and $n - q < n - (n - p/d) = p/d \leq (p + d - 1)/d$, so they could not like the requisite number of men. In the ordinary marriage problem with $m \leq n$ and $d = 1$, this just reduces to the requirement that there be no $p \times q$ block of zeros with $p + q > n = m$ (rearrangement of the men and/or women may be needed to create a consecutive block).

Proof of Theorem 1. The necessity of the generalized Hall condition follows from the observation that were it false, the p women who are compatible with at most r men, $r < (p + d - 1)/d$ cannot all be assigned to a compatible man since the r men can be assigned at most $rd < p + d - 1$ women. But since $p \equiv 1 \bmod d$, $(p - 1) + d$ is a multiple of d so $rd \leq (p + d - 1) - d = p - 1$. Hence $rd < p$ as asserted.

To prove the sufficiency we first construct a flow diagram that describes the general assignment problem (see FIGURE 1; see also [9]). The capacity on each branch from the source to a woman W_i is 1 while the capacity on each branch from a man M_j to the sink is d so that no man will be assigned more than d females. If a woman likes a man, a branch with capacity at least one should join them; the proof turns out to be simplest if we choose instead to assign capacity $m =$ the number of women to each such branch.

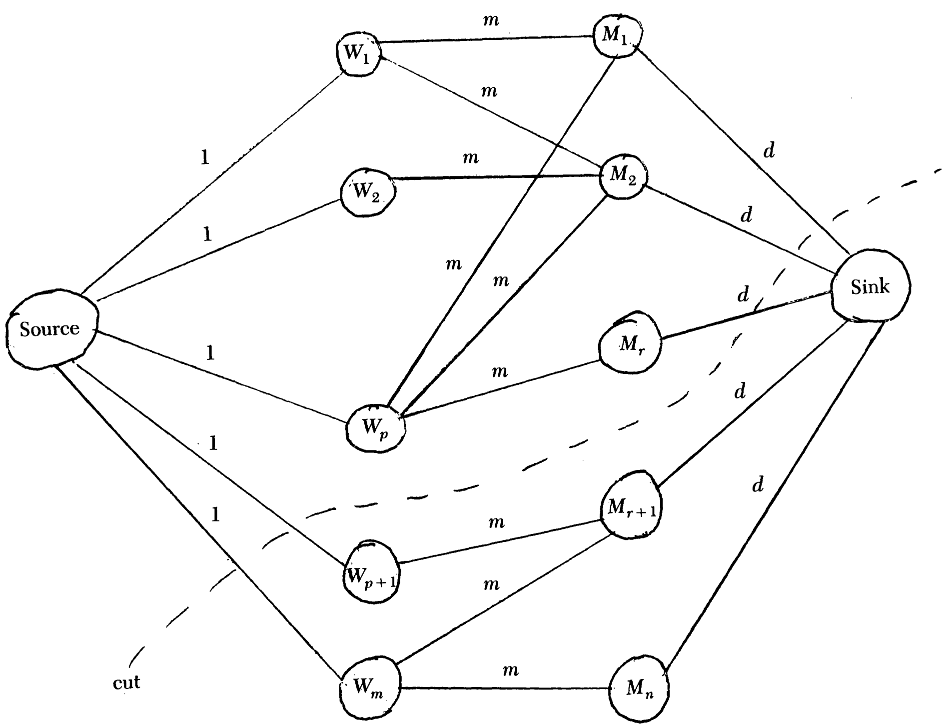


FIGURE 1
Cut capacity $< m$.

If not all the women can be assigned to a man whom they like, then the maximum flow in the network in FIGURE 1 is less than m since the existence of a flow having value m implies that marriage is possible. [Because in this case each branch from the source has unit flow, each woman is connected to at least one man. And, as already noted, each man can be connected to at most d woman. These connections provide the (polygamous) marriage assignment.] Therefore by the Ford-Fulkerson maximum flow equals minimum cut theorem [**1**, **8**, and **10**], the minimum cut also has capacity $< m$. This cut divides all the nodes into two disjoint non-empty sets, one of which contains the sink, the other, the source; moreover, it is such that its capacity, i.e., the sum of the capacities of all arcs or branches with one vertex in the first set, the other in the second, is less than or equal to that of all other cuts. Assume as shown in FIGURE 1 that nodes $W_1, W_2, \ldots, W_p$, $p < m$ and $M_1, M_2, \ldots, M_r$ are in the set that contains the source while nodes $W_{p+1}, \ldots, W_m$ and $M_{r+1}, \ldots, M_n$ are in one that contains the sink. The cut capacity is the sum of $m - p$ from the source to the women separated from it and dr from the sink to the men separated from it. No edge in FIGURE 1 would join any of the nodes $W_1, W_2, \ldots, W_p$ to any of the nodes $M_{r+1}, \ldots, M_n$ since such an edge would have capacity m, which by itself already would exceed the maximum flow in the network. Thus FIGURE 1 explicitly exhibits a set of p women that is compatible with only r men where the minimum cut capacity

$$(m - p) + rd \tag{1}$$

must be equal to the maximum flow and hence less than m. But

$$m - p + rd < m \tag{2}$$

yields

$$r < p/d. \tag{3}$$

If $p \equiv 1 \bmod d$ then the proof is complete since p/d is certainly $\leq (p + d - 1)/d$. If $p \not\equiv 1 \bmod d$, choose the largest integer $p' < p$, $p' \equiv 1 \bmod d$. Then we certainly have a subset of p' women compatible only with the same set of r men and since $p' \geq p - (d - 1)$,

$$r < p/d \leq (p' + d - 1)/d. \tag{4}$$

Application 1. Assume that m students are to be assigned to n tutorial groups with each group having a maximum capacity of 3 students. Because of potential time conflicts, students can not accept some assignments. If for $p = 1, 4, 7, \ldots, p \leq m$, each set of p students can without any time conflict attend at least $(p + 2)/3$ different tutorial sessions, then the generalized Hall condition with $d = 3$ is satisfied and hence by Theorem 1 all students can be assigned a tutorial group that they can attend without any time conflict. The actual assignment can be made by following the constructive Ford-Fulkerson labeling algorithm that leads to the maximum flow m in the network [**3**]. If the generalized Hall condition doesn't hold, we can still use this algorithm to find the maximum flow $p < m$ in the network and thereby make the maximum possible assignment of p students to tutorial groups.

Application 2. Assume that a university must schedule m classes during a day in n rooms, with each room having a capacity of 12 one-hour classes. A particular class will generally not be compatible with all the existing classrooms because of factors such as size, blackboards, laboratory facilities, etc. If for each integer $p = 1, 13, 25, \ldots, p \leq m$, every set of p classes is compatible with at least $(p + 11)/12$ classrooms, then all classes can be successfully scheduled. Then the actual scheduling may be determined by the Ford-Fulkerson algorithm to find the maximum flow m in the network. If the generalized Hall condition fails, the algorithm can still be used to schedule the maximum number $p < m$ of classes. Note that after 12 (or fewer) classes are assigned to a classroom, one then need determine only the order of the time slots for each of them.

The generalized assignment problem can be made more interesting by permitting the women to weigh the relative desirability of the set of men with whom they are compatible. For example, the students in Application 1 might prefer certain tutorial groups to others because of the time being more convenient or a certain classroom in Application 2 might be more suitable for a class than another even though both are acceptable. Can an assignment be made that accommodates these preferences? This problem amounts to nothing more than a maximum flow problem where a maximum flow that minimizes the total cost is sought. The details of determining a minimal cost maximum flow are described in Busacker-Saaty [2]. To be able to apply this algorithm, a woman's preferences should be quantified as positive real numbers, with magnitude related inversely to her preferences, i.e., the smallest weight is assigned to the man she likes the most. To prevent manipulation of the weights by a woman so as to increase her likelihood of getting her first choice it would seem reasonable that she be required to assign a weight of one to her first choice, two to her second etc. Thus in Application 1 a student would, in effect, prioritize his or her set of acceptable tutorial groups.

Theorem 2 is now a consequence of Theorem 1 and the theory of minimal cost flows. First we define an adjacency matrix in which each woman enters a zero for a man she doesn't like and positive integers $1, 2, 3, \dots$ for the men she does like in decreasing order (if there is a tie she may enter the average of the assigned integer weights).

THEOREM 2. *Assume that a weighted compatibility relationship between a set of m women and a set of n men is described by an $m \times n$ adjacency matrix and suppose that at most d women can be assigned to one man. If for every positive integer $p \equiv 1 \bmod d$, $p \leq m$, the adjacency matrix has no $p \times q$ block of zeros with $p + dq > nd$ (nor can such a block be created by rearrangement of the rows and/or columns), then a complete matching that minimizes the sum of the assigned weights from the adjacency matrix can always be achieved.*

The minimum "cost" matching can be found by the modified Ford-Fulkerson algorithm described in [2]. Thus one could assign the students to the tutorial groups so as to maximize their collective satisfaction.

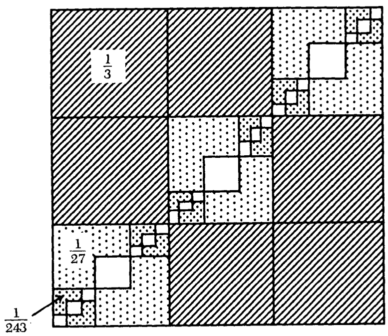
Conclusion The standard assignment problem, solved in operations research by either the transportation method or the faster Hungarian algorithm [7], involves matching one person with one task; several individuals are never assigned to the same task in this formulation. We have developed here the theory and methodology that can be employed to solve the generalized assignment problem where several individuals may be assigned to the same task, i.e., the so-called polygamous case.

REFERENCES

1. Behzad, Mehdi, Gary Chartrand and Linda Lesniak-Foster, *Graphs and Digraphs*, 1st edition, Prindle, Weber and Schmidt, Boston, 1979, pp. 159–162 and 378–382.
2. Busacker, Robert G. and Thomas L. Saaty, *Finite Graphs and Networks*, McGraw Hill, New York, 1965, pp. 255–260.

3. Gould, Ronald, *Graph Theory*, Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1988, pp. 99–112.
4. Halmos, P. R. and H. E. Vaughan, The marriage problem, *Amer. J. Math* 72 (1950), pp. 214–215.
5. Mirshy, L. *Transversal Theory*, Academic Press, New York, 1971.
6. Papadimitrion, C. H. and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, Englewood Cliffs, NJ.
7. Render, Barry and Ralph M. Stan Jr., *Quantitative Analysis for Management*, Allyn and Bacon, Boston, 1991, pp. 546–555.
8. Roberts, Fred S., *Applied Combinatorics*, Prentice Hall, Englewood Cliffs, NJ, 1984, pp. 549.
9. Strang, Gilbert, *Linear Algebra and Its Applications*, Harcourt, Brace, Javanovich, San Diego, 1988, pp. 424–429.
10. Tomescu, Joan and Robert A. Melter (translator) *Problems in Combinatorics and Graph Theory*, Wiley Interscience, New York, 1985, pp. 226–245.

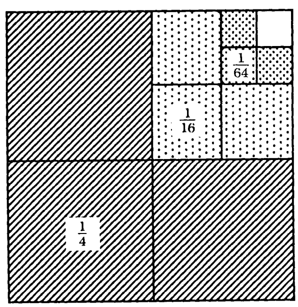
Proof without Words: Geometric Series



$$2\left(\frac{1}{3} + 3\cdot\frac{1}{27} + 9\cdot\frac{1}{243} + \cdots\right) = 1$$

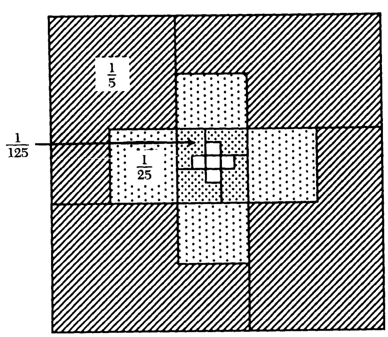
$$2\sum_{n=1}^{\infty}\frac{1}{3^n} = 1$$

$$\sum_{n=1}^{\infty}\frac{1}{3^n} = \frac{1}{2}$$



$$3\sum_{n=1}^{\infty}\frac{1}{4^n} = 1$$

$$\sum_{n=1}^{\infty}\frac{1}{4^n} = \frac{1}{3}$$



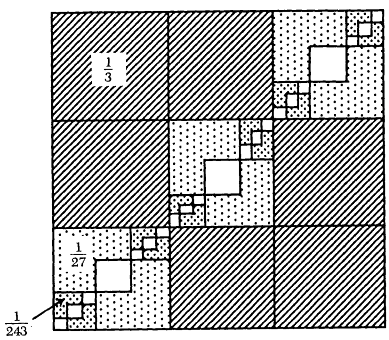
$$4\sum_{n=1}^{\infty}\frac{1}{5^n} = 1$$

$$\sum_{n=1}^{\infty}\frac{1}{5^n} = \frac{1}{4}$$

—Elizabeth M. Markham
Saint Joseph College
West Hartford, CT 06117

3. Gould, Ronald, *Graph Theory*, Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1988, pp. 99–112.
4. Halmos, P. R. and H. E. Vaughan, The marriage problem, *Amer. J. Math* 72 (1950), pp. 214–215.
5. Mirshy, L. *Transversal Theory*, Academic Press, New York, 1971.
6. Papadimitrion, C. H. and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, Englewood Cliffs, NJ.
7. Render, Barry and Ralph M. Stan Jr., *Quantitative Analysis for Management*, Allyn and Bacon, Boston, 1991, pp. 546–555.
8. Roberts, Fred S., *Applied Combinatorics*, Prentice Hall, Englewood Cliffs, NJ, 1984, pp. 549.
9. Strang, Gilbert, *Linear Algebra and Its Applications*, Harcourt, Brace, Javanovich, San Diego, 1988, pp. 424–429.
10. Tomescu, Joan and Robert A. Melter (translator) *Problems in Combinatorics and Graph Theory*, Wiley Interscience, New York, 1985, pp. 226–245.

Proof without Words: Geometric Series



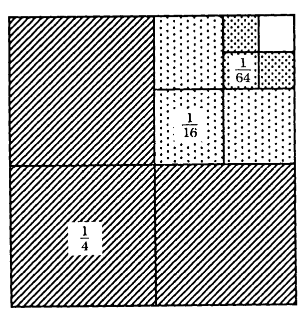
$$2\left(\frac{1}{3} + 3\cdot\frac{1}{27} + 9\cdot\frac{1}{243} + \cdots\right) = 1$$

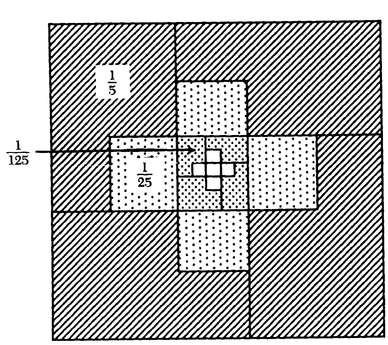
$$2\sum_{n=1}^{\infty}\frac{1}{3^n} = 1$$

$$\sum_{n=1}^{\infty}\frac{1}{3^n} = \frac{1}{2}$$

$$3\sum_{n=1}^{\infty}\frac{1}{4^n} = 1$$

$$\sum_{n=1}^{\infty}\frac{1}{4^n} = \frac{1}{3}$$





$$4\sum_{n=1}^{\infty}\frac{1}{5^n} = 1$$

$$\sum_{n=1}^{\infty}\frac{1}{5^n} = \frac{1}{4}$$

—Elizabeth M. Markham
Saint Joseph College
West Hartford, CT 06117

On the Limitations of a Well-Known Integration Technique

DENIS BELL
University of North Florida
Jacksonville, FL 32216

Most students of mathematics are familiar with the following classical technique for evaluating the improper integral

$$I=\int_0^\infty e^{-x^2}\,dx.$$

One writes the square of I as a double integral in the plane and transforms to polar coordinates to obtain

$$I^2=\int_0^\infty\int_0^\infty e^{-(x^2+y^2)}\,dx\,dy=\int_0^\infty\int_0^{\pi/2}e^{-r^2}r\,d\theta\,dr.$$

The integral on the right can now be evaluated to give $I^2=\pi/4$; thus $I=\sqrt{\pi}/2$. This is a beautiful argument, but one never encounters it outside of this one example. We address here the natural question of whether the same argument is applicable in a wider context.

Consider an integrable function f on $[0,\infty)$ and let $J=\int_0^\infty f(x)\,dx$. In order to integrate f by the above method there must exist functions g and h such that

$$f(x)f(y)=g(x^2+y^2)h(y/x),\qquad x>0,\quad y>0. \tag{1}$$

In this case, the foregoing argument will yield

$$J^2=\frac{1}{2}\int_0^\infty g(u)\,du\int_0^{\pi/2}h(\tan\theta)\,d\theta, \tag{2}$$

whence J can be determined, provided each of the above integrals can be evaluated.

One might hope that this scheme would allow for the integration of a large class of functions. However this turns out not to be the case. We analyze a class of functions f that admit a representation in the form (1) and derive the following:

THEOREM. *Let f, g, and h be continuous functions satisfying* (1). *Suppose there exists a positive real number d such that*

$$L=\lim_{x\to0^+}x^{-d}f(x)\neq0. \tag{3}$$

Then there exist constants A and c such that

$$f(x)=Ax^d\exp(cx^2).$$

Proof. Note that $h(1)\neq0$; otherwise (1) would give $f\equiv0$ contradicting (3). Since the form of g and h will not change if each is multiplied by a nonzero constant, we may assume that $h(1)=1$. Setting $y=x$ in (1) we obtain $f^2(x)=g(2x^2)$ and hence $g(z)=f^2(\sqrt{z/2})$ for $z\geqslant0$.

Substituting into (1) $f(x)f(y)=f^2(\sqrt{(x^2+y^2)/2})h(y/x)$. We define a function $\tilde{f}$ by $\tilde{f}(x)=x^{-d}f(x)$ if $x>0$ and $\tilde{f}(0)=L$. Note that $\tilde{f}$ is continuous at 0 by (3). Then

$$(2xy)^d\tilde{f}(x)\tilde{f}(y)=(x^2+y^2)^d\tilde{f}^2\left(\sqrt{(x^2+y^2)/2}\right)h(y/x). \tag{4}$$

Setting $y=tx$ for $t\geqslant 0$ and $x>0$ yields

$$(2t)^d\tilde{f}(x)\tilde{f}(tx)=(1+t^2)^d\tilde{f}^2\left(x\sqrt{(1+t^2)/2}\right)h(t).$$

We let x tend to zero and conclude that

$$h(t)=\left(\frac{2t}{1+t^2}\right)^d.$$

Substituting this into (4) gives

$$\tilde{f}(x)\tilde{f}(y)=\tilde{f}^2\left(\sqrt{(x^2+y^2)/2}\right) \quad \text{or equivalently}$$

$$r(x^2)r(y^2)=r^2((x^2+y^2)/2), \quad \text{where } \tilde{f}(x)=r(x^2).$$

Note that $r(0)\neq 0$; otherwise we would have $r\equiv 0$, implying $f\equiv 0$. Setting $y=0$, we get $r(x^2)=r^2(x^2/2)/r(0)$. Thus r satisfies

$$r^2(x^2/2)r^2(y^2/2)=r^2(x^2/2+y^2/2)r^2(0). \tag{5}$$

Since r^2 is continuous it follows from (5) that r^2 is an exponential function (see, e.g. [1]). Thus r is an exponential function, and this implies that f has the stated form.

We now return to our original question. The Theorem shows that the only functions (satisfying a condition of type (3)) amenable to integration via the foregoing technique are those of the form $Ax^d\exp(cx^2)$, where c will have to be negative to allow integrability. The corresponding functions g and h in (1) can be $g(x)=A^2x^d\exp(cx)$ and $h(x)=\{x/(1+x^2)\}^d$. From (2) we obtain the identity

$$J^2=A^2|2c|^{-(d+1)}\Gamma(d+1)\textstyle\int_0^{\pi/2}\sin^d\theta\,d\theta,$$

where Γ denotes the gamma function. However it is only possible to compute these integrals in closed form by elementary methods if d is an integer, and in this case J can be evaluated in terms of I by means of iterated integration by parts!

Final remark. An easy modification of the above argument shows that if f_1, f_2, g, and h are continuous functions satisfying the functional equation $f_1(x)f_2(y)=g(x^2+y^2)h(y/x)$ and f_1 and f_2 satisfy a condition of type (3) then f_1 and f_2 are both of the form $Ax^d\exp(cx^2)$, with the same values of c and d. Thus we prove that if under a change from Cartesian to polar coordinates, a product function $f_1(x)f_2(y)$ of x and y transforms to a product function $g(r)h(\theta)$ of r and θ then (modulo multiplicative constants) the functions have the form: $f_1(x)=f_2(x)=x^d\exp(cx^2)$, $g(r)=r^{2d}\exp(cr^2)$, $h(\theta)=(\cos\theta\sin\theta)^d$.

REFERENCE

1. S. Ross, *Introduction to Probability Models*, Academic Press, New York, 1972, p. 111.

A New Proof of the Arithmetic-Geometric Mean Inequality

MIHÁLY BENCZE
Braşov, Romania

NORMAN SCHAUMBERGER
Hofstra University
Hempstead, NY 11550

THEOREM. *If $x_k \geq 0$ for $k = 1, 2, \dots, n$ then we have the following inequality:*

$$x_n^n \geq x_1(2x_2 - x_1)(3x_3 - 2x_2) \cdots (nx_n - (n-1)x_{n-1})$$

with equality if and only if $x_1 = x_2 = \cdots = x_n$.

Proof. We first prove that if $x_k \geq 0$, then

$$x_k^k \geq (kx_k - (k-1)x_{k-1})x_{k-1}^{k-1} \tag{1}$$

with equality if, and only if, $x_k = x_{k-1}$. If $x_k \geq x_{k-1}$, then

$$x_k^{k-1} + x_k^{k-2}x_{k-1} + x_k^{k-3}x_{k-1}^2 + \cdots + x_{k-1}^{k-1} \geq kx_{k-1}^{k-1}$$

and if $x_k \leq x_{k-1}$ the inequality is reversed. In either case

$$x_k^k - x_{k-1}^k \geq kx_{k-1}^{k-1}(x_k - x_{k-1}), \tag{2}$$

with equality if and only if $x_k = x_{k-1}$. Clearly, (1) follows from (2). We now substitute $k = n, n-1, \dots, 2$, successively into (1), to get

$$x_n^n \geq x_1(2x_2 - x_1)(3x_3 - 2x_2) \cdots (nx_n - (n-1)x_{n-1}) \tag{3}$$

with equality if, and only if, $x_k = x_{k-1}$ $(k = 2, \dots, n)$.

Remark. If $a_k \geq 0$ for $k = 1, 2, \dots, n$ then the Arithmetic-Geometric Mean Inequality

$$1/n(a_1 + a_2 + \cdots + a_n) \geq \sqrt[n]{a_1 a_2 \dots a_n} \tag{4}$$

is obtained by substituting $x_k = 1/k(a_1 + a_2 + \cdots + a_k)$ into (3). Equation (4) holds with equality if and only if $a_1 = a_2 = \cdots = a_n$.

Did the Young Volterra Know About Cantor?

DAVID GAULD
University of Auckland
Auckland, New Zealand

In [3], Dunham recalls Volterra's proof [5] that any two functions from $\mathbb{R}$ to itself that are continuous on dense subsets of $\mathbb{R}$ have a common point of continuity. The only use of the completeness axiom for $\mathbb{R}$ in the proof is the nested intervals theorem. Actually Volterra showed more than just that there is one common point of continuity: He showed that every open interval contains a common point of continuity, i.e. that the set of common points of continuity forms a dense subset of $\mathbb{R}$. In this note even more is shown, viz. that the set of common points of continuity is an uncountable dense subset of $\mathbb{R}$ and again the only use of the completeness axiom is the nested intervals theorem.

It would be interesting to know whether Volterra, who was not yet 20 in February 1880 when he wrote [5], thought about whether the set of common points of continuity is uncountable. On the one hand, it might be guessed that he did not because the proof of the result requires little beyond his proof, just a little extra care. On the other hand, G. F. L. P. Cantor (1845–1918) had developed the notion of countability eight years before Volterra's work appeared. While he was still refining the notion of cardinality even up to 1881, already in 1874 Cantor had shown in [1] that the set of all real numbers between any two reals is not countable using a nested intervals argument. On balance it would seem likely that Volterra was not aware of Cantor's work on countability. A good early discussion of Cantor's work on the theory of transfinite numbers is [2].

It is now proved that if $f, g\colon \mathbb{R} \to \mathbb{R}$ are two functions and F, G are two subsets of $\mathbb{R}$ satisfying:

(i) $F = \{x \in \mathbb{R} \mid f \text{ is continuous at } x\}$;

(ii) $G = \{x \in \mathbb{R} \mid g \text{ is continuous at } x\}$;

(iii) F and G are dense in $\mathbb{R}$,

then $F \cap G$ is an uncountable dense subset of $\mathbb{R}$.

Let (a_0, b_0) be an interval in $\mathbb{R}$ and let C be any countable subset of $\mathbb{R}$. Since C is countable we may write $C = \{c_n \mid n = 1, 2, \dots\}$. It will be shown that

$$(a_0, b_0) \cap F \cap G - C \neq \varnothing,$$

from which the conclusion will then be deduced.

Use induction to define two sequences (a_n) and (b_n) with (a_n) strictly increasing and (b_n) strictly decreasing so that for all n,

(1) $a_n < b_n$;

(2) $c_n \notin (a_n, b_n)$;

(3) if n is odd then $|f(x) - f(y)| < 1/n$ for each $x, y \in (a_n, b_n)$;

(4) if n is even then $|g(x) - g(y)| < 1/n$ for each $x, y \in (a_n, b_n)$.

Induction has already been started at 0. Suppose m is such that a_n and b_n have been defined for each $n < m$. If m is odd, then since F is dense in $\mathbb{R}$ there is a number $a_m \in (a_{m-1}, b_{m-1}) \cap F$; it may be assumed that if $c_m < b_{m-1}$, then $a_m > c_m$. Since f is continuous at a_m there is a $\delta > 0$ so that if x is within δ of a_m then $|f(x) - f(a_m)| < 1/2m$. Choose b_m so that $a_m < b_m < b_{m-1}$ and $b_m < a_m + \delta$. Then for each $x, y \in (a_m, b_m)$ we have

$$|f(x)-f(y)| \leqslant |f(x)-f(a_m)|+|f(y)-f(a_m)| < \frac{1}{2m}+\frac{1}{2m}=\frac{1}{m}.$$

If m is even, then construct a_m and b_m similarly, replacing F by G and f by g.

The intervals $[a_n, b_n]$, $n=0,1,\dots$ are nested so there is $a \in \mathbb{R}$ so that $a \in [a_n, b_n]$ for each n. From $a_n < a_{n+1} \leqslant a \leqslant b_{n+1} < b_n$ it follows that $a_n < a < b_n$ for all n; in particular $a \in (a_0, b_0)$. The point a is in F, i.e. f is continuous at a, for suppose $\varepsilon > 0$. Let n be an odd integer with $n \geqslant 1/\varepsilon$. Let δ be the smaller of $a - a_n$ and $b_n - a$. By (3) if $|x-a| < \delta$ then $|f(x)-f(a)| < 1/n \leqslant \varepsilon$. Since $\varepsilon > 0$ is arbitrary, it follows that f is continuous at a so $a \in F$. Similarly $a \in G$. Finally, $a \neq c_n$ for each n as $a \in (a_n, b_n)$ but by (2) $c_n \notin (a_n, b_n)$. Thus $(a_0, b_0) \cap F \cap G - C \neq \varnothing$.

As $(a_0, b_0) \cap F \cap G \neq \varnothing$, it follows that $F \cap G$ is dense in $\mathbb{R}$. Furthermore, $F \cap G$ cannot be countable because if it were then we could take $C = F \cap G$ and obtain a contradiction as it has been shown that $F \cap G - C \neq \varnothing$.

Remark 1. As there is a function $\mathbb{R} \to \mathbb{R}$ that is continuous precisely at the irrational numbers, it may be concluded that every function $\mathbb{R} \to \mathbb{R}$ that is continuous at each rational number is also continuous on an uncountable dense set of irrational numbers.

Remark 2. By taking $f = g$ in the statement of our main result we can deduce that if the set of points at which a function $f\colon \mathbb{R} \to \mathbb{R}$ is continuous is dense in $\mathbb{R}$ then that set is uncountable. This gives an alternative proof of the conclusion of Remark 1. This result holds in a more general setting by appeal to the Baire category theorem; see for example problem 12 on pages 130–131 of [4].

Remark 3. We cannot hope to go any further with the argument presented here. For example, we cannot hope that the complement of $F \cap G$ or F is countable: A function that is 0 on an open dense subset of $\mathbb{R}$ and 1 elsewhere is continuous precisely on this open dense set but the complement may be uncountable (e.g. the Cantor ternary set).

Remark 4. Volterra's part of this argument translates to compact metric spaces: If two functions on a compact metric space have dense sets of continuity then the intersection of these two sets is also dense.

REFERENCES

1. G. Cantor, Ueber eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen, *Borchardt J.* 77 (1874), 258–263.
2. G. Cantor, Contributions to the founding of the theory of transfinite numbers, translated and with an introduction by P. Jourdain, *Open Court*, 1915. Reprinted by Dover Publications, Mineola, NY, 1952.
3. W. Dunham, A historical gem from Vito Volterra, this MAGAZINE 63 (1990), 234–237.
4. A. E. Taylor, *General Theory of Functions and Integration*, Blaisdell, New York, 1965.
5. V. Volterra, Alcune osservasioni sulle funzioni punteggiate discontinue, *Giornale di Matematiche* 19 (1881), 76–86. Reproduced in *Opere Matematiche di Vito Volterra*, Accademia Nazionale dei Lincei, Roma, 1954, Vol. 1, 7–15.

Classification of Finite Rings of Order p^2

BENJAMIN FINE
Fairfield University
Fairfield, CT 06430

Most treatments of elementary abstract algebra include a discussion of finite groups and some work on their classification. However, very little is done with finite rings. For example most beginning texts state and prove the theorem that for p a prime the cyclic group of order p is the only group of order p up to isomorphism. Yet the equally striking and easily proved result that for a prime p there are only two rings of order p up to isomorphism is either not mentioned at all or relegated to the exercises.

The purpose of this note is to give a complete classification of all finite rings of order p^2 with p a prime. In particular, we show that up to isomorphism there are exactly 11 rings of order p^2. The techniques are elementary and grew out of a project given to an undergraduate abstract algebra course. We use a concept called a *ring presentation*, which is an excellent computational tool for dealing with finite rings. After explaining this concept, we state our main result, Theorem 2, which can be given as a large project to a good undergraduate class with guidance provided by the instructor.

1. Presentation If R is a finite ring then its additive group is a finite abelian group and is thus a direct product of cyclic groups. Suppose these have generators $g_1, \dots, g_k$ of orders $m_1, \dots, m_k$. Then the ring structure is determined by the k^2 products

$$g_i g_j = \sum_{t=1}^{k} c_{ij}^t g_t \quad \text{with } c_{ij}^t \in \mathbf{Z}_{m_t} \tag{1}$$

and thus by the k^3 structure constants c_{ij}^t. We introduce a convenient notation, motivated by group theory, for giving the structure of a finite ring. A *presentation* for a finite ring R consists of a set of generators $g_1, \dots, g_k$ of the additive group of R together with *relations*. The relations are of two types:

(i) $m_i g_i = 0$ for $i = 1, \dots, k$ indicating the additive order of g_i, and

(ii) $g_i g_j = \sum_{t=1}^{k} c_{ij}^t g_t$ with $c_{ij}^t \in \mathbf{Z}_{m_t}$

for $i = 1, \dots, k$; $j = 1, \dots, k$; $t = 1, \dots, k$.

If the ring R has the presentation above we write

$$R = \left\langle g_1, \dots, g_k;\ m_i g_i = 0 \text{ for } i = 1, \dots, k,\ g_i g_j = \sum_{i=1}^{k} c_{ij}^t g_i \right\rangle.$$

For example the ring $\mathbf{Z}_2 + \mathbf{Z}_2 = \langle a, b;\ 2a = 2b = 0,\ a^2 = a,\ b^2 = b,\ ab = ba = 0 \rangle$, while the finite field of order 4 is $\langle a, b;\ 2a = 2b = 0,\ a^2 = a,\ ab = b,\ b^2 = a + b \rangle$. Notice that if the additive group is cyclic with generator g, the ring structure is completely determined by g^2. Therefore the ring $\mathbf{Z}_4 = \langle a;\ 4a = 0,\ a^2 = a \rangle$.

Finally if a relation follows by applying the ring properties to other relations, we delete it. For example suppose that a ring R is generated by a and b having prime additive orders p and q. If $a^2 = 0$ and $b^2 = 0$ it follows that $ab = 0$ and $ba = 0$, so these relations are deleted. To see this last fact notice that if $ab = ta + ub$ then $a^2b = 0 = ta^2 + uab = uta + u^2b$. Since a, b constitute an additive basis it follows that $u^2 \equiv 0 \pmod q$ and since q is a prime we must then also have $u \equiv 0 \pmod q$. Similarly, by calculating ab^2 we deduce that $t \equiv 0 \pmod p$.

2. Cyclic additive groups We first present a result that characterizes rings with cyclic additive groups. This appeared in [**2**].

THEOREM 1. *The number of rings R, up to isomorphism, with cyclic additive group C_m is given by the number of divisors of m. In particular, for each divisor d of m there is a ring $R_d = \langle g;\ mg = 0,\ g^2 = dg\rangle$ where g is an additive generator of C_m. For different d's these rings are nonisomorphic.*

Proof. Let R be a ring with additive group C_m and let g be an additive generator of C_m. Suppose $g^2 = ng$. If $(m, n) = 1$ then n has an inverse k modulo m so that $nk \equiv 1 \pmod m$. Let $g_1 = kg$. Since k is a unit in $\mathbf{Z}_m$, g_1 is also an additive generator. Now $g_1^2 = (kg)^2 = k^2g^2 = k^2(ng) = k(kng) = kg = g_1$. So the homomorphism from R to $\mathbf{Z}_m$ defined by $g_1 \to 1$ is an isomorphism. Therefore in this case $R = R_1$ is isomorphic to $\mathbf{Z}_m$.

Suppose $g^2 = 0$. Then all the multiplication is trivial and in this case $R = R_m$ is isomorphic to the ring with additive group C_m and trivial multiplication. These two possibilities correspond to the divisors 1 and m of m.

Now let d be a proper divisor of m with $m = kd$ and suppose $g^2 = kg$. Observe that kg generates a unique additive cyclic subgroup of order d. The generators of this subgroup are $(jk)g$ where $j = 0, 1, \ldots, d-1$ and $(j, d) = 1$. We show that for any $j = 0, 1, \ldots, d-1$ and $(j, d) = 1$, R is isomorphic to $\langle g_1;\ mg_1 = 0,\ g_1^2 = (jk)g_1\rangle$ for some generator g_1 of C_m. To do so, we show that there is an n with $(m, n) = 1$ such that if $g_1 = ng$ then $g_1^2 = (jk)g_1$. Since $(n, m) = 1$, g_1 is then an additive generator and the map $g \to g_1$ gives the desired isomorphism.

Suppose $g_1 = ng$ with n to be determined. Then $g_1^2 = n^2g^2 = (n^2k)g$. If this is to equal $(jk)g_1 = (jkn)g$ we are led to the congruence:

$$(1) \qquad kn^2 \equiv jkn \pmod m, \text{ with } n \text{ the variable.}$$

Assuming n is to be a unit mod m, we get

$$(2) \qquad kn \equiv jk \pmod m.$$

The solutions modulo m are $n = j + td$ with $t = 0, 1, \ldots, k-1$. Since $(j, d) = 1$, by Dirichlet's theorem there exists a solution such that $j + td$ is prime to m. This solution n of (2) is then a unit mod m and gives the indicated n.

Thus the rings with additive groups C_m and generator g such that $g^2 = (jk)g$ with $kd = m$ and $(j, d) = 1$ all fall in one isomorphism class.

Notice further that if $g^2 = kg$ and $(j, d) \neq 1$, then for any solution to (1) or (2) we would have $(j + td, m) \neq 1$ and therefore there is no unit solution. This implies that there exists no additive generator g_1 such that $g_1^2 = (jk)g_1$ and thus no isomorphism. Therefore the rings whose presentations are given in terms of divisors of m of different additive orders are precisely the isomorphism classes. This completes the proof of the theorem.

We immediately have the following corollaries classifying rings of prime order and rings of order pq where p and q are distinct primes. For an abelian group G we let $G(0)$ denote the ring with additive group G and trivial multiplication.

COROLLARY 1. *If p is a prime there are, up to isomorphism, exactly two rings of order p, namely $\mathbf{Z}_p$ and $C_p(0)$.*

COROLLARY 2. *If p and q are distinct primes there are, up to isomorphism, exactly four rings of order pq. These are $\mathbf{Z}_{pq}$, $C_{pq}(0)$, $C_p(0)+\mathbf{Z}_q$, and $\mathbf{Z}_p+C_q(0)$.*

More generally if n is a square-free positive integer and R is a ring of order n, then the additive group of R must be cyclic. The following corollary follows immediately.

COROLLARY 3. *If $n=p_1\dots p_k$ is a square-free positive integer with k distinct prime divisors then there are, up to isomorphism, exactly 2^k rings of order n.*

3. Rings of order p^2 We now give our main result. Notice that if a ring has order p^2 then its additive group is either C_{p^2} or $C_p\times C_p$.

THEOREM 2. *For any prime p there are, up to isomorphism, exactly 11 rings of order p^2. Specifically these are given by the following presentations:*

$$A=\langle a;\ p^2a=0,\ a^2=a\rangle=\mathbf{Z}_{p^2}$$

$$B=\langle a;\ p^2a=0,\ a^2=pa\rangle$$

$$C=\langle a;\ p^2a=0,\ a^2=0\rangle=C_{p^2}(0)$$

$$D=\langle a,b;\ pa=pb=0,\ a^2=a,\ b^2=b,\ ab=ba=0\rangle=\mathbf{Z}_p+\mathbf{Z}_p$$

$$E=\langle a,b;\ pa=pb=0,\ a^2=a,\ b^2=b,\ ab=a,\ ba=b\rangle$$

$$F=\langle a,b;\ pa=pb=0,\ a^2=a,\ b^2=b,\ ab=b,\ ba=a\rangle$$

$$G=\langle a,b;\ pa=pb=0,\ a^2=0,\ b^2=b,\ ab=a,\ ba=a\rangle$$

$$H=\langle a,b;\ pa=pb=0,\ a^2=0,\ b^2=b,\ ab=ba=0\rangle=\mathbf{Z}_p+C_p(0)$$

$$I=\langle a,b;\ pa=pb=0,\ a^2=b,\ ab=0\rangle$$

$$J=\langle a,b;\ pa=pb=0,\ a^2=b^2=0\rangle=C_p\times C_p(0)$$

$$K=GF(p^2)=\textit{finite field of order } p^2$$

$$=\begin{cases}\langle a,b;\ pa=pb=0,\ a^2=a,\ b^2=ja,\ ab=b,\ ba=b\rangle \\ \qquad \textit{where } j \textit{ is not a square in } \mathbf{Z}_p, \quad \textit{if } p\neq 2. \\ \langle a,b;\ 2a=2b=0,\ a^2=a,\ b^2=a+b,\ ab=b,\ ba=b\rangle, \quad \textit{if } p=2.\end{cases}$$

Proof. Let R be a ring of order p^2. Then the additive group is isomorphic to C_{p^2} or $C_p\times C_p$. If the additive group is C_{p^2} then from Theorem 1 there are three rings up to isomorphism whose presentations are given by:

$$A=\langle a;\ p^2a=0,\ a^2=a\rangle=\mathbf{Z}_{p^2}$$

$$B=\langle a;\ p^2a=0,\ a^2=pa\rangle$$

$$C=\langle a;\ p^2a=0,\ a^2=0\rangle=C_{p^2}(0).$$

We now concentrate on rings whose additive group is $C_p\times C_p$. In this case R is a vector space of dimension 2 over the finite field $\mathbf{Z}_p$. Therefore if a' and b' are additive generators of R and $a=xa'+yb'$, $b=wa'+zb'$, then a and b are also additive generators whenever $xw-zy\neq 0 \pmod p$.

To obtain the complete classification we show that given a set of additive generators a' and b' for R there exists a (possibly distinct) set of generators a and b such that R equals D, E, F, G, H, I, J, or K. At the same time we show that no two of these rings are isomorphic. This procedure involves an enumeration of cases involving a and b. These cases, in turn, break into two large groups. In the first R contains a set of additive generators a and b whose squares a^2 and b^2 are multiples of themselves—that is $a^2 = ma$ and $b^2 = nb$ with m, n in $\mathbf{Z}_p$. (m or n or both may be zero.) In the second set of cases R has no set of additive generators whose squares are multiples of themselves. We present part of the first set of cases in detail to illustrate the process and then sketch the remainder.

Suppose first that there exist generators a and b such that $a^2 = ma$, $b^2 = nb$, $m \not\equiv 0 \pmod p$, and $n \not\equiv 0 \pmod p$. Then a and b generate subrings of R isomorphic to $\mathbf{Z}_p$. Thus without loss of generality we may assume that $m = n = 1$ so that we have generators a and b with $a^2 = a$ and $b^2 = b$. Suppose then that $ab = ta + ub$. Then $a^2 b = ab = ta^2 + uab = (t + ut)a + u^2 b = ta + ub$. It follows that $u^2 \equiv u \pmod p$ so that $u \equiv 0 \pmod p$ or $u \equiv 1 \pmod p$.

Similarly by considering $ab^2 = ab$ we see that $t^2 \equiv t \pmod p$, hence either $t \equiv 0 \pmod p$ or $u \equiv 0 \pmod p$. This gives four possibilities for (t, u) namely $(0,0)$, $(1,0)$, $(0,1)$, or $(1,1)$.

If $(t, u) = (1,1)$ then $ab = a + b$. Then $a^2 b = ab = a(a+b) = a^2 + ab = 2a + b \neq a + b = ab$. Therefore the case $t = 1$, $u = 1$ is impossible and so there are only three possibilities for (t, u); namely $(0,0)$, $(1,0)$, and $(0,1)$.

By a symmetrical argument if $ba = xa + yb$ there are three possibilities for (x, y) again; namely $(0,0)$, $(1,0)$, and $(0,1)$. Thus if $a^2 = a$ and $b^2 = b$ there are nine possibilities for ab and ba.

Case 1. $a^2 = a$, $b^2 = b$, $ab = 0$, $ba = 0$. In this case $R = D = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = ba = 0 \rangle$, and so R is isomorphic to $\mathbf{Z}_p + \mathbf{Z}_p$ under the map $a \to (1,0), b \to (0,1)$.

Case 2. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = a, ba = a \rangle$. This is isomorphic to $\mathbf{Z}_p + \mathbf{Z}_p$, which we denoted D under the map $a \to (1,0), b \to (1,1)$.

Case 3. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = b, ba = b \rangle$. This case is symmetric to case 2 above and therefore this R is also isomorphic to $D = \mathbf{Z}_p + \mathbf{Z}_p$.

Case 4. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = a, ba = 0 \rangle$. In this case $aba = a^2 = a \neq 0$ since a is a generator. However $aba = a \cdot 0 = 0$ and so this case is impossible.

By the same arguments the following three cases are impossible.

Case 5. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = b, ba = 0 \rangle$.

Case 6. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = 0, ba = a \rangle$.

Case 7. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = 0, ba = b \rangle$.

We now consider

Case 8. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = a, ba = b \rangle$. This is a legitimate possibility in which R is not isomorphic to $\mathbf{Z}_p + \mathbf{Z}_p$ since R is noncommutative. We call this new ring E.

The final case is the following:

Case 9. $R = \langle a, b : pa = pb = 0, a^2 = a, b^2 = b, ab = b, ba = a \rangle$. This R is not isomorphic to $\mathbf{Z}_p + \mathbf{Z}_p$ since R is noncommutative. We claim that R is also not isomorphic to E. Call this ring F. We show that there are no elements in F that satisfy the relations of ring E. Let $A = ma + nb$ where at least one of m and n is nonzero. Suppose $A^2 = A$. Using the relations in F, we have $(m^2 + mn)a + (n^2 + mn)b = ma + nb$. This implies that $m^2 + mn \equiv m \pmod p$ and $n^2 + mn \equiv n \pmod p$. If $m = 0$ then $n^2 = n$ and so $n = 1$. Similarly, if $n = 0$ then $m = 1$. If $m \neq 0$ then $m(n + m) = M$, and so $n + m = 1$. Therefore if $A^2 = A$ we must have $A = a$ or $A = b$ or $A = na + (1 - n)b$ for some $n \neq 0$, 1. Similarly, if B is independent from A, we must have either $B = a$ or $B = b$ or $B = xa + (1 - x)b$ for some $x \neq 0$, 1. In case $A = a$ and $B = b$, $AB = ab = b \neq A$, and so A and B do not satisfy the relations of E. Similarly in case $A = b$ and $B = a$. In case $A = a$ and $B = xa + (1 - x)b$ with $x \neq 0$, 1, $AB = a(xa + (1 - x)b) = xa^2 + (1 - x)ab = xa + (1 - x)b = B \neq A$, and so A and B do not satisfy the relations of E. The result is similar in case $A = b$ and $B = xa + (1 - x)b$, in case $B = a$ and $A = na + (1 - n)b$, and in case $B = b$ and $A = na + (1 - n)b$. Therefore the only case in which we could get the ring E is the one in which $A = na + (1 - n)b$ for some $n \neq 0$, 1 and $B = xa + (1 - x)b$ for some $x \neq 0$, 1. Suppose then that $AB = A$ as it would be in E. By computing we find that AB also equals B, and so $A = B$, which contradicts the fact that A and B are independent. Therefore F is not isomorphic to E. Thus F is another ring with additive group $C_p \times C_p$.

Cases 1 through 9 describe the possibilities in which $a^2 = a$ and $b^2 = b$ and give us three additional nonisomorphic rings D, E, and F.

We now sketch the remainder of the proof. The details are carried out as in the above cases—possible presentations for R are identified and then shown to either equal a ring that is isomorphic to one already on the list or to be a new nonisomorphic ring.

For instance, consider the situation in which R has a set of additive generators a and b with one of their squares zero and the other a multiple of itself. If $a^2 = 0$ and $b^2 = b$, two new nonisomorphic rings, G and H, are obtained:

$$G = \langle a, b;\ pa = pb = 0,\ a^2 = 0,\ b^2 = b,\ ab = a,\ ba = a \rangle$$
$$H = \langle a, b;\ pa = pb = 0,\ a^2 = 0,\ b^2 = b,\ ab = 0,\ ba = 0 \rangle.$$

G is commutative and H is isomorphic to $\mathbf{Z}_p + C_p(0)$.

In the final situation R has no set of two independent generators whose squares are both nonzero multiples of themselves. If R has a generator a with $a^2 = b$ and b independent from a, then an enumeration of cases leads to two new additional nonisomorphic rings, I and K:

$$I = \langle a, b;\ pa = pb = 0,\ a^2 = b,\ ab = 0 \rangle$$
$$K = \begin{cases} \langle a, b;\ pa = pb = 0,\ a^2 = a,\ ab = b,\ b^2 = ja \text{ for some } j \text{ in } \mathbf{Z}_p \rangle, & \text{if } p \neq 2. \\ \langle a, b;\ 2a = 2b = 0,\ a^2 = a,\ b^2 = a + b,\ ab = b,\ ba = b \rangle, & \text{if } p = 2. \end{cases}$$

In both cases K is precisely the finite field $GF(p^2)$.

If R has two generators both of whose squares are trivial, then the multiplication is trivial and so $R = C_p \times C_p(0) = J$.

We mention in closing that the group ring $\mathbf{Z}_p(C_2)$, which also has order p^2, is isomorphic to $\mathbf{Z}_p + \mathbf{Z}_p$. Identifying 1 in $\mathbf{Z}_p$ with the generator a and the group generator of C_2 with b shows that the group ring has the presentation $\langle a, b;$ $pa = pb = 0,\ a^2 = a,\ b^2 = a,\ ab = b,\ ba = b \rangle$. The map $a \to (1, -1), b \to (1, -1)$ gives the desired isomorphism.

REFERENCES

1. J. A. Gallian, *Contemporary Abstract Algebra*, D. C. Heath and Company, Lexington, MA, 1990.
2. W. C. Waterhouse, Rings with cyclic additive group, *Amer. Math. Monthly* 71 (1964), 449–450.

A Relation Between the Jacobian and Certain Generalized Wronskians

KENNETH WOLSSON
Fairleigh Dickinson University
Teaneck, NJ 07666

The purpose of this note is to call attention to an interesting identity involving the Jacobian of a set of functions and certain of its generalized Wronskians. This identity provides a compact solution to an interesting geometric problem and has been helpful in extending the work in [1] (see [2], [3] for related results).

Let the first partial derivatives of $\phi = (\phi_1(t), \dots, \phi_n(t))$ be assumed to exist at a point $t = (t_1, \dots, t_n)$. Let J denote the Jacobian of ϕ—the determinant of the matrix $[J]$ whose kth row consists of the components of $\partial\phi/\partial t_k$, and let W_j denote the determinant of the matrix constructed from $[J]$ as follows: Its first row consists of the components of ϕ. Below this we append in order, the rows of $[J]$ omitting the jth one. These determinants belong to a class called *generalized Wronskians* of ϕ ([1] p. 74, [2] pp. 129–132, [3] pp. 138–141).

We claim that if $\phi_i(t) \neq 0$,

$$J(\phi_1, \dots, \phi_n) = \frac{1}{\phi_i} \sum_j \pm \frac{\partial \phi_i}{\partial t_j} W_j, \tag{1}$$

where "$\pm$" denotes sign alteration starting with "$+$."

To prove (1) we construct a matrix $[K]$ as follows: First form a matrix $[M]$ with the components of ϕ as its first row. The next n rows of $[M]$ are those of $[J]$ in order. To complete $[K]$ we prefix $[M]$ with its own ith column. For example, if $n = 3$ and $i = 2$,

$$[K] = \begin{bmatrix} \phi_2 & \phi_1 & \phi_2 & \phi_3 \\ \dfrac{\partial\phi_2}{\partial t_1} & \dfrac{\partial\phi_1}{\partial t_1} & \dfrac{\partial\phi_2}{\partial t_1} & \dfrac{\partial\phi_3}{\partial t_1} \\ \dfrac{\partial\phi_2}{\partial t_2} & \dfrac{\partial\phi_1}{\partial t_2} & \dfrac{\partial\phi_2}{\partial t_2} & \dfrac{\partial\phi_3}{\partial t_2} \\ \dfrac{\partial\phi_2}{\partial t_3} & \dfrac{\partial\phi_1}{\partial t_3} & \dfrac{\partial\phi_2}{\partial t_3} & \dfrac{\partial\phi_3}{\partial t_3} \end{bmatrix}.$$

Equation (1) follows by expanding the determinant of $[K]$ by cofactors of its first column and noting that it must vanish.

The result is unusual in that the undifferentiated ϕ_i's only *seem* to appear on the right side yet must cancel since they are absent from the left. Also (1) implies that the expression on the right is independent of which ϕ_i is used. We can exploit this to solve the problem of finding the direction in which the functions $\psi_1(t), \dots, \psi_n(t) \in C^1$ have a common derivative at each interior point of their common domain. If in (1) we put $\phi_i = \exp\psi_i$ (hence $\phi_i \neq 0$) and $\overline{W}_j = W_j(\exp\psi_1, \dots, \exp\psi_n)$, then (1) tells us that $\langle \overline{W}_1, -\overline{W}_2, \dots, (-1)^{n+1}\overline{W}_n \rangle$ is the required direction in which $d\psi_i/ds = J(\exp\psi_1, \dots, \exp\psi_n)/\sqrt{\Sigma \overline{W}_j^2}$ $(i = 1, 2, \dots, n)$. The only exceptions are of course points at which all $\overline{W}_j$ vanish simultaneously.

We can confirm this result for the case $n = 2$. Here the direction sought is $\langle \overline{W}_1, -\overline{W}_2 \rangle$ with

$$\overline{W}_1 = \begin{vmatrix} \exp \psi_1 & \exp \psi_2 \\ \dfrac{\partial}{\partial t_2} \exp \psi_1 & \dfrac{\partial}{\partial t_2} \exp \psi_2 \end{vmatrix} \quad \text{and} \quad \overline{W}_2 = \begin{vmatrix} \exp \psi_1 & \exp \psi_2 \\ \dfrac{\partial}{\partial t_1} \exp \psi_1 & \dfrac{\partial}{\partial t_1} \exp \psi_2 \end{vmatrix},$$

which by computation can be seen to have the same direction as $\langle \partial(\psi_2 - \psi_1)/\partial t_2, -\partial(\psi_2 - \psi_1)/\partial t_1 \rangle$. We recognize this as the tangent direction to the curve along which $\psi_2 - \psi_1$ is constant and hence along which the derivatives of ψ_1 and ψ_2 are equal. The above mentioned exception occurs only at a point where both components of this direction vanish. At such a point the derivative of $\psi_2 - \psi_1$ vanishes and hence the derivatives of ψ_1 and ψ_2 are equal—in *all* directions.

REFERENCES

1. K. Wolsson, Linear dependence of a function set of m variables with vanishing generalized Wronskians, *Linear Algebra and Its Applications* 117 (1989) 73–80.
2. W. J. LeVeque, *Topics in Number Theory*, Addison-Wesley Publishing Co., Reading, MA, 1956.
3. W. Schmidt, Diophantine Approximation, *Lecture Notes in Mathematics*, No. 785, Springer-Verlag, NY, 1980.

A Surprising Application of the Integral $\int_0^T (1 - F(x))\, dx$ to Revenue Projection

ALLAN KROOPNICK
Social Security Administration
Baltimore, MD 21235

The Social Security Administration annually estimates revenue into its trust funds in order to project present and future benefits and compare those benefits to present and future revenues received. The amount of money flowing into the trust funds depends on the taxable maximum T, which was \$51,300 for 1990. That is, the higher the T, the more money is exposed to taxation and hence more flows into the trust funds. For general overviews of the process see **[1]** and **[2]**. For a more mathematical treatment see **[3]**.

In this note, we show that the proportion of wages covered by Social Security and exposed to taxation may be represented by the integral $\int_0^T (1 - F(x))\, dx$, where F is the cumulative distribution function for wage and salaried workers and T is some taxable maximum (a similar approach works for self-employed workers; some minor adjustment is necessary, however, for workers with both wages and self-employment earnings). Thus, if f is a probability density function corresponding to F for such workers and the 1990 employer and employee tax rate was 7.65%, the amount of Social Security tax liability would be given by the expression

$$\text{Tax Liability} = 2(.0765) N \int_0^T (1 - F(x))\, dx \tag{1}$$

We can confirm this result for the case $n=2$. Here the direction sought is $\langle \overline{W}_1, -\overline{W}_2 \rangle$ with

$$\overline{W}_1 = \begin{vmatrix} \exp \psi_1 & \exp \psi_2 \\ \dfrac{\partial}{\partial t_2} \exp \psi_1 & \dfrac{\partial}{\partial t_2} \exp \psi_2 \end{vmatrix} \quad \text{and} \quad \overline{W}_2 = \begin{vmatrix} \exp \psi_1 & \exp \psi_2 \\ \dfrac{\partial}{\partial t_1} \exp \psi_1 & \dfrac{\partial}{\partial t_1} \exp \psi_2 \end{vmatrix},$$

which by computation can be seen to have the same direction as $\langle \partial(\psi_2 - \psi_1)/\partial t_2, -\partial(\psi_2 - \psi_1)/\partial t_1 \rangle$. We recognize this as the tangent direction to the curve along which $\psi_2 - \psi_1$ is constant and hence along which the derivatives of ψ_1 and ψ_2 are equal. The above mentioned exception occurs only at a point where both components of this direction vanish. At such a point the derivative of $\psi_2 - \psi_1$ vanishes and hence the derivatives of ψ_1 and ψ_2 are equal—in *all* directions.

REFERENCES

1. K. Wolsson, Linear dependence of a function set of m variables with vanishing generalized Wronskians, *Linear Algebra and Its Applications* 117 (1989) 73–80.
2. W. J. LeVeque, *Topics in Number Theory*, Addison-Wesley Publishing Co., Reading, MA, 1956.
3. W. Schmidt, Diophantine Approximation, *Lecture Notes in Mathematics*, No. 785, Springer-Verlag, NY, 1980.

A Surprising Application of the Integral $\int_0^T (1 - F(x))\, dx$ to Revenue Projection

ALLAN KROOPNICK
Social Security Administration
Baltimore, MD 21235

The Social Security Administration annually estimates revenue into its trust funds in order to project present and future benefits and compare those benefits to present and future revenues received. The amount of money flowing into the trust funds depends on the taxable maximum T, which was \$51,300 for 1990. That is, the higher the T, the more money is exposed to taxation and hence more flows into the trust funds. For general overviews of the process see [1] and [2]. For a more mathematical treatment see [3].

In this note, we show that the proportion of wages covered by Social Security and exposed to taxation may be represented by the integral $\int_0^T (1 - F(x))\, dx$, where F is the cumulative distribution function for wage and salaried workers and T is some taxable maximum (a similar approach works for self-employed workers; some minor adjustment is necessary, however, for workers with both wages and self-employment earnings). Thus, if f is a probability density function corresponding to F for such workers and the 1990 employer and employee tax rate was 7.65%, the amount of Social Security tax liability would be given by the expression

$$\text{Tax Liability} = 2(.0765) N \int_0^T (1 - F(x))\, dx \tag{1}$$

where we need the "2" to count both the employer's and employee's contributions to tax liability and N is the total number of workers. Also, if M is the mean wage, the proportion of wages subject to Social Security tax is thus given by the expression $N\int_0^T(1 - F(x))\,dx/NM = \int_0^T(1 - F(x))\,dx/M$ (note that $NM = N\int_0^\infty xf(x)\,dx$ represents total wages of the population for a given year). Without loss of generality, we may assume that the density f corresponding to F has a mean M of 1, which means that $M = \int_0^\infty xf(x)\,dx = 1$. This has the additional effect of normalizing projected taxable maxima in terms of the multiples (not necessarily integral) of the mean wage and simplifies the projection process considerably. Since we have normalized the mean wage M to 1, the calculations are independent of the value of M. The proportion of wages taxable thus becomes

$$\int_0^T(1 - F(x))\,dx. \tag{2}$$

The proof of formula (2) is quite straightforward. First, note that $f(x)$ is defined only for $x > 0$ (people with no earnings pay no tax). The amount of wages subject to Social Security is therefore given by the expression,

$$\text{Taxable Wages} = N\int_0^T xf(x)\,dx + TN(1 - F(T)). \tag{3}$$

The first term on the right-hand side represents the amount of money taxable for workers earning less than the taxable maximum T. In order to see this, we see that the proportion of workers earning between x_i and x_{i+1} is given by $F(x_{i+1}) - F(x_i) = F'(\bar{x})(x_{i+1} - x_i) = f(\bar{x})(x_{i+1} - x_i)$ for some $\bar{x}$ between x_i and x_{i+1}. Thus, the amount of wages of these workers that is taxable is approximately $N\bar{x}f(\bar{x})(x_{i+1} - x_i)$. Summing and taking limits from 0 to T gives the required integral since f is a bounded continuous function. The second summand is the amount of money taxable for workers earning the taxable maximum or more since the term $1 - F(T)$ is the proportion of wage and salaried workers earning the taxable maximum or more. Finally, a standard integration by parts of the first term yields

$$\begin{aligned} N\int_0^T xf(x)\,dx + NT(1 - F(T)) &= NTF(T) = N\int_0^T F(x)\,dx + NT(1 - f(T)) \\ &= N\int_0^T (1 - F(x))\,dx. \end{aligned} \tag{4}$$

Since $M = 1$, dividing (4) by N yields the required proportion. Also, we see that, in general, if M is the mean of a probability density function f defined on the positive real line and $F(x) = \int_0^x f(s)\,ds$ is its cumulative distribution function then using integration by parts and the fact that since F is a distribution function $\lim_{x\to\infty} F(x) = 1$,

$$\begin{aligned} \int_0^x (1 - F(s))\,ds &= x - \int_0^x F(s)\,ds = x - xF(x) + \int_0^x sf(s)\,ds \\ &= M \text{ as } x \text{ approaches } \infty. \end{aligned} \tag{5}$$

Equation (4) states the obvious fact that the average wage taxable by Social Security approaches the average wage in the population as the taxable maximum is increased.

Remark 1. The crucial advantage of formula (1) is that we need not know the density function f in order to calculate taxable ratios. Data, however, are available for the cumulative function F (see [3, pp. 142 ff.]). Finally, returning to equation (1), we easily see that if all wages are taxable then the amount of tax is $2(.0765)NM$.

Example 1. We illustrate this technique using a simple example. From [4, p. 137] the average wage for 1986 was \$16,361 and the taxable maximum was \$42,000. Letting $T = 42000/16361 = 2.567$, which normalizes the mean wage to 1, and using $F(x) = 1 - \exp(-x)$ as the cumulative distribution function, we easily compute the proportion of wages taxable to be $\int_0^{2.567}(1-(1-\exp(-x))\,dx = 1-\exp(-2.567) =$ 0.923 near to the published number of 0.913 for 1986. In practice, adding additional exponential terms of the form $\exp(-nx)$ where n is a positive integer and using the method of least squares provides an excellent fit of a cumulative distribution function to the data over a large range with the additional advantage of being easily integrable.

Example 2. The taxable maximum for 1991 was \$53,400. If we assume a mean wage of \$21,150 for 1991, then using the same distribution of example 1 yields a taxable ratio of 0.920. However, if we apply the correction factor of $0.913/.923 = 0.989$ then, perhaps, a better projection for 1991 would be 0.910. This follows if we were to assume a correction factor would be needed since F is not a perfect fit and assuming the degree of error is near that in the previous example.

Remark 2. Starting in 1991, the taxable maxima for Social Security and Medicare are different, i.e., \$53,400 and \$125,000 respectively. However, this presents no computational problem since the calculations may be treated separately. That is, the tax rate for Social Security is 6.2% while for Medicare it is 1.45%. All that needs to be done, then, is to apply the formula twice using the respective tax rate and taxable maximum.

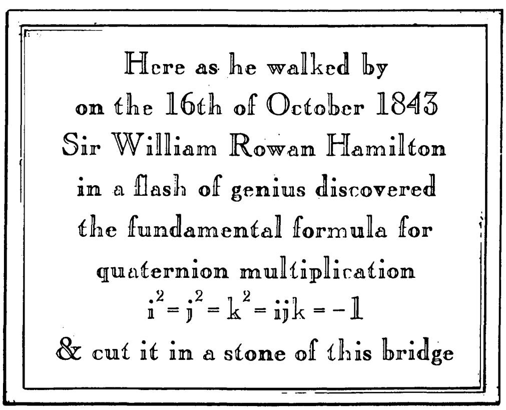
The author should like to thank the referees for their extremely helpful suggestions concerning this note.

REFERENCES

1. F. R. Bayo, M. P. Glanz, and C. L. Trowbridge, Components of trends in Social Security costs, *Transactions of the Society of Actuaries*, XXXVIII (1986), 7–42.
2. S. C. Goss, Long-range estimates of the financial status of the Old-Age, Survivors, and Disability Insurance program, *Actuarial Study No. 91*, Social Security Administration (SSA Pub. No. 11-11538), April 1984.
3. J. Hambor, An econometric model of OASDI, in : *Studies in Income Distribution*, No. 10, Social Security Administration, Office of Research and Statistics (HEW Pub. No. 79-11776), November 1978.
4. *Social Security Bulletin: Annual Statistical Supplement*, 1989, Social Security Administration (SSA Pub. No. 13-11700), December 1989.

---

150 years ago. . .



Broughm Bridge, Dublin

Monotonic Numbers

RICHARD BLECKSMITH
Northern Illinois University
DeKalb, IL 60115

CHARLES NICOL
University of South Carolina
Columbia, SC 29208

1. The Monotonic Product Property We say that a positive integer a is *monotonic base b* if the base b digits of a are monotonically increasing. Thus, 333667 is monotonic base 10, while 125167 is not. Sorting the digits of a number from smallest to largest, and then throwing away leading 0's always yields a monotonic integer. This is a key step in Conway's RATS (reverse, add, then sort) algorithm [3].

We are interested in the question: When is the product of two monotonic numbers also monotonic? For example, $11 \times 13 = 143$ is not monotonic, while 33366667×333667 has the *monotonic* product 11133355677889. Toying with a programmable calculator gives lots of small examples, such as the fact that

$$13 \times (112, 113, 123, 129, 188, 189, 223, 259, 266, 359)$$

are all monotonic. Another easy calculation shows that $17 \times 268 = 34 \times 134 = 67 \times 68 = 4556$ is monotonic. (The reader may ponder the question: What is the maximum number of such equal products?) We mention the old chestnut: $12345679 \times (9, 18, 27, 36, 45)$ are all monotonic. As the number of digits of the two factors increases, one might expect such examples to dwindle. It is fantastic to imagine that the product of, say, two 1 million digit monotonic numbers could possibly be monotonic itself. Yet this actually happens! We will find infinite classes of numbers for a fixed base b with just this "monotonic product property." The only requirement on the base b is that $b-1$ be composite.

We now introduce some notation. We shall use superscripts to indicate that a digit is repeated in the base b representation of a number. For example, $a = 3^m 6^n 7$ means that a consists of m 3's, followed by n 6's, and ends in the digit 7. In this notation, every monotonic number can be written as $1^{n_1} 2^{n_2} \cdots (b-1)^{n_{b-1}}$, where the n_i are nonnegative. (In forming a number, a digit to the 0th power represents the "null string", e.g. $3^2 6^0 7 = 337$. In particular, if all superscripts are 0, then the number is zero, so that $3^0 6^0 = 0$.)

We want to consider all monotonic numbers whose digits are multiples of a divisor of $b-1$ except that the rightmost (or units) digit is one greater than such a multiple. For each positive divisor d of $b-1$, define

$$M(d) = \{(a_t a_{t-1} \cdots a_1 a_0)_b + 1 \colon a_i = k_i d,\ 0 \le a_t \le \cdots \le a_0 < b-1\}.$$

Note that the inequality $a_0 < b-1$ ensures that the $+1$ does not increase a_0 to b and create a carry. In base 97, for example, the number $(6.6.30.30.66.66.90.91)_{97} \in M(6)$ has the following values of $k_i = a_i/6$: 1,1,5,5,11,11,15,15. For any base, $M(1)$ consists of all monotonic numbers base b whose units digit differs from the b's digit, while $M(b-1) = \{1\}$. In base 10, the set $M(3)$ consists of monotonic numbers of the form $3^m 6^n + 1, m, n \ge 0$, and we have the chain of subsets $\{1\} = M(9) \subset M(3) \subset M(1)$. In base 16, $M(3)$ comprises the hexadecimal numbers $3^m 6^n 9^p C^q + 1$, while $M(5)$

consists of the numbers $5^m A^n + 1$. (In hexadecimal notation, $A = 10$, $B = 11$, $C = 12, \ldots, F = 15$.) Observe, in general, that if d_1 and d_2 are divisors of $b - 1$ with $d_1 | d_2$, then $M(d_1) \supseteq M(d_2)$.

MONOTONIC PRODUCT THEOREM. *Let d_1 and d_2 be divisors of $b - 1$ such that $b - 1 | d_1 d_2$. Put $q = d_1 d_2/(b - 1)$. If $x \in M(d_1)$ and $y \in M(d_2)$, then $xy \in M(q)$. In particular, xy is monotonic in base b.*

When $b - 1$ is not composite, one of the divisors d_1 or d_2 in the Monotonic Product Theorem must be $b - 1$. Since $M(b - 1) = \{1\}$, the theorem is trivially true, but uninteresting, in this case. Unfortunately, for bases $b = 2$, 3, 4, 6, 8, 12, 20, 30, and 60, the number $b - 1$ is *not* composite, and this covers almost all bases that human beings have ever considered. The two most important exceptions are bases $b = 10$ and 16. In base 10, the Monotonic Product Theorem guarantees that the product of 33366667 and 333667 will be monotonic, while the hexadecimal product of $336669999CCD \in M(3)$ and $5AAAAB \in M(5)$ will be monotonic base 16. Here are these two products computed by the classical multiplication algorithm. (Multiplicative carries are indicated by superscripts.)

$$
\begin{array}{rrrrrrrrrrrrrr}
 & & & & & & 3 & 3 & 3 & 6 & 6 & 6 & 6 & 7 \\
\times & & & & & & & & 3 & 3 & 3 & 6 & 6 & 7 \\
\hline
 & & & & & 2^2 & 3^2 & 3^2 & 5^4 & 6^4 & 6^4 & 6^4 & 6^4 & 9 \\
 & & & & 2^2 & 0^2 & 0^2 & 2^4 & 0^4 & 0^4 & 0^4 & 0^4 & 2 & \\
 & & & 2^2 & 0^2 & 0^2 & 2^4 & 0^4 & 0^4 & 0^4 & 0^4 & 2 & & \\
 & & 1^1 & 0^1 & 0^1 & 1^2 & 0^2 & 0^2 & 0^2 & 0^2 & 1 & & & \\
 & 1^1 & 0^1 & 0^1 & 1^2 & 0^2 & 0^2 & 0^2 & 0^2 & 1 & & & & \\
1^1 & 0^1 & 0^1 & 1^2 & 0^2 & 0^2 & 0^2 & 0^2 & 1 & & & & & \\
\hline
1 & 1 & 1 & 3 & 3 & 3 & 5 & 5 & 6 & 7 & 7 & 8 & 8 & 9
\end{array}
$$

$$
\begin{array}{rrrrrrrrrrrrrrrrrr}
 & & & & & & 3 & 3 & 6 & 6 & 6 & 9 & 9 & 9 & 9 & C & C & D \\
\times & & & & & & & & & & & & 5 & A & A & A & A & B \\
\hline
 & & & & & 2^2 & 3^2 & 5^4 & 6^4 & 6^4 & 8^6 & 9^6 & 9^6 & 9^6 & B^8 & C^8 & C^8 & F \\
 & & & & 2^2 & 0^2 & 2^4 & 0^4 & 0^4 & 2^6 & 0^6 & 0^6 & 0^6 & 2^8 & 0^8 & 0^8 & 2 & \\
 & & & 2^2 & 0^2 & 2^4 & 0^4 & 0^4 & 2^6 & 0^6 & 0^6 & 0^6 & 2^8 & 0^8 & 0^8 & 2 & & \\
 & & 2^2 & 0^2 & 2^4 & 0^4 & 0^4 & 2^6 & 0^6 & 0^6 & 0^6 & 2^8 & 0^8 & 0^8 & 2 & & & \\
 & 2^2 & 0^2 & 2^4 & 0^4 & 0^4 & 2^6 & 0^6 & 0^6 & 0^6 & 2^8 & 0^8 & 0^8 & 2 & & & & \\
1^1 & 0^1 & 1^2 & 0^2 & 0^2 & 1^3 & 0^3 & 0^3 & 0^3 & 1^4 & 0^4 & 0^4 & 1 & & & & & \\
\hline
1 & 2 & 3 & 4 & 4 & 5 & 7 & 7 & 8 & 9 & A & B & C & D & D & E & E & F
\end{array}
$$

Observe that there are *no carries* when adding up the columns to produce the final answer. In general this "no carries" phenomenon always holds in applying the classical multiplication algorithm to compute the product xy for $x \in M(d_1)$ and $y \in M(d_2)$, where d_1, d_2 satisfy the hypothesis of the Monotonic Product Theorem.

MONOTONIC POWERS THEOREM. *Let $b = rs^n + 1$, $n \geq 2$. Then for every $x \in M(rs^{n-1})$, the numbers $x, x^2, \ldots, x^n$ are all monotonic base b.*

The Monotonic Powers Theorem is an easy corollary of the Monotonic Product Theorem. If we take $d_1 = d_2 = rs^{n-1}$ in the Monotonic Product Theorem, then for $x \in M(d_1)$, we have $x^2 \in M(rs^{n-2}), \ldots, x^n \in M(r)$.

Base 9 will serve in a modest way here to illustrate the Monotonic Powers Theorem. The square of 45_9 is 2267_9, while its cube is 114478_9, these three numbers being $\in M(4)$, $M(2)$, and $M(1)$ respectively. We can also obtain many interesting examples by considering various powers of the base. For example, the cube $(333667)^3$ doesn't immediately look monotonic, but when written in a traditional way,

$$37,148,370,629,851,963$$

is seen to be monotonic base 1000. This is guaranteed by the Monotonic Powers Theorem, since three 3's divide $1000 - 1$.

We look at some examples of the Monotonic Powers Theorem where the parameter s equals 2. Here's a base $b = 25$ calculation: $(12.12.13)_{25} \in M(12)$, its square, $(6.6.6.18.18.19)_{25} \in M(6)$, and its cube, $(3.3.3.12.12.12.21.21.22)_{25} \in M(3)$. In base $b = 97$, the j-th power of $(48.49)_{97}$ lies in $M(96/2^j)$ for $j = 1$ to 5. Starting with

$j = 2$, these powers are $(24.24.72.73)_{97} \in M(24)$, $(12.12.48.48.84.85)_{97} \in M(12)$, $(6.6.30.30.66.66.90.91)_{97} \in M(6)$, and $(3.3.18.18.48.48.78.78.93.94)_{97} \in M(3)$. The quotients k_i of the digits of these numbers divided by $(b-1)/2^j$—the same k_i's appearing in the definition of the set $M(d)$—form *Pascal-triangle-like* arrays. If we ignore repetition, these k_i are:

$$\begin{array}{ccccccccccc} &&&&& 1 &&&&& \\ &&&& 1 && 3 &&&& \\ &&& 1 && 4 && 7 &&& \\ && 1 && 5 && 11 && 15 && \\ & 1 && 6 && 16 && 26 && 31 & \\ 1 && 7 && 22 && 42 && 57 && 63 \\ & \dots && \dots && \dots && \dots && \dots . & \end{array}$$

Observe that the entries on the left edge of the triangle are all 1's, the entries on the right edge are one less than a power of 2, and each entry in the middle is the sum of the two "nearest" entries in the row above. The reader may consider what happens when $s = 3$ or when s takes on other values.

2. Polynomial representations and the proof of the Monotonic Product Theorem A very simple idea connects the base b representation of a number x with the polynomial whose coefficients are the base b digits of x: If $x = (a_r \dots a_1 a_0)_b$ and $f(t)$ is the polynomial $\sum_{i=0}^{r} a_i t^i$, then $x = f(b)$. We call $f(t)$ the *base b representation polynomial* of x. Of course, the structure of the ring of integers differs from that of the ring of polynomials over the integers. Some connections, however, do exist between x and its corresponding base b representation polynomial. For example, if x is prime in Z then the representation polynomial of x for any base $b \geq 2$ is irreducible in Z$[t]$. (See [**2**, Corollary 2].) On the other hand, if $f(t)$ is the representation polynomial of x and $g(t)$ is the representation polynomial of y, using the same base b, we would hardly expect the product $f(t)g(t)$ to be the representation polynomial of xy. Almost any (arbitrary) example the reader constructs will show that the coefficients of $f(t)g(t)$ can easily exceed the value $b-1$ of the largest permissible digit. It is surprising, therefore, that we can establish the Monotonic Product Theorem by obtaining the representation polynomial of the product xy from the representation polynomials of x and y, especially since no digit of x or y is 0.

Assume that d_1 and d_2 are divisors of $b-1$ satisfying the hypothesis of the Monotonic Product Theorem that $d_1 d_2 = q(b-1)$. Let $x \in M(d_1)$ and $y \in M(d_2)$ have the representation polynomials $f(t)$ and $g(t)$, respectively. It will be convenient for x and y to have the same number of digits, padding the shorter number with leading 0's, if necessary, so that $f(t)$ and $g(t)$ have the same number of coefficients. By the definitions of the sets $M(d_1)$ and $M(d_2)$, we can write

$$f(t) = \sum_{i=0}^{r} d_1 k_i t^i + 1 \quad \text{and} \quad g(t) = \sum_{j=0}^{r} d_2 l_j t^j + 1,$$

where $0 \leq k_r \leq \cdots \leq k_0 \leq (b-1)/d_1 - 1$ and $0 \leq l_r \leq \cdots \leq l_0 \leq (b-1)/d_2 - 1$. We can compute the product xy by multiplying the polynomials $f(t)$ and $g(t)$ and then substituting $t = b$:

$$\begin{aligned} xy = f(b)g(b) &= \left(\sum_{i=0}^{r} d_1 k_i b^i + 1 \right) \left(\sum_{j=0}^{r} d_2 l_j b^j + 1 \right) \\ &= q(b-1) \sum_{i,j=0}^{r} k_i l_j b^{i+j} + d_1 \sum_{i=0}^{r} k_i b^i + d_2 \sum_{j=0}^{r} l_j b^j + 1, \end{aligned} \tag{1}$$

using the hypothesis $d_1 d_2 = q(b-1)$. Moreover, q divides d_1 (since $d_1/q = (b-1)/d_2$) and q divides d_2, say, $d_1 = qu$ and $d_2 = qv$. From (1) it is clear that q divides $xy - 1$. In terms of polynomials,

$$\frac{xy-1}{q} = h(b)$$

where

$$h(t) = \sum_{m=0}^{2r+1} a_m t^m = (t-1)\sum_{i,j=0}^{r} k_i l_j t^{i+j} + u\sum_{i=0}^{r} k_i t^i + v\sum_{j=0}^{r} l_j t^j. \tag{2}$$

(Note that the factor $t-1$ in (2) becomes the factor $b-1$ in (1) when b is substituted for t.) As it turns out, the coefficients a_m lie between 0 and $(b-1)/q - 1$, so that $qh(t) + 1 = \sum_{m=0}^{2r+1} q a_m t^m + 1$ is precisely the base b representation polynomial of xy, a somewhat amazing fact considering the discussion above. To see this, our first goal is to simplify the summations in (2) into a form from which we can determine the coefficients a_m. By straightforward series manipulations,

$$h(t) = \sum_{j=0}^{r} k_r l_j t^{j+r+1} + \sum_{i=1}^{r}\sum_{j=0}^{r} (k_{i-1} - k_i) l_j t^{i+j} + \sum_{i=0}^{r} u k_i t^i + \sum_{j=0}^{r} (v - k_0) l_j t^j. \tag{3}$$

Re-index the first sum on the right-hand side of (3) by substituting $m = j + r + 1$, so that $j = m - r - 1$. In the second sum substitute $m = i + j$ and split the sum into two subsums for m in the ranges 1 to r and $r+1$ to $2r$. Finally, combine the last two sums in (3) to obtain

$$\begin{aligned} h(t) = &\sum_{m=r+1}^{2r+1} k_r l_{m-r-1} t^m + \sum_{m=1}^{r}\sum_{i=1}^{m} (k_{i-1} - k_i) l_{m-i} t^m \\ &+ \sum_{m=r+1}^{2r}\sum_{i=m-r}^{r} (k_{i-1} - k_i) l_{m-i} t^m + \sum_{m=0}^{r} (u k_m + (v - k_0) l_m) t^m. \end{aligned}$$

Thus the coefficients a_m of $h(t)$ are given by

$$a_m = \begin{cases} \displaystyle\sum_{i=1}^{m} (k_{i-1} - k_i) l_{m-i} + u k_m + (v - k_0) l_m, & \text{if } 0 \le m \le r \\ \displaystyle\sum_{i=m-r}^{r} (k_{i-1} - k_i) l_{m-i} + k_r l_{m-r-1}, & \text{if } r+1 \le m \le 2r+1. \end{cases} \tag{4}$$

(The summations in (4) are empty for $m = 0$ and $m = 2r+1$.)

The second part of the proof is to show that

$$0 \le a_{2r+1} \le \cdots \le a_0 \le \frac{b-1}{q} - 1. \tag{5}$$

This will complete the proof of the Monotonic Product Theorem since it implies that $xy = q\sum_{m=0}^{2r+1} a_m b^m + 1 \in M(q)$.

To begin establishing (5), note that $a_{2r+1} = k_r l_r \ge 0$. For m in the range $r+1 \le m \le 2r$, simple algebraic manipulation shows that

$$a_m - a_{m+1} = \sum_{\substack{i= \\ m+1-r}}^{r} (k_{i-1} - k_i)(l_{m-i} - l_{m+1-i})$$
$$+ k_r(l_{m-r-1} - l_{m-r}) + l_r(k_{m-r-1} - k_{m-r}),$$

which is nonnegative since $\{k_i\}$ and $\{l_j\}$ are nonincreasing sequences. For m in the range $0 \le m < r$, we have

$$a_m - a_{m+1} = \sum_{i=1}^{m} (k_{i-1} - k_i)(l_{m-i} - l_{m+1-i})$$
$$+ (u - l_0)(k_m - k_{m+1}) + (v - k_0)(l_m - l_{m+1}),$$

which is nonnegative since $l_0 < u$ and $k_0 < v$. Finally, for $m = r$,

$$a_r - a_{r+1} = \sum_{i=1}^{r} (k_{i-1} - k_i)(l_{r-i} - l_{r+1-i}) + (u - l_0)k_r + (v - k_0)l_r \ge 0.$$

This shows that $\{a_m\}$ is nonincreasing. It remains to prove in (5) that $a_0 \le (b-1)/q - 1$. Since $a_0 = uk_0 + (v - k_0)l_0$, this is equivalent to showing

$$qa_0 = d_1 k_0 - qk_0 l_0 + d_2 l_0 \le b - 1 - q. \tag{6}$$

We prove (6) by "reverse induction." Putting the largest possible values for k_0 and l_0, namely $k_0 = (b-1)/d_1 - 1$ and $l_0 = (b-1)/d_2 - 1$, into (6) gives the equality

$$d_1\left(\frac{b-1}{d_1} - 1\right) - q\left(\frac{b-1}{d_1} - 1\right)\left(\frac{b-1}{d_2} - 1\right) + d_2\left(\frac{b-1}{d_2} - 1\right)$$
$$= b - 1 - d_1 - \frac{d_1 d_2}{b-1}\left[\frac{b-1}{d_1}\frac{b-1}{d_2} - \frac{b-1}{d_1} - \frac{b-1}{d_2}\right]$$
$$- q + b - 1 - d_2 = b - 1 - q.$$

Next we reverse the usual direction of induction by showing that if (6) holds for a particular k_0, then it holds for the *previous* value $k_0 - 1$, since

$$d_1(k_0 - 1) - q(k_0 - 1)l_0 + d_2 l_0 = d_1 k_0 - qk_0 l_0 + d_2 l_0 + ql_0 - d_1$$

and $ql_0 < d_1$. Similarly, we can step from l_0 down to $l_0 - 1$. Since k_0 and l_0 are nonnegative integers, our backwards induction establishes (6) and thus completes the proof of the Monotonic Product Theorem.

3. Base 10 applications Consider the base $b = 10 = 3(3) + 1$. Here $M(3) = \{3^m 6^n + 1\colon m, n \ge 0\}$. By the Monotonic Product Theorem, the product of any two (monotonic) numbers x and y in $M(3)$ is again monotonic. In fact, using the classical multiplication algorithm, one can show that

$$(3^m 6^n + 1) \times (3^p 6^q + 1) = 1^p 2^{m-p} 3^{p+q-m} 4^{m+n-p-q} 5^{p-n} 6^q 7^{n-q} 8^q + 1. \tag{7}$$

This formula is valid provided all superscripts are nonnegative, i.e. $p + q \ge m \ge p \ge n \ge q \ge 0$. When all the inequalities in this chain are strict and when $q \ge 2$, then the right-hand side of (7) will end in $8^{q-1}9$ and thus every digit 1 through 9 appears in the decimal expansion of the product in (7).

By the Monotonic Powers Theorem, the square of each $x \in M(3)$ is monotonic.

Explicitly,

$$(3^m6^n+1)^2 = \begin{cases} 1^m3^m4^{n-m}6^m8^n+1 & \text{if } n \ge m \\ 1^m3^n5^{m-n}6^n8^n+1 & \text{if } m \ge n. \end{cases} \tag{8}$$

Note that the formula in (8) for $m \ge n$ follows immediately upon setting $m = p$ and $n = q$ in (7). (Formulas for $(6^n7)^2$ and $(3^m4)^2$, which are special cases of (8), appear in [1, p. 61] and [4, pp. 86–87].) These are not the only monotonic numbers whose squares are monotonic. There are two other infinite families:

$$(3^n5)^2 = 1^n2^{n+1}5 \tag{9}$$

and

$$(16^n7)^2 = 27^n8^{n+1}9. \tag{10}$$

(T. S. Motzkin communicated some of these formulas to J. L. Selfridge over 25 years ago. Apparently, Motzkin was unaware of the two-parameter family (8).) As it turns out, for numbers of four or more digits, the three families in (8), (9), and (10) are the only monotonic, base 10 numbers whose squares are monotonic. Proving this requires a complicated induction argument, which we will not present here. An examination of all monotonic numbers of three or fewer digits reveals that outside of the three classes (8), (9), and (10), the only numbers for which a and a^2 are both monotonic are 2, 3, 6, 12, 13, 15, 16, 38, 116, and 117.

One can naturally ask whether squares of nonmonotonic numbers are ever monotonic. Extensive computation shows that the largest nonmonotonic integer a, of 12 or fewer digits, such that a^2 is monotonic is 125167, whose square is 15666777889. We conjecture that for each nonmonotonic integer $a > 125167$, a^2 is not monotonic.

We would like to thank J. L. Selfridge for introducing us to monotonic numbers. We also thank Greg Manning, who wrote a symbol manipulation program that produced formulas (7) through (10). Finally, thanks to the referees for the numerous suggestions and improvements of the original manuscript. All examples for bases other than 10 and the interesting Pascal-like-triangle display are due to one of these referees.

REFERENCES

1. A. H. Beiler, *Recreations in the Theory of Numbers*, Dover Publications, Inc., Mineola, NY, 1966.
2. John Brillhart, Michael Filaseta & Andrew Odlyzko, On an irreducibility theorem of A. Cohn, *Canad. J. Math.* 33 (1981), 1055–1059.
3. Richard K. Guy, Conway's RATS and other reversals, *Amer. Math. Monthly* 96 (1989), 425–428.
4. F. B. Selkin, Number games bordering on arithmetic and algebra, *Teachers College Record*, 13 (1912), 68–111.

Geometric Proofs of Some Recent Results of Yang Lu

GRANT CAIRNS
MARGARET MCINTYRE
JOHN STRANTZEN
La Trobe University
Bundoora Victoria 3083, Australia

Computer-aided theorem-proving has been the object of recent research in China. The pioneering work of Wu Wentsun [5] described an algebraic algorithm that was successful in proving nontrivial theorems in fields including elementary geometry. More recently, Zhang Jingzhong and Yang Lu [8] have devised an algorithm that replaces most of the symbolic algebra used in the usual methods of automated theorem-proving with numerical computation. The Zhang-Yang method truthfully can be described as a "proof by sufficient specific examples."

The object of this note is to give elementary geometric proofs of the following two results recently proved by Yang Lu via his automated method [6], [7], [9].

THEOREM 1 [6]. *In any nondegenerate tetrahedron, if the four altitudes have lengths a_1, a_2, a_3, and a_4 and the three bialtitudes have lengths b_1, b_2, and b_3, then*

$$\frac{1}{a_1^2}+\frac{1}{a_2^2}+\frac{1}{a_3^2}+\frac{1}{a_4^2}=\frac{1}{b_1^2}+\frac{1}{b_2^2}+\frac{1}{b_3^2}.$$

THEOREM 2 [3], [6], [7], [9]. *If the angle sum of a spherical triangle is 2π, then its medial triangle is equilaterial and has angle sum $3\pi/2$ (i.e., it is an octant).*

These two results are certainly very surprising. They are by no means widely known results and are definitely absent from the standard texts on geometry. In fact, we have been unable to find any prior reference to Theorem 1 in the literature. This is particularly curious since, as we show below, Theorem 1 is just a simple exercise in vector algebra. Theorem 2 does appear in the literature however, though the result is extremely obscure. The only reference to it that we have managed to unearth occurs in the 1909 text on spherical trigonometry by M'Clelland and Preston [3], who record it as a London University exam paper question! As we show below, Theorem 2 can also be given an elegant geometric proof.

Let us first turn to the proof of Theorem 1. Consider a tetrahedron T in $\mathbb{R}^3$. Let us suppose that it has one of its vertices at the origin and the other three at the vectors t_1, t_2, and t_3 respectively. Now recall (see for instance [1]) that there is a unique parallelepiped P in $\mathbb{R}^3$ such that the edges of the tetrahedon T are diagonals of the faces of P. Indeed, P is the parallelepiped determined by the vectors

$$p_1=(-t_1+t_2+t_3)/2,\quad p_2=(t_1-t_2+t_3)/2,\quad p_3=(t_1+t_2-t_3)/2.$$

Now the bialtitudes of T are just the distances between opposite faces of P. So if v denotes the volume of P, then in vector notation one has $v=|p_1\cdot p_2\times p_3|$ and

$$\frac{1}{b_1^2}+\frac{1}{b_2^2}+\frac{1}{b_3^2}=\frac{|p_1\times p_2|^2}{v^2}+\frac{|p_2\times p_3|^2}{v^2}+\frac{|p_1\times p_3|^2}{v^2}.$$

On the other hand, for the sum of the squares of the reciprocals of the altitudes one has

$$\frac{1}{a_1^2}+\frac{1}{a_2^2}+\frac{1}{a_3^2}+\frac{1}{a_4^2}=\frac{|t_1\times t_2|^2}{|t_1\cdot t_2\times t_3|^2}+\frac{|t_2\times t_3|^2}{|t_1\cdot t_2\times t_3|^2}+\frac{|t_1\times t_3|^2}{|t_1\cdot t_2\times t_3|^2}+\frac{|(t_1-t_3)\times(t_2-t_3)|^2}{|t_1\cdot t_2\times t_3|^2}.$$

Substituting the values $t_1=p_2+p_3$, $t_2=p_1+p_3$ and $t_3=p_1+p_2$ gives the statement of Theorem 1.

We now turn to the proof of Theorem 2. The key idea is to observe that the 2-sphere is tiled by four copies of the spherical triangle $\triangle$. Indeed, by hypothesis, $\triangle$ has angle sum 2π, so at any point, o say, on the 2-sphere, one can position three copies of $\triangle$ so that the union of their regions defines a fourth triangular region, Ω say. Then the spherical triangle that bounds Ω is clearly congruent to $\triangle$. Hence S^2 is tiled by four copies of $\triangle$. Let us call these triangles $\triangle_1$, $\triangle_2$, $\triangle_3$, and $\triangle_4$.

Now, in each $\triangle_i$, draw the corresponding medial triangle Σ_i. We claim that the union of the sides of the Σ_i forms three great circles C_1, C_2, and C_3. Indeed, this follows from a straightforward congruence argument. Notice for example that, in FIGURE 1, the triangles opq and srq are congruent and hence the sides pq and qr, of Σ_1 and Σ_2 respectively, both lie on the same great circle. Identical arguments hold at each vertex of the Σ_i.

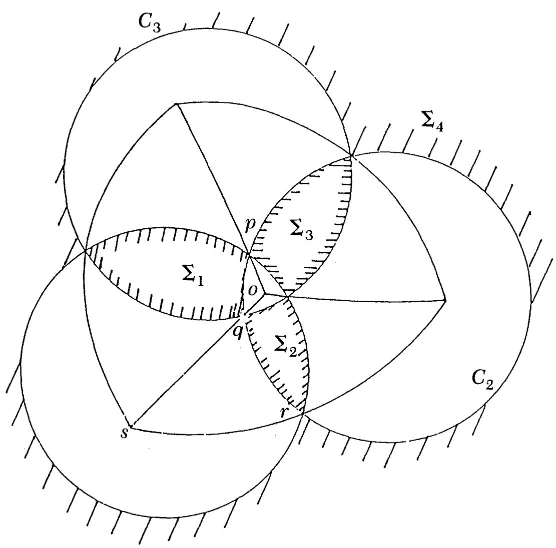


FIGURE 1

It is clear that each of the great circles C_i spends precisely one-quarter of its time in each of the triangles $\triangle_i$. Consequently the sides of the triangles Σ_i are all of length $\pi/2$ and each of the Σ_i is an octant. This completes the proof of Theorem 2.

There is a dual version of Theorem 2 (see [7]) and a converse to Theorem 2: If a spherical triangle $\triangle$ has equilateral medial triangle, then either $\triangle$ is equilateral or has angle sum 2π. We leave to the reader the pleasure of constructing geometric proofs of these results.

Finally, as commented above, Theorem 2 actually appears in a book on spherical trigonometry [3]. In fact, one can use the techniques of spherical trigonometry to obtain a much more complete picture. One has the following result.

THEOREM 3. *Let* $\triangle$ *be a spherical triangle with medial triangle* Σ. *Then the following conditions are equivalent*:

(i) $\triangle$ *has angle sum* 2π,
(ii) Σ *is an octant*,
(iii) Σ *has angle sum* $3\pi/2$,
(iv) Σ *has perimeter* $3\pi/2$,
(v) Σ *has one side of length* $\pi/2$,
(vi) Σ *has area one half the area of* $\triangle$,
(vii) *The sum of the cosines of the lengths of the sides of* $\triangle$ *equals* -1,
(viii) $\triangle$ *has a median whose length is supplementary to half the length of the side it bisects*,
(ix) *The sides of* Σ *bisect the great circle arcs from the vertices of* $\triangle$ *to the centroid of* $\triangle$.

Once again we leave the proof of this result to the reader. Let us simply say that the equivalence between the above conditions (i), (ii), (iii), (v), and (viii) has been previously observed [3] (see also [2] and [4]) and that the equivalence between the conditions (i), (ii), (iii), (iv), (v), (vii), and (viii) may all be derived easily from standard formulas such as Cagnoli's formula and Keogh's theorem. Conditions (vi) and (ix) are also not overly difficult, but do require a little extra work.

Our thanks go to Yang Lu for preprints of his papers and for many stimulating conversations.

REFERENCES

1. N. A. Court, *Modern Pure Solid Geometry*, Chelsea Publishing Co., New York, 1964.
2. T. M. MacRobert and W. Arthur, *Trigonometry; Part IV Spherical Trigonometry*, Methuen, London 1938.
3. W. J. M'Clelland and T. Preston, *A Treatise on Spherical Trigonometry with Applications to Spherical Geometry Part I*, Macmillan Publishing Co., New York, 1907, and *Part II*, 1909.
4. M. A. Todhunter, *Spherical Trigonometry*, Macmillan Publishing Co., New York, 1901.
5. Wu Wentsun, Some recent advances in mechanical proving of geometries, in *Automated Theorem Proving: After 25 Years*, W. Bledsoe and D. Loveland, eds, *A.M.S. Contemporary Math*. Vol. 29 (1984), 235–242.
6. Yang Lu, A new method of automated theorem proving, *The Mathematical Revolution Inspired by Computing*, Oxford University Press, Fairlawn, NJ, 1991, pp. 115–126.
7. Yang Lu, *Computer-aided Proving for New Theorems of Non-Euclidean Geometry*, Research Report no. 4, Dept. Math., Australian National University, 1989.
8. Zhang Jingzhong and Yang Lu, Principles of parallel numerical method and single-instance method of mechanical theorem proving (in Chinese), *Math. in Practice and Theory* 1 (1989) 34–43.
9. Zhang Jingzhong, Yang Lu, and Deng Mike, The parallel numerical method of mechanical theorem proving, *Theor. Comp. Sc.* 74 (1990) 253–271.

PROBLEMS

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

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by March 1, 1994.

1428. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Determine the remainder when $(x^2-1)(x^3-1)\cdots(x^{16}-1)(x^{18}-1)$ is divided by $1+x+x^2+\cdots+x^{16}$.

1429. *Proposed by Wee Liang Gan, student, Singapore.*

Let P be a point inside the convex n-gon $A_1A_2\cdots A_n$. Prove that at least one of the angles $\angle PA_iA_{i+1}$, $i=1,2,\dots,n$ is less than or equal to $(1/2-1/n)\pi$. (All subscripts are taken modulo n.)

1430. *Proposed by David Doster, Choate Rosemary Hall, Wallingford, Connecticut.*

Solve the recurrence $x_{n+1}=\dfrac{x_n^4+18x_n^2+9}{4x_n^3+12x_n}$, $\quad n\geqslant 0,\quad x_0=2.$

1431. *Proposed by Jiro Fukuta, Gifu-ken, Japan.*

In the given triangle ABC, let AD, AE be any cevians from A to BC. A circle drawn through A cuts AB, AC, AD, AE, or their extensions, at the points P, Q, R, S, respectively. (See FIGURE 1.)
Prove that

$$\frac{AP\cdot AB-AR\cdot AD}{AS\cdot AE-AQ\cdot AC}=\frac{BD}{EC},$$

where $AP, AB,\dots$ denote the lengths of the directed line segments $AP, AB,\dots$.

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

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

Solutions and new proposals should be mailed in duplicate to Loren Larson, Department of Mathematics, St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057-1098 or mailed electronically via fax: (507) 663-3549 or e-mail: larson@stolaf.edu.

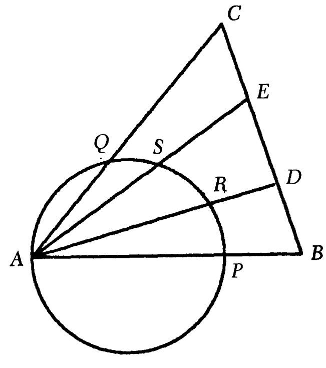


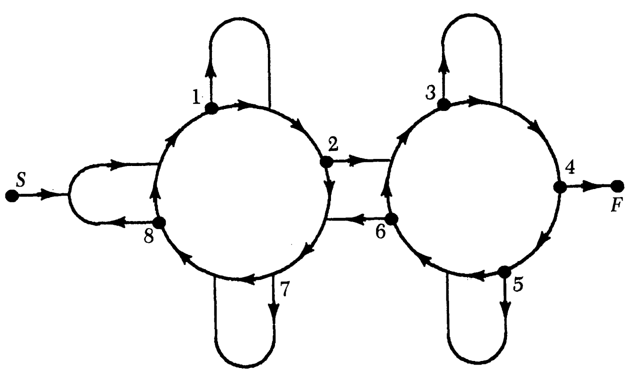
FIGURE 1

1432. *Proposed by Elliott A. Weinstein, Baltimore, Maryland.*

The map shown below is that of a notorious double-roundabout roadway in northern England. Consider a random "trip" on this roadway, which starts at S and finishes at F. Assume that the probability of choosing a particular continuation at any decision node is $1/2$. Define a *wrong turn* at any node as the selection that results in a trip to F longer than the minimal trip to F from that node. Let the word *route* mean the immediate sideroad that follows a wrong turn at node 1, 3, 5, 7, or 8.

a. Find the expected number of wrong turns.

b. Find the expected number of wrong turns, given that any particular route will have at most one wrong turn made onto it.



Quickies

Answers to the Quickies are on page 272

Q808. *Proposed by Wenbo V. Li, University of Wisconsin, Madison, Wisconsin.*

Does the convergence of $\sum_{n=1}^{\infty} a_n$ imply the convergence of $\sum_{n=1}^{\infty} a_n |a_n|$?

Q809. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

If the four altitudes of a tetrahedron are concurrent, prove that the six midpoints of the edges of the tetrahedron are cospherical.

Q810. *Proposed by Daniel Goffinet and Patrick Goffinet, Saint-Etienne, France.*

Assume that $f\colon \mathbf{R} \to \mathbf{R}$ has the property that $f(x+\lambda)-f(x)$ is continuous for all $\lambda \in \mathbf{R}$. Must f be continuous?

Solutions

Double Simpson's paradox **October 1992**

1403. *Proposed by Richard Friedlander, University of Missouri-St. Louis, and Stan Wagon, Macalester College, Saint Paul, Minnesota.*

Simpson's aggregation paradox admits a simple baseball interpretation. It is possible for there to be two batters, Veteran and Youngster, and two pitchers, Righty and Lefty, such that Veteran's batting average against Righty is better than Youngster's average against Righty, and Veteran's batting average against Lefty is better than Youngster's average against Lefty, but yet Youngster's combined batting average against the two pitchers is better than Veteran's. Question: Can there be a double Simpson's paradox? That is, is it possible to have the situation just described and, *at the same time*, have it be the case that Righty is a better pitcher than Lefty against either batter, but Lefty is a better pitcher than Righty against both batters combined?

Solution by D. M. Bloom, Brooklyn College of CUNY, Brooklyn, New York.

No. Let the matrix

	L	R	Overall
V	a	c	x
Y	b	d	y
Overall	z	w	

represent the batting averages of V(eteran) and Y(oungster) versus L(efty) and R(ighty), the overall averages of V and Y, and the overall averages against L and R. We are given

$$a>b, \qquad c>d, \qquad x<y \tag{1}$$

and we are asked whether at the same time we can have

$$a>c, \qquad b>d, \qquad z<w. \tag{2}$$

The key observation is that, in each row or column of the matrix, the 'overall' average must lie *between* the two subsidiary averages from which it comes ($a>z>b$, for instance). Thus, if it happens that $b>c$, then (1) implies $a>z>b>c>w>d$ and therefore $z>w$, contradicting (2). If instead $b \leqslant c$, then similarly, (2) implies $a>x>c \geqslant b>y>d$ and therefore $x>y$, contradicting (1).

Also solved by Alma College Problem Solving, Group, John Andraos (Canada), Michael H. Andreoli, S. F. Barger, David Jonathan Barrett, Brian D. Beasley and Stephen S. Ilardi, Walter Brady, David Callan, The Citadel Problem Solving Group, Curtis Cooper, Robert L. Doucette, Milton P. Eisner, David L. Farnsworth, Flintstones Problem Group, Arthur H. Foss, Kevin Ford (student), T. Robert Harris, Daniel

B. Hirschhorn, Dale Kilhefner, Emil F. Knapp, Billy Li, The Prestonsburg Community College Problem Solvers Group, Ken Rebman, Gordon Rice, George Schillinger, R. P. Sealy, John S. Sumner, John R. Thelin and Todd Crockrell, Trinity University Problem Group, and the proposers.

Difference of prime powers **October 1992**

1404. *Proposed by Hillel Gauchman and Ira Rosenholtz, East Illinois University, Charleston, Illinois.*

Find the smallest prime that is not the difference (in some order) of a power of 2 and a power of 3.

Solution by Reiner Martin (student), University of California at Los Angles, Los Angeles, California.

We claim that 41 is the smallest such prime.

First we observe that $2 = 3^1 - 2^0$, $3 = 2^2 - 3^0$, $5 = 3^2 - 2^2$, $7 = 3^2 - 2^1$, $11 = 3^3 - 2^4$, $13 = 2^4 - 3^1$, $17 = 3^4 - 2^6$, $19 = 3^3 - 2^3$, $23 = 3^3 - 2^2$, $29 = 2^5 - 3^1$, $31 = 2^4 - 3^0$, and $37 = 2^6 - 3^3$.

Now assume that $41 = 2^n - 3^m$, Clearly $n > 3$, so we have $1 \equiv -3^m \pmod 8$, which is impossible.

Finally assume $41 = 3^m - 2^n$. Now $m > 1$ and $n > 2$, so reducing modulo 3 and modulo 4 shows that n and m are even. But then reducing modulo 5 gives $1 \equiv (-1)^{m/2} - (-1)^{n/2}$, a contradiction.

Also solved by Brian D. Beasley, David Callan, Bill Correll Jr. (student), Fred Dodd, Milton P. Eisner, Kevin Ford (student), Arthur H. Foss, Lorraine L. Foster, Jane Friedman, Maria Fuente-Florencia, John Koker and Kandasamy Muthuvel and Robert Prielipp, Satoshi Kondo (student), David W. Koster, Peter W. Lindstrom, Jose Antonio Gomez Ortega and Bernardo Abrego and Silvia Fernandez (students, Mexico), Poo-Sung Park (student, Korea), Frank Schmidt, Heinz-Jürgen Seiffert (Germany), The Shippensburg University Mathematical Problem Solving Group, Shailesh Shirali (India), Daniel L. Stock, John S. Sumner and Kevin L. Dove, Trinity University Problem Group, University of Wyoming Problem Circle, Dave Witte, and the proposers. There were five incomplete solutions.

Koster noted that although 41 is not such a difference, it is a sum: $41 = 2^5 + 3^2$. Allowing for either sum or difference, he showed that 53 is the smallest prime that is neither a sum nor a difference (in some order) of a power of 2 and a power of 3.

Isogonally related circles **October 1992**

1405. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

Two circles inscribed in distinct angles of a triangle are *isogonally related* if the tangents from the third vertex not coinciding with the sides are symmetric with respect to the bisector of the third angle. Given three circles inscribed in distinct angles of a triangle, prove that if any two of the three pairs of circles are isogonally related then so is the third pair.

Solution by the proposer.

Let $\Gamma_1, \Gamma_2, \Gamma_3$ be circles inscribed in angles BAC, CBA, ACB, respectively, of the given triangle ABC. Let I_i, r_i be the center and the radius of Γ_i, $i = 1, 2, 3$. Let E and F denote the points on side AB, $E, F \notin \{A, B\}$, such that CE and CF are tangent to Γ_1 and Γ_2, respectively. Let $\gamma = \angle ACE$ and $\mu = \angle FCB$. As usual, let a, b, c denote the lengths of the sides BC, CA, AB, respectively.

By considering triangles AI_1C and I_2BC, respectively, we have

$$b = \left(\cot\frac{A}{2} + \cot\frac{\gamma}{2}\right)r_1$$

$$a = \left(\cot\frac{B}{2} + \cot\frac{\mu}{2}\right)r_2.$$

Now Γ_1 and Γ_2 are isogonally related if, and only if, $\gamma = \mu$, and, using the previous equations, this is the case if and only if

$$\frac{b}{r_1} - \cot\frac{A}{2} = \frac{a}{r_2} - \cot\frac{B}{2},$$

or equivalently,

$$\frac{b}{r_1} - \frac{s-a}{r} = \frac{a}{r_2} - \frac{s-b}{r},$$

where $2s = a + b + c$ and r is the inradius of triangle ABC. This can be regrouped into the form

$$\frac{a}{(1/r_1 - 1/r)} = \frac{b}{(1/r_2 - 1/r)}. \tag{1}$$

Similarly, Γ_2 and Γ_3 are isogonally related if, and only if,

$$\frac{b}{(1/r_2 - 1/r)} = \frac{c}{(1/r_3 - 1/r)}. \tag{2}$$

Combining (1) and (2), we obtain

$$\frac{a}{(1/r_1 - 1/r)} = \frac{c}{(1/r_3 - 1/r)},$$

which happens if, and only if, Γ_1 and Γ_3 are isogonally related.

Also solved by Richard Holzsager, Jiro Fukuta (Japan), and Francisco Bellot Rosado and María Ascensión López (Spain).

Difference equation **October 1992**

1406. *Proposed by Detlef Laugwitz, Fachbereich Mathematik Technische Hochschule, Darmstadt, Germany.*

Define a sequence $(a_n)_{n \geqslant 1}$ by

$$a_2 = \sqrt{3}, \qquad a_n = a_{n+1}(3 - a_{n+1}^2), \qquad 0 < a_n \leqslant a_1 \quad \text{for } n = 1, 2, 3, \ldots.$$

Show that $\lim_{n\to\infty} 3^n a_n$ exists and find its value.

Solution by David Callan, University of Wisconsin, Madison, Wisconsin.

The given conditions do not uniquely define a sequence of real numbers. However, for any $a_1 \in [0, 2]$, it is easy to see from the graph of the function $3x - x^3$, that there is a unique solution $\{a_n\}_{n=1}^{\infty}$ satisfying $0 \leqslant a_n \leqslant 1$, $n \geqslant 2$, and that this sequence monotonically approaches 0. Now define $\alpha_n \in [0, \pi/2]$ by $a_n = 2\sin\alpha_n$. Then, using the well-known trigonometic identity $\sin 3x = 3\sin x - 4\sin^3 x$, the given recurrence relation simplifies to $\sin\alpha_n = \sin 3\alpha_{n+1}$. It follows that $3\alpha_{n+1} = k\pi + (-1)^k\alpha_n$ for some integer k. The conditions $0 \leqslant a_1 \leqslant 2$ and $0 \leqslant a_n \leqslant 1$ for $n \geqslant 2$ certainly imply that for $n \geqslant 1$, $\alpha_{n+1} \in [0, \pi/6]$ and $\alpha_n \in [0, \pi/2]$, and it follows easily that $k \neq 0$ is impossible. Hence $3\alpha_{n+1}=\alpha_n$ and so $\alpha_n=\alpha_1/3^{n-1}$. This yields $a_n=2\sin(\alpha_2/3^{n-1})$ and $\lim_{n\to\infty}3^n a_n=6\alpha_1=6\sin^{-1}(a_1/2)$. For the given initial value $a_1=\sqrt{3}$, the limit is thus 2π.

Comment. An analogous result holds for $a_1\in[-2,0]$ since $3x-x^3$ is an odd function. For $|a_1|>2$ it is again easy to see from the graph that the recurrence $a_n=a_{n+1}(3-a_{n+1}^2)$ defines a unique real sequence (no additional conditions necessary) and that this sequence is alternating with absolute values exceeding 2. In fact, $(|a_n|)\downarrow 2$ and $9^{n-1}(|a_n|-2)\to(\cosh^{-1}(|a_1|/2))^2$. To see this, define $\alpha_n>0$ by $|a_n|=2\cosh\alpha_n$ and proceeding as before, use the identity $\cosh 3x=4\cosh^3 x-3\cosh x$ to obtain $|a_n|=2\cosh(\alpha_1/3^{n-1})$.

Also solved by John Andraos (Canada), Titu Andreescu, Seung-Jin Bang, Rich Bauer, J. C. Binz (Switzerland), David Doster, Robert L. Doucette, John F. Goehl, Richard Holzsager, Liang J. Huang, David W. Koster, Kee-Wai Lau (Hong Kong), Peter Lindstrom, Martin E. Muldoon, Andreas Müller (Germany), Waldemar Pompe (Poland), F. C. Rembis, Heinz-Jürgen Seiffert (Germany), Shailesh Shirali (India), Shreveport Problem Group, Glenn A. Stoops, John S. Sumner, Nora S. Thornber, Irvin Roy Hentzel, Trinity University Problem Group, Michael W. Vranos, WMC Problems Group, Adam Wolfe (student), Yan Loi Wong (Singapore), and the proposer. There was one unsigned solution.

Several solvers observed that a_n is the length of one side of a regular unit 3^n-gon, so that $(3^n a_n)$ increases and converges to the circumference of the unit circle, 2π. The Shreveport Problem Group made a connection between solutions of the difference equation and the Cantor set.

Multivariable optimization problem **October 1992**

1407. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Determine the maximum value of the sum

$$x_1^p+x_2^p+\cdots+x_n^p-x_1^q x_2^r-x_2^q x_3^r-\cdots-x_n^q x_1^r,$$

where p,q,r are given numbers with $p\geqslant q\geqslant r>0$ and $0\leqslant x_i\leqslant 1$ for all i.

Solution by David Jonathan Barrett, New York, New York.

We show that $\lfloor n/2\rfloor$ is the maximum value.

Let $f_{p,q,r}(x_1,\ldots,x_n)$ denote the given expression. Then for the x_i as given

$$f_{p,p,p}(x_1,\ldots,x_n)\geqslant f_{p,q,r}(x_1,\ldots,x_n). \tag{1}$$

We show that the left side reaches its maximum at some point where equality holds. For $n=2$ it is easy to check that $f_{p,p,p}(x_1,x_2)$ reaches its maximum of 1 at $(1,0)$ and $(0,1)$.

Assume $n\geqslant 3$. For fixed p,x_3,x_n, let

$$g(x_1,x_2)=x_1^p+x_2^p-x_1^p x_2^p-x_2^p x_3^p-x_n^p x_1^p.$$

Then $f_{p,p,p}(x_1,\ldots,x_n)$ is the sum of $g(x_1,x_2)$ and some function independent of x_1 and x_2. To maximize $f_{p,p,p}(x_1,\ldots,x_n)$ with respect to x_1,x_2, we need only to maximize $g(x_1,x_2)$. Since the latter has no relative extremum in the interior of the unit square, it must reach its maximum on the boundary, that is, where at least one of x_1,x_2 is 0 or 1. Examination reveals that either $(1,0)$ or $(0,1)$ must be a maximal point.

But the same argument goes through for any two adjacent variables in the function, so that some n-tuple of 0's and 1's (with never more than two consecutive 0's and 1's) must be a maximal point for $f_{p,p,p}(x_1,\ldots,x_n)$. In fact, by starting with $x_1=1$ and alternating 0's and 1's, we get the desired maximum at a point where equality in (1) holds.

Also solved by David Jonathan Barrett, Robert L. Doucette, Jiro Fukuta (Japan), WMC Problem Group, and the proposer.

Answers

Solutions to the Quickies on page 267

A808. No. For example, take

$$a_{3n+1}=\frac{1}{\sqrt{n+1}},\qquad a_{3n+2}=\frac{1}{\sqrt{n+1}},\qquad a_{3n+3}=\frac{-2}{\sqrt{n+1}}.$$

A809. Le t**A**, **B**, **C**, **D** denote vectors from the orthocenter to the vertices A, B, C, D, respectively, of the tetrahedron. Because **A** is orthogonal to **B** − **C** and **C** − **D**, etc., it follows that

$$\mathbf{A}\cdot\mathbf{B}=\mathbf{A}\cdot\mathbf{C}=\mathbf{A}\cdot\mathbf{D}=\mathbf{B}\cdot\mathbf{C}=\mathbf{B}\cdot\mathbf{D}=\mathbf{C}\cdot\mathbf{D}=\lambda.$$

We now show that the six midpoints of the edges $(\mathbf{A}+\mathbf{B})/2$, $(\mathbf{A}+\mathbf{C})/2$, etc., are all equidistant from the centroid $(\mathbf{A}+\mathbf{B}+\mathbf{C}+\mathbf{D})/4$. All we need to show is that 16 times the square of one of the six distances, say

$$16|(\mathbf{A}+\mathbf{B})/2-(\mathbf{A}+\mathbf{B}+\mathbf{C}+\mathbf{D})/4|^2=|\mathbf{A}+\mathbf{B}-\mathbf{C}-\mathbf{D}|^2,\dots.$$

is symmetric with respect to **A**, **B**, **C**, **D**. Expanding out and using the above identities, we get

$$|\mathbf{A}|^2+|\mathbf{B}|^2+|\mathbf{C}|^2+|\mathbf{D}|^2-4\lambda.$$

A810. Chose a basis for $\mathbf{R}$ over the rationals $\mathbf{Q}$ (by Zorn's Lemma), and define a nonconstant linear transformation $f\colon \mathbf{R}\to\mathbf{Q}$ by choosing values on the basis and extending to all of $\mathbf{R}$ by linearity. This function f is clearly discontinuous, and yet $f(x+\lambda)-f(x)=f(\lambda)$ is constant, hence continuous.

1992 Carl B. Allendoerfer Award Announced

The 1992 Carl B. Allendoerfer Award has been given to **Xun-Chen Huang** for his article "From Intermediate Value Theorem to Chaos" in the April 1992 issue of MATHEMATICS MAGAZINE (v65, n2, pp. 91–103). The citation reads,

"Huang's article is an outstanding achievement: exciting mathematics made accessible to a broad audience, written in an engaging expository style. As one member of the [award] subcommittee noted, 'this article is about the amazing fact that period 3 implies all periods for a continous function on the interval, and the author proves it using elementary means. He also manages to get across the shock wave that was felt by the mathematical community when this theorem was first proved (the introduction to mathematics of the concept of chaos), and the changes that these kinds of pursuits have engendered.' Another noted, 'the Xun-Chen Huang article is a standout. It is about real mathematics of current interest. And it is compelling and well written.' Indeed this is a model of what a MATHEMATICS MAGAZINE article should be, not least because it can be profitably studied by undergraduate students."

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.

Lemonick, Michael D., *Fini* to Fermat's Last Theorem: History's most celebrated math problem is solved at last; Kolata, Gina, At last, shout of "Eureka!" in an age-old math mystery, *New York Times* (National Edition) (24 June 1993) A1, A11. Kolata, Gina, Scientist at work: Math whiz who battled 350-year-old problem: Andrew Wiles, *New York Times* (National Edition) (29 June 1993) B5, B7. Swinnerton-Dyer, Peter, The end of the last theorem?, *Nature* 364 (1 July 1993) 13–14. Stewart, Ian, Thousand-page proof vindicates Fermat, *New Scientist* (3 July 1993) 14. Cipra, Barry, Fermat's last theorem finally yields, *Science* 261 (2 July 1993) 32–33. Ribet, Kenneth A., Wiles proves Taniyama's conjecture; Fermat's Last Theorem follows, *Notices of the American Mathematical Society* 40:6 (July-August 1993) 575–576.

If you've just gotten back from camping in the wilderness or vacationing on a Greek island for the past few months, you'll be sorry to learn that you missed all the dancing in the streets and parties given by your neighbors to honor Pierre de Fermat, Andrew Wiles (Princeton University), and local mathematicians such as yourself. Wiles announced a proof of Fermat's Last Theorem (FLT)—perhaps now better called the "Fermat-Wiles Theorem" (FWT)—in the third of a series of lectures in Cambridge, England, at the end of June. FWT is a corollary of Wiles's proof of the Taniyama conjecture, that every elliptic curve over Q is modular, for semistable elliptic curves. FWT has the great merit that the theorem can be explained easily to nonmathematicians, who perhaps can vicariously appreciate how over 350 years the problem has given rise to a great deal of number theory. Unless a much simpler proof is found, however, few mathematicians will understand why FWT is true; the rest of us will have to enjoy Wiles's triumph vicariously ourselves.

Rubenstein, Steve, 1,000 math buffs learn why theorem adds up, *San Francisco Chronicle* (30 July 1993). Schwartz, John, This equation figures to answer a 17th-century puzzle, *Washington Post* (2 August 1993) A3.

The Mathematical Sciences Research Institute in Berkeley put on a "Fermat Fest" at the Exploratorium in San Francisco, with lectures by Robert Osserman, Lenore Blum, Ken Ribet, Karl Rubin, and John Conway, as well as some new songs by Tom Lehrer. All 1,000 tickets sold out at $5; some folks paid scalpers $25 for a ticket, but 200 were turned away. A videotape will be available; for further information, send name, address, and email address to `video-request@msri.org`.

CORRECTIONS. In Vol. 66, No. 2 (April 1993), p. 137, l. 24, Paul Wallich's article on holographic proofs appeared on pp. 23–24 (thanks to Jay Rothman, University of Colorado). In Vol. 63, No. 3 (June 1990), p. 200, line 2, Gina Kolata's article "1 in a trillion coincidence, you say?..." appeared 27 February (thanks to David Hull, Valparaiso University).

Cipra, Barry, ***What's Happening in the Mathematical Sciences, Volume 1: 1993***, American Mathematical Society, 1993; 48 pp, free (thanks to Exxon Education Foundation and Sloan Foundation support); $7[!] shipping and handling. ISBN 0–8218–8999–0, ISSN 1065–9358

With its colorful and striking cover, clear and appealing writing, a multitude of photos and figures, and a splendid layout, this inaugural issue of a mathematics "annual" is absolutely marvelous! Its 10 essays treat important developments such as transparent proofs, environmental mathematics, and the infinitude of Carmichael numbers ("prime impostors"). It's about time the mathematical community made a serious effort at educating the public about contemporary achievements in mathematics, and this is an elegant and stylish way to let people know what we do. Perhaps the MAA can arrange to get future volumes into the hands of all MAA student members.

Jaffe, Arthur, and Frank Quinn, "Theoretical mathematics": Toward a cultural synthesis of mathematics and theoretical physics, *Bulletin of the American Mathematical Society* 29 (July 1993) 1–13. Horgan, John, Culture clash: Is mathematics becoming too much like physics?, *Scientific American* (August 1993) 26.

"Is speculative mathematics dangerous?" In physics, theory without proof can receive grand applause; consider, for example, string theory, for which there is no experimental evidence. Is a change in the relative rewards of conjecture and proof affecting the parts of mathematics that touch physics? How much credit should go to the conjecturer vs. the prover? For example, how much credit does Fermat deserve for conjecturing Fermat's Last Theorem vs. Wiles for proving it? The paper by Jaffe and Quinn raises questions and suggests prescriptions, while Horgan gives reactions from mathematicians.

Gibbs, W. Wayt, Practical fractal: Mandelbrot's equations compress digital images, *Scientific American* (July 1993) 107–108.

You came for the beauty, but now listen to the uses. Michael F. Barnsley's Iterated Systems firm now offers half a dozen image compression products based on fractals, achieving compression ratios greater than some competing products and offering resolution independence (the decompressed image can be reproduced in any size).

Gibbs, W. Wayt, Making wavelets: New math resurrects Brahms and compacts computer data, *Scientific American* (June 1993) 137–138.

At last we are able to hear without static and noise an 1889 recording of Brahms playing his First Hungarian Dance, thanks to adapted waveform analysis. This "deluxe model" of wavelet analysis uses a library of pulse shapes instead of just a single mother wavelet. Meanwhile, wavelets are getting set to challenge fractals in the arena of data compression.

Wilson, Elizabeth, Multiperfect numbers proliferate in Colorado, *New Scientist* (1 May 1993) 17.

A multiperfect number is one whose divisors add up to a multiple of the number; the multiple is called the *index*. For example, 360 is multiperfect with index 3. Until recently, only 700 multiperfects were known. An algorithm of Fred Helenius in Colorado has almost doubled the number of known multiperfects, including finding 14 with index 9. The algorithm starts with the product of powers of small primes and successively adjusts the number to find a multiperfect. Are there finitely many multiperfects of a fixed index? Are there finitely many altogether?

Engel, Arthur, *Exploring Mathematics with Your Computer*, New Mathematical Library 35, MAA, 1993; ix + 301 pp + diskette, $38. ISBN 0-88385-636-0

This is an ideal companion for high-school and college students (and their teachers) who have decided to learn how to program in Pascal in order to use it for mathematical purposes. It is a mathematics book, not a programming book, full of mathematical ideas and problems for which computing can suggest or check conjectures. Featured are algorithms in number theory, probability, statistics, combinatorics, and numerical analysis. An accompanying diskette contains the text of all of the programs (in Turbo Pascal).

Benjamin, Arthur, and Michael Brant Shermer, *Mathemagics: How to Look Like a Genius without Really Trying*, Lowell House, 1993; xx + 218 pp, $22.95. ISBN 0-929923-54-5

Art Benjamin has entertained magicians, mathematicians, schoolchildren, and TV audiences with his abilities as a lightning calculator. In this book he reveals his "secrets," which anyone can learn and practice.

Cook, Charles K., *Index of the Fibonacci Quarterly, Vols. 1-30, 1963-1992*, Fibonacci Association, 1993; 3.5-inch high-density diskette, $40 + postage (order from Prof. Charles K. Cook, Dept. of Mathematics, University of South Carolina at Sumter, 1 Louise Circle, Sumter, SC 29150; do not send payment with order—an invoice will be sent).

Are paper indexes obsolete? This IBM-compatible diskette contains WordPerfect files providing indexes to the *Fibonacci Quarterly* by title author, subject, and keyword, with special indexes for elementary, advanced, and special problems.

Peterson, Ivars, Party numbers: Deploying a host of computers to sort out a mathematical puzzle, *Science News* 144 (17 July 1993) 46-47.

Stanislaw P. Radziszowski (Rochester Institute of Technology) and Brendan D. McKay (Australian National University) have calculated the smallest unsolved Ramsey number, showing that "25 is the minimum number of guests needed to guarantee that a party includes a group of at least four people who all know one another or a group of at least five people who are strangers to one another." The calculations required a total of 11 years of processor time on as many as 110 desktop computers.

Steen, Lynn Arthur, Will everybody ever count?, in *Developments in School Mathematics Education Around the World*, vol. 3, edited by Izaak Wirszup and Robert Streit, National Council of Teachers of Mathematics, 1992, pp. 9-13.

Responding to the volume *Everybody Counts* by the Mathematical Sciences Education Board (1989), Lynn Steen asks, "Should everybody count?" Some observers suggest that we set different national targets for each level of mathematical proficiency, in terms of the percentage of students who achieve at that level as assessed by the National Assessment of Educational Progress. But Steen objects that "differentiated targets will do little but reinforce the entrenched economic and class differences of society." He urges that "each year each school should increase the percentage of its children who achieve at different levels," while additional resources go to equalizing "opportunity for effective instruction." Assessment of progress toward learning goals which must be both "narrative and numerical" ("innovative" assessment is much in vogue at the NSF these days); it will be even more difficult than teaching toward them.

NEWS AND LETTERS

LETTERS TO THE EDITOR

Dear Editor:

Professors Bailey and Lautzenheiser (February 1993) gave a proof (for $n \geq 6$) of the characterization of the so-called *curious sequences* $(a_0, a_1, \ldots, a_n)$ such that each a_i is the number of i's in the sequence. Here is a shorter proof of this result. Following the authors, we start by noting that $a_i > 0$ and

$$n + 1 = \sum a_i = \sum i a_i. \quad (*)$$

LEMMA. *Exactly one of the terms $a_1, \ldots, a_n$ is a 2, and the rest are 0's and 1's.*

Proof. Let p be the number of nonzero terms among $a_1, \ldots, a_n$. Since $a_0 > 0$, the total number of nonzero terms is $p + 1$, i.e., $a_1 + \ldots + a_n = p + 1$. Thus the nonzero terms among $a_1, \ldots, a_n$ are p positive integers whose sum is $p + 1$.

It follows that at most one term of the sequence (namely a_0) can be greater than 2, i.e., $a_3 + \ldots + a_n \leq 1$. Now the proof splits into two cases.

First, suppose $a_j = 1$ for some $j \geq 3$. Then we must have $a_0 = j$, $a_2 = 1$, and $a_1 = 2$; thus the curious sequence is $(j, 2, 1, 0, \ldots, 0, 1, 0, 0)$, where the first block of 0's has length $j - 3$.

On the other hand, suppose $a_3 = \ldots = a_n = 0$. Then $n + 1 = a_0 + a_1 + a_2 \leq 5$ so there are no further curious sequences with $n > 4$. Moreover, (*) implies that $a_0 = a_2$; thus the only possibilities for (a_0, a_1, a_2) are (1,2,1), (2,0,2), and (2,1,2), leading to the three "sporadic" curious sequences (1,2,1,0), (2,0,2,0), and (2,1,2,0,0).

Fred Gavin
The University of Kansas
Lawrence, KA

Editor's Note: Several readers wrote in reference to the above-mentioned Note.

James Roche of AT&T Laboratories, Murray Hill, NJ also presented a short proof.

Steven Kahan of Queens College, NY points out that his article, with the same title, appeared in the Nov-Dec 1975 issue of this MAGAZINE. It stated and proved (differently) the same theorem. He points out that two of his later articles, Mutually counting sequences (*The Fibonacci Quarterly*, 18.1, 1980, pp.47-50) and Cyclic counting trios (*The Fibonacci Quarterly*, 25.1, 1987, pp. 11-20) have established generalizations of this result.

Lee Sallows writes that a general formula for *self-descriptive strings* appears in a paper by Michael D. McKay and Michael S. Waterman (*Mathematical Gazette*, v66, n435, 1982). A rediscovery of the same result due to Quyen Giang, appeared in 1983. An entire chapter of *Discovering Mathematics* by Anthony Gardiner is devoted to this topic, as Bailey and Lautzenheiser point out. In addition, he continues:

"As a point of general interest, Co-descriptive strings (*Mathematical Gazette*, v70, n451, 1986) by Lee Sallows and Victor L. Eijkout went beyond self-descriptors to investigate closed cycles of strings in which each string decribes its predecessor in the loop. The main result is that for length N ≥ 8 there exists but one co-descriptive cycle of period 2; i.e., each string describes the other:

0	1	2	3	4	5	...	N-5	N-4	N-3	N-2	N-1
N-4	3	0	0	0	0	...	0	0	1	0	0
N-3	1	0	1	0	0	...	0	1	0	0	0

Self- and co-descriptive matrices have also been studied (*Cubism for Fun*, No.25, Part 4, 1991, pp.21-23; newsletter of the Dutch Cube Club). In a self-descriptive matrix M, each entry, m_{ij}, enumerates the occurrences of the ordered pair i,j in the closed string formed from the concatentated rows of *M*. An example illustrates:

i \ j	0	1	2	3
0	3	2	0	1
1	1	3	1	1
2	2	0	0	0
3	0	1	1	0

..00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33..
3	2	0	1	1	3	1	1	2	0	0	0	0	1	1	0

Lee Sallows
Buurmansweg 30
6525 RW Nijmegen,
The Netherlands

Dear Editor:

In the article "Roads and Wheels," by Leon Hall and Stan Wagon (December, 1992), it is mentioned that the parabola $y = x^2 - 1/4$ is a road which is its own wheel (p. 291). As it happens, this result appeared as problem A-5 on the 35th William Lowell Putnam Mathematical Competition, which was given on December 7, 1974. I just thought it might be of interest.

Richard Stone
Burnsville, MN

A Math Wizard, Hero to His Family

ERIC ZORN

I don't pretend to understand Zorn's Lemma—it is a statement of principle in higher mathematical set theory, and I never got smart enough to take a class where it came in handy.

And although it's not as common or useful as, say, the Pythagorean Theorem, it does appear in many standard dictionaries as well as in the title of the 1969 popular reference book "Whose What? Aaron's Beard to Zorn's Lemma." Math types always please me when they ask, "Are you related to *the* Zorn?"

I am. My grandfather, Max, published the lemma in 1935. I had occasion to think a lot about the man and his lemma Monday afternoon when, in response to an urgent call from my father, I drove to Bloomington, Ind., hoping to get to his bedside before he died.

He was 86 and had suffered unexpected and severe congestive heart failure. His lungs were filling with fluid, Dad said, his heart was nearly dead and nothing could or would be done to save him. His doctor had given him between three hours and two days to live.

It takes roughly five hours to drive from here to Bloomington, and on the way I thought back on what he had accomplished. I was always proud to be his only grandson, but what I was proudest of was not that he had written the lemma, but that he had fought against the emerging Nazi party in his native Germany before World War II.

He spoke with a raspy, airy voice most of his life. Few people knew why, because he only told the story after significant prodding, but he talked that way because pro-Hitler thugs who objected to his politics had battered his throat in a 1933 street fight.

He and his wife, Alice, and their young son, my father, fled to the United States in 1934.

He was not yet 30 when he made his first and, as it turned out, only lasting mark on his profession. Zorn's Lemma gave him international recognition, but ended up haunting him, as early glory so often does.

closed string formed from the concatentated rows of *M*. An example illustrates:

i \ j	0	1	2	3
0	3	2	0	1
1	1	3	1	1
2	2	0	0	0
3	0	1	1	0

..00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33..
3	2	0	1	1	3	1	1	2	0	0	0	0	1	1	0

Lee Sallows
Buurmansweg 30
6525 RW Nijmegen,
The Netherlands

Dear Editor:

In the article "Roads and Wheels," by Leon Hall and Stan Wagon (December, 1992), it is mentioned that the parabola $y = x^2 - 1/4$ is a road which is its own wheel (p. 291). As it happens, this result appeared as problem A-5 on the 35th William Lowell Putnam Mathematical Competition, which was given on December 7, 1974. I just thought it might be of interest.

Richard Stone
Burnsville, MN

A Math Wizard, Hero to His Family

ERIC ZORN

I don't pretend to understand Zorn's Lemma—it is a statement of principle in higher mathematical set theory, and I never got smart enough to take a class where it came in handy.

And although it's not as common or useful as, say, the Pythagorean Theorem, it does appear in many standard dictionaries as well as in the title of the 1969 popular reference book "Whose What? Aaron's Beard to Zorn's Lemma." Math types always please me when they ask, "Are you related to *the* Zorn?"

I am. My grandfather, Max, published the lemma in 1935. I had occasion to think a lot about the man and his lemma Monday afternoon when, in response to an urgent call from my father, I drove to Bloomington, Ind., hoping to get to his bedside before he died.

He was 86 and had suffered unexpected and severe congestive heart failure. His lungs were filling with fluid, Dad said, his heart was nearly dead and nothing could or would be done to save him. His doctor had given him between three hours and two days to live.

It takes roughly five hours to drive from here to Bloomington, and on the way I thought back on what he had accomplished. I was always proud to be his only grandson, but what I was proudest of was not that he had written the lemma, but that he had fought against the emerging Nazi party in his native Germany before World War II.

He spoke with a raspy, airy voice most of his life. Few people knew why, because he only told the story after significant prodding, but he talked that way because pro-Hitler thugs who objected to his politics had battered his throat in a 1933 street fight.

He and his wife, Alice, and their young son, my father, fled to the United States in 1934.

He was not yet 30 when he made his first and, as it turned out, only lasting mark on his profession. Zorn's Lemma gave him international recognition, but ended up haunting him, as early glory so often does.

Even after his retirement from the Indiana University mathematics department in 1972, he continued to write in his notebooks and go to his office every day hoping, it seemed, to come up with something equally lasting or more profound.

I won the race to Bloomington, Max (I always called him Opa) was conscious when I arrived shortly after 9 p.m., and greeted me with a surprisingly strong handshake. He asked, in a voice muffled by the oxygen mask through which he was drawing horrible, wet breaths, if my 3-year-old son was able to dress himself yet.

Small talk. Earlier that day he'd spoken to his doctor and to the family about the gravity of his condition and the impossibility of recovery. There was no hope for a miracle here, no doubt of the outcome. So he and my grandmother had taken time to embrace and reminisce about the old days when they had been university students together in Germany.

After I had been there a while and the room turned quiet, he said to all of us, "Thank you," then took a breath, "for coming to see me off," he took another breath, "in a certain way." He shook my hand again and gave a stiff, half-wave solute that was his trademark.

It was a bravura performance, one that he was unable to sustain as his condition worsened. By 11 p.m., he could only gasp out one word at a time, usually a request for water. Sometimes a simple cry for help.

Shortly before midnight, the nurse told us now was the time to summon anyone who wanted to see him for the last time. My dad left quickly to fetch my aunt and my grandmother from home nearby, and left me alone with Opa.

He could not respond with pressure when I squeezed his hand, so I stroked his arm lightly, soothingly, I hoped. I wet down a rag and daubed at his forehead, and I adjusted the breathing mask over the thick, careless white beard he's grown in retirement.

I held him and spoke loudly and directly into his right ear. I promised him I would tell his great-grandson all about him one day, I told him he was a good man, something I'm not sure he ever truly believed.

There are sad things and there are tragedies, and this was just a sad thing. Tragedies are when people are cut down in or even before their prime with hosts of promises unfulfilled. But Opa had lived in nine decades, achieved a measure of professional success, raised a family, lived to be able to walk down a street with four generations of his own family and never lost the edge from his sharp and unusual mind.

He was dying sooner than any of us wanted or expected, but he'd avoided the interminable decline that afflicts so many of his age, and most of the prolonged suffering that often attends death. We should all last so long, we should all go so quickly, we should all be able to hear and understand the parting sentiments of those we love.

I was lucky. I got to him in time to say to him words that, next time, with the next person, I swear I will not wait so long to say "We're proud of you," I said into his ear as I bent over him. "Your family loves you."

He struggled to echo me, one faint word at a time. "My/family/loves/me."

It was the last sentence he ever said—not as far reaching or famous a proposition as Zorn's Lemma, but equally lasting and, I think, more profound. It, too, will be his legacy.

JOIN THE MATHEMATICAL ASSOCIATION OF AMERICA NOW

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036-1385

Name ______________________________

Mailing Address ______________________________

______________________________ Zip ____________

Employer/School ______________________________

Position (Rank) ______________________________

Employer's City/State ______________________________

Highest Earned Degree ____________________ Year Degree Earned ____________________

Institution Awarding Degree ______________________________

Month/Year of Birth ______________________________

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Find the column for your desired combination of MAA journals in the table below. All members receive FOCUS. **Select** the row appropriate for your initial membership period (1 year or 1.5 years) and your status (student or regular). **Circle** your dues in the table. **Write** the amount on the form below.

Journal codes used in the columns of the dues table are: THE AMERICAN MATHEMATICAL MONTHLY = **M**
MATHEMATICS MAGAZINE = **G** THE COLLEGE MATHEMATICS JOURNAL = **J**

Student Membership	M	G	J	M + G	M + J	G + J	M + G + J
1 year (Jan.–Dec. 1994)	$ 37.00	$31.00	$32.00	$ 43.00	$ 44.00	$ 38.00	$ 50.00
1.5 years (July 1993–Dec. 1994)	$ 55.00	$46.00	$47.50	$ 64.00	$ 65.50	$ 56.50	$ 74.50
Regular Membership							
1 year (Jan.–Dec. 1994)	$ 74.00	$59.00	$62.00	$ 89.00	$ 92.00	$ 77.00	$107.00
1.5 years (July 1993–Dec. 1994)	$110.50	$88.00	$93.00	$132.50	$137.50	$115.00	$159.50

These are **specially discounted rates for new members** provided to help **MAA** reach a wider audience. They are not available to those who have been members for two years. **Student membership** is available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Student rates apply to unemployed persons who are seeking employment. **Annual dues include subscription prices as follows:** Regular Member $30 **M,** $18 **J,** $15 **G,** $10 **Focus.** Student members $12 **M,** $7 **J,** $6 **G,** $5 **Focus.** These rates are guaranteed for the indicated periods only.

Payment Enclosed $ ____________ *U.S. Funds Only* (see below)

METHOD OF PAYMENT ☐ Check Enclosed ☐ VISA ☐ MasterCard

Card Number ______________________________ Expiration Date |__|__|__|__|
(month/year)

Interbank Number |__|__|__|__| *(MasterCard only—located above name on card)*

SIGNATURE ______________________________

March 1993

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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