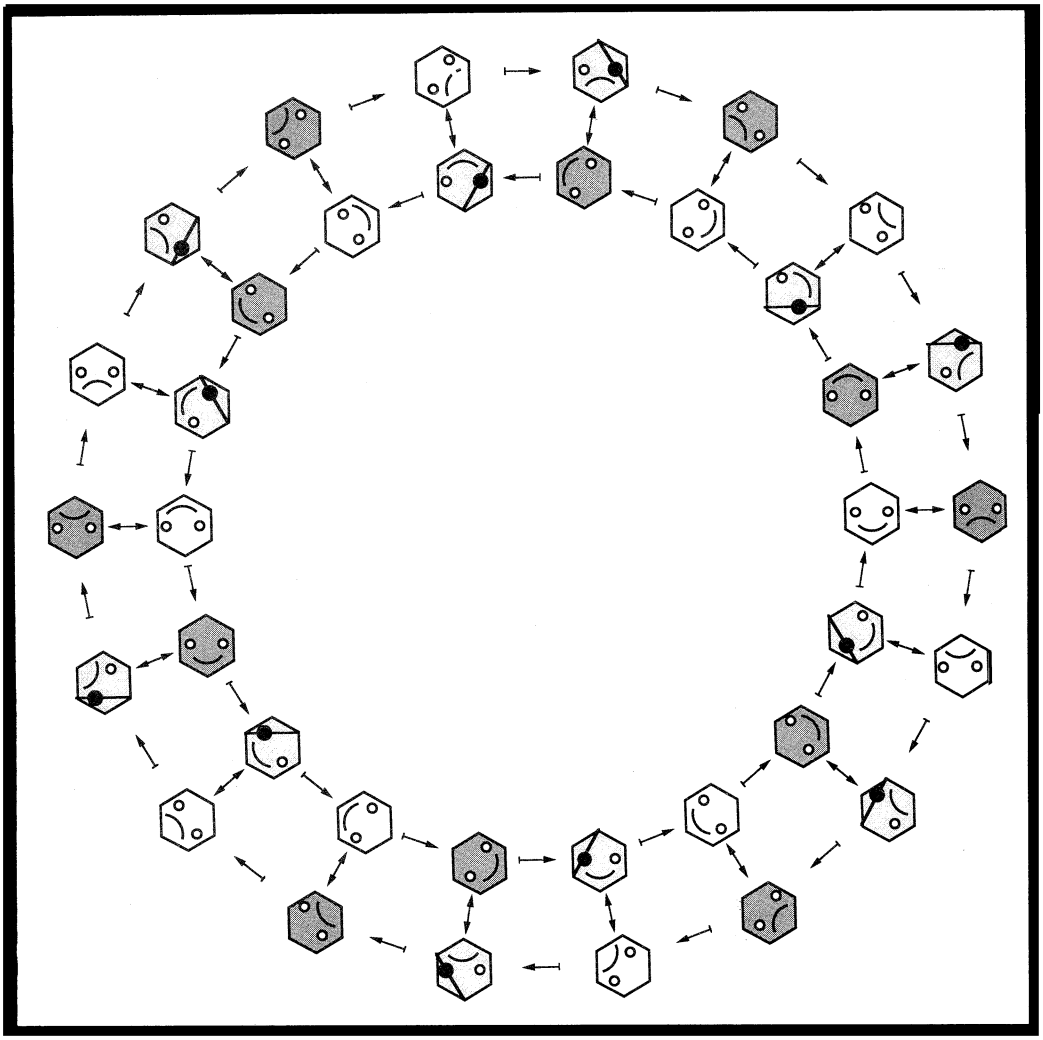
Vol. 70, No. 4 October 1997

# MATHEMATICS MAGAZINE



- Faces of the Tri-Hexaflexagon
- Centers of Triangles
- A Family Portrait of Primes
- Halley's Comment

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

# EDITORIAL POLICY

*Mathematics Magazine* aims 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. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 69, pp. 78–79, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Send new manuscripts to Paul Zorn, Editor, Department of Mathematics, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three 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. Copies of figures should be supplied on separate sheets, both with and without lettering added.

# AUTHORS

**Barry Cipra** is a freelance mathematics writer based in Northfield, Minnesota. He is currently a contributing correspondent for *Science* magazine and also writes regularly for *SIAM News*. He received the 1991 Merten M. Hasse Prize from the MAA for an expository article on the Ising model, published in the American Mathematical Monthly in December 1987. His book Misteaks... and how to find them before the teacher does... is published by Academic Press.

**Chuck Groetsch** is Professor of Mathematics at the University of Cincinnati. He received his Ph.D. from LSU in 1971; his research interests center on the theory of approximation methods for ill-posed inverse problems. Groetsch's tender years were misspent in the Irish Channel neighborhood of New Orleans, where he was known as a pretty decent rock thrower. Besides an unpleasant memory or two, this experience may have also provided an early stimulus to study the topic of this paper.

**Peter Hilton** is Distinguished Professor of Mathematics Emeritus at the State University of New York and Distinguished Professor of Mathematics at the University of Central Florida. He is an algebraic topologist. **Jean Pedersen** is Professor of Mathematics at Santa Clara University, in California. She is a geometer. **Hans Walser** is a teacher at the State College of Thurgau and lecturer at the ETH Zürich and the University of Basel. His primary research interests are in geometry.

The three authors found themselves together at Santa Clara University in May, 1996. Their paper on the tri-hexaflexagon was one of several results of this coincidence.

**William Mueller** received his bachelor's degree from MIT in 1981. Then he did some other things. He received his Ph.D., from Duke University, in 1991. His research interests include recursive function theory and educational technology. He is currently working on projects to create interactive computer modules on topics in differential equations, linear algebra, and engineering mathematics for distribution on the World Wide Web.

**Shailesh Shirali** received his Bachelor's degree in Physics from St. Stephen's College, in 1974, and his doctorate in Operations Research from the University of Texas, in 1981. He taught briefly at the University of New Brunswick before returning to India, where he has been since 1983. He is currently Principal of Rishi Valley School in Southern India. He is deeply involved in Olympiad-style problem solving and problem posing, and this paper evolved from an investigation inspired by a problem from the well-known book by Pólya and Szegő. He is particularly interested in geometry, combinatorics, and number theory, and in doing expository work in mathematics.

Vol. 70, No. 4, October 1997

# 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

# The Faces of the Tri-Hexaflexagon

PETER HILTON
State University of New York
Binghamton, NY 13902-6000
and
University of Central Florida
Orlando, FL 32816

JEAN PEDERSEN
Santa Clara University
Santa Clara, CA 95053

HANS WALSER
ETH Zürich
CH-0892 Zürich
Switzerland
and
Santa Clara University
Santa Clara, CA 95053

## Introduction

Hexaflexagons were invented at Princeton in 1939 by Arthur H. Stone, then a graduate student, now Professor Emeritus of Mathematics at the University of Rochester. Martin Gardner gives an interesting account (see [**1**]) of Stone's work and his collaboration with Bryant Tuckerman, then a graduate student and now a retired research mathematician from IBM (Yorktown Heights, NY), the late Richard P. Feynman, then a graduate student in physics and later a Nobel Laureate, and John W. Tukey, then a young mathematics instructor and now an Emeritus Professor at Princeton. It is interesting to remark that the diagrams Feynman devised for analyzing 6-faced hexaflexagons were forerunners of the famous *Feynman Diagrams* in modern atomic physics. A description of how to construct a 3-faced hexaflexagon may be found in any of the references [**1**] through [**4**]. Further, a detailed description is given in [**3**, pp. 63–74] of how to construct hexaflexagons with $3n$ faces without the use of straightedge or compass.

The particular hexaflexagon we will consider in this article is the tri-hexaflexagon,[1] so named because it has 3 *faces*; that is, in any given state of the flexagon, one face (consisting of 6 equilateral triangles) will be up, one face will be down, and one face will be hidden. Although the orientation of the faces will vary from state to state, the same 6 triangles will always appear together on a face.

We will show in this article how, by drawing a human visage on each face of the flexagon, and using a different color for each face, we can keep track of all the possible positions of the flexagon as it lies in a plane. We are thereby able to discover

[1] We may refer to the tri-hexaflexagon as simply "the flexagon" if no confusion would result.

that the set of motions of this flexagon which bring it into coincidence with itself constitutes the dihedral group $D_{18}$.

## 1. How to Build the Tri-hexaflexagon

The tri-hexaflexagon is constructed from a strip of paper containing 10 equilateral triangles[2] as shown in FIGURE 1. In order that the final model will flex easily the fold lines between the triangles should be creased firmly in *both* directions. Now we

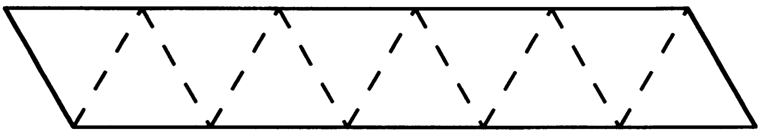


**FIGURE 1**
The strip

decorate the strip as shown in FIGURE 2, where we make the bottom surface of the strip visible by flipping the entire pattern piece over a *horizontal* axis as indicated by the figure (where the vertices $A$, $B$, $C$, $D$ should correspond with $A'$, $B'$, $C'$, $D'$,

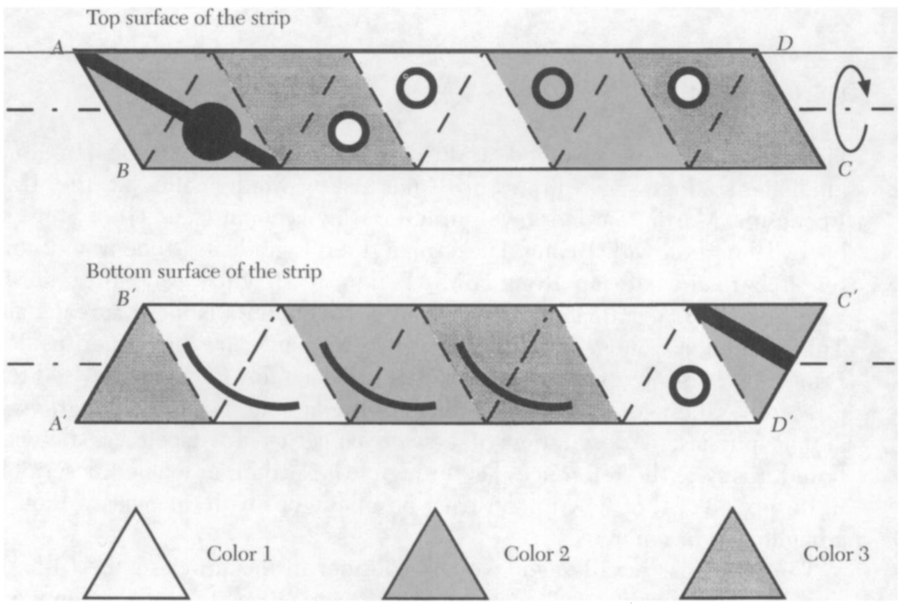



**FIGURE 2**
Decorating the strip

respectively, after you have flipped the piece over). *Caution:* Be careful here! Flipping the pattern piece over a *vertical* axis, and then decorating it as shown does *not* produce the desired flexagon.

[2] Notice that, from the point of view of decorating this piece, we have available a total of 20 triangles (because the strip of paper has two surfaces, the top surface and the bottom surface). When two of the triangles are glued to each other there remain 18 triangles with which to form the 3 faces.

Now we suggest that you view the construction of the flexagon as a puzzle. Here are some hints for constructing the flexagon with smiling (and frowning) faces.

(1) The first triangle on the upper portion of the strip is ultimately glued to the last triangle on the bottom portion (and it doesn't matter which one is on top of which). We suggest that you attach these triangles with a paper clip at first, and save the actual gluing until you are certain about the correctness of the construction.
(2) The completed flexagon should show the visage of a smiling face, entirely of color 1, as you lay it down as shown in FIGURE 3(a). And when you flip the flexagon over, about a horizontal axis, it should show the visage of a frowning face, entirely of color 3, oriented as displayed in FIGURE 3(b).

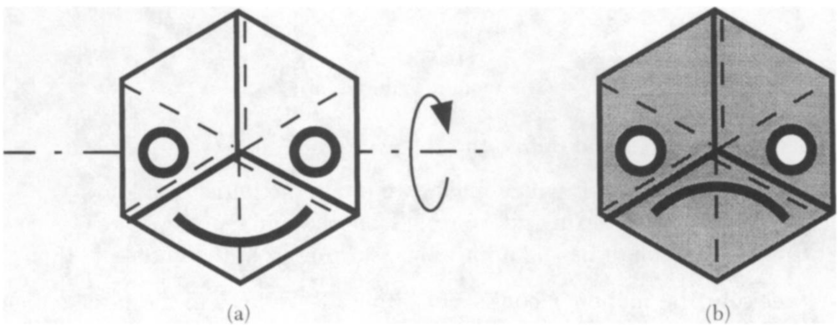



**FIGURE 3**
The tri-hexaflexagon

(3) The strip that created the hexagon contains three half-twists; thus, like the Möbius band, it has only one surface (or side). Geometrically this means there will be three slits on any face of this flexagon, symmetrically located at 120° intervals about its center. These slits are created by edges of the strip that go from alternate vertices of the hexagon to its center as shown in both parts of FIGURE 3.

From the last hint above we know that the flexagon now has only one surface. After you become proficient at manipulating your flexagon you may wish to verify with your own model that the repetitive pattern of three mouths, three right eyes, and three left eyes, in the colors 1, 2, 3, respectively, occurs as shown in FIGURE 4.

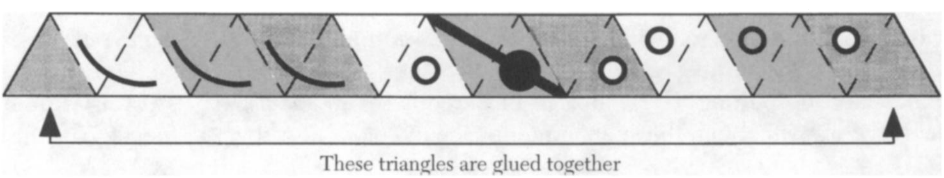



**FIGURE 4**
The entire surface of the strip

## 2. The Happy Group

First we will always need to start with the flexagon in a *standard initial position*, that is, with the smiling face of color 1 up and oriented precisely as shown in FIGURE 5(a).

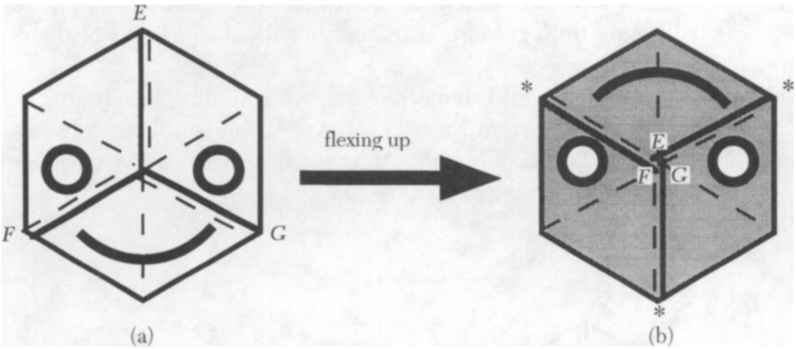



**FIGURE 5**
The motion $f$, flexing up

Now we assume $n \geq 0$ and define the following motions:

$$\begin{cases} \text{the identity motion } 1, \text{ which means we retain the initial position,} \\ f = \text{the motion of flexing up, starting from the initial position,} \\ f^n = \text{the motion of flexing up } n \text{ times, starting from the initial position.} \end{cases}$$

More precisely, the motion $f$ consists of lifting the vertices of the hexagon labelled $E$, $F$, and $G$ (in FIGURE 5(a)) above the flexagon until they meet, when the flexagon will come apart at the bottom and fall into the shape of a new hexagon with the vertices $E$, $F$, and $G$ at its center. If this is done correctly (it is important not to rotate the flexagon in either direction), you will see the upside-down smiling face of color 3 shown in FIGURE 5(b). Notice that the slits in the flexagon have revolved $\frac{1}{6}$ of a turn. Thus, when you flex up the second time you will have to bring the vertices marked with asterisks (*) together above the flexagon. A simple way to remember what to do is that, in each case, the vertex at the forehead of the human visage gets lifted to the center (and it disappears as the motion is completed).

We now follow the usual, obvious procedure of identifying a motion with its effect on the initial position. When we do this we see that $f^{18} = 1$, the identity motion.

Once you have mastered the motions $f^n$, you may verify the sequence of motions which produce the Happy Group shown in FIGURE 6; here we have adopted the identification indicated above.

Since $f^{18}$ is the identity, we see that the Happy Group is the cyclic group $C_{18}$, generated by $f$.

Next we define *flexing down*. To describe this motion $\bar{f}$, we begin, as before, with the flexagon in the standard initial position shown[3] in FIGURE 7(a). Then $\bar{f}$ means that we push the vertices of the hexagon labelled $H$, $J$, and $K$ downwards until they meet; at that stage the flexagon will come apart at the *top* and fall into the shape of a new hexagon with the vertices $H$, $J$, and $K$ at its center, but *underneath* the hexagon (this is indicated by putting $H$, $J$, and $K$ in parentheses in FIGURE 7(b)). If this is done correctly, we will obtain the smiling pirate face of color 2 as shown in FIGURE 7(b). Just

[3]This is, of course, the same initial position as that in FIGURE 5(a), but the labeling has changed.

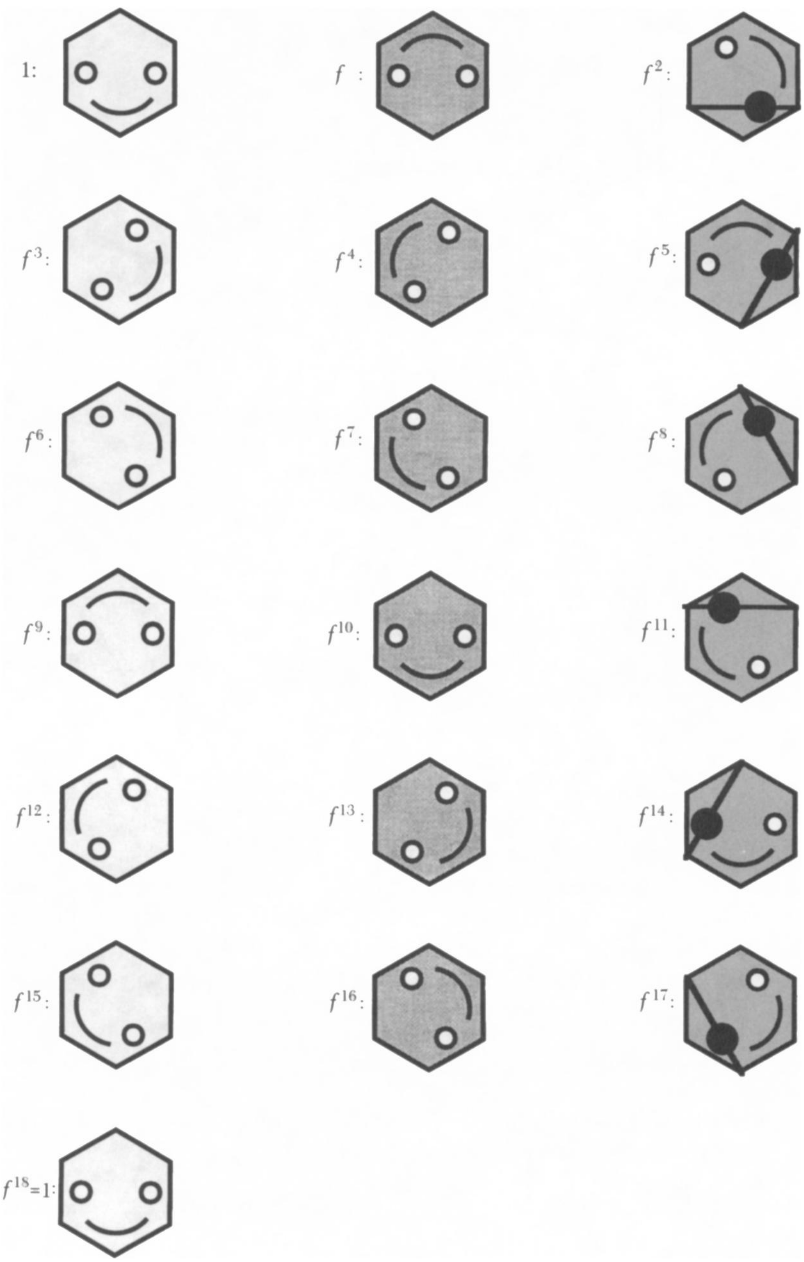



**FIGURE 6**
The Happy Group

as with the up-motions, it is important not to rotate the flexagon in either direction as we flex it. To obtain $\bar{f}^n$, we simply repeat the process of flexing down $n$ times (notice that when we flex down the second time it is the vertices labelled with the asterisk (*) that come together beneath the flexagon). It is interesting that, in flexing down, the vertex at the forehead of the human visage moves up (as when flexing up), but in this case the flexagon visibly splits across the forehead before it falls flat, revealing the pirate.

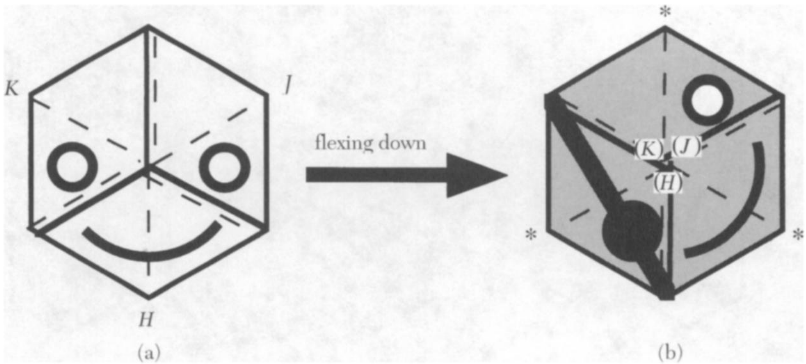



**FIGURE 7**
The motion $\bar{f}$, flexing down

Beginning with the flexagon in the standard starting position, you may verify that $\bar{f}$ yields the same face as $f^{17}$ of FIGURE 6. Thus $\bar{f} = f^{17} = f^{-1}$. This means that, if you start with the position indicated on the right of FIGURE 7 and flex down, you get the initial position. In other words, flexing up is the inverse of flexing down (and vice versa), as you might expect.

If you're enjoying this you may check your flexing skill by reversing all of the steps of the Happy Group in FIGURE 6.

## 3. The Entire Group

We realize that the full group for this flexagon must be larger than $C_{18}$ because no frowning faces ever appeared under the motions $f^n$. Cheerful as this situation is, it is plainly not complete. Like everything in this world this flexagon has good (happy) and bad (unhappy) features. In order to get the entire group we certainly need to have a motion that makes the unhappy faces visible. To achieve this we introduce a new motion,

$$t = \text{turn over (so the rotation is about a } \textit{horizontal} \text{ axis)}.$$

Thus, if we begin with the flexagon in the standard initial position and perform the motion $t$ we will see a frowning face of color 3 (see FIGURE 8).

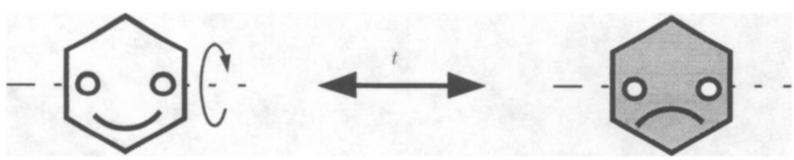



**FIGURE 8**
The motion $t$, turning over

Obviously $t$ is an involution, that is, $t^2 = 1$. FIGURE 9 shows that the motion $ft$ (meaning first do $t$, then do $f$) is not the same as $tf$ (meaning first do $f$, then do $t$). Check this (remembering that the flexagon should be in the standard initial position, in both cases, when you start). Thus we see that our new motion $t$ does *not* commute

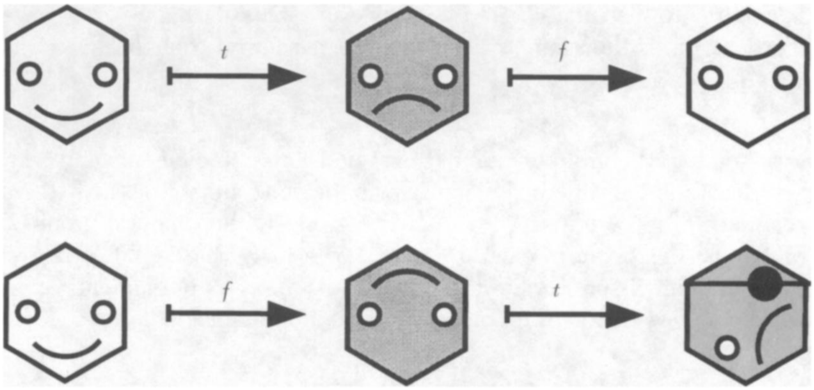



**FIGURE 9**
$ft \neq tf$

with $f$. We also notice that when the pirate frowns his patch covers his left eye, instead of the right one![4]

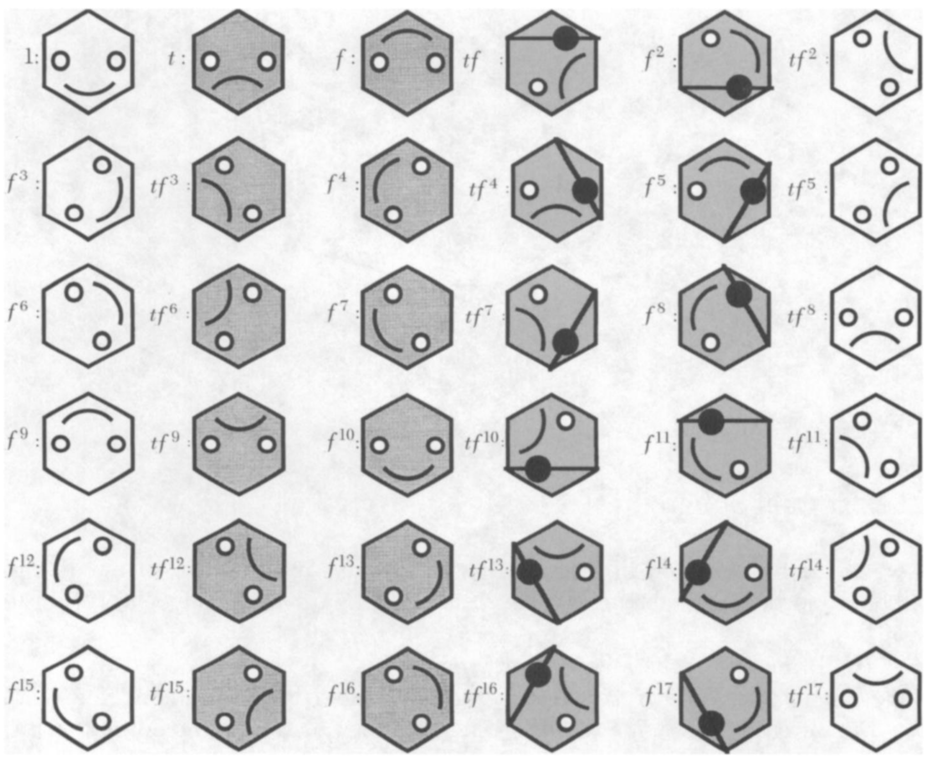



**FIGURE 10**
The entire group

FIGURE 10 shows all the possibilities for $f^n$ and $tf^n$. Notice that the first, third, and fifth columns are just the smiling faces from FIGURE 6. This observation may give you an idea of an easy way to confirm that the visages in FIGURE 10 are correct.

[4]Although we could do without any eye patches in analyzing this particular flexagon, it is clear that this feature may provide a better way of keeping track of the faces on more complicated flexagons—and we thought the pirate made an interesting addition to this group (visually and socially!).

We have already seen that $tf^n \neq f^n t$. However, since flexing *up*, as viewed from *above* the flexagon, is the same as flexing *down*, as viewed from *below* the flexagon, we have,

$$f^n t = tf^{-n}.$$

Thus we see that the group generated by $f$ and $t$ has 36 elements and is therefore the full group of motions of our flexagon. Since the generators $f$, $t$ satisfy the defining set of relations $f^{18} = 1$, $t^2 = 1$, $ft = tf^{-1}$, the group is the dihedral group $D_{18}$, the group of symmetries of the regular 18-gon (shown in FIGURE 11). The figure on the front cover of this issue shows the effects of the group elements, where the single-headed arrows denote the $f$ action and the double-headed arrows denote the $t$ action.

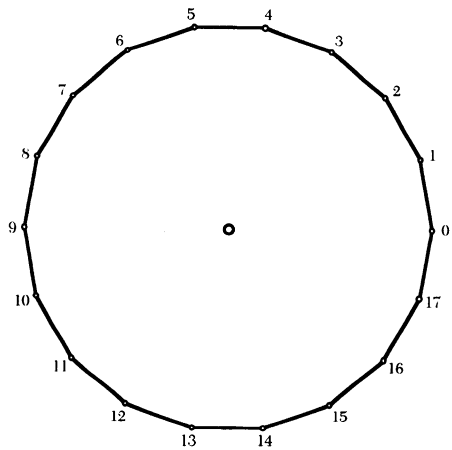



**FIGURE 11**
The regular 18-gon

## 4. A Normal Subgroup

If we are only interested in the different expressions on the faces of our flexagon, without respect to orientation, we have only 6 cases (as seen in FIGURE 12), instead of 36.

FIGURE 10 motivates the following argument.

We obtain the group of motions of the unoriented faces by adding the relation $f^3 = 1$ to our group $D_{18}$. The resulting quotient group of $D_{18}$ by the normal subgroup generated by $f^3$ is then generated itself by $F$ and $t$, subject to $F^3 = 1$, $t^2 = 1$,

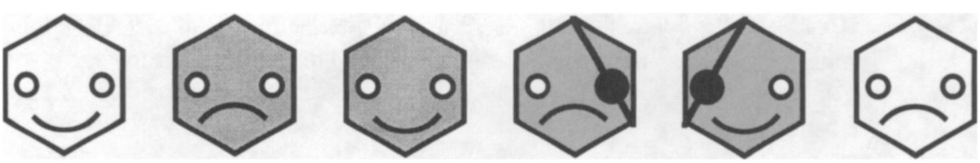


**FIGURE 12**
Expressions of the flexagon

$Ft = tF^{-1}$. Here, $F$ is, of course, the image in the quotient group of $f$; and the quotient group is just the symmetric group $S_3$.

## 5. A Challenge to the Reader

In **[4]** the tri-hexaflexagon was discussed and the group $S_3$ was obtained by using a flexagon where each of the 3 faces simply had different colors. In **[2]** the group $D_9$ was obtained by a systematic labeling of the vertices of the 6 triangles on each of the 3 faces of the tri-hexaflexagon. However, in order to obtain the entire group $D_{18}$, it was necessary to introduce a finer method of distinguishing between the different orientations of the faces; distinguishing between smiling and frowning visages did the trick. The obvious next question to explore is whether or not this, or some refinement of it, will help to identify the mathematical structure of the hexa-hexa-flexagon (with 6 faces). If we had a quick, or easy, answer to this question we wouldn't be stopping here!

### REFERENCES

1. Martin Gardner, *The Scientific American Book of Mathematical Puzzles and Diversions*, Simon and Schuster, New York, NY, 1959.
2. Michael Gilpin, Symmetries of the trihexaflexagon, this MAGAZINE 49, No. 4 (1976), 189–192.
3. Peter Hilton and Jean Pedersen, *Build Your Own Polyhedra*, Addison-Wesley, Menlo Park, CA, 1994.
4. Jean Pedersen, Sneaking up on a group, *Two-year College Mathematics Journal* 3 (1972), 9–12.

# Centers of Triangles of Fixed Center: Adventures in Undergraduate Research

WILLIAM MUELLER
University of Arizona
Tucson, AZ 85721

## 1. Introduction

There are many popular misconceptions about undergraduate research projects in mathematics: that they are only for advanced students; that they require a tremendous amount of extracurricular time and effort; that they must be guided by an advisor with a vast and detailed knowledge of the subject; that good topics, simultaneously rich and accessible, are difficult to find. The truth is much simpler. The joys and frustrations of mathematical research can be experienced in any subject, at any level. One must only learn to look closely, with an openness to the possibility of discovery.

And "looking" is easier than it has ever been before. Fast computers and friendly software allow students to manipulate their intuitions with a naturalness that most research mathematicians only dreamed of during their own undergraduate days. In the new age of electronic Surrealism,[1] students may wander freely through the landscape of mathematics, finding their own way to "advanced" destinations that were previously inaccessible to them. Moreover, and more importantly for the development of undergraduate research projects, it is now possible for students to stop along the way and scrutinize the objects in the landscape in all of their extraordinary detail. The computer, when guided by a genuine curiosity, is capable of revealing "neglected associations" in even the simplest things. The research advisor, on journeys such as these, need only act as a gentle chaperon, reminding students that the "control exercised by reason" (look carefully and systematically; write down what you see; prove it) is the key to turning their adventures into legitimate mathematical formulations.

## 2. Motivations

The investigation described here was suggested by some initial computer explorations carried out by students in a multivariable calculus class. Although the results of the investigation are of interest in themselves (in fact, despite their provenance in a subject that has been explored for centuries, I believe that the results are essentially new), we also mean to present the process of the investigation, as an example of how naturally an undergraduate research program may develop. The topic of the investigation comes along in the course of other studies and, at least initially, the investigation

[1]SURREALISM, *n.* Psychic automatism in its pure state, by which one proposes to express—verbally, by means of the written word, or in any other manner—the actual functioning of thought. Dictation of thought, in the absence of any control exercised by reason, exempt from any aesthetic or moral concern. Surrealism is based on the belief in the superior reality of certain forms of previously neglected associations, in the omnipotence of dreams, in the disinterested play of thought. It tends to ruin once and for all all other psychic mechanisms and to substitute itself for them in solving all the principal problems of life.

André Breton, *Manifesto of Surrealism* (1924) is an elementary extension of those studies. Careful, systematic observation, however, reveals the topic's beautiful, unexpected complexities.

Early on in multivariable calculus, students must be introduced to vectors. Before vectors are used to describe something else, which is itself new to students (such as space curves), it is instructive to use them in a more familiar setting. Force vectors are a traditional choice for this—intuitive if not always familiar—and their applicability is readily accepted. This ready acceptance comes at a price, however. The algebra of force vectors is so elementary that the notation's powerful simplicity is obscured.

The advantages of vector notation are much more apparent in another familiar setting: plane geometry. What most undergraduates seem to remember about this subject, from their high school experience, is a tortuous series of theorems and corollaries, and not a great deal of actual geometry. Presented with vector methods, however, the theorems of high school geometry take surprisingly simple form. Expressed elegantly and succinctly by the notation, the geometry itself becomes apparent.

The principal idea of vector-based proofs is to replace selected line segments in geometric figures with vectors, and then substitute the unifying notions of scalar multiplication and (vanishing) dot product for the sundry Euclidean propositions on parallel and perpendicular lines, respectively. For example, consider the theorem that says that the medians of a triangle are concurrent. If such theorems are proved at all in high school texts, they are usually treated algebraically, as a demonstration of analytic geometry, or, alternatively, as a kind of grand crescendo at the end of the text, in which all of the previous propositions are allowed to blow their horns. A tribute to the rigors of mathematical reason, in either presentation, but the geometry is usually lost. By vector methods, the proof is more straightforward:

THEOREM. *The medians of a triangle are concurrent.*

*Proof.* Let medians $AE$ and $BD$ meet at $P$, as in FIGURE 1. Then for some scalar $s$,

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{A} + s\boldsymbol{AE} = \boldsymbol{A} + s(\boldsymbol{E} - \boldsymbol{A}) = \boldsymbol{A} + s(((1/2)(\boldsymbol{B} + \boldsymbol{C})) - \boldsymbol{A}) \\ &= (1-s)\boldsymbol{A} + (s/2)\boldsymbol{B} + (s/2)\boldsymbol{C}. \end{aligned}$$

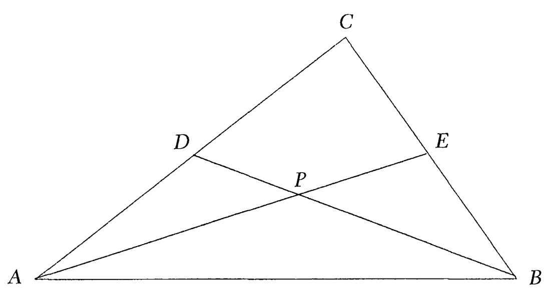



**FIGURE 1**
Two medians.

For some scalar $t$,

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{B} + t\boldsymbol{BD} = \boldsymbol{B} + t(\boldsymbol{D} - \boldsymbol{B}) = \boldsymbol{B} + t(((1/2)(\boldsymbol{A} + \boldsymbol{C})) - \boldsymbol{B}) \\ &= (1-t)\boldsymbol{B} + (t/2)\boldsymbol{A} + (t/2)\boldsymbol{C}. \end{aligned}$$

Equating the two expressions for $\boldsymbol{P}$, we find $s = t = 2/3$. So $\boldsymbol{P} = (1/3)(\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{C})$. By the symmetry of this expression, we see that each pair of medians would lead to the same intersection.

Not only is the concurrence immediately apparent in this proof, but the fact that it occurs at a trisection point of each of the medians seems to come for free. The notational advantages displayed in such proofs are much more impressive than any of the vector demonstrations that, for example, two people can pull a car out of the mud. Usually, in fact, a certain number of students (often the ones with the most painful memories of high school geometry) are impressed enough to want to do more; to see what other theorems can be proved this way. At which point the advisor-to-be smiles gently.

## 3. Questions and Answers

Readers of this MAGAZINE were reminded, in **[1]**, that there are, in fact, many interesting linear triples associated with a triangle, meeting at many other *centers*. The center where the medians meet is called the *centroid* of the triangle. Other classic centers include the *orthocenter*, where the altitudes are concurrent; the *incenter*, where the angle bisectors are concurrent; and the *circumcenter*, where the perpendicular bisectors of the sides are concurrent. That each of these centers exists is easily proved with vector methods, and these proofs make good follow-up exercises to the theorem presented above.

Perhaps the most basic question about these centers is this: where do they lie? That is, if one considers triangles of many different shapes and sizes, what is the locus of a particular center? It isn't difficult to demonstrate that, allowed arbitrary rotations and dilations of a triangle (without translation), any of these centers can be made to fall at any point in the plane—a less than marvelous result. Still, the question remains interesting for specific natural classes of triangles, such as those for which one of the centers is held fixed. In this case, the results are indeed quite marvelous.

We will consider two such classes: triangles of a fixed incenter and triangles of a fixed circumcenter. These particular classes are chosen because the auxiliary existence of a fixed incircle or circumcircle, respectively, leads to a parametrization of each class that is especially simple. For each of these two classes, we will consider the locus of the remaining three centers listed above.

**Triangles of fixed incenter** How does one consider, e.g., all triangles of a fixed incenter? Similar triangles obtained through rotations and dilations generate only copies of the essential loci, differing in orientation and in scale, and so it suffices to consider triangles with a unit incircle and a base $AB$ parallel to the $x$-axis, as in FIGURE 2.

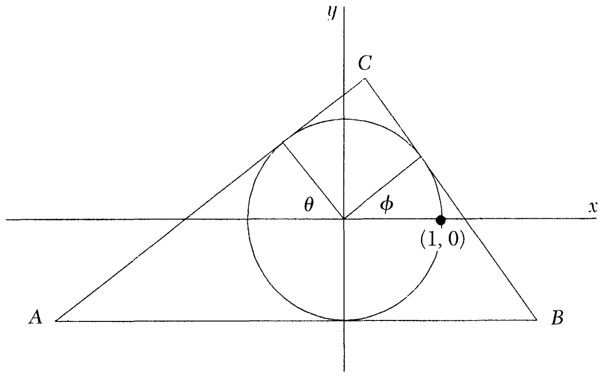



**FIGURE 2**
Triangles of fixed incircle.

This class may be parametrized by the two angles $\theta$ and $\phi$, measured upward from the $x$-axis, giving the inclination of the sides $AC$ and $BC$, respectively. To assure that the sides meet at $C$, we will need to restrict the parameters to $\theta \in (-\pi/2, \pi/2)$ and $\phi \in (-\theta, \pi/2)$. A discretization of this domain is shown in FIGURE 3.

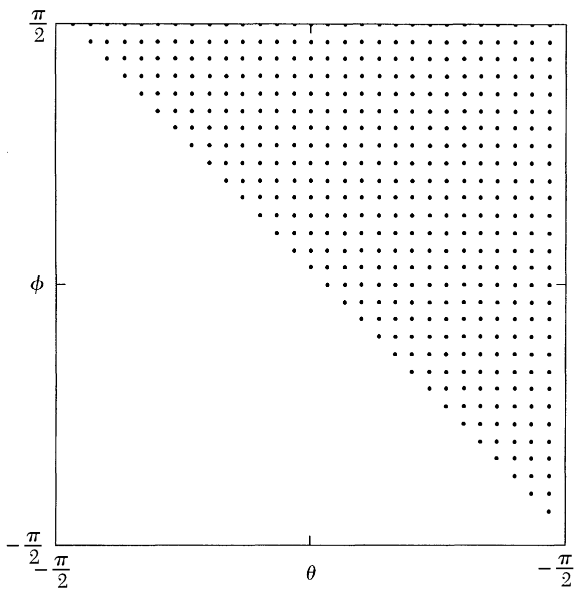



**FIGURE 3**
Domain of triangles of fixed incircle.

Each point in this domain represents a particular triangle with its incenter at the origin. For purposes of computation, the domain must be treated discretely; but, of course, the mesh size may be varied to obtain arbitrary detail (at the cost of increased computation time).

In order to find the other centers of the triangles in FIGURE 3, it is first necessary to express the coordinates of the triangle's vertices in terms of the parameters $\theta$ and $\phi$. This is another good vector exercise, involving parametrization of the three lines representing the sides, followed by a computation of the intersections by solving the resulting systems of equations. A symbolic processor, such as *Maple*, is useful for the latter task. ("Useful" has both a computational and a pedagogical sense. *Maple*, for example, often returns trigonometric expressions that are exceedingly complicated, and it is useful to simplify (and confirm) these expressions with trigonometric identities.) The results, referring to FIGURE 2, are:

$$\boldsymbol{A} = (-(\tan\theta + \sec\theta), -1),$$

$$\boldsymbol{B} = (\tan\phi + \sec\phi, -1),$$

and $$\boldsymbol{C} = ((\sin\theta - \sin\phi)/\sin(\theta+\phi), (\cos\theta + \cos\phi)/\sin(\theta+\phi)).$$

It is instructive to plot these vertices as $\theta$ and $\phi$ step through the domain at equal intervals, as in subsequent computations. The plot is shown in FIGURE 4, together with the corresponding edges of some of the triangles.

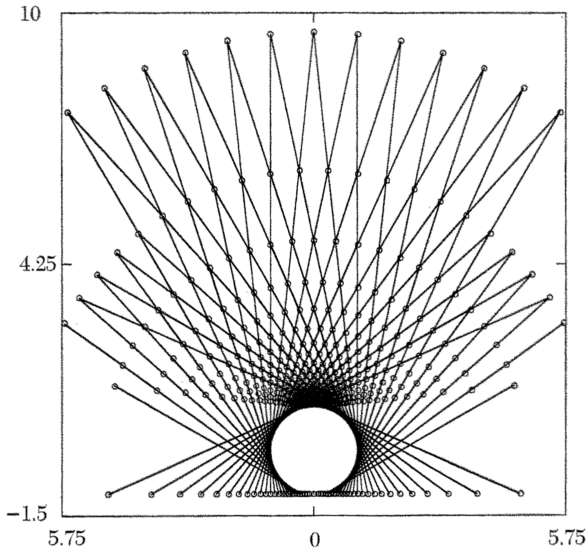



**FIGURE 4**
Vertices of triangles of fixed incircle.

We see that, for equal increments of the parameters, the triangles themselves do not fill the plane uniformly. Technically, the transformations we consider from the $\theta\phi$-plane to the $xy$-plane will not be linear ones. We will keep this in mind when examining later plots.

Before considering other centers, a good limbering-up exercise is to compute the values of a more familiar function of $\theta$ and $\phi$: the area. Using vector algebra to determine the base and the height, we find

$$A(\theta,\phi)=\frac{(\sin(\theta+\phi)+\cos(\theta)+\cos(\phi))^2}{\sin(\theta+\phi)\cos(\theta)\cos(\phi)}.$$

This surface and its contours are plotted in FIGURE 5.

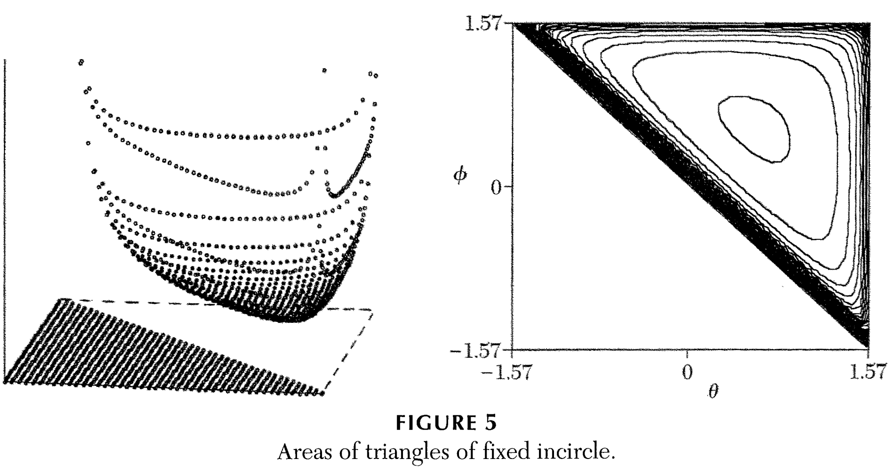



**FIGURE 5**
Areas of triangles of fixed incircle.

There is an obvious minimum, which occurs when $\theta$ and $\phi$ are both $\pi/6$. This result should be familiar to anyone who has worked the exercises in old calculus books: the triangle with fixed incircle of minimum area is equilateral.

**Centroids** Now let us return to the centroid, the point at which the medians are concurrent. The medians have already been expressed in terms of the vertices, in the proof that we gave above, and computing the coordinates of the centroid at $P$ amounts to finding the value of either of the parameters $s$ or $t$ at the point of intersection, exactly as in the proof. Pedagogically, it is nice that, for each of the centers we consider, the proof of existence provides just such a method of computation. In the case of the centroid, it is great fun to have a symbolic processor strenuously compute the solution to the system of equations in $s$ and $t$ for medians given in terms of $\theta$ and $\phi$, only to have it finally announce that $\theta$ and $\phi$ have disappeared and $s = t = 2/3$, which we already knew (but may have forgotten).

With this information in hand, we can then ask the computer to step through the domain in FIGURE 3, plotting the centroid for each of the triangles in the class. The result, as shown in FIGURE 6, is spectacular:

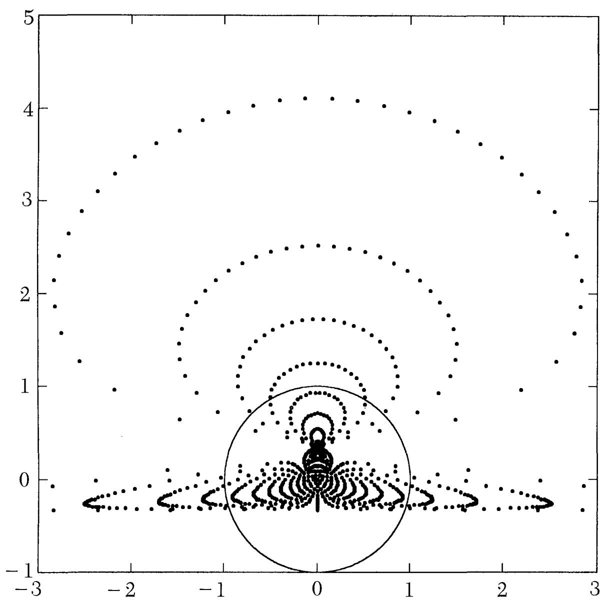



**FIGURE 6**
Centroids of triangles of fixed incircle.

The circle shown is the incircle. The complexity of this figure seems to persist at smaller mesh sizes and greater resolutions. Apparently, the centroids fall only within a very specific region of the plane, the boundaries of which are far from obvious. Providing an accurate description of this region, or of the curves of constant $\theta$ and $\phi$, would make an interesting research project, although it is by no means clear what level of mathematical sophistication would be required.

**Orthocenters** Computing the loci of other centers while the incenter is held fixed produces equally intriguing results. Consider, next, the orthocenter, the computation of which involves the solution of a similar system of equations.

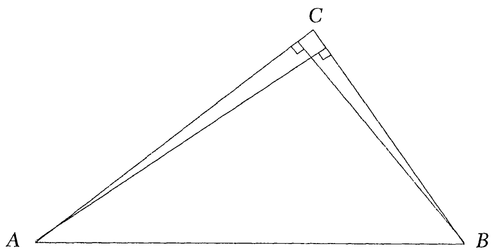



**FIGURE 7**
Computing the orthocenter.

In this case, we parametrize lines whose direction must have a vanishing dot product with the direction of other lines as in FIGURE 7. The resulting system of equations is predictably not pretty, but it may be solved, once again, with the aid of a symbolic processor (followed by the requisite checking and simplification). The results are plotted in FIGURE 8.

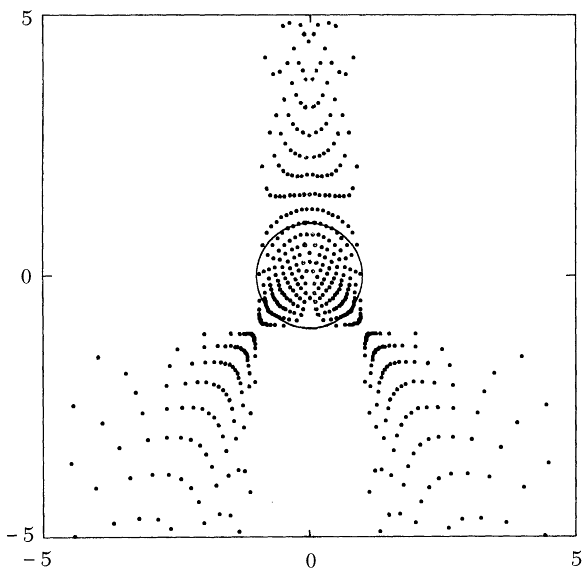



**FIGURE 8**
Orthocenters of triangles of fixed incircle.

Again, the centers appear to be confined to a region of the plane with unknown boundaries.

**Circumcenters** We save for last the problem whose statement contains the most appealing balance of Euclidean yin and yang: circumcenters of triangles with a fixed incenter. As FIGURE 9 shows, this problem is computationally similar to the problem of the orthocenters. The resulting locus is shown in FIGURE 10.

**Triangles of fixed circumcenter** Investigating the class of triangles with a fixed circumcenter presents no new difficulties; in fact, the parametrization of this class is a

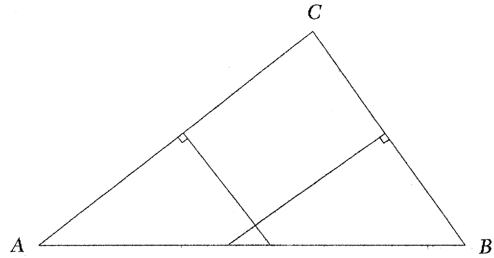



**FIGURE 9**
Computing the circumcenter.

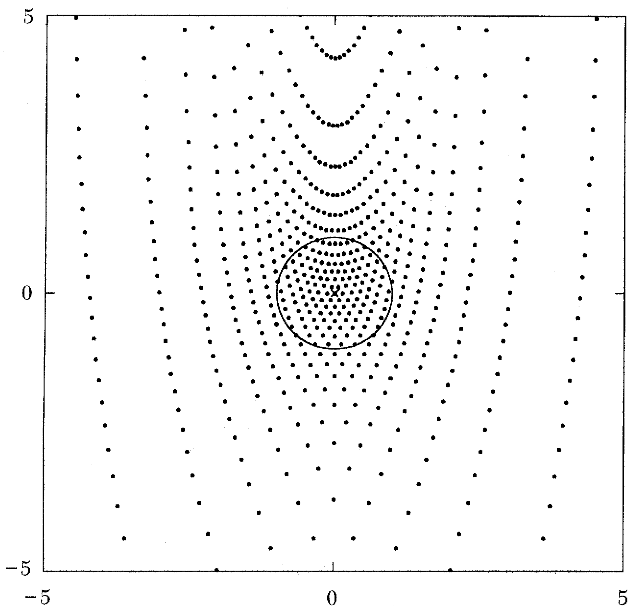



**FIGURE 10**
Circumcenters of triangles of fixed incircle.

bit simpler. As before, it suffices to consider triangles with a unit circumcircle and a base $AB$ parallel to the $x$-axis. The members of this class are conveniently described by the angle $\theta$ and the intercept $b$, shown in FIGURE 11.

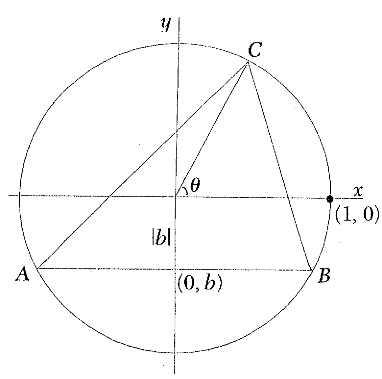



**FIGURE 11**
Triangles of fixed circumcircle.

A discretization of the parameter domain is shown in FIGURE 12.

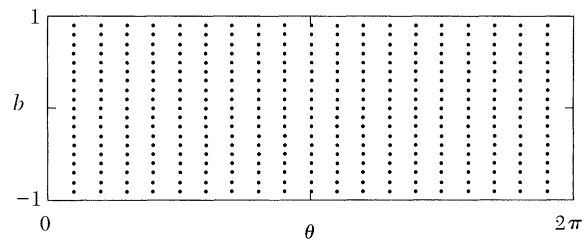



**FIGURE 12**
Domain of triangles of a fixed circumcenter.

The coordinates of the vertices are:

$$\mathbf{A} = \left(-\sqrt{1-b^2}\,, b\right), \quad \mathbf{B} = \left(\sqrt{1-b^2}\,, b\right), \quad \text{and} \quad \mathbf{C} = (\sin\theta, \cos\theta).$$

**Surface and countour plots** of the areas of these triangles, given by $A(\theta, b) = \sqrt{1-b^2}\,|\sin\theta - b|$, is shown in FIGURE 13.

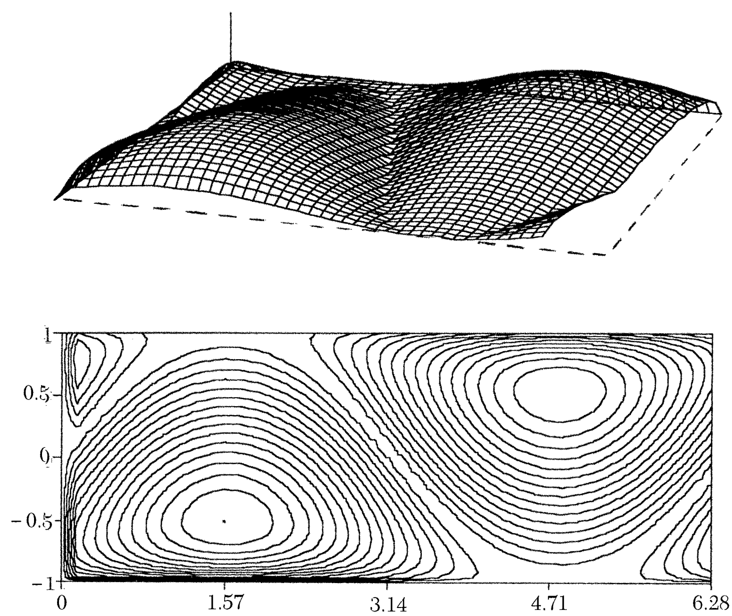



**FIGURE 13**
Areas of triangles of fixed circumcircle.

The two equilateral triangles in the domain produce the maximum areas. Notice, as well, the hollow that snakes around the peaks, corresponding to the two $\theta$'s that will lead to zero area for any $b$.

**Centroids, orthocenters** Taken together, the plots of the centroids and the orthocenters are quite striking in this case, as shown in FIGURES 14 and 15:

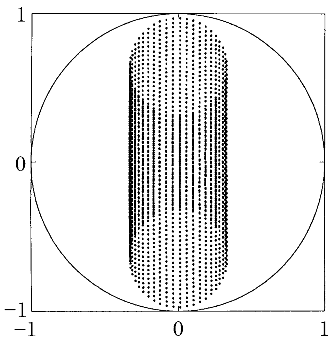



**FIGURE 14**
Centroids of triangles of fixed circumcircle.

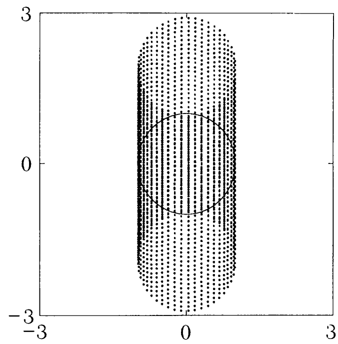



**FIGURE 15**
Orthocenters of triangles of fixed circumcircle.

The circumcircle is shown in each plot; without the circumcircles, the two would be indistinguishable (except for scale). Apparently, the centroids fall only within the central third of the circumcircle, and the orthocenters fall within a diameter on either side. Surely, theorems await here.

**Incenters** Finding the incenters for this class offers a new computational challenge: we must parameterize lines whose direction from two other lines is at equal angles. (See FIGURE 16.) The dot product will again allow us to describe this situation, but the resulting system of equations is more complex than in any of the previous cases.

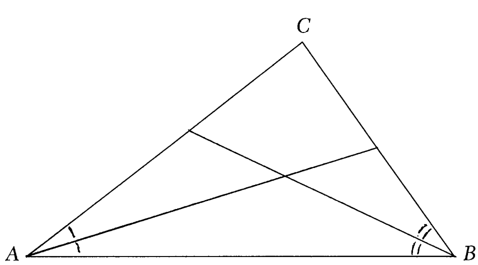



**FIGURE 16**
Computing the incenter.

After much clicking and whirring, the computer produces FIGURE 17.

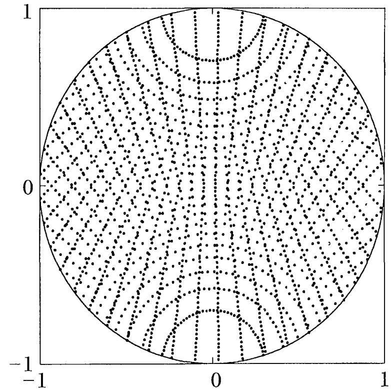



**FIGURE 17**
Incenters of triangles of fixed circumcircle.

We close with the observations of a student of fixed points, who was seen lingering thoughtfully at the International Surrealist Exhibition in London in 1936.

*At the still point of the turning world. Neither flesh nor fleshless;*
*Neither from nor towards; at the still point, there the dance is,*
*But neither arrest nor movement. And do not call it fixity,*
*Where past and future are gathered. Neither movement from nor towards,*
*Neither ascent nor decline. Except for the point, the still point,*
*There would be no dance, and there is only the dance.*

T.S. Eliot, *Burnt Norton*

**Acknowledgment.** The author wishes to thank Betty Mayfield and the referees for many helpful suggestions.

## REFERENCE

1. C. Kimberling, Central points and central lines in the plane of a triangle, this MAGAZINE 67 (1994), 163–187.

# A Family Portrait of Primes—A Case Study in Discrimination

SHAILESH A. SHIRALI
Rishi Valley School, Rishi Valley-517352
Chittoor District, Andhra Pradesh,
India

## Introduction

We often come across examples of how nature discriminates among its children—bestowing wonderful favors on some while others seem less fortunate. The following note documents one such case study, namely, the discrimination between primes of the type 1(mod 4) and of the type $-1$(mod 4). We shall find that there are surprisingly many differences in behavior between the primes of these two families. Our main aim will be to present a proof of the following result (see **[3]**, page 113), which does not seem to be too well-known: if $p$ is a prime number of the form 1(mod 4), then

$$\sum_{i=0}^{i=[p/4]} \left[\sqrt{ip}\right] = \frac{p^2-1}{12}.$$

(As usual, $[z]$ denotes the greatest integer $\leq z$.)

## 1. First Observations

We start with an elementary observation: *If $n$ is a positive integer of the form $-1$(mod 4), then $n$ possesses a prime factor of the same form.* For instance, the number 63 possesses the prime factor 7, and both 63 and 7 are of the form $-1$(mod 4). No such observation can be made for numbers of the form 1(mod 4); for instance, $21 \equiv 1 \pmod 4$ but neither of the two prime factors of 21 is of this form.

The proof is easy. Since $n$ is odd, its prime factors are all odd. If its prime factors were all of the form 1(mod 4), then $n$ too would be of this form. This follows because $1 \times 1 = 1$ in modulo 4 arithmetic. Since $n \not\equiv 1 \pmod 4$, it follows that there is at least one prime factor $p$ of $n$, $p \not\equiv 1 \pmod 4$. (By an easy extension of the proof, it follows that there must be an *odd* number of such prime factors.)

This observation leads to an elementary proof that there are infinitely many prime numbers of the form $-1$(mod 4). We define the integer sequence $\{A_n\}_{n=1}^{\infty}$ in the following recursive manner: $A_1 = 3$, and for $n > 1$,

$$A_n = 4(A_1 A_2 A_3 \ldots A_{n-1})^2 - 1.$$

Then $A_n \equiv -1 \pmod 4$, so $A_n$ possesses at least one prime factor of the form $-1$(mod 4). Next, observe that the $A_n$ are coprime to one another. To see why, note that for each pair of indices $m$ and $n$, $m > n$, $A_m \equiv -1 \pmod{A_n}$, so a common divisor of $A_m$ and $A_n$ would have to be a divisor of 1. This means that the common divisor is 1, thus confirming that $A_m$ and $A_n$ are coprime to one another and implying that for each pair of indices $m$ and $n$, $m \neq n$, the sets of prime factors of $A_m$ and $A_n$ are mutually disjoint. Since there are infinitely many such sets, and since each of these sets contains at least one prime of the form $-1 \pmod 4$, it follows that the number of such primes is infinite.

One can prove in a somewhat analogous manner that there are infinitely many primes of the form $1 \pmod 4$, but the proof is, not surprisingly, rather more subtle. We start by proving an important result about primes of the form $1 \pmod 4$.

THEOREM. *Let $x$ be a positive integer. Then every odd prime divisor of $x^2+1$ is of the form* $1 \pmod 4$.

*Example.* Consider the numbers $8^2+1=65=5\times 13$ and $13^2+1=170=2\times 5\times 17$. Note that the primes 5, 13, and 17 are all of the form $1 \pmod 4$.

*Proof.* Let $p$ be an odd prime divisor of $x^2+1$; then $x^2 \equiv -1 \pmod p$, and by raising each side of this congruence to the $(p-1)/2$-th power, we obtain:

$$x^{p-1} \equiv (-1)^{(p-1)/2} \pmod p.$$

Now $p$ and $x$ are coprime (since $p \mid x^2+1$), so by the little Fermat theorem, $x^{p-1} \equiv 1 \pmod p$. This means that $(-1)^{(p-1)/2} \equiv 1 \pmod p$. The quantity on the left side is $\pm 1$, and the minus sign cannot hold (a minus sign would imply that $-1 \equiv 1 \pmod p$ or $p \mid 2$, which is absurd). We conclude that $(-1)^{(p-1)/2} = 1$ and therefore that $p \equiv 1 \pmod 4$. □

We now show how this result can be used to show that there are infinitely many primes of the form $1 \pmod 4$. The proof is organized along the same lines as the one presented in §1. We consider the Fermat numbers $\{F_n\}_{n=1}^{\infty}$, defined thus:

$$F_n = 2^{2^n} + 1.$$

Thus $F_1 = 5$, $F_2 = 17$, $F_3 = 257, \dots$ . An alternative definition is as follows: $F_1 = 5$, $F_{n+1} = (F_n - 1)^2 + 1$ for $n \geq 1$. As is readily seen, for each $n \geq 1$, $F_n$ is of the form $x^2+1$ and so each prime factor of $F_n$ is of the form $1 \pmod 4$. It only remains to note that the $F_n$ are coprime to one another. To show this, consider the numbers $F_8$ and $F_5$. We have, $F_8 = (F_7-1)^2+1 = (F_6-1)^4+1 = (F_5-1)^8+1$, or, writing $y = F_8$, $x = F_5$ for notational convenience,

$$y = (x-1)^8 + 1 = x^8 - 8x^7 + \cdots - 8x + 2 \equiv 2 \pmod x.$$

Therefore a common prime factor $p$ of $x, y$ would have to be a divisor of 2. Since $x, y$ are both odd, we have $p > 2$; but this is absurd. It follows that $x$ and $y$ are coprime. We see in a similar manner that the Fermat numbers are all coprime to one another (our proof has been specific to the indices 5 and 8 but the general idea should be clear) and so the sets of prime factors of these numbers are mutually disjoint. Since there are infinitely many Fermat numbers, it follows that there are infinitely many prime numbers, indeed that there are infinitely many primes of the form $1 \pmod 4$, since the prime factors are all of this form.

Yet another proof in this direction can be devised by considering the integer sequence $\{B_n\}_{n=1}^{\infty}$ defined thus: $B_1 = 5$, and for $n > 1$,

$$B_n = 4(B_1 B_2 B_3 \dots B_{n-1})^2 + 1.$$

The details are left to the reader.

## 2. Quadratic Residues

We move next to the concept of quadratic residues. For a given prime number $p$, a non-zero integer $a$ is called a *quadratic residue* of $p$ if there exists an integer $x$ such that $x^2 \equiv a \pmod p$; if not, $a$ is a *quadratic non-residue*. For instance, 2 is a quadratic residue of 17, since $6^2 = 36 \equiv 2 \pmod{17}$. Let $QR(p)$ denote the set of quadratic residues of the prime $p$. ("The set" makes sense only when we remember that we are working modulo $p$. This implies that we do not include in $QR(p)$ elements that are congruent to one another, because a set is not permitted to contain repeated elements. Therefore from each residue class modulo $p$, at most one element is placed in $QR(p)$.) Obviously $QR(p)$ is congruent modulo $p$ to the set

$$\{1^2, 2^2, 3^2, \dots, (p-2)^2, (p-1)^2\}.$$

Since $a^2 \equiv (p-a)^2 \pmod p$ for any integer $a$, we can equally well write:

$$QR(p) = \left\{1^2, 2^2, 3^2, \dots, \left(\frac{p-1}{2}\right)^2\right\} \pmod p.$$

*Example.* $QR(11) = \{1^2, 2^2, 3^2, 4^2, 5^2\}$, so $QR(11) = \{1, 4, 9, 5, 3\}$; that is, the quadratic residues of 11 are 1, 3, 4, 5 and 9. (Since we are working modulo 11, it is equally valid to write: $QR(11) = \{1, 15, 9, 16, 3\}$.) It follows from this observation that for each prime $p$, there are as many quadratic residues as quadratic non-residues.

The definition extends in a natural manner to non-prime moduli: the non-zero integer $a$ is said to be a quadratic residue of the integer $n$ if there exists an integer $x$ for which $x^2 \equiv a \pmod n$, and $QR(n)$ is the set of quadratic residues of $n$. (As earlier, we work modulo $n$; that is, we ensure that pairs of numbers congruent to one another modulo $n$ are not both in $QR(n)$.)

*Example.* For the modulus 10,5 is a quadratic residue while 7 is a quadratic non-residue, and $QR(10) = \{1, 4, 5, 6, 9\}$.

The theorem proved earlier can now be stated compactly as follows: *If* $p \equiv -1 \pmod 4$, *then* $-1 \notin QR(p)$. For, by definition, if $-1 \in QR(p)$, then there exists an integer $x$ such that $x^2 \equiv -1 \pmod p$, that is, such that $p \mid x^2 + 1$. However, we have already established that the odd prime factors of $x^2 + 1$ are all of the form $1 \pmod 4$.

We now prove a still stronger statement.

THEOREM. *If $p$ is a prime number of the form* $1 \pmod 4$, *then* $-1 \in QR(p)$; *that is, there exists an integer $x$ for which $x^2 + 1$ has $p$ as a divisor.*

*Proof.* We use Wilson's theorem (see **[2]**, page 68): *For any prime $p$,*

$$1 \times 2 \times 3 \times \cdots \times (p-2) \times (p-1) \equiv -1 \pmod p.$$

Now $a \equiv -(p-a) \pmod p$, so the factors $a$ and $(p-a)$ can be paired off in this product ($a = 1, 2, 3, \dots, (p-1)/2$). We obtain the following relation:

$$\left(1 \times 2 \times 3 \times \cdots \times \left(\frac{p-1}{2}\right)\right)^2 \times (-1)^{(p-1)/2} \equiv -1 \pmod p.$$

Since $p \equiv 1 \pmod 4$, $(p-1)/2$ is even and $(-1)^{(p-1)/2} = 1$. Writing $x$ for $1 \times 2 \times 3 \times \cdots \times (p-1)/2$, we obtain $x^2 \equiv -1 \pmod p$ or $p \mid x^2 + 1$. We have thus found an integer $x$ with the stated property. Indeed, we have shown the following: *For an odd prime $p$,* $-1 \in QR(p) \Leftrightarrow p \equiv 1 \pmod 4$. □

*Example.* If $p = 17$, then $(p-1)/2 = 8$ and $1 \times 2 \times 3 \cdots \times (p-1)/2 = 40320$. The analysis above assures us that $17|40320^2 + 1$. To find a smaller $x$ for which $17|x^2 + 1$, we use the congruence $40320 \equiv 13 \pmod{17}$ to conclude that $17|13^2 + 1$, an assertion readily verified by hand.

An even stronger statement can be proved using these general techniques in conjunction with some ideas from lattice point geometry: *An odd prime p is a sum of two squares if and only if* $p \equiv 1 \pmod 4$. The 'only if' part of this assertion is trivial, because every square is congruent to 0 or 1(mod 4), so a sum of two squares is congruent to 0, 1 or 2(mod 4). Therefore if $p \equiv -1 \pmod 4$, then $p$ is not a sum of two squares. The reverse part of the implication is harder to prove and we shall not do so. (See **[6]** for a proof different from the usual ones, and **[5]** for some comments on this proof.)

## 3. A Remarkable Equality

We now present the statement and proof of a remarkable theorem (**[3]**, page 113).

THEOREM. *If p is a prime number of the form* 1(mod 4), *then*

$$\sum_{i=1}^{[p/4]} \left[\sqrt{ip}\right] = \frac{p^2-1}{12}.$$

*Example.* To illustrate the meaning of the theorem, let $p = 17$. Here is the result:

$$\left[\sqrt{17}\right] + \left[\sqrt{34}\right] + \left[\sqrt{51}\right] + \left[\sqrt{68}\right] = 4 + 5 + 7 + 8 = 24 = \frac{17^2 - 1}{12},$$

in agreement with the claim. No such result appears to hold for primes of the form $-1 \pmod 4$. For example, with $p = 19$ we obtain the sum $4 + 6 + 7 + 8 = 25$ on the left side, and $(19^2 - 1)/12 = 30$ on the right side.

*Proof.* For notational convenience, we write

$$F(p) = \sum_{i=1}^{[p/4]} \left[\sqrt{ip}\right].$$

Now $[\sqrt{ip}\,] = k$ if and only if $k^2 < ip < (k+1)^2$, that is, $k^2/p < i < (k+1)^2/p$. (Equality cannot hold, because $p$ is prime and $i < p$.) For a fixed value of $k$, the number of values of $i$ for which this relation holds is

$$\left[\frac{(k+1)^2}{p}\right] - \left[\frac{k^2}{p}\right].$$

Therefore,

$$F(p) = \sum_{k=1}^{[p/2]} k \cdot \left(\left[\frac{(k+1)^2}{p}\right] - \left[\frac{k^2}{p}\right]\right).$$

(To see where the upper limit $[p/2]$ comes from, note that the largest value of $k$

corresponds to $i=[p/4]$.) This sum is rewritten as follows:

$$F(p)=\sum_{k=1}^{[p/2]}\left((k+1)\cdot\left[\frac{(k+1)^2}{p}\right]-k\cdot\left[\frac{k^2}{p}\right]-\left[\frac{(k+1)^2}{p}\right]\right)$$
$$=\left(\frac{p+1}{2}\right)\cdot\left[\frac{(p+1)^2}{4p}\right]-1\cdot\left[\frac{1}{p}\right]-\sum_{k=1}^{[p/2]}\left[\frac{(k+1)^2}{p}\right].$$

Now $[(p+1)^2/4p]=[p/4+1/2+1/4p]=(p-1)/4$ and $[1/p]=0$, so we obtain,

$$F(p)=\frac{p^2-1}{8}-\sum_{k=1}^{[p/2]}\left[\frac{(k+1)^2}{p}\right]$$
$$=\frac{p^2-1}{8}-\sum_{k=1}^{[p/2]+1}\left[\frac{k^2}{p}\right],$$

making use once more of the fact that $[1/p]=0$. It follows that

$$F(p)=\frac{p^2-1}{12}\Leftrightarrow\sum_{k=1}^{[p/2]+1}\left[\frac{k^2}{p}\right]=\frac{p^2-1}{24}.$$

We write $G(p)$ for the sum $\sum_{k=1}^{[p/2]+1}[k^2/p]$. Note that

$$G(p)=\sum_{k=1}^{[p/2]+1}\left[\frac{k^2}{p}\right]=\sum_{k=1}^{[p/2]+1}\frac{k^2}{p}-\sum_{k=1}^{[p/2]+1}\frac{\mathrm{Rem}(k^2\div p)}{p},$$

where the symbol $\mathrm{Rem}(a\div b)$ denotes the non-negative remainder obtained from the indicated division. (*Example*: $\mathrm{Rem}(17\div 5)=2$.) Next,

$$\sum_{k=1}^{[p/2]+1}\frac{k^2}{p}=\frac{1}{p}\,\frac{((p+1)/2)((p+3)/2)(p+2)}{6}$$
$$=\frac{(p+1)(p+2)(p+3)}{24p},$$

making use of the formula for the sum $\sum_{i=1}^{n}i^2$. Therefore the relation $G(p)=(p^2-1)/24$ holds if and only if the following equality holds:

$$\sum_{k=1}^{[p/2]+1}\mathrm{Rem}(k^2\div p)=p\left(\frac{(p+1)(p+2)(p+3)}{24p}\right)-p\left(\frac{p^2-1}{24}\right),$$

that is, if and only if

$$\sum_{k=1}^{[p/2]+1}\mathrm{Rem}(k^2\div p)=\left(\frac{p+1}{2}\right)^2.$$

Now the sum $\sum_{k=1}^{[p/2]+1}\mathrm{Rem}(k^2\div p)$ is equal to

$$\sum_{k=1}^{(p-1)/2}\mathrm{Rem}(k^2\div p)+\mathrm{Rem}\left(\left(\frac{p+1}{2}\right)^2\div p\right).$$

The first term, namely $\sum_{k=1}^{(p-1)/2} \mathrm{Rem}(k^2 \div p)$, is the sum of the quadratic residues $x$ of $p$, $1 \leq x \leq p$. The second term simplifies to $(3p+1)/4$, because

$$\left(\frac{p+1}{2}\right)^2 - \left(\frac{3p+1}{4}\right) = \frac{p(p-1)}{4},$$

a multiple of $p$. Next, $-1$ is a quadratic residue of $p$, because $p \equiv 1 (\mathrm{mod}\ 4)$, so if $x \in QR(p)$ then $p - x \in QR(p)$. It follows that $\sum_{k=1}^{(p-1)/2} \mathrm{Rem}(k^2 \div p)$ is equal to $(1+2+3+\cdots+p-1)/2 = p(p-1)/4$ and therefore that

$$\sum_{k=1}^{[p/2]+1} \mathrm{Rem}(k^2 \div p) = \frac{p(p-1)}{4} + \frac{3p+1}{4}$$
$$= \left(\frac{p+1}{2}\right)^2,$$

as required, and the stated result stands proved. □

## 4. Extensions to Other Moduli

It is interesting to try to extend the result to other moduli. Noting that the sum $\sum_{k=1}^{(p-1)/2} \mathrm{Rem}(k^2 \div p)$ occurs in several places in the analysis above, we define the function $R(n)$ for positive integers $n$ as follows:

$$R(n) = \sum_{k=1}^{[(n-1)/2]} \mathrm{Rem}(k^2 \div n).$$

For which values of $n$ is $R(n)$ equal to $n(n-1)/4$? We already know that equality holds when $n$ is a prime of the form $1 (\mathrm{mod}\ 4)$. A further result is provided by the following theorem.

THEOREM. *Let $n$ be a product of distinct primes, each congruent to* 1 *modulo* 4. *Then $R(n) = n(n-1)/4$.*

*Proof.* The observation that $-1$ is a quadratic residue of each prime dividing $n$ is crucial; it implies that $-1$ is a quadratic residue of $n$ itself. Therefore if $x$ is a summand of $\sum_{k=1}^{[(n-1)/2]} \mathrm{Rem}(k^2 \div n)$, then so is $n - x$. (Note that there are no zero terms in the summation.) Since there are $(n-1)/2$ terms in all, it follows that

$$R(n) = (x + n - x) \cdot \frac{n-1}{4} = \frac{n(n-1)}{4}.$$

To see why $-1$ is a quadratic residue modulo $n$, it suffices to show that if $-1$ is a quadratic residue of each of two coprime moduli $u$ and $v$, then $-1$ is also a quadratic residue of the product $uv$. Here is a proof of this statement. We note that by definition there exist integers $a$ and $b$ such that

$$a^2 \equiv 1\ (\mathrm{mod}\ u), \quad b^2 \equiv -1\ (\mathrm{mod}\ v).$$

We seek an integer $s$ such that

$$(a + su)^2 \equiv -1\ (\mathrm{mod}\ v).$$

If we locate such an integer, our task is done, for the congruence $(a+su)^2 \equiv -1 (\mathrm{mod}\ u)$ holds trivially, and if $(a+su)^2 + 1$ is divisible by both $u$ and $v$, then it is divisible by $uv$, because $u, v$ are coprime. It now suffices to give $s$ a value such that

$$a + su \equiv b \pmod{v},$$

that is, such that

$$su \equiv b - a \pmod{v}.$$

This is clearly possible because $u, v$ are coprime. (An example may help at this point. Let $u = 5$, $v = 17$. Both $u$ and $v$ are of the form 1(mod 4), so there exist integers $a$ and $b$ such that $a^2 \equiv -1 \pmod{5}$ and $b^2 \equiv -1 \pmod{17}$. We take $a = 2$, $b = 13$. We now seek an integer $s$ such that $2 + 5s \equiv 13 \pmod{17}$. This holds for $s = 9$, which gives $2 + 5s = 47$. It follows that $47^2 \equiv -1 \pmod{85}$, an assertion that is easily verified. (Proof: $47^2 + 1 = 2210 = 85 \times 26$.)) □

Consider next the case $n = p^2$, where $p$ is a prime congruent to 1(mod 4). Note that $-1$ is a quadratic residue modulo $n$. (Proof: Let $a^2 \equiv -1 \pmod{p}$; then $a^2 + 1 = kp$, say, for some integer $k$. Proceeding as we did above, we seek an integer $s$ such that $(sp + a)^2 \equiv -1 \pmod{p^2}$. This is the same as requiring that $2sap + (a^2 + 1) \equiv 0 \pmod{p^2}$ or $2sa + k \equiv 0 \pmod{p}$. The last congruence is clearly solvable in the unknown $s$.) Therefore in the summation

$$\sum_{k=1}^{[(p^2-1)/2]} \operatorname{Rem}(k^2 \div p^2),$$

for each index $k$, each non-zero term $x$ has a matching term $n - x$ occurring elsewhere in the sum. The zero terms occur precisely when $k$ is a multiple of $p$. The number of such values of $k$ is clearly equal to $(p-1)/2$; also, for each such $k$, there is a 'loss' of $p^2/2$. Therefore

$$R(p^2) = \left(\frac{p^2-1}{4}\right) \cdot p^2 - \frac{p^2}{2} \cdot \left(\frac{p-1}{2}\right) = \frac{p^3(p-1)}{4}.$$

More generally, we have the following result: *Let $n > 1$ be composed only of primes (possibly repeated) of the form* 1(mod 4). *Write $n$ as $a^2 b$, where $b$ is square-free; then* $R(n) = n(n-a)/4$. The argument used is practically identical to the one used above. We have

$$R(n) = \frac{n(n-1)}{4} - \frac{n}{2} \cdot c,$$

where $c$ is the number of integers $k$ in the interval $[1, (n-1)/2]$ for which $n | k^2$. Since $n = a^2 b$, the condition $n | k^2$ is equivalent to $ab | k$, so

$$c = \left[\frac{a^2 b - 1}{2} \div ab\right] = \left[\frac{a}{2} - \frac{1}{2ab}\right] = \frac{a-1}{2}.$$

It follows that

$$R(n) = \frac{n(n-1)}{4} - \frac{n(a-1)}{4} = \frac{n(n-a)}{4}.$$

This completes the proof. (Observe that when $n$ is square-free we have $a = 1$ and $R(n) = n(n-1)/4$, in agreement with the earlier result.)

## 5. A Remarkable Connection—or a Fantastic Coincidence?

We can write the result obtained in §4 in the following form: *If $n$ is a square-free product of primes all congruent to* 1(mod 4), *then* $R(n)/n = [n/4]$. The following question now poses itself very naturally: *For what other odd values of $n$ is* $R(n)/n = [n/4]$? A computer-assisted check shows that the only integers less than $10^5$ and of the form $-1 \pmod 4$ for which this relation holds are the following:

$$7, 11, 19, 43, 67, 163.$$

Note that the numbers are all prime. The law of formation of the sequence remains unclear, as does the question of whether the sequence continues at all.

To students of algebraic number theory, who will be familiar with the ideas of unique factorization, irreducibility, quadratic fields and so on, the astonishing fact that one observes at this point is that these numbers are part of a *very* famous list—the integers $d > 0$ for which the ring of integers of the field $\mathbb{Q}(\sqrt{-d})$ has the unique factorization property. This list is known to be the following:

$$d = 1, 2, 3, 7, 11, 19, 43, 67, 163.$$

The fact that these $d$'s have the stated property has been known since the time of Gauss and only recently has it been proved that the list cannot be extended. The methods used to obtain these results are, however, far beyond the scope of this article; see **[4]** for further details. It remains unclear whether there is any immediate connection between the property being studied in this section and that of unique factorization, or whether we have at hand a rather fantastic coincidence.

## 6. Further Explorations

Observe that $R(n) = n(n-1)/4 - n$ if $n$ is of the form $25b$, where $b$ is square-free and composed only of primes of the form 1(mod 4). In particular, $R(n) = n(n-1)/4 - n$ for each of the following values of $n$:

$$25, 125, 325, 425, 725, 925, \ldots .$$

However, it is curious that the relation $R(n) = n(n-1)/4 - n$ holds for many odd values of $n$ other than those on this list, for instance:

$$77, 133, 209, 301, 469, 473, 737, 817, 1141, \ldots .$$

It is instructive to examine the factorization patterns for these numbers:

$$\begin{array}{lll} 77 = 7 \times 11, & 133 = 7 \times 19, & 209 = 11 \times 19, \\ 301 = 7 \times 43, & 469 = 7 \times 67, & 473 = 11 \times 43, \\ 737 = 11 \times 67, & 817 = 19 \times 43, & 1141 = 7 \times 163. \end{array}$$

The primes that occur repeatedly in this list are precisely those that occurred in the earlier list! One would anticipate that the following numbers too should have the property that $R(n) = n(n-1)/4 - n$:

$$11 \times 163, \quad 19 \times 163, \quad 43 \times 67, \quad 43 \times 163, \quad 67 \times 163,$$

and it turns out that such indeed is the case. A simple `Mathematica` program for doing the calculations is given at the end of this article, in the Appendix. However, we find that the relation $R(n) = n(n-1)/4 - n$ does *not* hold for the numbers $7 \times 11 \times 19 \times 43$, nor for $7 \times 11 \times 19 \times 67$. The pattern is rather obscure at this stage and the author has as yet no further results in this direction, nor any independent criterion for identifying the primes that belong to the list. Clearly, there is much scope for further investigation here. Readers are requested to communicate any additional results that they discover to the author.

## 7. Application to Diophantine Equations

As another somewhat unexpected application of the fact that $-1 \notin QR(p)$ if $p$ is a prime of the form $-1 \pmod 4$, we show that the diophantine equation

$$y^2 = x^3 + 7$$

has no integral solutions. We assume the contrary and suppose that $(x, y) = (u, v)$ is a solution. Now $u$ must be odd, for if not this would lead to:

$$v^2 = u^3 + 7 \equiv 7 \equiv -1 \pmod 4,$$

which is not possible, as no square is of the form $-1 \pmod 4$. We now write the relation $v^2 = u^3 + 7$ in the form

$$v^2 + 1 = u^3 + 8 = (u+2)(u^2 - 2u + 4).$$

Since $u$ is odd, $u^2 - 2u + 4 \equiv 1 - 2 + 0 \equiv -1 \pmod 4$, which implies that $u^2 - 2u + 4$ possesses a prime factor of the form $-1 \pmod 4$. This in turn implies that $v^2 + 1$ possesses a prime factor of this form. However, we already know this to be impossible, so we reach a contradiction. We conclude that the equation $y^2 = x^3 + 7$ has no integral solutions.

There are many integral values of $k$ for which the Diophantine equation $y^2 = x^3 + k$ has no integral solution, and for which unsolvability is demonstrated by the fact that $-1$ is a quadratic non-residue of every prime of the form $-1 \pmod 4$. Apostol documents several such cases in **[1]**. He proves, for instance, that the equation $y^2 = x^3 + k$ has no integral solutions when $k$ has the form $(4a-1)^3 - 4b^2$, where $a, b$ are integers such that $b$ is not divisible by any prime of the form $-1 \pmod 4$. The proof is organized along much the same lines as the one presented above.

**Conclusion** It is really quite remarkable that so many differences should exist between the properties of primes of the two families $\pm 1 \pmod 4$, and that the differences should go so deep. It would be instructive to examine whether such a correspondingly rich catalog of differences can be built up for the primes of the two families $\pm 1 \pmod 6$.

**Acknowledgment.** I thank the referees for numerous valuable comments and suggestions.

## Appendix

The following sample `Mathematica` commands do the calculations referred to in the text.

```
In[1]:= ClearAll[R, a, b]
In[2]:= R[n_]:=Sum[Mod[i^2, n], {i, 1, Floor[(n-1)/2]}]
In[3]:= a[n_]:=R[n]/n==Floor[n/4]
In[4]:= b[n_]:=R[n]==n(n-1)/4-n
In[5]:= SetAttributes[{R, a, b}, Listable]
In[6]:= Select[Range[3, 100000, 4], a]
Out[6] = {7, 11, 19, 43, 67, 163}
In[7]:= b[{11*163, 19*163, 43*67, 43*163, 67*163}]
Out[7] = {True, True, True, True, True}
In[8]:= b[{7*11*19*43, 7*11*19*67}]
Out[8] = {False, False}
```

## REFERENCES

1. Tom Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, NY, 1976
2. G. H. Hardy and E. M. Wright, *Introduction to the Theory of Numbers*, 5th edition, Oxford Univ. Pr., Oxford, UK, 1979
3. George Polya and Gabor Szegő, *Problems and Theorems in Analysis*, Volume 2, Springer-Verlag, New York, NY, 1970
4. Ian Stewart and David Tall, *Algebraic Number Theory*, Chapman and Hall, London, UK, 1979
6. Shailesh Shirali, On Fermat's two-square theorem, to appear in *Resonance*, Indian Acad. of Sciences, Bangalore, India
5. Don Zagier, A one-sentence proof that every prime $p \equiv 1 (\text{mod } 4)$ is a sum of two squares, *Amer. Math. Monthly* 97 (1990), 144

# Mathematics Teachers

Dedicated to the memories of Laura Church and T. K. Pan

Chalk in hand,
she tosses down the book,
strides across the room,
excited by trigonometry,
excited that we,
restless in our rows,
caught some of it.
Flamboyant, silver,
imperial woman,
opening minds.

He leans
toward the blackboard,
drawing quick sketches
to show us why the one-form
is the heart of the matter,
the seed for intuition
to put forth new ideas.
Thoughtful, graceful,
passionate man,
understanding hearts.

Though their words are lost in years,
they opened windows through which I gaze,
grateful and dissatisfied.

—JoAnne Growney
Bloomsburg University
Bloomsburg, PA 17815

REFERENCES

1. Tom Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, NY, 1976
2. G. H. Hardy and E. M. Wright, *Introduction to the Theory of Numbers*, 5th edition, Oxford Univ. Pr., Oxford, UK, 1979
3. George Polya and Gabor Szegő, *Problems and Theorems in Analysis*, Volume 2, Springer-Verlag, New York, NY, 1970
4. Ian Stewart and David Tall, *Algebraic Number Theory*, Chapman and Hall, London, UK, 1979
6. Shailesh Shirali, On Fermat's two-square theorem, to appear in *Resonance*, Indian Acad. of Sciences, Bangalore, India
5. Don Zagier, A one-sentence proof that every prime $p \equiv 1 (\bmod 4)$ is a sum of two squares, *Amer. Math. Monthly* 97 (1990), 144

# Mathematics Teachers

Dedicated to the memories of Laura Church and T. K. Pan

Chalk in hand,
she tosses down the book,
strides across the room,
excited by trigonometry,
excited that we,
restless in our rows,
caught some of it.
Flamboyant, silver,
imperial woman,
opening minds.

He leans
toward the blackboard,
drawing quick sketches
to show us why the one-form
is the heart of the matter,
the seed for intuition
to put forth new ideas.
Thoughtful, graceful,
passionate man,
understanding hearts.

Though their words are lost in years,
they opened windows through which I gaze,
grateful and dissatisfied.

—JoAnne Growney
Bloomsburg University
Bloomsburg, PA 17815

# Halley's Comment—Projectiles With Linear Resistance

C.W. GROETSCH
University of Cincinnati
Cincinnati, OH 45221-0025

BARRY CIPRA
305 Oxford St.
Northfield, MN 55057

If only gravity were working, the path would be symmetrical

it is the wind resistance that produces the tragic curve

Norman Mailer
*The Naked and the Dead*

## 1. Introduction

Modern dynamics began with Galileo's investigations of the motion of projectiles in a nonresisting medium. Galileo ingeniously compounded the accelerated vertical motion of a projectile (obtained from his law of falling bodies) with its unaccelerated horizontal motion (an expression of his principle of inertia) to conclude that the path of the projectile is *parabolic*. Projectile motion is now a prime application in virtually every calculus text. Indeed, the power of analytic geometry and calculus, along with the elements of Newton's dynamics, has reduced Galileo's great achievement to a mere exercise. If a projectile of unit mass is projected from the origin at an angle $\theta$ with the horizontal with initial speed $v$, under the influence of a uniform gravitational acceleration $g$, then the equations of motion

$$\begin{aligned} \ddot{y}(t) &= -g, \quad \dot{y}(0) = v\sin\theta, \quad y(0) = 0 \\ \ddot{x}(t) &= 0, \quad \dot{x}(0) = v\cos\theta, \quad x(0) = 0 \end{aligned} \tag{1}$$

may be immediately integrated to yield:

$$y(t) = -\frac{g}{2}t^2 + (v\sin\theta)t; \quad x(t) = (v\cos\theta)t.$$

The parabolic nature of the trajectory is then revealed when the time parameter is eliminated:

$$y(x) = -\frac{g}{2v^2\cos^2\theta}x^2 + (\tan\theta)x.$$

The *range* of the projectile, $R(\theta)$, is the positive $x$-intercept of the trajectory:

$$R(\theta) = \frac{v^2}{g}\sin 2\theta, \tag{2}$$

and therefore the maximum range is attained when the angle of elevation is $\pi/4$ radians, a fact that was proved by Galileo **[5]** and observed much earlier by Tartaglia **[3]**.

Air resistance has so far been neglected. The precise nature of air resistance is a very complicated matter, but it is the common experience of those who have ridden motorcycles that resistance increases with velocity. The linear model, in which resistance is taken to be proportional to velocity, is the accepted first approximation to resistive behavior **[9]**, **[10]** and under some circumstances gives predictions that square quite well with observation **[1]**, **[4]**. The equations of motion in the linear resistance model are

$$\begin{aligned} \ddot{y}(t) &= -g - k\dot{y}(t), & \dot{y}(0) &= v\sin\theta, & y(0) &= 0 \\ \ddot{x}(t) &= -k\dot{x}(t), & \dot{x}(0) &= v\cos\theta, & x(0) &= 0 \end{aligned} \tag{3}$$

where $k$ is the *resistance constant*. These equations, while more complicated than those in the nonresistive case (1), are *linear* and therefore may be routinely solved to give

$$\begin{aligned} x(t) &= (v\cos\theta)(1 - e^{-kt})/k \\ y(t) &= \left(\frac{v\sin\theta}{k} + \frac{g}{k^2}\right)(1 - e^{kt}) - \frac{gt}{k}. \end{aligned}$$

The range $R(\theta)$ is again that positive value of $x$ giving $y = 0$. As before, this may be found by eliminating $t$, setting $y = 0$, and solving for the positive root $x$. A little rearranging shows that

$$R(\theta) = \frac{\cos\theta}{a}\left(1 - e^{-A(\theta)R(\theta)}\right) \tag{4}$$

where $A(\theta) = a\sec\theta + b\tan\theta$, $a = k/v$, $b = k^2/g$ **[7]**. Unlike in the nonresistive case, where the range function is defined explicitly by (2), now it is characterized as a fixed point of the transformation defined in (4). In **[7]** it is shown that for each $\theta \in (0, \pi/2)$ fixed-point iteration converges monotonically to $R(\theta)$. Modern numerical and graphical software can therefore be used to quickly and reliably calculate and display the range function for a projectile in a linearly resisting medium.

In this note we show how computational studies of Equation (4) and some associated relationships can be used to discover interesting features of the angle giving maximum range in a linearly resisting medium. Rigorous analytical proofs for the conjectures suggested by the computations are provided. We also prove a theorem on the range function that was suggested by a rather cryptic remark of Edmond Halley in 1686 and is supported by the graphical evidence in the next section. Our point is that strong graphical evidence can do more than just suggest results; it can also spur the search for analytical proofs. Galileo said as much in the words of his character Salviati **[6**, p. 60]:

> Pythagoras, a long time before he found the demonstration for the Hecatomb, had been certain that the square of the side subtending the right angle in a rectangular triangle was equal to the square of the other two sides; the certainty of the conclusion helped not a little in the search for a demonstration.

## 2. Graphical Observations

FIGURE 1 shows the range function for $k = 1$ and for various muzzle velocities ($v = 100$, 200, 500, bottom to top). The plot was produced by a simple MATLAB routine that computed $R(\theta)$ by fixed-point iteration for 100 equally-spaced angles in $[0, \pi/2]$. Similar plots, for fixed muzzle velocity and various resistance constants, can be found in [**7**].

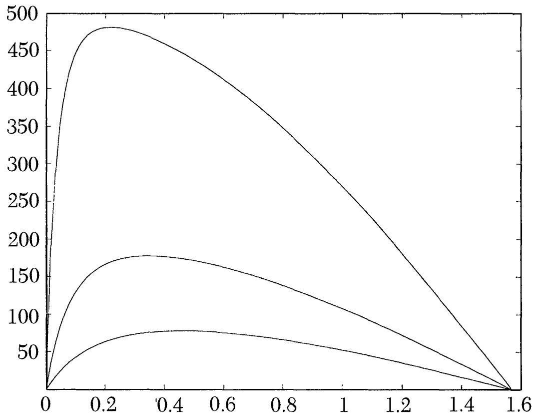



**FIGURE 1**
Range versus angle of elevation

In a nonresistive medium the maximum range occurs at $\theta = \pi/4$, by (2). The plots in FIGURE 1 (as well as those in [**7**] and the trajectory curves traced in [**4**] and [**9**]) all indicate that in a linearly resisting medium the maximum range *always* occurs at an angle of elevation $\theta < \pi/4$. The angle at which the maximum range occurs has no known explicit representation, but it can be characterized in terms of the unique fixed point of a certain function. To see this, we note that the maximum range occurs at an angle $\theta$ satisfying $R'(\theta) = 0$ (see [**7**] for a proof that $R$ is differentiable). Differentiating (4) and setting $R'(\theta) = 0$ we obtain

$$0 = -\frac{\sin\theta}{a}\left(1 - e^{-A(\theta)R(\theta)}\right) + \frac{\cos\theta}{a}\left(a\sec\theta\tan\theta + b\sec^2\theta\right)R(\theta)e^{-A(\theta)R(\theta)}.$$

However, by (4)

$$1 - e^{-A(\theta)R(\theta)} = a\sec\theta R(\theta).$$

Substituting this above and rearranging, we find that

$$\sin\theta = (\sin\theta + c)e^{-A(\theta)R(\theta)} \tag{5}$$

where we have set $c = b/a = vk/g$. By (4) we also have

$$e^{-A(\theta)R(\theta)} = 1 - a\sec\theta\, R(\theta)$$

and therefore by (5)

$$R(\theta) = \frac{(c/a)\cos\theta}{\sin\theta + c}$$

and hence

$$A(\theta)R(\theta) = (a\sec\theta + b\tan\theta)R(\theta) = \frac{c + c^2\sin\theta}{\sin\theta + c}. \tag{6}$$

Setting $s = \sin\theta$, and substituting (6) into (5) we find that $s$ satisfies the fixed point equation

$$s = (s+c)e^{-\frac{c+c^2 s}{s+c}}. \tag{7}$$

Equation (7) allows the possibility of computing $s$, and hence the optimal angle of elevation, $\theta = \sin^{-1}s$, directly by a numerical method, rather than tracing trajectories and selecting visually that which seems to have maximum range. More importantly, the equation suggests the possibility of an analytical study of the behavior of $s = \sin\theta$ with regard to $k$ and $v$, the main physical parameters of the problem.

A plot of the function values $s(c)$, defined by Equation (7), versus $c$ in which the values $s(c)$ were computed for 1,000 equally-spaced values of $c$ in $(0, 50)$ by a simple MATLAB routine (based on fixed-point iteration) is given below.

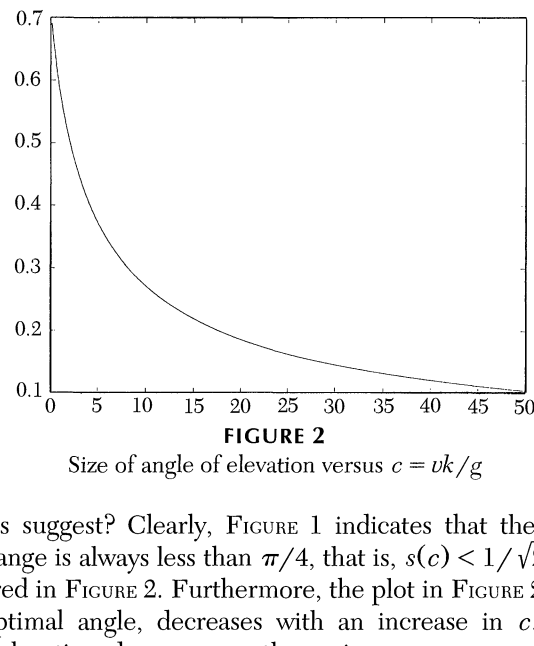



**FIGURE 2**

Size of angle of elevation versus $c = vk/g$

What do the plots suggest? Clearly, FIGURE 1 indicates that the angle of elevation giving maximum range is always less than $\pi/4$, that is, $s(c) < 1/\sqrt{2}$ for all $c > 0$. This evidence is bolstered in FIGURE 2. Furthermore, the plot in FIGURE 2 suggests that $s(c)$, and hence the optimal angle, decreases with an increase in $c$. In particular, the optimal angle of elevation decreases as the resistance constant increases, for fixed muzzle velocity; it also decreases as the muzzle velocity increases, for fixed resistance constant. All of these properties have interpretations as statements about the variables $s$ and $c$ satisfying the relationship (7). In the next section we use a parametric representation derived from (7) to give analytic proofs of these statements.

Recall that Tartaglia and Galileo pointed out that in a nonresisting medium the optimal angle of elevation is $\pi/4$. Furthermore, Galileo **[5]** proved that equal deviations above and below the optimal angle produce the same range, that is,

$$R(\pi/4 - \phi) = R(\pi/4 + \phi) \quad \text{for} \quad 0 \le \phi \le \pi/4,$$

a fact which is evident from (2). In 1686 Edmond Halley **[8]** noted that in certain experimental firings of small shot by a crossbow a consistent asymmetry, due evidently to air resistance, occurs in this relationship. To quote Halley: "*... in smaller brass-schott ... constantly and evidently, the under ranges outwent the upper.*" What Halley meant is not entirely clear, but we interpret his statement to say that

$$R(\pi/4 - \phi) > R(\pi/4 + \phi) \quad \text{for} \quad 0 < \phi < \pi/4. \tag{8}$$

This naturally raises the question whether (8) holds in a linearly resisting medium. A typical plot of $R(\pi/4 - \phi) - R(\pi/4 + \phi)$, computed using the same routine that produced FIGURE 1, is given in FIGURE 3. This plot bears out Halley's observation in a linearly resisting medium; in the final section of this note we prove this fact analytically.

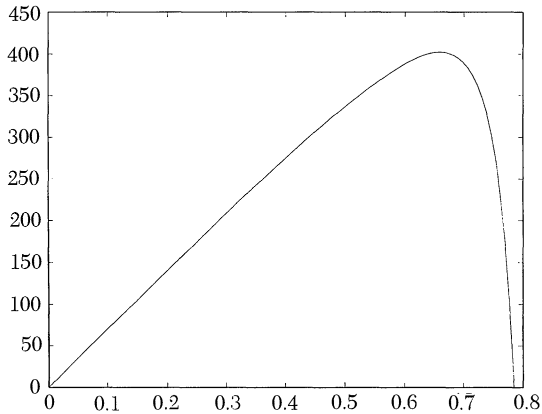



**FIGURE 3**
$R(\pi/4-\phi)-R(\pi/4+\phi)$ for $0 \le \phi \le \pi/4$

## 3. The Angle Giving Maximum Range

We will show that the angle of elevation giving maximum range is less than $\pi/4$ radians and that this angle decreases with an increase in the parameter $c = vk/g$. For this it suffices to show that $s = s(c)$ given by (7) satisfies $s(c) < 1/\sqrt{2}$, and that $s(c)$ decreases as $c$ increases. To avoid dealing directly with (7), we administer a dose of symmetry, rewriting (7) as

$$\frac{es}{s+c} = e^{\frac{es}{s+c}\frac{(1-c^2)}{e}}$$

or, equivalently,

$$x = e^{hx} \tag{9}$$

where

$$x = \frac{es}{s+c} \quad \text{and} \quad h = \frac{1-c^2}{e}. \tag{10}$$

As a further simplification we set $u = c/s$, then $x = e/(1+u)$, where $0 < u < \infty$. From (9) and (10) we have

$$h = \frac{\ln x}{x} = (1+u)(1-\ln(1+u))/e$$

and hence

$$c^2 = 1 - eh = -u + (1+u)\ln(1+u) \tag{11}$$

Differentiating this with respect to $c$ (the fact that $u$ is a differentiable function of $c$ is an easy consequence of the Implicit Function Theorem [2] applied to (11)), we obtain

$$2c = \ln(1+u)\frac{du}{dc}. \tag{12}$$

Since $s = c/u$, we then have by (12) and (11)

$$\begin{aligned}\frac{ds}{dc} &= \frac{1}{u^2}\left(u - c\frac{du}{dc}\right)\\ &= \frac{1}{u^2}\left(u - \frac{2c^2}{\ln(1+u)}\right)\\ &= \frac{1}{u^2}\left(-2 - u - \frac{2u}{\ln(1+u)}\right) < 0,\end{aligned}$$

since $u > 0$, and hence $s$ decreases with respect to $c$. Showing that the optimal angle is less than $\pi/4$ radians is equivalent to showing that $s^2 < 1/2$. Since $s$, and hence $s^2$, decreases with respect to $c$, it is enough to show that

$$\lim_{c \to 0^+} s^2 = \frac{1}{2}.$$

By (11) we have

$$s^2 = \frac{c^2}{u^2} = \frac{-u + (1+u)\ln(1+u)}{u^2}; \tag{13}$$

by l'Hôpital's rule,

$$\begin{aligned}\lim_{c \to 0^+} s^2 &= \lim_{u \to 0^+} s^2\\ &= \lim_{u \to 0^+} \frac{-1 + 1 + \ln(1+u)}{2u}\\ &= \lim_{u \to 0^+} \frac{1}{2(1+u)} = \frac{1}{2}.\end{aligned}$$

Therefore, $s < 1/\sqrt{2}$ for all $c > 0$ and hence the angle giving maximal range is always less than $\pi/4$. We also note that equations (11) and (13) give an alternative, parametric way of producing the graph in FIGURE 2.

## 4. The Halley Problem

In this section we prove analytically that the relation (8), suggested by FIGURE 3, must hold in a linearly resisting medium. We showed in the previous section that the maximum of the range function $R$ occurs at some angle in $(0, \pi/4)$; hence (8) holds for *some* angle $\phi \in (0, \pi/4)$. To show that (8) holds for *all* $\phi \in (0, \pi/4)$, it is therefore sufficient to show that the equation

$$R(\pi/4 - \phi) = R(\pi/4 + \phi)$$

has no solution $\phi \in (0, \pi/4)$. We accomplish this by reformulating (4) in terms of the function $Q(\theta) = aR(\theta)/\cos\theta$. In fact, routine manipulations convert (4) to

$$Q(\theta) = 1 - e^{-B(\theta)Q(\theta)}, \tag{14}$$

where

$$B(\theta) = 1 + c\sin\theta$$

and $c = b/a$. Note that $0 < Q(\theta) < 1$. Given $\phi \in (0, \pi/4)$, let $\theta = \pi/4 - \phi$ and $\theta' = \pi/4 + \phi$. Then $\sin\theta = \cos\theta'$, $\cos\theta = \sin\theta'$ and we wish to show that $R(\theta) = R(\theta')$ is impossible. Now, if $R(\theta) = R(\theta')$, then

$$\frac{c\sin\theta}{Q(\theta)} = \frac{c\sin\theta\cos\theta}{aR(\theta)} = \frac{c\cos\theta'\sin\theta'}{aR(\theta')} = \frac{c\sin\theta'}{Q(\theta')}. \tag{15}$$

However, $c\sin\theta = B(\theta) - 1$, and by (13)

$$B(\theta) = -\frac{\ln(1 - Q(\theta))}{Q(\theta)}. \tag{16}$$

Therefore,

$$c\sin\theta = -\frac{Q(\theta) + \ln(1 - Q(\theta))}{Q(\theta)},$$

and a similar formula holds for $\theta'$.

Substituting these results into (15), we obtain

$$\frac{Q(\theta) + \ln(1 - Q(\theta))}{Q(\theta)^2} = \frac{Q(\theta') + \ln(1 - Q(\theta'))}{Q(\theta')^2}.$$

However, for $0 < x < 1$, the function

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

is clearly strictly decreasing and hence one-to-one. Therefore, $Q(\theta) = Q(\theta')$ and hence, by (16), $B(\theta) = B(\theta')$ and, $\sin\theta = \sin\theta'$. Since $\theta, \theta' \in (0, \pi/2)$, $\theta = \theta'$, so

$$\pi/4 - \phi = \theta = \theta' = \pi/4 + \phi,$$

which is impossible for $\phi \in (0, \pi/4)$. We conclude that in a linearly resisting medium Halley's observation (8) holds for all $\phi \in (0, \pi/4)$.

**Acknowledgment.** CWG is grateful for support received from the National Science Foundation.

## REFERENCES


1. A. Battye, Modelling air resistance in the classroom, *Teaching Mathematics and Its Applications* 10 (1991), 32–34.
2. R. Courant, *Differential and Integral Calculus*, Vol. II, Interscience, New York, NY, 1961.
3. S. Drake and I. Drabkin, *Mechanics in Sixteenth-Century Italy*, University of Wisconsin Press, Madison, WI, 1969.
4. H. Erlichson, Maximum projectile range with drag and lift, with particular application to golf, *Amer. Jour. of Physics* 51 (1983), 357–361.
5. G. Galilei, *Two New Sciences* (Elzevirs, Leyden, 1638), translated with a new introduction and notes, by Stillman Drake, Second Edition, Wall and Thompson, Toronto, Canada, 1989.
6. G. Galilei, *Dialogue on the Great World Systems* (Florence, 1630), translated by Giorgio de Santillana, University of Chicago Press, Chicago, IL, 1953.

7. C.W. Groetsch, Tartaglia's inverse problem in a resistive medium, *Amer. Math. Monthly* 103 (1996), 546–551.
8. E. Halley, A discourse concerning gravity, and its properties wherein the descent of heavy bodies, and the motion of projects is briefly, but fully handled: together with the solution of a problem of great use in gunnery, *Philosophical Transactions of the Royal Society of London* 16 (1686), 3–21.
9. K. Symon, *Mechanics*, Addison-Wesley, Reading, MA, 1953.
10. S. Timoshenko and D. Young, *Advanced Dynamics*, McGraw-Hill, New York, NY, 1948.

# Carl B. Allendoerfer Awards – 1997

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in *Mathematics Magazine.* Carl B. Allendoerfer, a distinguished mathematician at the University of Washington, served as President of the Mathematical Association of America, 1959–60. This year's awards were presented at the August 1997 Mathfest, in Atlanta. The citations follow.

**Colm Mulcahy, "Plotting and Scheming with Wavelets," Mathematics Magazine 69 (December 1996).** "Wavelets are acquiring a visibility and popularity that may soon be on the scale first enjoyed by fractals a few years back." So begins Colm Mulcahy's "Plotting and Scheming with Wavelets." In other words, wavelets are a hot topic, recently developed, widely applied, and increasingly discussed and written about by the mathematics research community. Mulcahy's article is an admirable piece of work, presenting this topic now at a level fully accessible to undergraduate students and to the uninitiated mathematicians of all backgrounds. Starting from first principles and using clear and increasingly sophisticated graphics, the author leads the reader from simple data manipulation, into problems and questions on 1- and 2-dimensional graphics, and finally to the efficient and powerful data storage and retrieval effected by the techniques of wavelets. Mulcahy writes with clarity, charm, and humor so that the reader is irresistibly drawn into fundamental problems and solutions for both humans and computers; in addition he has researched and provided a thorough bibliography for further wavelet reading. This paper is a significant contribution to the expository literature of mathematics of the twenty-first century.

**Lin Tan, "The Group of Rational Points on the Unit Circle," Mathematics Magazine 69 (June 1996).** The mathematics community rejoices in the solution of Fermat's Last Theorem and in the powerful mathematics there developed and applied. What happens if instead of focusing on elliptic curves and cubic equations as was necessary for the latter work, one asks similar questions about conics and the unit circle in particular? In the innovative paper "The Group of Rational Points on the Unit Circle," Lin Tan simply and naturally addresses some related questions about rational points on more manageable and familiar curves. Taking a group-theoretic point of view, the author clearly and enticingly develops the mathematics needed to determine the rational-point group structure for the circle and the related hyperbola, $x^2 - y^2 = 1$. The reader is skillfully led from elementary theory and trigonometry into profound and important mathematics, pointing toward and stimulating interest in the relationship between groups, quadratic forms, and cubic curves, among other applications. The scholarly bibliography stretches from Euler to Wiles, supporting the historical and motivational tone of the paper that succeeds admirably in moving the reader from a nicely understandable and simple problem to the frontiers of modern mathematics at its best.

7. C.W. Groetsch, Tartaglia's inverse problem in a resistive medium, *Amer. Math. Monthly* 103 (1996), 546–551.
8. E. Halley, A discourse concerning gravity, and its properties wherein the descent of heavy bodies, and the motion of projects is briefly, but fully handled: together with the solution of a problem of great use in gunnery, *Philosophical Transactions of the Royal Society of London* 16 (1686), 3–21.
9. K. Symon, *Mechanics*, Addison-Wesley, Reading, MA, 1953.
10. S. Timoshenko and D. Young, *Advanced Dynamics*, McGraw-Hill, New York, NY, 1948.

## Carl B. Allendoerfer Awards – 1997

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in *Mathematics Magazine*. Carl B. Allendoerfer, a distinguished mathematician at the University of Washington, served as President of the Mathematical Association of America, 1959–60. This year's awards were presented at the August 1997 Mathfest, in Atlanta. The citations follow.

**Colm Mulcahy, "Plotting and Scheming with Wavelets," Mathematics Magazine 69 (December 1996).** "Wavelets are acquiring a visibility and popularity that may soon be on the scale first enjoyed by fractals a few years back." So begins Colm Mulcahy's "Plotting and Scheming with Wavelets." In other words, wavelets are a hot topic, recently developed, widely applied, and increasingly discussed and written about by the mathematics research community. Mulcahy's article is an admirable piece of work, presenting this topic now at a level fully accessible to undergraduate students and to the uninitiated mathematicians of all backgrounds. Starting from first principles and using clear and increasingly sophisticated graphics, the author leads the reader from simple data manipulation, into problems and questions on 1- and 2-dimensional graphics, and finally to the efficient and powerful data storage and retrieval effected by the techniques of wavelets. Mulcahy writes with clarity, charm, and humor so that the reader is irresistibly drawn into fundamental problems and solutions for both humans and computers; in addition he has researched and provided a thorough bibliography for further wavelet reading. This paper is a significant contribution to the expository literature of mathematics of the twenty-first century.

**Lin Tan, "The Group of Rational Points on the Unit Circle," Mathematics Magazine 69 (June 1996).** The mathematics community rejoices in the solution of Fermat's Last Theorem and in the powerful mathematics there developed and applied. What happens if instead of focusing on elliptic curves and cubic equations as was necessary for the latter work, one asks similar questions about conics and the unit circle in particular? In the innovative paper "The Group of Rational Points on the Unit Circle," Lin Tan simply and naturally addresses some related questions about rational points on more manageable and familiar curves. Taking a group-theoretic point of view, the author clearly and enticingly develops the mathematics needed to determine the rational-point group structure for the circle and the related hyperbola, $x^2 - y^2 = 1$. The reader is skillfully led from elementary theory and trigonometry into profound and important mathematics, pointing toward and stimulating interest in the relationship between groups, quadratic forms, and cubic curves, among other applications. The scholarly bibliography stretches from Euler to Wiles, supporting the historical and motivational tone of the paper that succeeds admirably in moving the reader from a nicely understandable and simple problem to the frontiers of modern mathematics at its best.

# NOTES

## Scheduling a Bridge Club Using a Genetic Algorithm

R. J. SIMPSON
Curtin University of Technology
Perth, Western Australia

**Introduction** In 1992 Elenbogen and Maxim **[1]** published an article in this journal about a problem in scheduling a bridge tournament. The problem was interesting in itself, and difficult in the sense that there did not seem to be a method for solving it that would avoid some type of search of the space of potential solutions. The article **[1]** aimed not just to solve the problem but also to illustrate how different techniques of combinatorial optimization might be applied. The techniques discussed were exhaustive search, greedy algorithm, branch and bound, steepest descent, and simulated annealing. The last four methods were implemented, but not the first, the excuse being that it would take about $1.34 \times 10^{13}$ years. We will describe how a sixth technique, genetic algorithms, can be applied to this problem. (A seventh method, tabu search, is described in Luis Morales' paper in this issue of the MAGAZINE.) We will review Elenbogen and Maxim's scheduling problem, describe what a genetic algorithm is and how it works, and show how the problem can be incorporated into a genetic algorithm. Finally, we report our results.

**The problem** Elenbogen and Maxim describe the bridge club problem as follows:

> *A bridge club consists of 12 couples. At each club meeting, the club is divided into 3 groups of 4 couples. Each of the four couples then competes against the remaining 3 couples in its group, which requires 6 games per group, 18 for the meeting. These meetings happen 8 times a year, resulting in a total of 144 games. The club has requested a schedule for the year with the following property: the number of times any couple competes against any opposing couple is the same for all couples. If no such schedule exists, then the club requests a schedule where the number of times any couple plays any other couple is most nearly the same for all couples.*
>
> *Since each couple competes 8 times against 3 opposing couples, there are 24 opposing couples (teams). Since a couple does not play itself, there are only 11 possible opponents. Hence an optimal schedule would have each couple playing 9 other couples twice and 2 other couples three times. However, there is no guarantee that such an "optimal" schedule exists.*

We will describe a schedule for the competition as a vector $(x_1, \ldots, x_{96})$, each element of which is a number from 1 to 12. The first 4 elements show the first group of teams for the first meeting, the second 4 elements are the second group, and the third is the third group for this meeting. Thus $(x_1, \dots, x_{12})$ is a permutation of the numbers $1, 2, \dots, 12$. Then $(x_{13}, \dots, x_{24})$ is another permutation of $1, 2, \dots, 12$ and shows the arrangement for the second meeting with teams $(x_{13}, \dots, x_{16})$ being the first group, $(x_{17}, \dots, x_{20})$ the second and so on.

Altogether there are 24 groups of 4 teams. Two teams, say team $i$ and team $j$, will play together whenever $i$ and $j$ belong to the same group of 4.

We need to set this problem up so that the optimal solution corresponds to minimizing some function, the domain of which is the set of possible schedules. Let $S$ be a schedule and let $n_{ij}$ be the number of times pair $i$ and $j$ fall in the same group in this schedule, which is the same as the number of times $i$ and $j$ play against each other. Now define

$$C(S) = \sum_{i<j} (n_{ij} - 2)^2.$$

This is the *cost function* we wish to minimize. An "optimal" solution, if one exists, will involve 12 pairs $i, j$ for which $n_{ij} = 3$. For all other pairs $n_{ij} = 2$. If such a schedule exists it will have a cost of 12. It is not hard to see that any non-optimal schedule will have a greater cost, and only slightly harder to see that for any schedule the value of $C$ will be an even integer. Using their four optimization techniques Elenbogen and Maxim found schedules with costs of 22, 20, 14, and 14. These were the best schedules their algorithms found, but this does not imply that no schedule with cost 12 exists. In fact such a schedule was later found by Kreher, Royle, and Wallis **[6]**, **[7]** using a variation of one of Elenbogen and Maxim's techniques and a more powerful computer.

The difficulty in solving the problem is that there are so many possible schedules. A raw count gives $1.24 \times 10^{30}$, though many of these are essentially equivalent. Taking advantage of this equivalence reduces the search space (we will do this in a later section) but still leaves too many cases to be separately considered.

**Genetic algorithms** Genetic algorithms are optimization techniques first popularized by Holland **[4]** in 1975; they have since generated a great deal of research (see, for instance, **[3]**). An accessible introduction is Holland's article in *Scientific American* **[5]**. Each year an international conference on the area is held **[2]**. The idea is that natural evolution seems to be an effective way of finding good, if not optimal, solutions to the problems of survival, and that by mimicking this natural process we may be able to find good or optimal solutions to mathematical problems.

Much of the terminology is borrowed from biology. Suppose we wish to maximize a function $f(x_1, x_2, \dots, x_k)$. In typical applications the number of variables $k$ is large but the set of values $x_i$ can take for each $i$ is small; often $x_i$ takes only the values 0 and 1. A vector of values for $(x_1, x_2, \dots, x_k)$ is called a *chromosome*. The value of $f(x_1, x_2, \dots, x_k)$ is called the *fitness* of this chromosome, so we want a chromosome of maximum fitness. The algorithm begins with a *population* of $n$ randomly chosen chromosomes, and proceeds through a sequence of *generations*. In each generation a new population is created using three *genetic operators*, which produces a population with a higher average fitness and higher maximum fitness than the previous one.

The first operator is *selection*. A new population, also of size $n$, is created by selecting members from the previous one, with substitution, in such a way that the probability that a chromosome is selected increases with the fitness of the chromosome. Thus there are likely to be several copies of very fit individuals in the new population, and none of very unfit ones. In biological terms this is survival of the fittest. The actual probability function used depends on the application; often, something like the following is used:

$$P(\text{chromosome } x \text{ is chosen}) = \frac{f(x) + c}{n(\bar{f} + c)}.$$

Here $\bar{f}$ is the average fitness of the previous population, and $c$ is a constant. A large value of $c$ will mean that the selection process slightly favors the fit individuals, whereas a small or negative value will mean fit individuals are strongly favored. Sometimes genetic algorithms converge quickly to a local optimum, so that the population consists of identical chromosomes which are good but not the best possible. This is like finding a local maximum rather than a global maximum for a function of one variable. To avoid this we try to maintain considerable genetic diversity, that is, we try to maintain a population in which most chromosomes are different. This can be done by choosing the parameter $c$ sufficiently large so that selecting a good chromosome is only slightly more probable than selecting an average one.

We now have a new population, which can be manipulated using the other two genetic operators: *crossover* and *mutation*. In crossover the population is randomly partitioned into pairs of chromosomes. The population size $n$ is always even so that we have a whole number of pairs of chromosomes. For each pair we either leave it unchanged with probability $1 - p_c$ or, with probability $p_c$, apply the crossover operator. Suppose our chromosome pair (the "parents") is $(x_1, x_2, \dots, x_k)$ and $(y_1, y_2, \dots, y_k)$. We randomly choose an integer $i$ from the set $1, 2, \dots, k-1$ and "crossover" the chromosomes immediately after their $i$th elements. This means we produce two new chromosomes (the "offspring")

$$(x_1, x_2, \dots, x_i, y_{i+1}, \dots, y_k) \quad \text{and} \quad (y_1, y_2, \dots, y_i, x_{i+1}, \dots, x_k).$$

The situation resembles sexual reproduction. Each of the "parents" passes on some of its genetic makeup to the "offspring." The idea is that a fairly fit individual may owe its fitness to a particular combination of $x_i$ values in one part of its chromosome, while another individual may owe its fitness to a happy combination of values in another part. Crossover may combine these "good" sub-chromosomes, and thus confer an advantage. The crossover probability $p_c$ is chosen by the user. Experienced users of genetic algorithms often recommend a value of 0.15. For this value, and for the values of other parameters that appear in the algorithm, there is not much mathematics to rely on and practitioners normally choose the values using trial and error or experience.

The final genetic operator is *mutation*: the algorithm goes through each of the $kn$ elements of the $n$ chromosomes in the population and changes it, with (fixed) probability $p_m$, to some other legal value. If the only allowed values are 0 and 1, the new value is determined; otherwise a probability distribution is imposed on the set of legal values, and the new value chosen from that. The point of the mutation operator is that it changes a chromosome to another which is "nearby." If the function $f$ is not too badly behaved, then its value at the new chromosome will not be too different from that at the previous one. If we get an improvement then things are going well; if the new chromosome is worse then we're likely to throw it out at the next round of selection anyway.

Having applied these three operators we've completed a generation, and are ready for the next round of selection, crossover, and mutation. The algorithm proceeds for as many generations as the user chooses. It's often convenient to allow the process to run until it seems to have stopped making progress, perhaps when it has gone through 100 generations without significant improvement of the average or maximum fitness.

*Variations.* We have described the "vanilla" version of genetic algorithms. Sometimes a problem cannot be described nicely as a function of independent variables—the bridge club problem is like this—and we need to adapt the basic version. There are also many variations on the way the operators are applied. Some practitioners will change the values of $c$, $p_c$ and $p_m$ as the algorithm proceeds. The crossover operator described above is "one-point crossover." We can instead have multiple-point crossover or some other variation. Sometimes the mutation operator is applied in such a way that mutation only occurs when it is increasing the fitness of the chromosome or at least not greatly reducing it. Another operator sometimes used is *elitism*, whereby one copy of the best individual passes unchanged to the next generation. This protects against the danger that a very good individual may be lost forever because it was not selected for the next generation, or selected but then corrupted by crossover or mutation.

**A genetic algorithm for bridge club scheduling** Recall that for each of 8 meetings we must partition the numbers 1 to 12 into 3 sets of size 4, and that such an arrangement is described by a vector $(x_1, \dots, x_{96})$. These vectors will be the chromosomes for our algorithm. The situation is now a bit more complicated than that in the standard genetic algorithm, since the values of the elements are not independent : the values of $x_1, \dots, x_{12}$ must be distinct, as must the values of $x_{13}, \dots, x_{24}$, and so on. This means that our cross over and mutation operators must be a little more sophisticated. We need also to *minimize* rather than maximize our function, though this requires only minor tinkering.

Before describing the operators let us reduce the size of our search space by invoking some of the equivalences among different schedules. First, note that our labeling of the teams is arbitrary. No matter how the human players are assigned at the first meeting, we can label the teams in the first group $1, 2, 3, 4$, the second group $5, 6, 7, 8$, and the third $9, 10, 11, 12$. Thus we need only consider chromosomes beginning $1, 2, \dots, 12$. Next, notice that the labeling of the groups in later meetings is arbitrary: there are 3 groups playing at meeting 2, and we can call any one of them group 1. We will say that the group that includes team 1 is always group 1. This means that we need only consider schedules in which $x_{13} = 1, x_{25} = 1, \dots, x_{85} = 1$. This reduces the size of the search space from $(12!)^8 \approx 2.8 \times 10^{69}$ to $(11!)^7 \approx 1.6 \times 10^{53}$. There are other equivalences we can use; for instance, the order in which teams appear within a group is immaterial. Our genetic operators will not make changes to such orders and so not waste time examining schedules which are equivalent in this way.

*Choosing the initial population.* The first step in the algorithm is to create $n$ schedules that satisfy the above constraints, but are otherwise random. The constraints mean that for each schedule we need 7 permutations of the 11 numbers $2, 3, \dots, 12$, one permutation to be the elements $x_{14}, \dots, x_{24}$, one to be $x_{26}, \dots, x_{36}$, and so on.

We can create such a permutation by using a random number generator that outputs a random integer from the set $\{2, \dots, 12\}$. In the Basic programming language this can be done using `INT(11*RND+2)`. The first random integer we generate will be the first element of our permutation. If the second random integer is different from the first, we'll use it as the second element of the permutation. If it's the same as the first, we keep using the generator until we get one that is different. We continue like this till we've generated the full set of 11 numbers.

There are no fixed rules for deciding the best population size. A large population gives more variety but slows down the algorithm. A smaller one gives less variety in each generation, but because less processing time is used it's possible to have more generations. After some experimentation a population size of 150 was chosen.

*Selection.* We let the probability of selecting schedule $S_i$ be given by

$$P(S_i) = \frac{b + C_{min} - C(S_i)}{n\left(b + C_{min} - \overline{C}\right)}$$

where $\overline{C}$ is the average cost, $C_{min}$ is the minimum over all schedules in the previous generation, and $b$ is a parameter. (Here we are using $b + C_{min} - C(S_i)$ as the fitness of the chromosome.) After some experimentation it was found that a value of $b$ close to 10 gave satisfactory results. Note that this formula ensures that the lower the cost of a schedule, the higher its probability of being selected. The process can be thought of as rolling a die with $n$ sides labeled $S_1, \ldots, S_n$; the probability of side $S_i$ turning up is $P(S_i)$. The die is rolled $n$ times and after each roll the side that comes up names the schedule that is put into the next generation.

*Crossover.* If an arbitrary crossover point is chosen, we would probably produce a schedule in which some teams play in two groups at the same meeting and some do not play at all. To avoid this we will allow crossover to occur only at points separating parts of the chromosome that correspond to different meetings of the schedule, that is, between element $12i$ and element $12i + 1$, for $i = 2, \ldots, 7$. (Crossover between elements 12 and 13 is not used because the first 12 elements of every chromosome are $1, 2, \ldots, 12$, so the offspring from such a crossover would be identical to their parents.) Thus there are only 6 possible crossover points, far fewer than usual for genetic algorithms. (Standard genetic algorithms allow crossover between any two elements, and there may be several hundred elements.) The crossover probability was set, after some experimentation at 0.1—also lower than usual.

*Mutation.* Changing only one element of a chromosome would produce a schedule with some team not playing at all at some meeting. To avoid this, mutation swaps the position of two elements in some meeting of the schedule. Similarly, there is no point in changing a team's position within a group, so the two teams to be swapped are chosen from different groups in the same meeting. It was found that high mutation probabilities were required; after some experimentation, $p_m = 0.7$ was found suitable. A lower mutation rate does not produce enough variety in the chromosomes, while a higher rate is too likely to destroy the fitter chromosomes of the previous generation.

*Results.* The algorithm was run 10 times with parameter values $p_c = 0.1$, $p_m = 0.7$, and $b = 10$, using a population of size 150, and 250 generations. The cost of the best schedule in each case varied from 12 to 20; optimal schedules (cost = 12) were found twice. One of these is displayed in Table 1. It is encouraging that the genetic algorithm found an optimum schedule while Elenbogen and Maxim's techniques did not. However, 10 runs of our genetic algorithm meant evaluating the cost function considerably more often than their algorithms needed, and the genetic algorithm required a lot of fine tuning to find appropriate parameter values. Table 2 shows the number of cost evaluations and best cost found by these different methods. (For ours the number of evaluations is for 10 runs, with population 150 and number of generations 250.) The genetic algorithm is also more complicated than its competitors, which adds to execution time and to programming time. The successful program of Kreher et al. **[6]**, for example, used more evaluations but a simpler algorithm.

TABLE 1. A minimal-cost bridge club schedule.

| *Table 1* | | | | *Table 2* | | | | *Table 3* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 7 | 5 | 11 | 9 | 12 | 6 | 4 | 2 | 8 | 3 | 10 |
| 1 | 6 | 3 | 12 | 4 | 5 | 11 | 8 | 9 | 10 | 2 | 7 |
| 1 | 5 | 10 | 6 | 8 | 12 | 9 | 2 | 3 | 11 | 7 | 4 |
| 1 | 12 | 8 | 7 | 10 | 2 | 11 | 6 | 5 | 9 | 3 | 4 |
| 1 | 11 | 8 | 9 | 3 | 5 | 12 | 10 | 7 | 4 | 2 | 6 |
| 1 | 3 | 5 | 9 | 11 | 3 | 8 | 6 | 10 | 7 | 12 | 4 |
| 1 | 4 | 10 | 8 | 5 | 11 | 2 | 12 | 3 | 9 | 6 | 7 |

TABLE 2. Comparison of algorithms.

| *Algorithm* | *Number of Evaluations* | *Lowest Cost* |
|---|---|---|
| Greedy Algorithm **[1]** | 40,397 | 22 |
| Partial Branch and Bound **[1]** | 100,000 | 20 |
| Steepest Descent **[1]** | 75,000 | 14 |
| Simulated Annealing **[1]** | 50,000 | 14 |
| Kreher, et al. **[6]** | 1,000,000 | 12 |
| Genetic Algorithm | 375,000 | 12 |

Our experiences seem typical of genetic algorithms. Successes are not dramatic and any advantage over competing techniques is not obvious. There are many unanswered questions about the optimal method of implementation. What values should $p_c$ and $p_m$ take? Should their values change as the algorithm proceeds and, if so, how? Is it better to have a large population and few generations, or a small population and many generations? Should crossover be one-point, or of some other variety? Is elitism useful? What selection function should be used? Until such questions are answered it will probably not be possible to decide whether genetic algorithms have real advantages over their competitors.

## REFERENCES

1. B. S. Elenbogen and B. R. Maxim, Scheduling a bridge club (a case study in discrete optimization), this MAGAZINE 65 (1992) 18–26.
2. S. Forrest (Program Chair), *Proceedings of the Fifth International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA (1993).
3. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, (1989).
4. J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, (1975).
5. J. Holland, Genetic algorithms, *Scientific American*, July 1992, 44–50.
6. D. L. Kreher, G. F. Royle, and W. D. Wallis, A family of resolvable regular graph designs, *Discrete Mathematics*, 156 (1996), 269–275.
7. G. F. Royle and W. D. Wallis, Constructing bridge club designs, *Bull. International Combinatorics Assoc.*, 11 (1994), 122–125.

# Scheduling a Bridge Club by Tabu Search

LUIS B. MORALES
IIMAS, Universidad Nacional Autónoma de México
Apdo. Postal 70-221, México, DF, 04510,
Mexico

**Introduction** A bridge club scheduling problem was presented in [**1**], as follows:

> A bridge club consists of 12 couples. At each meeting, the club is divided into 3 groups of 4 couples. Each of the 4 couples then plays against the remaining 3 couples in its group, which requires 6 games per group. Thus 18 games are played at each meeting, of which 8 occur per year, resulting in a total of 144 games. The club needs to prepare a schedule for the year such that the number of times any couple plays any other couple should be the same for all couples. If no such schedule exists, then the club needs to prepare a schedule for the year such that the number of times any couple plays any other couple is most nearly the same for all couples.

The authors of [**1**] formulated this scheduling problem as a discrete optimization problem, and used four techniques of combinatorial optimization: greedy algorithm, branch and bound, steepest descent, and simulated annealing (see, e.g., [**3**]) to seek an optimal schedule. None of their methods produced an optimal schedule.

**An optimal solution** In the course of the year's eight meetings each couple competes against 24 couples (teams). Since a couple has only 11 different possible opponents, an "optimal" schedule would have each couple playing 9 other couples twice and 2 other couples three times. However, there is no *a priori* guarantee that such an "optimal" schedule exists.

Define the competition matrix of a schedule as the matrix $(C_{ij})$, where $C_{ij}$ is the number of times team $i$ competes against team $j$ in that schedule. This matrix is symmetric $(C_{ij} = C_{ji})$, with 0's on the diagonal, so only the 66 elements above (or below) the main diagonal need to be examined. If an optimal schedule exists, then each row of $C$ would consist of nine 2's and two 3's (with 0's on the diagonal).

A *cost function* $f(x)$ of a particular schedule $x$ can be computed by summing the square of the difference between the ideal number of times two teams should meet, namely 2, and each of the entries above the main diagonal of the competition matrix. That is,

$$f(x) = \sum_{i<j}^{12} (C_{ij} - 2)^2.$$

Clearly, $f(x)$ decreases when the competition matrix $C_{ij}$ has more entries that are closer in value to 2; the minimum value for $f(x)$ is thus 12. Using a simple tabu search method (TS) we will find an optimal schedule of cost 12 (see [**2**] for more details on TS).

**Tabu search** TS is a general search procedure for solving combinatorial optimization problems of the following general type:

$$\text{Minimize } f(x), \text{ subject to } x \in X,$$

where $f$ is a cost function and $X$ is a set of feasible solutions. The procedure has its roots in the intelligent problem solving approach, found in artificial intelligence. (By contrast, simulated annealing is based in the second law of thermodynamics, from physics.)

The TS method is an iterative process; it starts from an initial feasible solution and tries to reach a global minimum by moving step by step. To accomplish this, we must define a set $M$ of simple modifications that can be applied to a given solution in order to move to another. These modifications are called *moves*. The notation $x' = m(x)$, $m \in M$, indicates that $m$ transforms $x$ into $x'$. This leads us to the definition of a *neighborhood*, an ingredient common to most heuristic and algorithmic procedures for optimization. For each feasible solution $x$, the neighborhood $N(x)$ is the set of all feasible solutions directly reachable from $x$ by a single move $m$ in $M$. At each step of the iterative process, we generate a subset $V^*$ of $N(x)$, and we move from $x$ to the best solution $x^*$ in $V^*$, whether or not $f(x^*)$ is better than $f(x)$. (If $N(x)$ is not large, it is possible to take $V^* = N(x)$.)

Up to this point, the algorithm is close to a local improvement technique, except that we may move from $x$ to a worse solution $x^*$, and thus may escape from local minima of $f$. (This situation occurs in simulated annealing, where a non-improving move may be accepted, with an exponential probability distribution that decreases as the number of iterations grows.) To prevent cycling, any move that returns to any local optimum recently visited is forbidden. This is accomplished in a *short-term memory* framework by storing the forbidden (tabu) move in a set $T$, called a *tabu list*, of length $|T| = t$ (the length may be fixed or variable). A move remains forbidden during $t$ iterations, so the tabu list can be represented by a queue: At each iteration, the opposite move from $x$ to $x^*$ is added at the end of $T$, while the oldest is removed from $T$.

Unfortunately, the tabu list may forbid certain interesting moves, such as moves that lead to a better solution than the best one found so far. An *aspiration criterion* is introduced to cancel the tabu status of a move when this move is judged useful.

Stopping rules must also be defined; in many cases (including the bridge club problem) a lower bound $f^*$ of $f$ is known in advance. As soon as we are close enough or we have reached this bound, we may interrupt the algorithm. In general, however, $f^*$ is not available with sufficient accuracy, so a fixed maximum number of iterations is given.

**Adaptation of tabu search** In our approach a feasible solution (or schedule) is an arrangement of eight columns into which the 12 couples are assigned to the three groups $A$, $B$, and $C$; each group contains four couples. By the preceding definition, $x'$ is a neighboring solution of $x$ if $x$ and $x'$ are identical for every meeting but one, and in the exceptional meeting, exactly two teams switch groups when moving from $x$ to $x'$. A move is entirely defined by one meeting and two teams in that meeting, which belong to different groups. Each feasible solution $x$ has $8 \times 3 \times 4 \times 4 = 384$ neighbors, because there are 8 different meetings in which a neighbor may differ, and in each meeting, there are 3 ways to choose the groups involved in the switch and 4 ways to choose a team from each of the groups.

An initial feasible solution is generated as follows: For each column of the schedule, we randomly split the 12 couples into the groups $A$, $B$, and $C$, with four couples each.

To construct the tabu list, we forbid the exchange of any two teams that were exchanged between two groups at a meeting during the preceding $t$ iterations, where $t$ is the tabu size. Formally, the tabu list consists of triples $(i, j, k)$, where $i$ and $j$ are teams that cannot be exchanged in meeting $k$. The value of $t$ is crucial for the performance of TS, as we will later see.

For the problem at hand we define the aspiration criterion as a function $A(z)$ defined for every value $z$ of the cost function. The criterion allows a tabu move to be selected if it leads to a solution better than the best found so far. This means that for any solution $x$, $A(f(x)) = f(x^o)$ where $x^o$ is the best solution found so far; the tabu status of a move from $x$ to $x'$ may be dropped if $f(x') < A(f(x))$.

Heuristic techniques, including TS, do not always guarantee reaching an optimal solution. So the process is stopped if the given maximum number of iterations has been reached. On the other hand, we must stop the process if the cost function $f$ has reached the optimal value, in this case 12.

**Results** In order to assess good values for the tabu list size $t$, we performed TS 100 times with a random initial schedule and a maximum of 500 iterations, for each $t$ between 7 and 42. In Figure 1 we have plotted (a) the number of times an optimal solution was found out of 100 (success percentage); and (b) the mean number of iterations needed to find it for each value of $t$ in [7, 42].

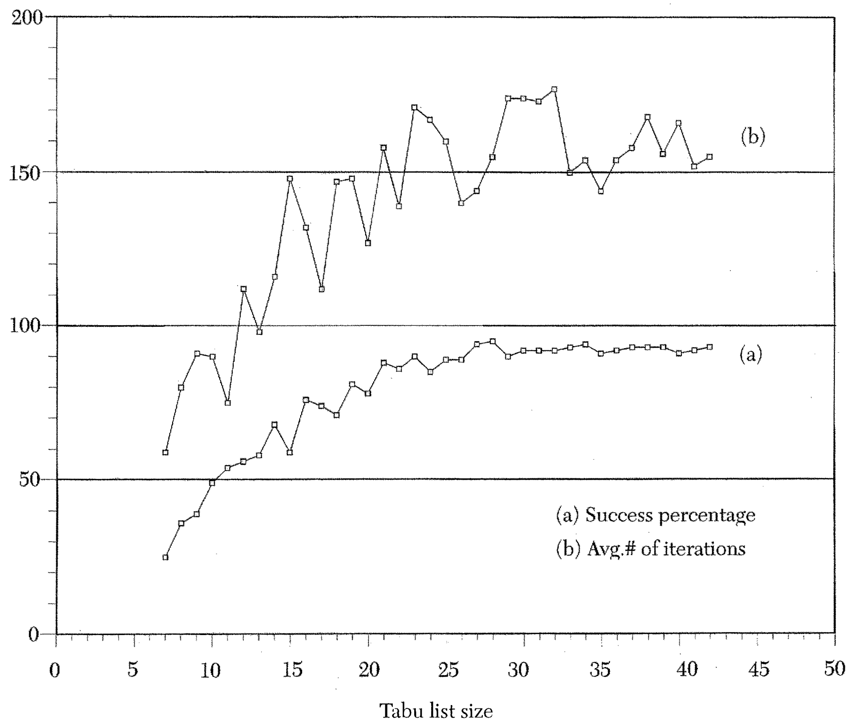



FIGURE 1

(a) A plot of the number of times that TS found an optimal solution of 100 runs vs. tabu sizes. (b) A plot of the mean number of iterations to obtain an optimal solution vs. tabu size.

Note that in the interval [7, 29] both the success percentage of TS and the mean number of iterations (time complexity to find optimality) increase when the tabu list size increases. However, for tabu size values between 30 and 42, the success percentage seems to stay constant, and the time to solve optimality tends to grow. Hence, the best tabu sizes seem to be integers somewhere between 26 and 29. With these sizes TS always found the solution of cost 14 reported in [1]. For a small tabu size ($t < 16$), cycling was detected with a high probability. But if the tabu size was too large, the number of iterations to find the optimal schedule became larger too.

Also in our experiments we have seen that for a fixed tabu list size, the number of iterations needed to find an optimal solution strongly depends on the initial schedule. For example, a schedule of cost 12 was found in 9 iterations with one initial solution, while about 400 iterations were required to reach an optimal solution using a different initial schedule.

At each step of the TS process, we examined the entire neighborhood; that is $V^* = N(x)$ for each current solution $x$. This method needs more calculation time than the partial inspection of the set $N(x)$, but the results are better. However, the complete examination becomes very expensive as the problem size increases. For large problems, $V^*$ can be restructured as follows: We examine the neighbors and take the first that improves the current solution $x$. If no move improves $x$, or if every improving move is tabu, then we examine $N(x)$. Another approach consists in randomly generating $V^*$ with a fixed number of neighbors of $x$.

An optimal schedule obtained by TS and its competition matrix are given in FIGURE 2.

Schedule

(entries are group label assigned to team)

| Team # | Meeting 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | B | C | C | B | B | A | B | B |
| 2 | A | B | C | C | C | B | B | A |
| 3 | C | A | C | B | A | B | C | C |
| 4 | A | A | A | B | C | C | A | B |
| 5 | B | A | A | A | C | A | C | A |
| 6 | C | B | B | B | B | A | A | A |
| 7 | C | C | B | C | C | C | C | B |
| 8 | A | A | B | C | A | A | B | C |
| 9 | B | C | A | C | B | B | A | C |
| 10 | B | B | B | A | A | B | A | B |
| 11 | A | B | C | A | B | C | C | C |
| 12 | C | C | A | A | A | C | B | A |

Competition Matrix

(entries are # of times team plays opponent)

| Team # | Opponent 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 2 | 2 | 2 |
| 2 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 2 |
| 3 | 2 | 2 | 0 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 2 |
| 4 | 2 | 2 | 2 | 0 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 2 |
| 5 | 2 | 2 | 2 | 3 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 3 |
| 6 | 3 | 2 | 2 | 2 | 2 | 0 | 2 | 2 | 2 | 3 | 2 | 2 |
| 7 | 2 | 2 | 2 | 3 | 2 | 2 | 0 | 2 | 2 | 2 | 2 | 3 |
| 8 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 0 | 2 | 2 | 2 | 2 |
| 9 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 3 | 2 | 2 |
| 10 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 0 | 2 | 2 |
| 11 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 |
| 12 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 0 |

**FIGURE 2**
An optimal schedule and its competition matrix.

**Conclusion** The tabu list size is important for the performance of the tabu search heuristic method. We have analyzed the performance of TS with different tabu list sizes; likewise, we found the best range of adequate tabu list sizes for our implementation of TS. For some of these tabu sizes our procedure obtained, with high probability (up to 95%), an optimal solution for the scheduling problem.

## REFERENCES


1. B. S. Elenbogen, and B. R. Maxim, Scheduling a bridge club (a case study in discrete optimization), this MAGAZINE 65 (1992), 18–26.
2. F. Glover, Tabu Search Part I, *ORSA Journal on Computing* 1 (1989), 190–206.
3. S. Kirkpatrick, C. D. Gelatt Jr., and M. Vecchi, Optimization by simulated annealing, *Science* 220 (1983), 671–680.

# Arg

JOHN A. BAKER
WILLEM SLUIS
University of Waterloo
Waterloo, ON
Canada N2L 3G1

A nonzero complex number $z$ has precisely one argument in the interval $(-\pi, \pi]$; call it the *principal argument* of $z$ and denote it by Arg $z$. If $z = x + iy$ with $x > 0$ then

$$\text{Arg } z = \arctan(y/x). \qquad (*)$$

Thus, in the right half plane (but nowhere else), Arg is given by formula $(*)$. Since Arg is discontinuous on the negative real axis, there can be no "nice" formula for it on the nonzero complex numbers; we will, however, exhibit such a formula for the *slit* plane $\mathbb{C}^{\#} := \mathbb{C}\backslash\{x \in \mathbb{R} : x \leq 0\}$.

Suppose $z = x + iy \in \mathbb{C}^{\#}$. Note that $x + \sqrt{x^2 + y^2} > 0$. Let

$$u = \sqrt{\frac{x + \sqrt{x^2 + y^2}}{2}}, \quad v = \frac{y}{2u}, \quad \text{and} \quad w = u + iv.$$

Then $z = w^2$ and Arg $w = \arctan(v/u) \in (-\pi/2, \pi/2)$, since $u > 0$. Hence Arg $z = 2$Arg $w$. But $v/u = y/2u^2$. We conclude that for all $z = x + iy$ in $\mathbb{C}^{\#}$,

$$\text{Arg } z = 2\arctan\left(\frac{y}{x + \sqrt{x^2 + y^2}}\right).$$

Note that this formula clearly exhibits the real analyticity of Arg on $\mathbb{C}^{\#}$.

---

## Math Bite: Proof of an Empirical Observation Made by a Character of Amos Oz

In Amos Oz's latest novel *Panther in the Basement* (in Hebrew, Keter, Jerusalem, 1995), Proffie, the book's twelve-and-a-quarter-year-old precocious hero, makes the following observation regarding the ratio of his age to that of the age of his twenty-year-old neighbor, Yardena, on whom he has a crush (p. 127):

> *As the years will increase, the distance between me and her will decrease (in percentage), but the sad side is that this decreasing gap will decrease more and more slowly. Like a marathon-runner that got tired.*

And indeed, the ratio of their ages, $f(t) = t/(7.75 + t)$, is such that $f'(t) = (7.75)/(7.75 + t)^2 > 0$, but $f''(t) = (-15.5)/(7.75 + t)^3 < 0$.

—DORON ZEILBERGER
TEMPLE UNIVERSITY
PHILADELPHIA, PA 19122

# Understanding the Extra Power of the Newton-Cotes Formula for Even Degree

KENNETH J. SUPOWIT
The Ohio State University
Columbus, OH 43210

The closed $(n+1)$-point Newton-Cotes method for numerical integration uses the approximation

$$\int_a^b f(x)\,dx \approx \int_a^b P_n(x)\,dx,$$

where the left-hand integral is to be approximated, and $P_n(x)$ is the (unique) $n$th degree polynomial that agrees with $f(x)$ at $n+1$ evenly-spaced points from $a$ to $b$, inclusive (see, e.g., [**2**]). That is, $P_n(a+ih)=f(a+ih)$ for $i=0,1,\dots n$ and $h=(b-a)/n$. (The method is "closed" because it involves the endpoints.) Two famous special cases are the trapezoid rule (with $n=1$):

$$\int_a^b f(x)\,dx \approx \int_a^b P_1(x)\,dx = \frac{h}{2}[f(a)+f(b)],$$

and Simpson's rule (with $n=2$):

$$\int_a^b f(x)\,dx \approx \int_a^b P_2(x)\,dx = \frac{h}{3}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right].$$

If the interval $[a,b]$ is large, it is common practice to partition it into many subintervals and to approximate the integral of $f$ on each sub-interval by a Newton-Cotes estimate.

It turns out that the error $R=\int_a^b f(x)\,dx-\int_a^b P_n(x)\,dx$ in the approximation satisfies

$$R=\begin{cases}\alpha_n h^{n+1}f^{(n+1)}(\xi), & \text{if } n \text{ is odd and } f\in C^{n+1}[a,b]\\ \alpha_n h^{n+2}f^{(n+2)}(\xi), & \text{if } n \text{ is even and } f\in C^{n+2}[a,b]\end{cases}, \tag{1}$$

where $\alpha_n$ is a constant and $\xi\in[a,b]$. For the trapezoid rule and Simpson's rule these formulas specialize to

$$R=-\frac{h^3}{12}f''(\xi) \quad\text{and}\quad R=-\frac{h^5}{90}f^{(4)}(\xi),$$

respectively. Error bounds of this type are proved in numerical analysis texts (see, e.g., [**1, 2**]) using Taylor's theorem.

There is a striking difference between the odd and even cases in the formula (1): for even $n$, the method seems to give one extra degree of accuracy, for no obvious reason. In Simpson's rule, for example, the error bound is proportional to the *fourth* derivative; hence the formula is exact if $f$ is any *third*-degree polynomial. Geometrically, this means that the area from $a$ to $b$ under a parabola is the same as the area under *any* cubic curve that meets the parabola at points $a$, $a+h$, and $b$ (see FIGURE 1). (Equivalently, the two enclosed areas are equal.)

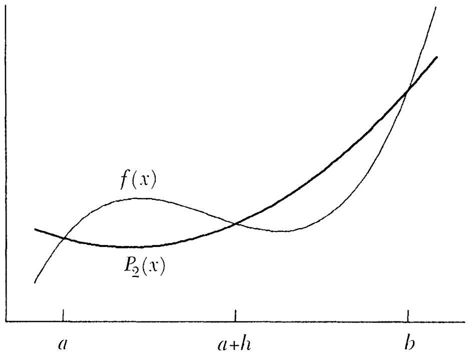



**FIGURE 1**
Simpson's rule where $f$ is a third-degree polynomial.

Indeed, for all even $n$, if $f$ is a polynomial of degree $n+1$, then the $n$th-degree Newton-Cotes formula commits zero error. The following proposition makes this precise; the proof is based on high school algebra.

PROPOSITION. *Let $f(x)$ be a polynomial of degree $n+1$, and let $P_n(x)$ be the nth degree polynomial that agrees with $f(x)$ at $n+1$ evenly-spaced points, starting at a and ending at b. If n is even, then $\int_a^b f(x)\,dx = \int_a^b P_n(x)\,dx$.*

*Proof.* Let $n = 2m$. For notational simplicity, assume that $[a, b]$ is centered at 0 (the general case follows easily). Consider the error function $E(x) = f(x) - P_n(x)$. Now $E(x)$ is a polynomial of degree $n+1$, with roots at 0, $\pm h, \pm 2h, \dots, \pm mh$. Thus, for some constant $c$,

$$\begin{aligned} E(x) &= cx(x-h)(x+h)\cdots(x-mh)(x+mh) \\ &= cx(x^2-h^2)(x^2-2^2h^2)\cdots(x^2-m^2h^2). \end{aligned}$$

Thus $E$ is an odd function, so $\int_a^b E(x)\,dx = 0$. This completes the proof.

*Remark.* The reasoning above fails for odd $n$. In this case, assuming as above that $b = -a$, the roots of the error function $E(x)$ come in positive/negative pairs; hence $E(x)$ is an even function, and so does not integrate to zero.

## REFERENCES

1. R. L. Burden and J. D. Faires, *Numerical Analysis*, 5th ed., PWS-Kent, Boston, MA (1993), 176–181.
2. D. Kincaid and W. Cheney, *Numerical Analysis*, Brooks/Cole, Pacific Grove, CA (1991), 443–452.

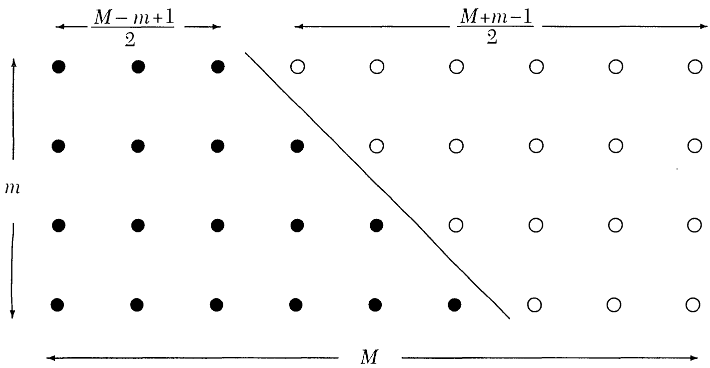
# Proof Without Words: Sums of Consecutive Positive Integers

Every integer $N > 1$, not a power of two, can be expressed as the sum of two or more consecutive positive integers.

$$N = 2^n(2k+1) \quad (n \geq 0,\ k \geq 1),$$

$$m = \min\{2^{n+1}, 2k+1\},$$

$$M = \max\{2^{n+1}, 2k+1\}.$$




$$M = \max\{2^{n+1}, 2k+1\}.$$

$$2N = mM,$$

$$N = \left(\frac{M-m+1}{2}\right) + \left(\frac{M-m+1}{2} + 1\right) + \cdots + \left(\frac{M+m-1}{2}\right).$$

## REFERENCES

1. Peter Ross, Problem 1358, this MAGAZINE 63 (1990) 350.
2. Jack V. Wales, Jr., Solution to Problem 1358, this MAGAZINE 64 (1991) 351. [See especially the comment following the solution.]

—C. L. FRENZEN
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

# The Volume of a Cone, Without Calculus

M. HIRSCHHORN
University of New South Wales
Sydney 2052, NSW
Australia

The volume $V$ of a cone with base area $A$ and height $h$ is well known to be given by $V = \frac{1}{3}Ah$. The factor $\frac{1}{3}$ arises from the integration of $x^2$ with respect to $x$. The object of this note is to start by supposing $V = cAh$, and to show—*without calculus*—that $c = \frac{1}{3}$. Using the cone formula, we'll also deduce the volume and the surface area of a sphere of radius $R$.

Consider the frustum of height $h$, top area $a$, and base area $A$, cut from a cone of height $e + h$ ("$e$" is for "extra") and base area $A$. The volume of the frustum is

$$V = cA(e+h) - cae.$$

Now, the area of a cross-section of the cone is proportional to the square of its distance from the vertex, so

$$\frac{\sqrt{a}}{e} = \frac{\sqrt{A}}{e+h}.$$

It follows that

$$e = \frac{\sqrt{a}}{\sqrt{A} - \sqrt{a}}h, \quad e + h = \frac{\sqrt{A}}{\sqrt{A} - \sqrt{a}}h$$

and the volume of the frustum is

$$V = cA\left(\frac{\sqrt{A}}{\sqrt{A} - \sqrt{a}}h\right) - ca\left(\frac{\sqrt{a}}{\sqrt{A} - \sqrt{a}}h\right)$$
$$= c\left(\frac{A\sqrt{A} - a\sqrt{a}}{\sqrt{A} - \sqrt{a}}\right)h = c(A + \sqrt{Aa} + a)h.$$

Now consider what happens as $a$ tends to $A$. The frustum becomes a cylinder, and we find that $V = 3cAh$. But we know that, for a cylinder, $V = Ah$, so $c = \frac{1}{3}$, and we conclude that the volume of a cone is

$$V = \frac{1}{3}Ah.$$

As a bonus, we obtain the volume of a frustum:

$$V = \frac{1}{3}(A + \sqrt{Aa} + a)h.$$

We conclude with two simple applications of the formula.

**The volume of a sphere** FIGURE 1 shows a sphere radius $R$, together with a cylinder of radius $R$ and length $2R$; cones are drilled out from each end of the cylinder to its center.

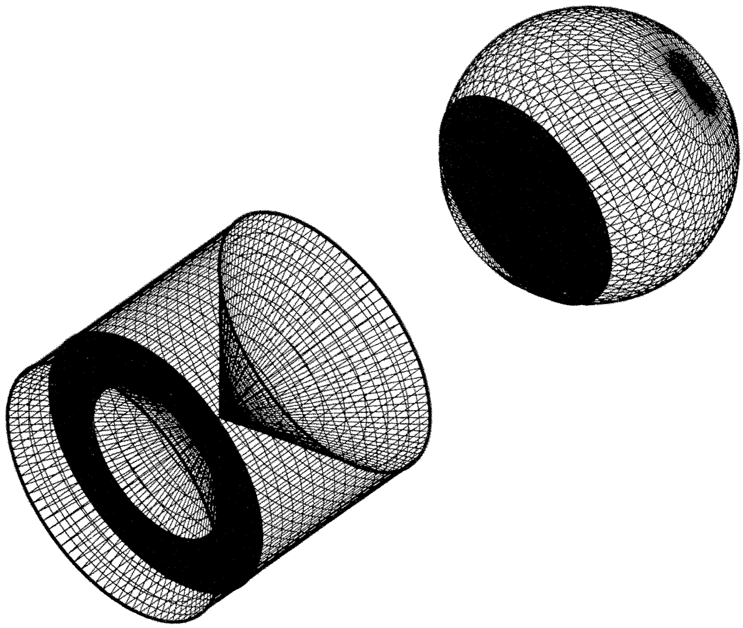


**FIGURE 1**
Two objects with the same volume

If we slice each object at a distance $x$ from its center, the area of the slice is, in each case, $\pi(R^2 - x^2)$. Thus the two solids have the same volume, and we conclude that

$$V = \pi R^2 \cdot 2R - 2 \cdot \frac{1}{3} \cdot \pi R^2 \cdot R = \frac{4}{3}\pi R^3.$$

**The surface area of a sphere** Given a sphere, we divide the surface into very many small (flat) pieces of area $A_i$, $i = 1, \dots, n$. We join each to the center, forming sharp cones.

The volume of a typical cone is $V = \frac{1}{3}A_i R$, and the total volume of all the cones is

$$V = \frac{1}{3}R\sum_{i=1}^{n} A_i = \frac{1}{3}RS,$$

where $S$ is the surface area of the sphere. Thus $\frac{1}{3}RS = \frac{4}{3}\pi R^3$, and so

$$S = 4\pi R^2.$$

# A Basis for the Intersection of Subspaces

KUNG-WEI YANG
Western Michigan University
Kalamazoo, MI 49008

If $U$ and $V$ are vector subspaces of $\mathbb{R}^n$, then so are $U+V$ and $U \cap V$. A well-known identity relates the dimensions of all four subspaces (see, e.g., [**1**], p. 46):

$$(*) \qquad \dim(U) + \dim(V) = \dim(U+V) + \dim(U \cap V).$$

This note describes a simple way of finding an explicit basis of $U \cap V$. Suppose that $U$ has basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r\}$ and $V$ has basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s\}$. Form the $n$-by-$(r+s)$ matrix $A = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r, \boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s]$, with the $\boldsymbol{u}$'s and $\boldsymbol{v}$'s as columns. Reduce $A$ to its reduced row echelon form, rref($A$). Then the number $t$ of non-pivotal columns in rref($A$) is the dimension of the intersection $U \cap V$, and the $t$ linear combinations $a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \cdots + a_r\boldsymbol{u}_r$, where $[a_1, a_2, \dots, a_r, *, *, \dots, *]$ is the transpose of a non-pivotal column of rref($A$), form a basis of the intersection.

**Example** Let $\boldsymbol{u}_1 = [1,1,1,1]$, $\boldsymbol{u}_2 = [1,2,3,4]$, $\boldsymbol{u}_3 = [0,1,2,2]$, $\boldsymbol{v}_1 = [2,3,4,7]$, $\boldsymbol{v}_2 = [1,0,1,0]$, $\boldsymbol{v}_3 = [0,1,0,3]$, and $U = \langle \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \rangle$, $V = \langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle$. Then we have

$$A = \begin{bmatrix} 1 & 1 & 0 & 2 & 1 & 0 \\ 1 & 2 & 1 & 3 & 0 & 1 \\ 1 & 3 & 2 & 4 & 1 & 0 \\ 1 & 4 & 2 & 7 & 0 & 3 \end{bmatrix} \quad \text{and} \quad \operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 3 & 0 & 2 \\ 0 & 0 & 1 & -2 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

The dimension of the intersection $U \cap V$ is therefore 2, and a basis of the intersection is

$$\{(-1)[1,1,1,1] + 3[1,2,3,4] + (-2)[0,1,2,2], (-1)[1,1,1,1] + 2[1,2,3,4] + (-2)[0,1,2,2]\} = \{[2,3,4,7],[1,1,1,3]\}.$$

To see why the method works, note that the rank of $A$ is clearly equal to $\dim(U+V) = r+s-t$. By (*), we have $t = \dim(U \cap V)$. If the non-pivotal column $[a_1, a_2, \dots, a_r, *, *, \dots, *]^{\mathrm{T}}$ is the $(r+k)$-th column, then

$$(a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \cdots + a_r\boldsymbol{u}_r) + (\text{a linear combination of "previous" pivotal } \boldsymbol{v}_j\text{'s}) = \boldsymbol{v}_k.$$

It follows that $(a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \cdots + a_r\boldsymbol{u}_r) \in U \cap V$ and that these vectors together with the $(s-t)$ pivotal $\boldsymbol{v}_i$'s span $V$. Hence these $t$ vectors form a basis of $U \cap V$.

## REFERENCE

1. K. Hoffman and R. Kunze, *Linear Algebra*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1971.

# PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

ZE-LI DOU, KEN RICHARDSON, and SUSAN G. STAPLES, *Assistant Editors*
Texas Christian University

## Proposals

*To be considered for publication, solutions should be received by March 1, 1998.*

**1529.** *Proposed by David C. Kay, University of North Carolina at Asheville, Asheville, North Carolina.*

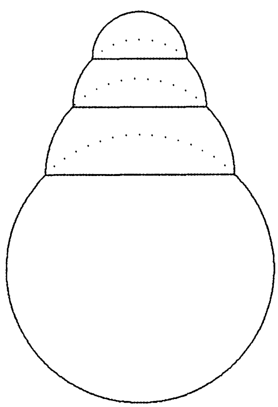
For what positive numbers $a$ is

$$\sqrt[3]{2+\sqrt{a}}+\sqrt[3]{2-\sqrt{a}}$$

an integer?

**1530.** *Proposed by Allen J. Schwenk, Western Michigan University, Kalamazoo, Michigan.*

A spherical bubble of radius 1 is surmounted by a smaller, hemispherical bubble, which in turn is surmounted by a still smaller hemispherical bubble, and so forth, until



$n = 4$

*We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution.*

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

*Solutions and new proposals should be mailed to George T. Gilbert, Problems Editor, Department of Mathematics, Box 298900, Texas Christian University, Fort Worth TX 76129, or mailed electronically (ideally as a LATEX file) to* g.gilbert@tcu.edu. *Readers who use e-mail should also provide an e-mail address.*

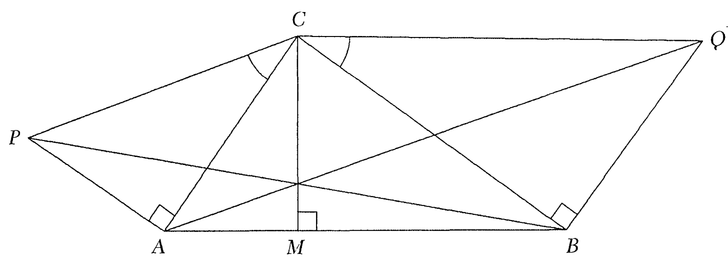
$n$ chambers including the initial sphere are formed. What is the maximum height of any bubble tower with $n$ chambers?

**1531.** *Proposed by Claus Mazanti Sorensen, student, Aarhus University, Aarhus, Denmark.*

For which positive integers $n$ does there exist a permutation $\sigma$ in the symmetric group $S_n$ such that the map $k \mapsto |\sigma(k) - k|$, $k \in \{1, 2, \ldots, n\}$ is injective?

**1532.** *Proposed by Herbert Gülicher, Westfälische Wilhelms—Universität, Münster, Germany.*

Let $\triangle ABC$, $\triangle ACP$ and $\triangle BCQ$ be non-overlapping triangles in the plane with $\angle CAP$ and $\angle CBQ$ right angles. Let $M$ be the foot of the perpendicular from $C$ to $AB$. Prove that lines $AQ$, $BP$, and $CM$ are concurrent if and only if $\angle BCQ = \angle ACP$.




**1533.** *Proposed by Joaquín Gómez Rey, I. B. "Luis Buñuel," Alcorcón, Madrid, Spain.*

Solve the recurrence relation

$$a_{n+1} = \sum_{k=0}^{n} \binom{n}{k} a_k a_{n-k}$$

in terms of $a_0$.

# Quickies

*Answers to the Quickies are on page 306.*

**Q868.** *Proposed by Hoe Teck Wee, Lengkok Bahru, Singapore.*

For any positive integer $n$, express the number of ordered triples $(k, r, s)$ of positive integers for which $krs = n$ with $r$ and $s$ relatively prime in terms of a known "arithmetical" function of $n$.

**Q869.** *Proposed by Larry Hoehn, Austin Peay State University, Clarksville, Tennessee.*

Solve

$$\sqrt{\frac{t^3 + a^3}{t + a}} + \sqrt{\frac{t^3 + b^3}{t + b}} = \sqrt{\frac{a^3 - b^3}{a - b}},$$

where $0 < a < b$.

**Q870.** *Proposed by John Brillhart and J. S. Lomont, University of Arizona, Tucson, Arizona.*

Let $(u_j)_{j=0}^{\infty}$ be an arbitrary sequence of complex numbers. For $n$, $r$, and $s$ nonnegative integers, prove that

$$\sum_{j=0}^{n}\binom{rn+2s+2}{rj+s+1}u_ju_{n-j}=2\sum_{j=0}^{n}\binom{rn+2s+1}{rj+s+1}u_ju_{n-j}.$$

# Solutions

**Sets Whose Elements Divide Their Sum** **October 1996**

**1504.** *Proposed by Erwin Just, emeritus, Bronx Community College, Bronx, New York.*

For which positive integers $n$ does there exist a set of $n$ distinct positive integers such that

(i) each member of the set divides the sum of all members of the set, and
(ii) none of its proper subsets with two or more elements satisfies (i)?

*Solution by John Christopher, California State University, Sacramento, California.*

The answer is all positive integers except $n = 2$.

A set $\{a, b\}$ of two distinct positive integers cannot satisfy (i) since both $a$ and $b$ must divide $(a+b)$ which in turn implies $a \mid b$ and $b \mid a$, an impossibility if $a \neq b$. A set with one element certainly satisfies (i) and also (ii) by default. For $n \geq 3$, let $T_n$ be the following set of distinct positive integers:

$$T_n := \{1, 2, 2\cdot 3, 2\cdot 3^2, \ldots, 2\cdot 3^{n-3}, 3^{n-2}\}.$$

To prove that $T_n$ satisfies (i) and (ii) for $n \geq 3$, let $S(T_n)$ denote the sum of the members of $T_n$. Then

$$S(T_n) = 1 + 2(1 + 3 + 3^2 + \cdots + 3^{n-3}) + 3^{n-2} = 2\cdot 3^{n-2},$$

and it is apparent that each member of $T_n$ divides $S(T_n)$ and thus $T_n$ satisfies (i). To show that (ii) is also satisfied, let $U_k$ be a proper subset of $T_n$ containing $k \geq 2$ elements and denote the sum of its members by $S(U_k)$. If $3^{n-2} \in U_k$, then $3^{n-2} < S(U_k) < 2\cdot 3^{n-2}$ and so $3^{n-2} \nmid S(U_k)$. If $2\cdot 3^j$, for some integer $j \in [0, n-3]$, is the largest member of $U_k$, then

$$2\cdot 3^j < S(U_k) \leq 1 + 2(1 + 3 + 3^2 + \cdots + 3^j) = 3^{j+1} < 2(2\cdot 3^j)$$

and so $2\cdot 3^j \nmid S(U_k)$. Hence (ii) will be satisfied for all proper subsets $U_k$ of $T_n$ with $k \geq 2$.

*Comments.* Other sets constructed include

$$T_n := \{1, 3, 3\cdot 2^2, 3\cdot 2^4, \ldots, 3\cdot 2^{2n-6}, 2^{2n-5}\}$$

and

$$T_n := \{1, 2^{n-2}, 2^{n-2}+1, 2(2^{n-2}+1), \ldots, 2^{n-3}(2^{n-2}+1)\}.$$

*Also solved by Matt Baker (student), J. C. Binz (Switzerland), David M. Bloom, Marc A. Brodie, Calvin Catt, Robert L. Doucette, Ron Ensey, Lenny Jones and Jedd Beall, John Koker, Victor Y. Kutsenok, Reginald Laursen, Jack McCown, University of Central Florida Problems Group, and the proposer. There was one incorrect solution.*

**An Extremal Problem for a Family of Functions** **October 1996**

**1505.** *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Canada, and Cecil C. Rousseau, The University of Memphis, Memphis, Tennessee.*

Let $a$ and $b$ be positive numbers satisfying $a+b \geq (a-b)^2$. Prove that

$$x^a(1-x)^b + x^b(1-x)^a \leq \frac{1}{2^{a+b-1}}$$

for $0 \leq x \leq 1$, with equality if and only if $x = 1/2$.

I. *Solution by Robert L. Doucette, McNeese State University, Lake Charles, Louisiana.*

With the change of variables $x = \frac{1}{2} - \frac{1}{2}y$, the given inequality becomes

$$(1-y)^a(1+y)^b + (1-y)^b(1+y)^a \leq 2 \tag{1}$$

for $|y| \leq 1$. Let $g(y)$ be the left-hand side of (1).

If $a = b$ then (1) becomes $(1-y^2)^a \leq 1$, which is clearly true for $|y| \leq 1$.

We assume in the following that $b > a$. Note that the maximum of $g$ on the interval $[-1,1]$ must occur at some interior point of the interval. A routine calculation shows that $g'(y) = 0$ if and only if

$$\left(\frac{1+y}{1-y}\right)^{b-a} = \frac{1+\dfrac{b+a}{b\text{-}a}y}{1-\dfrac{b+a}{b-a}y}. \tag{2}$$

Observe that the left-hand side of (2) is always positive on $(-1,1)$, while the right-hand side is positive only for $|y| < (b-a)/(b+a)$. Using the series representation

$$\ln\left(\frac{1+y}{1-y}\right) = 2\sum_{n=0}^{\infty} \frac{1}{2n+1} y^{2n+1}, \quad |y| < 1,$$

to expand each side of (2), we seek $|y| < (b-a)/(b+a)$ such that

$$y\sum_{n=0}^{\infty} \frac{1}{2n+1}\left[\left(\frac{b+a}{b-a}\right)^{2n+1} - (b-a)\right] y^{2n} = 0. \tag{3}$$

From the conditions imposed on $a$ and $b$, it follows that

$$\left(\frac{b+a}{b-a}\right)^{2n+1} - (b-a) \geq \left(\frac{b+a}{b-a}\right) - (b-a) \geq 0,$$

where the first inequality is strict for $n > 0$. This means (3), and hence the equation $g'(y) = 0$, has the unique solution $y = 0$ in the interval $(-1,1)$. It follows easily that $g(y) \leq g(0) = 2$ on $[-1,1]$, with strict inequality for $y \neq 0$.

II. *Solution by Joseph G. Gaskin, SUNY College at Oswego, Oswego, New York.*

Let $f(x) = x^a(1-x)^b + x^b(1-x)^a$, where $0 \leq x \leq 1$. If $a = b$, then $f(x) \leq f(1/2) = 2(1/4)^a$, with equality if and only if $x = 1/2$. So, supposing that $a \neq b$, we may assume $a < b$.

Since $f(0) = f(1) = 0$ and since $f$ is differentiable and positive on $(0,1)$, it follows that $f$ is maximized at a critical point in $(0,1)$. From

$$f'(x) = x^{a-1}(1-x)^{a-1}\left[(a-(a+b)x)(1-x)^{b-a} + (b-(a+b)x)x^{b-a}\right],$$

we see that $f'(x)=0$ for $x\in(0,1)$ if and only if

$$g(x) := \frac{(a+b)x-a}{b-(a+b)x}\left(\frac{1-x}{x}\right)^{b-a} = 1.$$

Note that $g(1/2)=1$ and that if $g(x)=1>0$ on $(0,1)$, then $a/(a+b)<x<b/(a+b)$.

After a bit of algebra, we find that

$$g'(x) = \frac{b-a}{(b-(a+b)x)^2}\left(\frac{1-x}{x}\right)^{b-a}\frac{\left[(a+b)-(a+b)^2\right](x-x^2)+ab}{x(1-x)}.$$

Every factor is clearly positive on $(a/(a+b), b/(a+b))$ except possibly for

$$\left[(a+b)-(a+b)^2\right](x-x^2)+ab.$$

This term is clearly positive if $(a+b)-(a+b)^2\geq 0$. Otherwise, the hypothesis $(a+b)\geq(a-b)^2$ implies

$$\begin{aligned}\left[(a+b)-(a+b)^2\right](x-x^2)+ab &\geq \left[(a+b)-(a+b)^2\right]/4+ab\\ &\geq \left[(a-b)^2-(a+b)^2\right]/4+ab=0,\end{aligned}$$

with strict inequality for $x\neq 1/2$. We conclude that $g$ is strictly increasing on $(a/(a+b),b/(a+b))$, thereby proving that the only critical point of $f(x)$ is when $x=1/2$. The inequality $f(x)\leq f(1/2)=2^{1-a-b}$ follows immediately.

*Also solved by Con Amore Problem Group (Denmark), Refik Keskin (Turkey), Phil McCartney and Jordan Stoyanov, Can A. Minh (student), Heinz-Jürgen Seiffert (Germany), and the proposers. There were six incorrect solutions.*

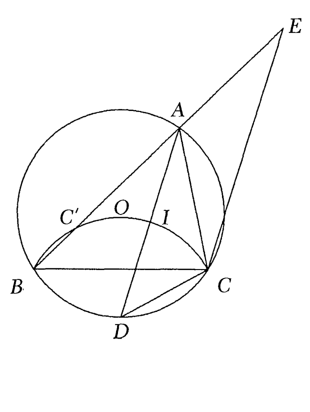
**An Angle at the Incenter of a Triangle** **October 1996**

**1506.** *Proposed by Wu Wei Chao, He Nan Normal University, Xin Xiang City, He Nan Province, China.*

Let $I$ and $O$ denote the incenter and circumcenter, respectively, of $\triangle ABC$. Assume $\triangle ABC$ is not equilateral. Prove that $\angle AIO\leq 90°$ if and only if $2BC\leq AB+CA$, with equality holding only simultaneously.

I. *Solution by Con Amore Problem Group, Copenhagen, Denmark.*

Without loss of generality, assume $AB\geq AC$. As shown in the figure below, let $D$ be the intersection of $AI$ and the circumcircle of $\triangle ABC$. Construct the circle with



center $D$ that passes through $B$ and $C$. By the symmetry of $AB$ and $AC$ in the angle bisector $AD$, this circle intersects segment $AB$ in a point $C'$ such that $AC' = AC$. Let $P$ be the intersection of this circle with $AD$. Then chords $CP$ and $C'P$ have equal length. If $AB > AC$, this directly implies $P$ is on the angle bisector of $\angle ABC$. In the special case of $AB = AC$, $\angle ABC = \angle ADC = 2\angle PBC$. Therefore, in either case, $P = I$. Finally, let $E$ be on ray $\overrightarrow{BA}$ such that $AE = AC$.

It is easy to see that $\triangle ADC$ is similar to $\triangle EBC$, and so

$$\frac{ID}{AD} = \frac{CD}{AD} = \frac{BC}{BE} = \frac{BC}{AB + AC}.$$

But $\angle AIO \leq 90°$ if and only if $ID/AD \leq 1/2$, and so is equivalent to $BC/(AB + AC) \leq 1/2$, or $2BC \leq AB + AC$, with equality holding only simultaneously.

II. *Solution by Can A. Minh, student, University of Southern California, Los Angeles, California.*

We have $\angle AIO \leq 90°$ if and only if $\cos\angle AIO \geq 0$, if and only if $AO^2 \leq OI^2 + IA^2$.

Let $a = BC$, $b = AC$, $c = AB$, $R$ be the circumradius, and $r$ be the inradius. Substitution from $OI^2 = R^2 - 2Rr$ (Euler's formula) and the easily-obtained $IA = r/\sin(A/2)$ yields that the above is equivalent to $2R \leq r/\sin^2(A/2) = 2r/(1 - \cos A)$. From $R = a/(2\sin A)$, $r = bc\sin A/(a + b + c)$, and the law of cosines, we deduce the equivalent condition

$$2a^2 + a(b + c) - (b + c)^2 = (2a - b - c)(a + b + c) \leq 0,$$

or $2a \leq b + c$.

*Also solved by Roy Barbara (Lebanon), Jiro Fukuta (Japan), Murray S. Klamkin (Canada), Victor Y. Kutsenok, George Tsapakidis (Greece), and the proposer.*

**Convergence of an Iterated Exponential** **October 1996**

**1507.** *Proposed by Howard Morris, Ridgeland, Mississippi.*

For what real values of $a$ and $b_0$ does the sequence $(b_n)_{n \geq 0}$ defined by $b_{n+1} = e^{ab_n}$ converge?

*Solution by Michael Woltermann, Washington and Jefferson College, Washington, Pennsylvania.*

It is convenient to view the recursion $b_{n+1} = e^{ab_n}$ as a first-order dynamical system. Its equilibrium values or fixed points are solutions to the equation $x = e^{ax}$ or $\ln(x)/x = a$. Since $\ln(x)/x$ has an absolute maximum value of $1/e$, $x = e^{ax}$ has no fixed points for $a > 1/e$, and $(b_n)$ diverges.

If $a = 1/e$, the recursion has $e$ as its unique fixed point. For $b_0 \leq e$, a simple induction argument shows that $(b_n)$ is nondecreasing and bounded above by $e$, hence converges to $e$. If $b_0 > e$, $(b_n)$ is an increasing sequence each of whose terms exceeds $e$, and therefore $(b_n)$ cannot converge.

If $0 < a < 1/e$, the recursion has two distinct fixed points $x_0$ and $x_1$, $1 < x_0 < e < x_1$. If $x < x_0$ or $x > x_1$, $e^{ax} > x$, and thus $(b_n)$ is strictly increasing for $b < x_0$ or $b > x_1$. When $b_0 < x_0$, a simple induction argument shows that all $b_n$ are less than $x_0$, and thus $(b_n)$ converges to $x_0$. When $b_0 > x_1$, $(b_n)$ is an increasing sequence, each of whose terms exceeds $x_1$, and thus $(b_n)$ cannot converge. If $x_0 < x < x_1$, $x_0 < e^{ax} < x$, and thus for $x_0 < b_0 < x_1$, $(b_n)$ is strictly decreasing and bounded below by $x_0$; hence $(b_n)$ converges to $x_0$.

If $a=0$, $b_n=1$ for $n>0$ regardless of the value of $b_0$, and thus $(b_n)$ converges to 1 for all real $b_0$.

Now consider $-e \le a < 0$. Let $x_0 < 1$ be the unique fixed point of $e^{ax}$, and let $f(x)=e^{ae^{ax}}$. A straightforward computation shows that

$$f'(x) \le f'(\ln(-1/a)/a) = -a/e \le 1,$$

with equality if and only if $a=-e$ and $x=1/e$. It follows that $f(x)-x$ is a decreasing function. Since $f(x_0)=x_0$, $f(x)>x$ for $x<x_0$ and $f(x)<x$ for $x>x_0$. In addition, $f$ is increasing, so for $x<x_0$, $x<f(x)<f(x_0)=x_0$ and for $x>x_0$, $x_0=f(x_0)<f(x)<x$. When $b_0<x_0$, it follows that $(b_{2n})$ increases to $x_0$, while $(b_{2n+1})$ decreases to $x_0$. Similar reasoning applies when $b_0>x_0$. Thus $(b_n)$ converges to $x_0$ for all real $b_0$ when $-e \le a \le 0$.

If $a<-e$, the recursion has a unique positive fixed point, $x_0<1/e$. With $f(x)=e^{ax}$, $f'(x_0)=ae^{ax_0}=ax_0<\ln(1/e)=-1$, so $|f'(x_0)|>1$. In the terminology of dynamical systems, $x_0$ is a repelling fixed point of the recursion, and thus $(b_n)$ converges to $x_0$ if and only if $b_0=x_0$.

*Also solved by John Christopher, Parviz Khalili, Western Maryland College Problems Group, and the proposer. There were five incorrect solutions.*

**Determinants of Matrices with Bernoulli Entries** **October 1996**

**1508.** *Proposed by Saul Stahl, University of Kansas, Lawrence, Kansas.*

Let $\det_n$ denote the determinant of the $n \times n$ matrix whose entries are independent random variables each of which has value 1 with probability $p$ and value 0 with probability $1-p$. Compute the mean and variance of $\det_n$ for each positive integer $n$.

*Composite of solutions due to Matt Baker, student, University of California at Berkeley, Berkeley, California, and Nicholas C. Singer, Annandale, Virginia.*

We prove the more general result that when the entries are independent, identically distributed random variables with mean $\mu_1$ and finite second moment $\mu_2$ (so that the mean and variance of $\det_1$ are $\mu_1$ and $\mu_2-\mu_1^2$, respectively), $\det_n$ has mean 0 and variance, for $n \ge 2$,

$$n!\left[(n-1)\mu_1^2+\mu_2\right]\left(\mu_2-\mu_1^2\right)^{n-1}.$$

In the special case stated in the problem, $\mu_1=\mu_2=p$, this simplifies to

$$n!p^n(1-p)^{n-1}[(n-1)p+1].$$

To avoid trivial cases, assume $\mu_2>0$ and $n \ge 2$. We compute the formula for the mean as follows. Let $(X_{ij})$, $1 \le i,j \le n$ denote the entries of our matrix. Let $S_n$ denote the group of permutations of $n$ elements. By standard properties of the determinant and of expectation,

$$E[\det_n] = E\left[\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) X_{1\sigma(1)} \cdots X_{n\sigma(n)}\right]$$

$$= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} E\left[X_{i\sigma(i)}\right] = \mu_1^n \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) = 0,$$

since there are exactly as many odd permutations as even ones for $n \ge 2$.

Because $E[\det_n] = 0$,

$$\operatorname{Var}[\det_n] = E\left[\det_n^2\right] = E\left[\left(\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) X_{1\sigma(1)} \cdots X_{n\sigma(n)}\right)^2\right]$$
$$= \sum_{\sigma, \tau \in S_n} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) E\left[X_{1\sigma(1)} \cdots X_{n\sigma(n)} X_{1\tau(1)} \cdots X_{n\tau(n)}\right].$$

By symmetry, this is just

$$n! \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) E\left[X_{1\sigma(1)} X_{n\sigma(n)} X_{11} \cdots X_{nn}\right].$$

From the given moments and assumption of independence,

$$E\left[X_{1\sigma(1)} \cdots X_{n\sigma(n)} X_{11} \cdots X_{nn}\right] = \mu_2^{\operatorname{Fix}(\sigma)} \cdot \mu_1^{2n - 2\operatorname{Fix}(\sigma)} = \mu_2^n \left(\mu_1^2/\mu_2\right)^{n - \operatorname{Fix}(\sigma)},$$

where $\operatorname{Fix}(\sigma)$ is the number of elements in $\{1, \ldots, n\}$ fixed by $\sigma$.

Now define $\langle n, k \rangle = \sum_{\sigma \in S_n, \operatorname{Fix}(\sigma) = k} \operatorname{sgn}(\sigma)$ for $n \geq 1$, $0 \leq k \leq n$. Then

$$\operatorname{Var}[\det_n] = n! \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \mu_2^n \left(\mu_1^2/\mu_2\right)^{n - \operatorname{Fix}(\sigma)} = n! \mu_2^n \sum_{k=0}^{n} \langle n, k \rangle \left(\mu_1^2/\mu_2\right)^{n-k}.$$

We claim that

$$\langle n, k \rangle = \binom{n}{k} (-1)^{n-k-1} (n - k - 1) \quad \text{for } n \geq 1, 0 \leq k \leq n.$$

To prove the claim, first observe that $\langle n, k \rangle = \binom{n}{k} \langle n - k, 0 \rangle$. So we just need to show that $\langle n, 0 \rangle = (-1)^{n-1}(n-1)$ for $n \geq 1$. We do this by induction. The result is clear for $n = 1$. Let $n \geq 2$, and assume the result for $1, \ldots n - 1$. Now

$$\sum_{k=0}^{n} \langle n, k \rangle = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) = 0.$$

Thus

$$\langle n, 0 \rangle = -\sum_{k=1}^{n} \langle n, k \rangle = -\sum_{k=1}^{n} \binom{n}{k} \langle n - k, 0 \rangle$$
$$= -\sum_{k=1}^{n} \binom{n}{k} (-1)^{n-k-1} (n - k - 1) = -\sum_{k=0}^{n-1} \binom{n}{k} (-1)^{k-1} (k - 1).$$

Differentiating the identity

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

with respect to $x$ and then multiplying by $x$ yields the identity

$$\sum_{k=0}^{n} \binom{n}{k} (k - 1) x^{k-1} = \frac{((n-1)x - 1)(1 + x)^{n-1}}{x}. \tag{1}$$

Substituting $x = -1$ into (1) completes the proof of the claim.

Therefore

$$\begin{aligned}\operatorname{Var}[\det_n] &= n!\mu_2^n \sum_{k=0}^{n} \binom{n}{k}(-1)^{n-k-1}(n-k-1)\left(\mu_1^2/\mu_2\right)^{n-k}\\ &= n!\mu_2^n \sum_{k=0}^{n} \binom{n}{k}(-1)^{k-1}(k-1)\left(\mu_1^2/\mu_2\right)^{k}.\end{aligned}$$

Substituting $x = -\mu_1^2/\mu_2$ into (1) (the special case $\mu_1 = 0$ is easily checked) yields

$$\operatorname{Var}[\det_n] = n!\left[(n-1)\mu_1^2 + \mu_2\right]\left(\mu_2 - \mu_1^2\right)^{n-1},$$

as we set out to prove.

*Comment.* David Callan cleverly evaluated

$$\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)\, \mu_2^{\mathrm{Fix}(\sigma)} \mu_1^{2n-2\mathrm{Fix}(\sigma)}$$

by recognizing it as the determinant of the $n \times n$ matrix $\mu_1^2 J + (\mu_2 - \mu_1^2) I$, where $J$ is the matrix of ones and $I$ is the identity matrix. Observing that the eigenvalues of $\mu_1^2 J$ are $n\mu_1^2$ and, with multiplicity $n-1$, 0, this determinant equals

$$\left(n\mu_1^2 + (\mu_2 - \mu_1^2)\right)\left(\mu_2 - \mu_1^2\right)^{n-1} = \left((n-1)\mu_1^2 + \mu_2\right)\left(\mu_2 - \mu_1^2\right)^{n-1}.$$

*Also solved by David Callan, Thomas Jager, Kee-Wai Lau (Hong Kong), Roger Pinkham, Western Maryland College Problems Group, and the proposer.*

# Answers

*Solutions to the Quickies on page 299.*

**A868.** *I.* Let $A$ denote the set of ordered triples $(k, r, s)$ with the required property, and let $B$ denote the set of all positive divisors of $n^2$. We show that $\phi$: $A \to B$ defined by $\phi(k, r, s) = nr/s$ is a bijection.

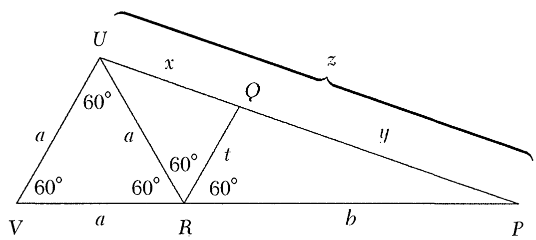
Clearly $\phi$ maps from $A$ to $B$. Because $r$ and $s$ are relatively prime, it is also clear that $\phi$ is injective. For $d \in B$, let $d/n = r/s$, where $r$ and $s$ are relatively prime. Note that $r$ divides not just $n^2$, but actually $n$. Therefore, $(n/(rs), r, s)$ is in $A$ and maps to $d$, hence $\phi$ is surjective, completing the proof.

*II. Provided by the Editors.* Let $f(n)$ denote the number of triples of the required form. Observe that $f$ is multiplicative, i.e., $f(mn) = f(m)f(n)$ whenever $m$ and $n$ are relatively prime. Thus, we need only evaluate $f$ on prime powers. For $n = p^a$, $p$ prime, we may choose $k$ dividing $n$ in $a+1$ ways. For each $k$, either $r = p^{a-k}$ and $s = 1$, or $r = 1$ and $s = p^{a-k}$. These two possibilities coincide when $k = a$. Hence $f(p^a) = (a+1)\cdot 2 - 1 = 2a+1$. We note that $2a+1$ is the number of divisors of $p^{2a}$, or $\sigma_0(p^{2a})$. Recalling that the divisor function $\sigma_0$ is also multiplicative, we conclude that $f(n) = \sigma_0(n^2)$, the number of divisors of $n^2$.

**A869.** *I.* Let $x = \sqrt{t^2 - at + a^2}$, $y = \sqrt{t^2 - bt + b^2}$, and $z = \sqrt{a^2 + ab + b^2}$. Then the given equation is equivalent to $x + y = z$. Furthermore,

$$\begin{aligned} x^2 &= t^2 + a^2 - 2at\cos 60^\circ, \\ y^2 &= t^2 + b^2 - 2bt\cos 60^\circ, \\ z^2 &= a^2 + b^2 - 2ab\cos 120^\circ. \end{aligned}$$

From the law of cosines, we see that $x$, $y$, and $z$ are sides of triangles in the figure below. Triangles $PQR$ and $PUV$ are similar, hence $t/a = b/(a+b)$ or $t = ab/(a+b)$.




*II. Provided by the Editors.* After canceling common factors in the fractions and some algebra, the given equation is equivalent to

$$\sqrt{\left(t-\frac{a}{2}\right)^2+\left(-\frac{\sqrt{3}}{2}a\right)^2}+\sqrt{\left(t-\frac{b}{2}\right)^2+\left(\frac{\sqrt{3}}{2}b\right)^2}=\sqrt{\left(\frac{a}{2}-\frac{b}{2}\right)^2+\left(\frac{\sqrt{3}}{2}a+\frac{\sqrt{3}}{2}b\right)^2}$$

The three radicals represent the distances between $(t,0)$, $(a/2, \sqrt{3}\,a/2)$, and $(b/2, -\sqrt{3}\,b/2)$, so that $(t,0)$ must be on the line segment between $(a/2, \sqrt{3}\,a/2)$ and $(b/2, -\sqrt{3}\,b/2)$. It follows that

$$t = \frac{b}{a+b}\cdot\frac{a}{2} + \frac{a}{a+b}\cdot\frac{b}{2} = \frac{ab}{a+b}.$$

**A870.** For $0 \le j \le n$, we have

$$f_{r,s}(n,j) := \binom{rn+2s+2}{rj+s+1} - 2\binom{rn+2s+1}{rj+s+1} = -\frac{r(n-2j)}{rn+2s+2}\binom{rn+2s+2}{rj+s+1}.$$

Then

$$\begin{aligned} f_{r,s}(n,n-j) &= \frac{r(n-2j)}{rn+2s+2}\binom{rn+2s+2}{rn-rj+s+1} \\ &= \frac{r(n-2j)}{rn+2s+2}\binom{rn+2s+2}{rj+s+1} = -f_{r,s}(n,j). \end{aligned}$$

Thus, re-indexing by $j \to n-j$, we have

$$S := \sum_{j=0}^{n} f_{r,s}(n,j)u_j u_{n-j} = \sum_{j=0}^{n} f_{r,s}(n,n-j)u_j u_{n-j} = -S.$$

Thus, $S = 0$, which implies the claim.

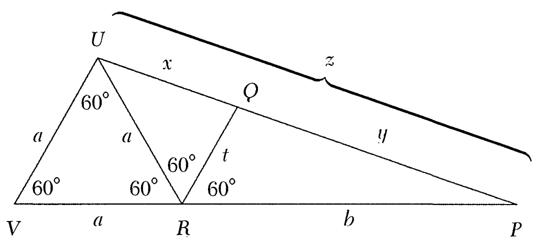
# Correction.

**S1495, April 1997.** Joel Rosenberg was inadvertently omitted from the list of those correctly solving the problem.

**A869.** *I.* Let $x=\sqrt{t^2-at+a^2}$, $y=\sqrt{t^2-bt+b^2}$, and $z=\sqrt{a^2+ab+b^2}$. Then the given equation is equivalent to $x+y=z$. Furthermore,

$$x^2=t^2+a^2-2at\cos 60°,$$
$$y^2=t^2+b^2-2bt\cos 60°,$$
$$z^2=a^2+b^2-2ab\cos 120°.$$

From the law of cosines, we see that $x$, $y$, and $z$ are sides of triangles in the figure below. Triangles $PQR$ and $PUV$ are similar, hence $t/a=b/(a+b)$ or $t=ab/(a+b)$.




*II. Provided by the Editors.* After canceling common factors in the fractions and some algebra, the given equation is equivalent to

$$\sqrt{\left(t-\frac{a}{2}\right)^2+\left(-\frac{\sqrt{3}}{2}a\right)^2}+\sqrt{\left(t-\frac{b}{2}\right)^2+\left(\frac{\sqrt{3}}{2}b\right)^2}=\sqrt{\left(\frac{a}{2}-\frac{b}{2}\right)^2+\left(\frac{\sqrt{3}}{2}a+\frac{\sqrt{3}}{2}b\right)^2}.$$

The three radicals represent the distances between $(t,0)$, $(a/2,\sqrt{3}\,a/2)$, and $(b/2,-\sqrt{3}\,b/2)$, so that $(t,0)$ must be on the line segment between $(a/2,\sqrt{3}\,a/2)$ and $(b/2,-\sqrt{3}\,b/2)$. It follows that

$$t=\frac{b}{a+b}\cdot\frac{a}{2}+\frac{a}{a+b}\cdot\frac{b}{2}=\frac{ab}{a+b}.$$

**A870.** For $0\le j\le n$, we have

$$f_{r,s}(n,j):=\binom{rn+2s+2}{rj+s+1}-2\binom{rn+2s+1}{rj+s+1}=-\frac{r(n-2j)}{rn+2s+2}\binom{rn+2s+2}{rj+s+1}.$$

Then

$$f_{r,s}(n,n-j)=\frac{r(n-2j)}{rn+2s+2}\binom{rn+2s+2}{rn-rj+s+1}$$
$$=\frac{r(n-2j)}{rn+2s+2}\binom{rn+2s+2}{rj+s+1}=-f_{r,s}(n,j).$$

Thus, re-indexing by $j\to n-j$, we have

$$S:=\sum_{j=0}^{n}f_{r,s}(n,j)u_ju_{n-j}=\sum_{j=0}^{n}f_{r,s}(n,n-j)u_ju_{n-j}=-S.$$

Thus, $S=0$, which implies the claim.

# Correction.

**S1495, April 1997.** Joel Rosenberg was inadvertently omitted from the list of those correctly solving the problem.

# REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College
1997–98: University of Augsburg,
Germany

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

Drosnin, Michael, *The Bible Code*, Simon & Schuster, 1997. Bruni, Frank, New book claims code in bible predicts world events, *New York Times* (29 May 1997), B6. Witztum, Doron, Eliyahu Rips, and Yoav Rosenberg, Equidistant letter sequences in the Book of Genesis, *Statistical Science* 9 (3) (1994) 429–438. Reviewed in *Chance News* 4.15 (21 October 1995—16 November 1995) and 6.07 (11 May 1997—7 June 1997) at `http://www.geom.umn.edu/locate/chance` .

Does the original Hebrew of the Book of Genesis tell us about events of today? Do names of present-day politicians occur there in coded form? Does it foretell the assassination of Yitzak Rabin? Yes, claims the book in question, which has been featured in full-page ads in the *New York Times*. I heard its author discuss it on TV on (don't laugh) "The Oprah Winfrey Show" (thanks to an alert from Assistant Editor Rosenthal). Since I have not seen the book itself, I am not reviewing it, just drawing it to your attention. Author Drosnin's claim to fame is that he warned Rabin a year in advance. He did his "data-mining" using the equidistant letter sequence "code" that was explored by Witztum et al., thus giving his work a cloak of respectability. Witztum et al. represented the Book of Genesis as a long string of characters; a name is "found" there if its letters appear in order, equally spaced apart, anywhere in the string. They found that names and dates of famous rabbis occur and occur in closer proximity than chance should allow ($P = .00002$). The *Chance News* articles mention further studies, by serious statisticians, of this "code," who apply it to other books (biblical and otherwise) in trying to explain the results of Witztum et al. [The Old Statistician comments: "Rare events happen all the time."]

Grossman, Jerrold W., The Erdős Number Project Web Site, `http://www.oakland.edu/~grossman/erdoshp/html` .

Paul Erdős wrote more joint mathematics papers than anyone else in history. If you co-authored a paper with Paul Erdős, your *Erdős number* is 1. If the smallest Erdős number of the co-authors of your papers is $n-1$, then your Erdős number is $n$. This Web site deals with mathematical collaboration in general and lists people with Erdős numbers 1 and 2, plus links to other information about Erdős himself. [Thanks to Jerry Grossman.]

Monastyrsky, Michael, *Modern Mathematics in the Light of Fields Medals*, A K Peters, 1997; + 160 pp, $35. ISBN 1–56881–065–2.

This short book is a surprisingly readable account of the work of each of the winners of the Fields Medal, "a road map to the territory of mathematics, with the Fields Medals as a convenient set of nodal points" (from the Foreword by Freeman Dyson).

Seife, Charles, New test sizes up randomness, *Science* 276 (25 April 1997) 532; `http://www.sciencemag.org` . Stewart, Ian, Putting randomness in order, *New Scientist* (10 May 1997); `http://www.newscientist.com/` . Pincus, Steve, and Rudolf E. Kalman, Not all (possibly) "random" sequences are created equal, *Proceedings of the National Academy of Sciences* 94, 3513–3518; `http://www.pnas.org/` . Falk, Ruma, and Clifford Konold, Making sense of randomness, *Psychological Review* 104 (2) (1997) 301–318. Reviewed in *Chance News* 6.07 (11 May 1997—7 June 1997) at `http://www.geom.umn.edu/locate/chance` .

What makes a sequence random? A sequence of heads and tails is *truly random* if each successive outcome is independent of the previous ones. But if you get HTHTHTHT ... , the result doesn't *seem* random. A sequence is *(apparently) random* if it has no "obvious" patterns, that is, if it is not predictable. Mathematicians can measure this randomness objectively in terms of the shortest computer program to produce the sequence, and this definition turns out to be equivalent to the sequence not being rejected as random by certain statistical tests. However, by this definition, the first million digits of $\pi$ do not form an (apparently) random sequence. Pincus and Kalman propose a new definition in terms of *approximate entropy*. Let $p_H$ and $p_T$ be the proportions of heads and tails, and let $p_{HH}$, $p_{HT}$, $p_{TH}$, and $p_{TT}$ be the proportions of ordered pairs, and so forth. Define the entropy for each successive distribution by

$$\begin{aligned} E(p_1) &= p_H \log p_H + p_T \log p_T, \\ E(p_2) &= p_{HH} \log p_{HH} + p_{HT} \log p_{HT} + p_{TH} \log p_{TH} + p_{TT} \log p_{TT}, \end{aligned}$$

and so forth. Pincus and Kalman then define the *kth order approximate entropy* of the sequence as

$$ApEn(k) = E(p_k) - E(p_{k-1}).$$

A sequence of length $n$ is (apparently) random only if it has maximum approximate entropy of all orders $k < \log(\log n) + 1$; hence, it must be as uniform as possible in the distribution of heads and tails, pairs, triples, and so forth. For example, according to this definition, the only (apparently) random sequences of length 5 are HHTTH, HTTHH, TTHHT, and THHTT; and while $\pi$ is (apparently) random, the algebraic number $\sqrt{2}$ is more random than the transcendental number $e$. Meanwhile, psychologists have concluded that people base their assessment of how random a sequence is based on the difficulty of memorizing or copying it. But should experimenters redo a randomized assignment of subjects to treatments just because it turns out to appear nonrandom?

Rauff, James V., Sad songs: The formal languages of Warlpiri iconography, *Humanistic Mathematics Network Journal* #15 (July 1997) 17–27.

The Warlpiri are aborigines in the Central Australian desert. They tell "sand stories" about ancestral events by drawing figures in the sand, along with songs and narration. There is a basic "alphabet" of figures, combined in certain particular ways. Author Rauff formulates a formal grammar of sand stories, with about a dozen rules, as well as grammars for other kinds of iconographic designs drawn by the Warlpiri. He then asks, "Is Warlpiri iconography mathematics? ... The Warlpiri recognize the components and rules of their iconography ... What they don't seem to have is a 'theory of iconography' that abstracts general patterns ... Perhaps [they have] 'mathematical ideas' rather than mathematics."

*Philosophy of Mathematics Education Newsletter/Journal*, `http:/www.ex.ac.uk/~PErnest/`.

This electronic journal aims to "foster awareness of philosophical aspects of mathematics education and mathematics," "to disseminate news of events and new thinking in these topics," and to encourage communication among teachers and scholars. The journal's home is at the University of Exeter, and most articles are by British mathematics educators.

# NEWS AND LETTERS

**The root and ratio tests.** Several readers, including Ted Alper (Stanford University), Prem N. Bajaj (Wichita State University), and Yongzhi (Peter) Yang (University of St. Thomas), pointed out an error in David Cruz-Uribe's note, *The relation between the root and ratio tests*, this MAGAZINE 70, pp. 214-215. In response, Professor Cruz-Uribe writes as follows:

In the article *The relation between the root and ratio tests*, we gave as an example the series $\sum a_n$, where

$$a_0 = 1, \qquad a_n = a_{n-1} \cdot \frac{\log(1 + 1/(n+1))}{\log(n+1)\log(n+2)}, \quad n \geq 1,$$

and claimed that this series was such that the ratio test could not be used to show convergence, the root test was not easily applied, but the so-called arithmetic mean test could be used to show convergence. However, this example is incorrect. The limit

$$\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

equals 0, not 1 as was stated in the article, so the ratio test does indeed show convergence.

Beyond the calculation error, the failure of this example follows from the Cesàro summability theorem:

**Cesàro Summability:** *If the sequence* $\{x_n\}$ *converges to* $\alpha$ *then the sequence of arithmetic means*

$$\sigma_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

*also converges to* $\alpha$.

(For a proof, see R.G. Bartle, *The Elements of Real Analysis*, 2nd Ed., Wiley, New York, 1976, pp. 128-30.) This theorem implies that if the limit in the ratio test exists, then the limit in the arithmetic mean test also exists, and has the same value.

Therefore, to find an example of a series on which the ratio test fails but the arithmetic mean test shows convergence, we must find a series such that the limit of the consecutive ratios does not exist. One example is the re-arranged geometric series given in the article:

$$1/2 + 1 + 1/8 + 1/4 + 1/32 + 1/16 + \cdots.$$

Following is an example to which the root test is not easily applied. For $k \geq 1$, let $i_k = 1$ if 5 divides $k$ and let $i_k = -1$ otherwise. Define the series $\sum a_n$ by

$$a_0 = 1, \qquad a_n = \prod_{k=1}^{n} \left(1 + \frac{i_k}{k}\right)^k.$$

Since

$$\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e \quad \text{and} \quad \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = 1/e,$$

the limit of the consecutive ratios does not exist. To apply the root test we would have to evaluate the limit

$$\lim_{n\to\infty} \prod_{k=1}^{n} \left(1 + \frac{i_k}{k}\right)^{k/n}.$$

To apply the arithmetic mean test, note that by Cesàro summability,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{5n} \sum_{k=1}^{5n} \frac{a_n}{a_{n-1}} &\le \frac{1}{5} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left(1 + \frac{1}{5k}\right)^{5k} + \lim_{n\to\infty} \frac{1}{5n} \sum_{k=1}^{5n} \left(1 - \frac{1}{k}\right)^{k} \\
&= \frac{1}{5} \lim_{n\to\infty} \left(1 + \frac{1}{5n}\right)^{5n} + \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{n} \\
&= e/5 + 1/e \\
&< 1.
\end{aligned}$$

It follows immediately that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{a_n}{a_{n-1}} < 1$$

and so, by the arithmetic mean test, the series converges.

(I would like to thank Ted Alper for pointing out my error and for calling my attention to the relevance of Cesàro summability.)

David Cruz-Uribe, SFO
Trinity College
Hartford, CT 06106-3100

**Repeated powers: a comment in verse.** Another reader, John Derbyshire (Huntington, NY) commented—in the form of a Petrarchan sonnet—on *The limit of* $x^{x^{\cdot^{\cdot^{\cdot^{x}}}}}$ *as* $x$ *tends to zero*, J. Marshall Ash, this MAGAZINE 69, June 96, as follows:

When $x$ is raised to power of $x$ we see
The first step of an iteration which
Can then be carried on without a glitch
For ever. In the range of powers of $e$
From minus $e$ itself to $e$ inverted
These endless tottering stairs of shrinking $x$'s
Converge! And yet one question still perplexes:
Beyond that range, what facts can be asserted?
J. Marshall Ash, a scholar from DePaul
Has shown us that, when $x$ is microscopic,
The even steps climb up without a stall
To one; the odd steps, likewise asymptotic,
Decline to zero. Thanks go out from all
For shedding light upon this curious topic.

# CONTENTS

## ARTICLES



## NOTES



## PROBLEMS



## REVIEWS



## NEWS AND LETTERS



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