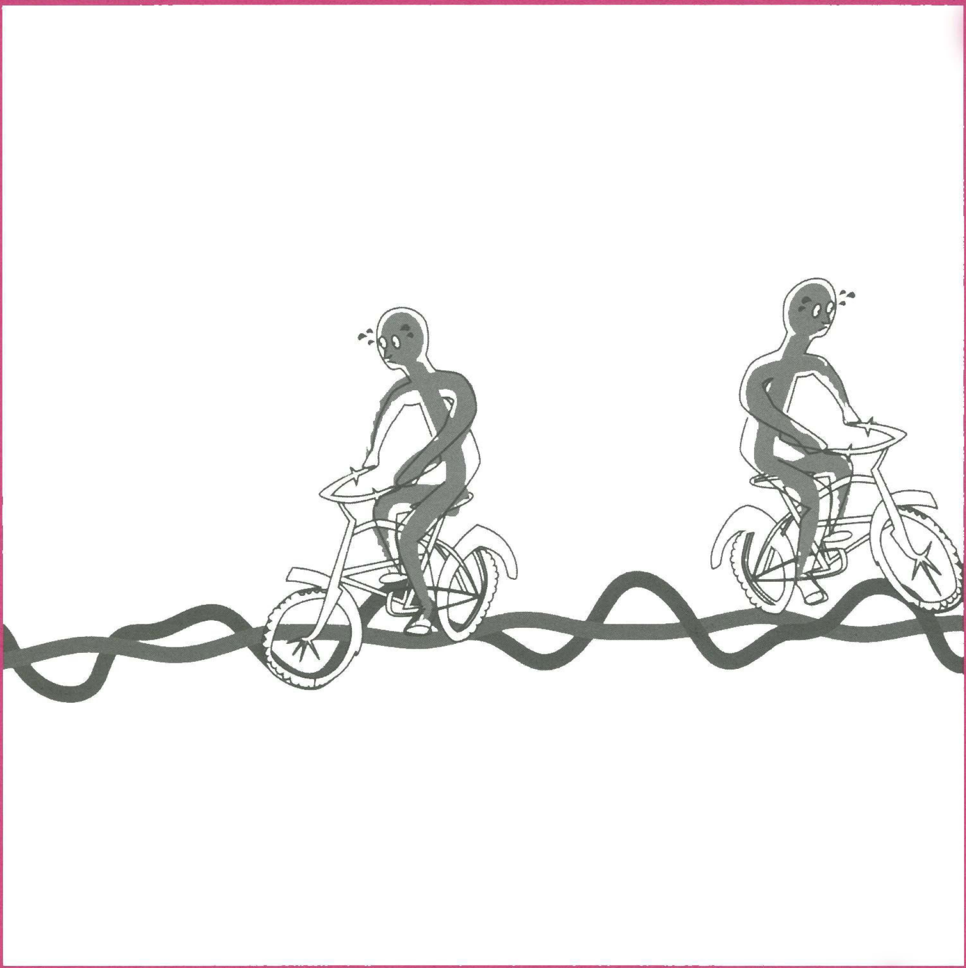
Vol. 77, No. 5, December 2004

MATHEMATICS MAGAZINE



- It's Okay to Be Square If You're a Flexagon
- Loxodromes: A Rhumb Way to Go
- Which Way Did You Say That Bicycle Went?

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. The *Magazine* is not a research journal, so the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate 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 among various branches of mathematics and between mathematics and other disciplines.

A more detailed statement of author guidelines appears in this *Magazine,* Vol. 74, pp. 75–76, and is available from the Editor or at www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Submit new manuscripts to Frank A. Farris, Editor, Mathematics Magazine, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053-0373. Manuscripts should be laser printed, with wide line spacing, and prepared in a style consistent with the format of *Mathematics Magazine.* Authors should mail three copies and keep one copy. In addition, authors should supply the full five-symbol 2000 Mathematics Subject Classification number, as described in *Mathematical Reviews.*

Cover image: *The Bicycle Went **Which** Way?,* by Jason Challas. In an article in this issue, David Finn shows how to construct bicycle tracks that thwart any attempt to answer, "Which way did the bicycle go?" One rider on the cover successfully navigates the path from right to left, the other from left to right, but it seems to be a tricky business.

Jason Challas is successfully navigating a new job as Art Instructor this fall at West Valley College in Saratoga, CA.

AUTHORS

Ethan Berkove has been interested in folding things ever since he was 5 years old, when a family friend taught him to fold the Japanese crane. He eventually also developed an interest in mathematics, and received his Ph.D. from the University of Wisconsin, Madison in 1996. He has very much enjoyed finding a topic that brings both of these interests together. Although his field of research is algebraic topology, he is also interested in mathematical recreations and applications of mathematics. This article grew out of an independent study Jeff Dumont worked on during his junior and senior years at Lafayette College.

Jeffrey Dumont received his B.S. degree in mathematics from Lafayette College in 2001. After a brief stop teaching high school mathematics, he is now employed as an international meeting planner. Currently, his field of interest is the analysis of baseball box scores with a specialty in Red Sox historical trivia and that "infamous" number 1918. While not working, he loves being outside with his wife and dog.

James Alexander is Levi Kerr Professor of Mathematics at Case Western Reserve University, where he moved after many years at the University of Maryland. He also serves as Chair of the Department and Interim Chair of the Department of Statistics. His mathematical interests are eclectic, and not particularly selective, ranging from algebraic topology to various kinds of applied mathematics. As a result, he is often delighted by the stories, mathematical and otherwise, associated with the discipline.

David Finn is an associate professor of mathematics at Rose-Hulman Institute of Technology. He received his Bachelor's degree from Stevens Institute of Technology and his Master's degree and Ph.D. from Northeastern University. His mathematical interests lie at the intersection and interplay of analysis (differential equations), geometry, and physics. His nonmathematical interests are devoted to his family, specifically his wife and young son, and rooting for the Red Sox to finally win the World Series.

Vol. 77, No. 5, December 2004

MATHEMATICS MAGAZINE

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 *MATHEMATICS MAGAZINE* to an individual member of the Association is $131. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 20% dues discount for the first two years of membership.)

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 Frank Peterson (*FPetersonj@aol.com*), Advertising Manager, the Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

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

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

Printed in the United States of America

It's Okay to Be Square If You're a Flexagon

ETHAN J. BERKOVE
Lafayette College
Easton, PA 18042
berkovee@lafayette.edu

JEFFREY P. DUMONT
35 Forest St.
Claremont, NH 03743

It has been said that a mathematician can be content with only paper and pencil. In fact, there are times when one doesn't even need the pencil. From a simple strip of paper it is possible to make a surprisingly interesting geometric object, a *flexagon*. The flexagon can credit its creation to the difference in size between English-ruled paper and American binders. The father of the flexagon, Arthur Stone, was an English graduate student studying at Princeton University in 1939. To accommodate his smaller binder, Stone removed strips of paper from his notebook sheet. Not being wasteful, he creased these lengths of paper into strips of equilateral triangles, folded them in a certain way, and taped their ends. Stone noticed that it was possible to flex the resulting figure so that different faces were brought into view—and the flexagon was born [**4**]. Stone and his colleagues, Richard Feynman, Bryant Tucker, and John Tukey, spent considerable time cataloging flexagons but never published their work.

Like many geometric objects, flexagons can be appreciated on many levels of mathematical sophistication (the first author remembers folding flexagons in elementary school). With so adaptable a form, it is not surprising that flexagons have been studied from points of view that vary from art to algebra. Our interest in flexagons was sparked by a question posed in a paper by Hilton, Pedersen, and Walser [**9**]. They studied one of the *hexaflexagons*, so-named because the finished model has the shape of a hexagon. They calculated the group of motions for a certain hexaflexagon, then inquired about other members of the hexaflexagon family. We have determined that the trihexaflexagon is exceptional, as it is the *only* member of the hexaflexagon family whose collection of motions forms a group.

Working to generalize this result, we shifted our attention to *tetraflexagons*, which are constructed from strips of squares folded into a 2×2 square final form. We discovered that tetraflexagons are, if anything, more complicated and interesting than their hexagonal cousins. These results convinced us that tetraflexagons, often only mentioned in passing in the literature, deserve to be brought into the limelight. In this paper, we will summarize the results of our investigations. Although most of the material on hexaflexagons is known, the material on tetraflexagons includes new results and open questions. It is our intent to give interested readers enough material to start their own explorations of these fascinating objects.

The hexaflexagon family

It is easy to fold a flexagon, and we highly recommend making one of your own as this experience will be helpful in following the results in this section (and it's fun). Construct a strip of nine equilateral triangles and a tab as in FIGURE 1; this strip is called the *net* of the flexagon. Each triangle in the strip, and in general each polygon in a flexagon net, is called a *leaf* of the flexagon. You may want to label both sides of

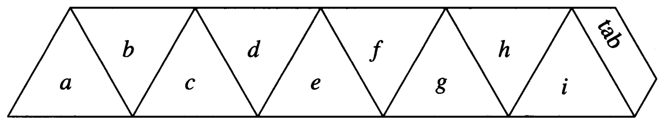


Figure 1 Trihexaflexagon net

each leaf and precrease all edges in both directions. Hold the leaf marked a in your hand, fold leaf c over leaf b, f over e, and i over h. Finish the flexagon by gluing or taping the tab onto leaf a. The final model should look like the flexagon depicted in FIGURE 2. Clockwise from the top, one can read off the leaves f, d, c, a, i, g. We call (f, d, c, a, i, g) a *face* of the flexagon.

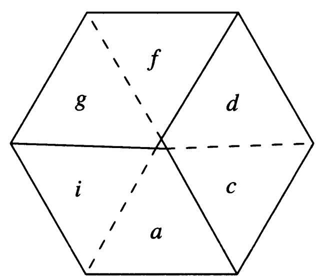


Figure 2 The (f, d, c, a, i, g) face of the trihexaflexagon

To flex your new creation, bring the three corners at the dashed lines down together so they meet. The hexagon will form a Y, at which point it will be possible to open the configuration at the middle. (This is the only possible way to perform a *flex-down* for this flexagon. There is also an inverse operation, a *flex-up*.) The result is a different face (f, e, c, b, i, h) of the flexagon. The flex can be repeated to get a third face (g, e, d, b, a, h), and one more flex returns the flexagon to (g, f, d, c, a, i), the original face rotated clockwise through an angle of $\pi/3$. Since it has three distinct faces, this flexagon is known as the *trihexaflexagon*—it is the simplest member of the hexaflexagon family. The three faces can be seen more easily if they are marked somehow: Wheeler **[15]** shows how to color the net so each flex brings out a new color, and Hilton et al. **[9]** give a way of marking the net so flexes bring out happy and sad pirate faces.

We would like a way to keep track of all the faces of a flexagon while we flex, which we can do using a graph: vertices represent the faces of the flexagon, and an edge joins two vertices if there is a flex that takes one face of the flexagon to the other. On occasion we will use a directed graph, where an arrow points towards the face that is the result of a flex-down. We choose to ignore the orientation of a face in the graph as this has a tendency to make the graph overly complicated. The completed graph is called a *structure diagram.* The cycle in FIGURE 3 is the structure diagram for the trihexaflexagon. It shows the three distinct faces, as well as their relationship via flexing.

The flex-down we described is a *motion* of the flexagon, a transformation that takes one hexagonal face of the flexagon to another hexagonal face. We require that our flexagons have no faces containing loose flaps that can be unfolded or moved so the hexagonal shape is lost. This becomes a significant issue as the number of triangles in the net increases. Indeed, in larger nets it is increasingly likely that a random folding of the net yields a face containing a loose flap, which in turn causes the entire flexagon to fall apart into a Möbius band with multiple twists. Therefore, we only consider flexagons that are folded in such a way that every flex is a motion.

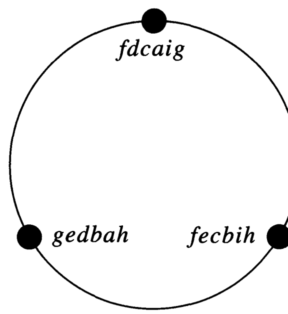


Figure 3 The structure diagram for the trihexaflexagon: vertices denote faces, edges denote flexes between faces

To determine the structure of the set of motions, let f denote a flex-down (hence f^{-1} denotes a flex-up), and denote by τ a flip along the x-axis of the flexagon. Then $f^3 = \tau^2 = id$ and it is easy to confirm that $\tau f \tau = f^{-1}$. The group with this presentation is well known as S_3, the symmetric group on 3 letters, which has 6 elements.

This analysis ignores the fact that f^3 is not strictly the identity flex, but is instead a rotation through $\pi/3$ degrees. The complete set of motions of the trihexaflexagon, including rotations, is analyzed in **[9]**; the only difference from the argument above is that $f^{18} = id$, and the resulting group is D_{18}, the dihedral group with 36 elements.

Motions of the hexahexaflexagon We perform a similar analysis for the member of the hexaflexagon family with six faces. This *hexahexaflexagon* can be constructed from the net of 18 triangular leaves with a tab as in FIGURE 4. To create the flexagon, label all leaves front and back and precrease all edges as before. Fold leaf a under the rest of the strip. Then fold the edge between leaves c and d so that a and d are adjacent. Next fold the edge between e and f so that c and f are adjacent. Continue rolling the strip in this manner until o and r are adjacent. The finished roll should look like the initial strip for the trihexaflexagon with leaf b on the far left. Fold this like the trihexaflexagon, then tape the tab to a to complete the model.

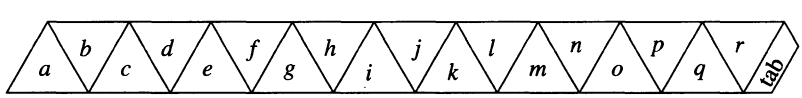


Figure 4 The hexahexaflexagon net

In contrast to the trihexaflexagon, from the initial face of the hexahexaflexagon either alternating set of corners flexes down. We can distinguish the two flexes by looking at what they do to the number of leaves in a triangular segment of the hexagon. Following Oakley and Wisner **[11]**, we call the entire triangular segment a *pat*. In our newly folded hexahexaflexagon, pats alternately contain 2 and 4 leaves, or (2, 4) for short. One of the flexes, which we call f, preserves thicknesses so is *pat-preserving*. The other flex, g, is *pat-changing*, from (2, 4) to (1, 5) and vice versa. The pat thicknesses are invariant under τ, a flip along the x-axis.

Starting with the initial face of the hexahexaflexagon, one finds that f^3 rotates the hexagon clockwise through an angle of $\pi/3$, so $f^{18} = id$. In addition, as $\tau f \tau = f^{-1}$, f and τ generate a copy of D_{18} in the collection of hexahexaflexagon motions. On the other hand, g leads to a face where the only possible flex is the pat-preserving f, which leads to a face where the only possible flex is the pat-changing g. The three-flex combination gfg rotates the hexagon clockwise through an angle of $\pi/3$, and g

satisfies $\tau g \tau = g^{-1}$. Therefore, there is at least one other copy of D_{18} in the motions of the hexahexaflexagon.

We turn this information into the structure diagram in FIGURE 5. Solid arcs between faces denote pat-preserving flexes, while dotted arcs denote pat-changing flexes. There is a primary cycle that can be traversed using pat-preserving flexes, and three subsidiary cycles that can be entered at faces where two flexes are possible. From the cycle structure we learn an important fact: the collection of motions of the hexahexaflexagon must have at least two generators, f and g. However, there are faces where only one of f and g can be applied, so it is not always possible to apply f or g twice in a row. In fact, g^3 does not even make sense. We now have an answer to the question posed in [9]: the hexahexaflexagon's motions do not form a group!

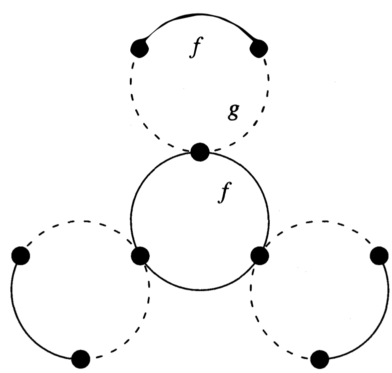


Figure 5 Structure diagrams for the hexahexaflexagon: dotted edges are pat-changing flexes, solid edges are pat-preserving flexes

This conclusion surprised us, as the collection of motions for every other geometric object we know of has a group structure. Furthermore, all hexaflexagons except the trihexaflexagon share this characteristic, as they contain pat-changing flexes. Pat-changing flexes occur at faces with two possible flexes, and the only hexaflexagon structure diagram without intersecting cycles is the trihexaflexagon's. Structure diagrams are necessarily finite, and results in Wheeler [15] imply that cycles cannot form a closed link. As a result, in any hexaflexagon there are only so many times the pat-changing flex can be applied. Thus, for every hexaflexagon with a pat-changing flex g there is a k such that g^k is undefined, which implies that the collection of motions for that hexaflexagon cannot form a group.

The astute reader might have noted that the structure diagram in FIGURE 5 seems to have nine faces, not six, as the hexahexaflexagon's name suggests. This discrepancy can be explained by carefully studying the hexahexaflexagon's faces. Upon closer inspection, three identical sets of triangles occur in two separate faces, but in different orders. If a face depends on both the triangles and their order, then indeed the hexahexaflexagon has nine faces. If order is disregarded, however, there are only six faces. The latter is the standard accounting, hence the name (although Oakley and Wisner [11] distinguish between the six *physical* faces and the nine *mathematical* faces).

We can construct other members of the hexaflexagon family by increasing the number of triangles in the initial net. For example, starting with a straight strip of 36 leaves, the strip can be rolled, then rerolled to yield the net in FIGURE 1, then folded as in the trihexaflexagon case to yield the dodecahexaflexagon. One can fold other members of the family by starting with nets that are not straight. The article by Wheeler [15] and Pook's book [14] contain some nice directions for folding the tetrahexa- and pentahexa- cases. There is also a HexaFind program [**3**], which generates all nets for hexaflexagons with any given number of faces.

We have only discussed a bit of what is known about hexaflexagons. In [**11**], Oakley and Wisner introduce the concept of the pat, then use it to count the total number of hexaflexagons with a particular number of faces. Madachy [**10**], O'Reilly [**12**], and Wheeler [**15**] describe connections between hexaflexagons and their structure diagrams. McIntosh [**7**], Madachy [**10**], and Pook [**14**] provide lengthy bibliographies to other work on hexaflexagons. Gilpin [**8**], Hilton et al. [**9**], and Pedersen [**13**] study the motions of the trihexaflexagon. Furthermore, the group of motions of the trihexaflexagon is identified in [**8**] and [**9**].

The distant cousins: tetraflexagons

We were introduced to tetraflexagons in a Martin Gardner *Mathematical Games* column in *Scientific American* [**6**]. We immediately noticed some differences; the direct analog of a strip of triangles, a straight strip of squares, makes for a poor flexagon; when you fold square over square you end up with a roll that does not flex at all (and has a trivial structure diagram). Therefore, we allow the nets to have right-angled turns. These turns occur at what we call *corner squares*, those attached to their neighbors on adjacent edges rather than opposite ones. Because so much is known about hexaflexagons, we felt that the tetraflexagon family would be readily analyzed. Our actual experience mirrored Stone's, as reported by Gardner [**6**]: "Stone and his friends spent considerable time folding and analyzing these four-sided forms, but did not succeed in developing a comprehensive theory that would cover all their discordant variations." However, we have established some notation and proved some results; our investigations have convinced us that tetraflexagons are as fascinating as, and more subtle than, their hexaflexagon relatives.

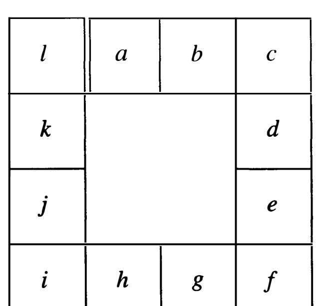


Figure 6 A "4 × 4 ring" net

Here is one way to fold a tetraflexagon using a 4 × 4 net with a cut adjacent to a corner square. Label a net as in FIGURE 6 and lay it flat. Mark the cut edges of a and l so you can tape them together after you have folded the tetraflexagon.

1. Fold leaf c over leaf d. Leaves a and b will remain to the left of c, but will flip over.
2. Fold c and d over e; the back of leaf c will touch the front of leaf e, and again, a and b are left free.
3. Fold f over g; move the flap with a and b so it points to the right.
4. Fold f and g over h so f and h are touching. Then slip the flap with a and b under j.

5. Fold j over i (without moving a and b), and flip the partially folded tetraflexagon upside down. Position the tetraflexagon so k and l are at the top of the model and a and b are upside down and facing to the left.
6. Fold i and j over k so i and k are touching.
7. Slide the flap with a and b under l, then fold a over l. Finally, tape a and l together on the right, along the edges that were originally cut.

The finished tetraflexagon will have 90-degree rotational symmetry, and should look like FIGURE 8a.

Now that you have a tetraflexagon in front of you, let's introduce some notation (see also FIGURE 7). The tetraflexagon has four pats, which conveniently look like the four quadrants in Cartesian coordinates. Therefore, we will refer to pats by the quadrant they are in: pat I, pat IV, etc. If you look carefully at your tetraflexagon, you'll notice that adjacent pats are connected by a layer of paper, called a *bridge*. The two leaves, one in each quadrant, that make up this layer are *bridge leaves*. Leaves b and c are bridge leaves, as are k and l. Bridges will figure prominently in our analysis.

Define a *pocket* as a strict subset of leaves within a pat attached to the rest of the tetraflexagon on two adjacent sides. Without the strictness condition, an entire pat is a pocket, connected to the rest of the flexagon by the two bridges; this is an extreme case we wish to avoid. On the other hand, a pocket may consist of a single leaf, in which case the pocket is a bridge leaf. As an example, squares f and g form a pocket in the tetraflexagon you just folded and this is the only pocket in its pat. The attached sides of the pocket are *connected*, whereas the other two sides are *free*. When a free side of a pocket lies along the outside edge of the tetraflexagon (like the outside edge of h and i), we call the entire tetraflexagon edge a *foldable edge*. The foldable edge will be folded in half during the flex.

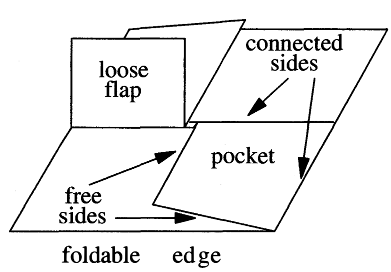


Figure 7 Tetraflexagon notation

There are two conditions that a tetraflexagon must satisfy in order to flex: there must be a pocket with a component bridge leaf, and the bridge in the pocket must be at right angles to the foldable edge. (Since the bridge goes from a pocket to an adjacent pat it cannot cross the pocket's free side.) To perform a flex, orient the tetraflexagon so the foldable edge is forward and the pocket lies on top of the tetraflexagon, as in FIGURE 8a. Fold the tetraflexagon in half perpendicular to the foldable edge so that the pocket remains on the outside. Put your thumbs into the pocket and its kitty-corner companion (FIGURE 8b), then pull outwards in the direction of the arrows. The pocket layers will separate from the rest of the pat, rotating outwards 180 degrees but staying in the same quadrant. The rest of the pat layers will maintain their orientation but move to the top of the adjacent pat. Simultaneously, the layers in the adjacent pats will rotate outwards 180 degrees. When you flatten the tetraflexagon, you will see a new face as in FIGURES 8c and 8d.

We call the flex with the pockets up and the foldable edge forward a *book-down flex*. It is only one of four possibilities, depending on the position of the pocket and foldable edge. We therefore distinguish between four types of flexes: flexes up or down along the y-axis (*book-up* and *book-down*) and flexes up or down along the x-axis (*laptop-up* and *laptop-down*). From FIGURE 8, we see that an up flex is the inverse of a down flex and vice versa. As before, the top layer of leaf of the tetraflexagon, like (a, d, g, j), is called a *face*, and a flex is a *motion* if it takes one tetraflexagon face to another.

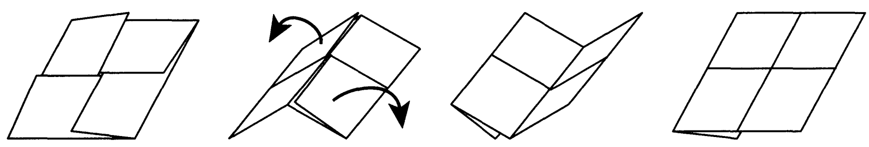


Figure 8 Flexing a tetraflexagon

Before we go further into our exploration, we introduce some assumptions about the tetraflexagons we consider to aid in our analysis.

1. *Net* assumption. All tetraflexagon nets have exactly four corner squares. The two vertical strips are each composed of m leaves, and the two horizontal strips of n leaves. All nets are folded into four-pat (2×2) tetraflexagons.
2. *Winding* assumption. There is only one bridge between adjacent pats of the tetraflexagon. In other words, as the net is folded, the strip of paper enters and exits each quadrant exactly once. This assumption is easily justified; we will see in the Flexing Lemma below that subsequent flexes maintain all four bridges as single layers.
3. *Rigidity* assumption. No tetraflexagon face contains a loose flap. A *loose flap* is a collection of two or more leaves connected to the rest of the tetraflexagon on only one side (see FIGURE 7). Since a loose flap determines one of two possible tetraflexagon faces depending on its position, movement of a loose flap can be considered a motion. The rigidity assumption ensures that the only tetraflexagon motions are flexes.
4. *Symmetry* assumption. Folded tetraflexagons have 180-degree rotational symmetry (in contrast to the hexaflexagons' 120-degree symmetry).

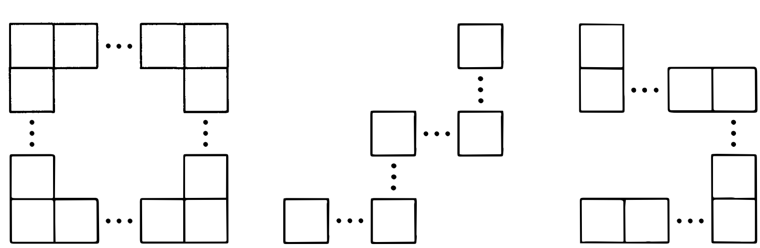


Figure 9 Rings, bolts, and snakes

The net assumption implies that nets are one of three shapes, depending on parity. All nets are shown in FIGURE 9: when m and n are both even, the only possible net is a *ring*; when both m and n are odd, the only possible net is a *lightning bolt*; and when exactly one of m and n are odd, the only possible net is a *snake*. We remark that analogs of the other three assumptions hold for most hexaflexagons.

Basic analysis: folding 101 and structure diagrams with dead ends Flexing for tetraflexagons is a bit more complicated than for hexaflexagons. We carefully describe what happens during a flex in the following

FLEXING LEMMA. The following hold for tetraflexagons under our assumptions:

1. Corner squares remain in their pats during flexes.
2. A pocket that has a component bridge leaf is either a single layer (that is, it is the bridge leaf), or consists of every leaf in the pat except for either the top or bottom one. In the latter case, the leaf not belonging to the pocket is also a bridge leaf.
3. A flex maintains all four bridges as single layers.

Proof. To simplify the arguments, we will assume for this proof that the pocket is always on top of pat IV with the foldable edge forward, as in FIGURE 8a. The other cases will follow by rotation and mirror images. Moreover, by the symmetry assumption we only need focus on pats III and IV.

We start with the first claim. When we perform the flex shown in FIGURE 8, all leaves in pat III remain in pat III, which means that the only corner square that can change pats lies in pat IV. However, if the book-down flex moves this corner square to pat III, then the flex is also forced to move the strip of leaves connecting the corner squares in pats I and IV. This is impossible as the corner square in pat I is fixed by the symmetry assumption.

For the second claim, assume that the pocket with a bridge leaf consists of more than one layer. We claim that the corner square in pat IV must be a leaf of the pocket. By way of contradiction, assume the corner square is below the pocket. Notice that the corner squares in pats I and IV are connected by a vertical strip. As this strip is woven from the corner square in pat IV to the corner square in pat I, it must at some point become the unique bridge to pat I (by the winding assumption). One of the bridge leaves is a leaf of the pocket, so the vertical strip of squares must also cross from the bottom of pat IV to the pocket. Because the vertical strip is woven back and forth, it can only cross to the pocket on the front or back edge of pat IV. The vertical strip cannot cross the gap at the front edge of pat IV because that edge is free by assumption. The vertical strip cannot cross the gap at the back edge of the pocket either, as that would lock the pocket to the rest of pat IV, making a flex impossible. We conclude that the corner square must be a leaf of the pocket.

An analogous argument shows that the vertical strip cannot cross from the corner square to the bottom layer either. By the rigidity assumption, this bottom layer must therefore consist of a single leaf. Now if the bridge leaf to pat III were any layer but the bottom, it would force the left side of the pocket in pat IV to be connected, rather than free. This is impossible as all pockets have two free sides.

For the final claim, the winding assumption implies that, when first folded, the tetraflexagon has single layer bridges between pats. Consider the effect of a book-down flex. This flex flips the bridge leaves that lie in the "pages" of the book upside down, so those bridges remain single layers. For the two bridges that cross the spine of the book, focus on pats III and IV, and assume first that the pocket is a single layer; then the book-down flex leaves a single layer in pat IV, which must be the bridge leaf. Otherwise, the pocket has more than one leaf and the bottom leaf of pat IV is a single layer. In this case, the completed flex moves the single layer from pat IV to pat III, again maintaining the bridge as a single layer. ■

The Flexing Lemma gives us insight about what happens to a tetraflexagon during a single flex. What about the larger picture? What can we say about the set of all the motions of the tetraflexagon we folded? Again, we answer these questions with a structure diagram, but now use the orientation of the graph's edges to distinguish among the four possible types of flexes. Denote a book flex by a horizontal line segment in the structure diagram, a laptop flex by a vertical one, and let arrows point to the faces that are the result of a flex down. With a little flexing and experimentation, we find that the structure diagram for the tetraflexagon we folded is an L shape, with five vertices (corresponding to faces), and four edges, two in each part of the L (corresponding to flexes). On all edges, the arrows point away from the corner of the L. This structure diagram, which appears in the bottom left of FIGURE 11, has a feature we have not seen before: there are vertices incident to a single edge. We call the faces associated to these vertices *dead ends*, because once these faces are reached via a flex, the only motion that can be applied is the flex's inverse.

For the hexaflexagon cases, a flex of finite order appeared in the structure diagram as a cycle. The structure diagram we just constructed has no cycles, hence no elements of finite order. In addition, any flex can be applied at most twice, at which point some other flex must be applied. In other words, there is really no group structure at all in the motions of the tetraflexagon we constructed. An immediate question is whether there is a tetraflexagon whose collection of motions forms a group.

There is, and it can be folded from the 3×3 lightning bolt net shown in FIGURE 10. Start with a copy of this net—you may want to specially mark the shaded sides of a and h as shown in the diagram, as those are the edges that get taped together at the end. Fold leaf b over c, d (and a, b, and c) over e, and g over f. Make sure h lies on top of a, then tape a and h together on the right, along their shaded sides. Your finished tetraflexagon should have 180-degree rotational symmetry, as per the symmetry assumption. From this initial face, the only possible down flex is a laptop-down. If this is followed with book-down, laptop-down, and book-down flexes (again, the only possible down flexes) you get back to the initial face. The resulting structure diagram is a square cycle, corresponding to the symmetry group $\mathbb{Z}/4\mathbb{Z}$, the cyclic group with 4 elements. We believe that this is the *only* tetraflexagon whose collection of motions forms a group (although a rigorous proof eludes us).

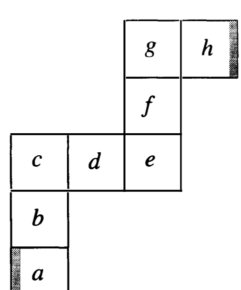


Figure 10 A 3×3 bolt net

You may have noticed that although we call the net in FIGURE 10 a 3×3 lightning bolt, there are only two leaves, g and h, in one row. Actually, that row does contain a third leaf, a, which becomes part of the row after we finish taping. In general, when we refer to a net as m by n, we include the corner squares in the count. For the cases of the bolt and snake nets, this means that either one row or one column will be a leaf short when the net is first constructed.

By reducing larger nets to the 3×3 case, one can construct many tetraflexagons whose structure diagrams contain cycles. For example, starting with the net in FIGURE 6, folding b over c and i over h turns the 4×4 ring net into a 4×3 snake net.

Another pair of folds results in a 3×3 lightning bolt net, which can then be folded as above. More generally, one can turn an $m \times n$ net into either an $m - 1 \times n$ net or an $m \times n - 1$ net by making appropriate folds. Here, appropriate means that the pair of folds results in a net that satisfies the symmetry assumption. Also, a given sequence of edge choices must result in a tetraflexagon that satisfies our four assumptions. Once the net is a 3×3 lightning bolt, it is folded to guarantee at least one cycle in the structure diagram. This process is reminiscent of how a hexaflexagon is folded from a strip containing $9(2^n)$ triangles by rolling it n times to form the trihexaflexagon net with 9 triangles. A difference, though, is that there are many ways of choosing edges in pairs to reduce an $m \times n$ net to the 3×3 net, so a generic tetraflexagon net can yield a large number of tetraflexagons with distinct structure diagrams.

As an example, through exhaustive folding, both figurative and literal, we have identified in FIGURE 11 most (all?) of the possible structure diagrams that result from a 4×4 ring. Some of the diagrams are very basic, but notice that three of them contain cycles, all but one contain dead ends, and two contain both. This variety in structure diagrams from the same net demonstrates how difficult it is to classify even simple tetraflexagons.

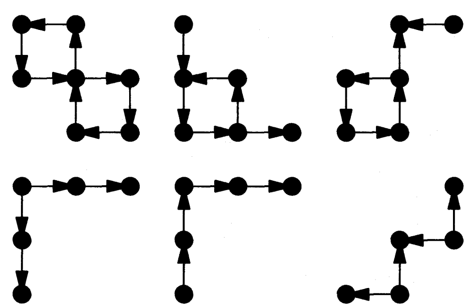


Figure 11 Tetraflexagon structure diagrams folded from a 4×4 ring

Despite their "discordant variations," the tetraflexagons we investigated have some features in common. In a number of the examples we folded, we came across faces where all four flexes were possible. We call such faces *crossroads*. For the reader who wants to see a crossroad in action, start with the net in FIGURE 6, labeled front and back with each letter. Fold b under c, c over d, e (and a–d) under f, g (and the rest of the net) over f, i over h, j over i, and l over k. Lift a over l, then tape a and l together on the right (outside) edge. The final tetraflexagon is shown in FIGURE 12d. (What is the resulting structure diagram?) Crossroads are among the most interesting features in a structure diagram, and we wondered how many crossroads a structure diagram could contain. We noticed that after we flexed a crossroad, we never saw another crossroad, which led us to prove the

CROSSROAD THEOREM. Crossroad faces are never adjacent in tetraflexagons built from ring, lightning bolt, or snake nets.

Proof. Let's analyze the shape of a tetraflexagon at a crossroad face. In order to perform all four types of flexes, every edge of the tetraflexagon must be a foldable edge and there must be two pockets per edge to allow both types of flexes. It is easy to show that the pockets cannot be on the same side of the foldable edge, as in FIGURE 12a. If the pockets were on the same side, the remaining leaves of the two pats would be the bridge between the pats, by the second claim of the Flexing Lemma. By the symmetry assumption, the same would be true for the other two pats of the tetraflexagon, and the entire tetraflexagon would simply fall apart.

We conclude that one pat along the foldable edge must have a pocket on the top of the pat to allow a flex down, while the adjacent pat along the edge must have a pocket on the bottom of the pat to allow a flex up. There are essentially two cases to consider, shown in FIGURES 12b and 12c (FIGURE 12d is FIGURE 12c's mirror image). We can show that FIGURE 12b is impossible by a careful analysis of pat IV. Since one of the two pockets in pat IV contains all the layers but one, assume, without loss of generality, that it is the pocket associated to the foldable edge on the right. Then the bridge to pat III crosses the edge marked by the arrow, and the component bridge leaf in pat IV is one of the leaves in the pocket by the second claim of the flexing lemma. However, this implies that the edge marked by the arrow is a connected edge, which makes a book-down flex impossible. Therefore, the only possible crossroad faces have configurations like FIGUREs 12c and its mirror image 12d.

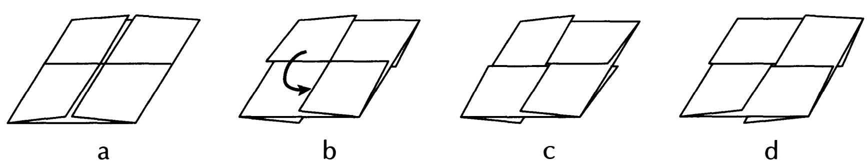


Figure 12 Two impossible configurations and two crossroads

We next determine when a flex from a crossroad yields another crossroad. By symmetry arguments, it is sufficient to consider FIGURE 12c. We claim that is impossible to perform a book-down flex from FIGURE 12c and get the crossroad in FIGURE 12d. This time, consider pat III. Note that by the second claim of the Flexing Lemma one of the two pockets in pat III must consist of every layer in the pat except one; assume (by symmetry again) that it is the bottom pocket. After a book-down flex is applied, the original layers in pat III are flipped upside down, and are covered by a single layer from pat IV. This single layer must be part of the new bridge to pat IV. Therefore, it is not possible for the right edge of pat III to be free, as in FIGURE 12d.

Thus, if two crossroads are connected by a flex, they must be identical. We rule out this possibility too. Consider the enlargement of pat IV of FIGURE 12c, shown in FIGURE 13. In order for a book-down flex to yield the same configuration, pat IV must contain a pocket as part of the bridge. But the bridge must be a single layer, by the second claim of the Flexing Lemma. ■

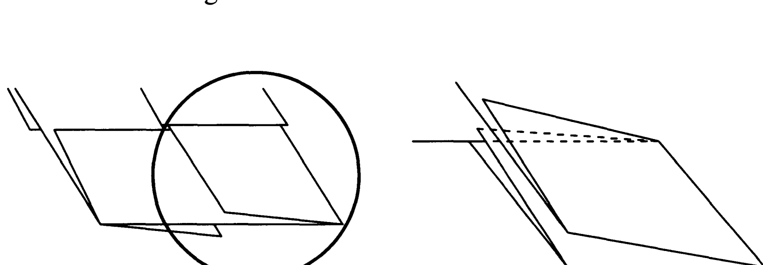


Figure 13 Closeup of pat IV

The Crossroad Theorem tells us that cycles cannot be too dense in a structure diagram. For example, we have a corollary to the theorem, named after the symbol that appears on all Purina pet foods.

PURINA COROLLARY. Under our assumptions, there is no "Purina Tetraflexagon" with the structure diagram shown in FIGURE 14.

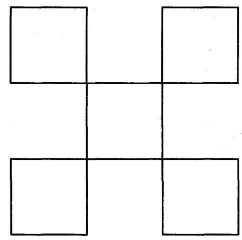


Figure 14 Purina structure diagram

Surgery: adding new parts to an old flexagon So far, we have a very general idea of what can and cannot be part of a structure diagram. On the other hand, it would be desirable to add a given component to a structure diagram, allowing us to tailor-make tetraflexagons with interesting properties. There is a technique we learned from Harold McIntosh's online notes [7] that allows us to do exactly that at an appropriate spot in a structure diagram. We call this technique *surgery*. One performs surgery on a tetraflexagon by symmetrically grafting two strips of squares to the model to add the new feature.

An instance of surgery is shown in FIGURE 15. Start with a tetraflexagon at a face where one of the pats is a single leaf, that is, a corner square. Place this tetraflexagon so the single leaves are in pats II and IV with the foldable edge forward, as in the top left picture in FIGURE 15. By considering mirror images, if necessary, we may assume that the pocket in pat III is on the bottom of the pat. Tape a 2×1 strip to the right edge at pat IV and make a cut in the tetraflexagon along the positive x-axis as in the top right picture. Fold the strip on top of pat IV, then fold the strip up and in half as in the bottom left picture. Finally, tape the top edge of the strip (in pat IV) to the cut edge in pat I as in the bottom right picture. Perform the same procedure symmetrically

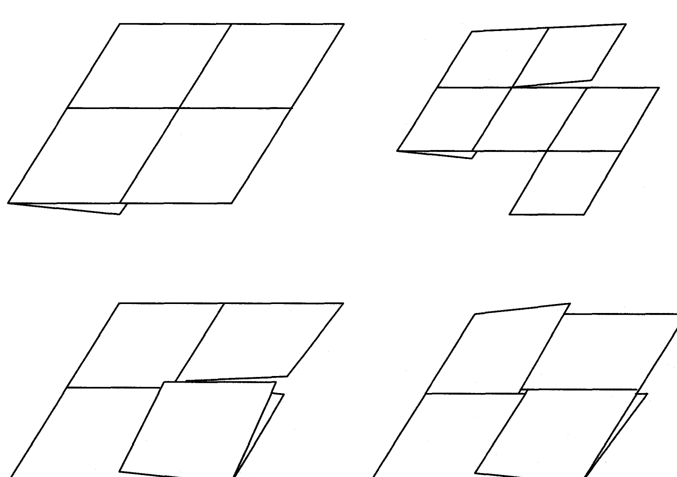


Figure 15 How to add a cycle to a tetraflexagon

on pat II; when you are done, the front and back foldable edges will have pockets on both top and bottom. The resulting tetraflexagon's structure diagram will be the same as before, but with a cycle tacked on at the appropriate face.

Following this technique, it is straightforward, at least in theory, to build a tetraflexagon whose structure diagram consists of an arbitrarily long row of four-cycles connected one to the other at opposite corners. (In light of the Crossroad Theorem, in some sense this is as crowded as collections of cycles can become.) In practice, the tetraflexagons quickly become too thick to flex easily as the size of the net increases. We challenge the reader to use surgery to construct a tetraflexagon with the structure diagram shown in FIGURE 16.

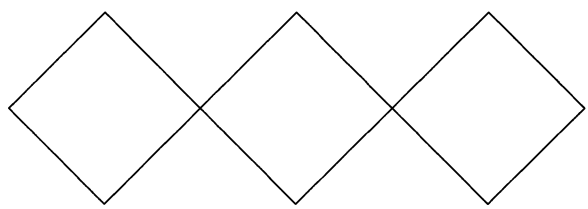


Figure 16 A challenging structure diagram

One can also use surgery to add dead ends to a structure diagram. Start with a tetraflexagon in the same position as before and make the same cut along the positive x-axis, but this time add a $1 \times n$ strip horizontally to the right edge of pat IV. Roll the strip counterclockwise until it lies on top of pat IV, then tape the top edge of the middle square of the roll to the cut edge of pat I. Repeat this symmetrically on pat II. This procedure will add n horizontal segments to the structure diagram at the appropriate place.

To finish this section we mention that besides appearing in Gardner articles [**5**] and [**6**], "square" flexagons are the topic of a short note by Chapman [**2**] where he uses primary and secondary colors to distinguish tetraflexagon faces. This note also contains directions for constructing tetraflexagons whose structure diagrams are a cycle and two linked cycles. For more information on surgery and tetraflexagons, the reader should consult McIntosh's notes [**7**] or Pook's book [**14**].

Where do we go from here?

At this point, we know something about tetraflexagons—how they flex, how and where cycles can occur, and why some structure diagrams are impossible—but there are many questions we haven't answered. What are the possible tetraflexagon structure diagrams? How many tetraflexagons, up to rotation and mirror image, can be made from an $m \times n$ net with four corners? We can get a rough upper bound by recalling that the corner squares in a tetraflexagon stay fixed in their quadrant. Flex a tetraflexagon so that one pat consists of a single leaf and its adjacent pats contain $n + m - 1$ leaves. With the exception of the bottom (or top) leaf, these might be in any order, so there are at most $(n + m - 2)!$ possible tetraflexagons that can be folded from a given $m \times n$ net. Many of these tetraflexagons will not satisfy our four assumptions; is there a better upper bound? Finally, there is the question that provided our initial motivation to study flexagons: Are there any other tetraflexagons besides the one built from the net in FIGURE 10 whose collection of motions forms a group?

We developed most of our tetraflexagon results before looking through McIntosh's material [7], and you can imagine our surprise when we saw directions for the Purina Tetraflexagon in his notes! McIntosh introduced this creature in a discussion on surgery, and we were no less surprised when we constructed it and confirmed that it worked. The apparent discrepancy between beautiful theory and ugly counterexample was explained once we took a scissors to the flexagon and opened it up; in the net, every one of the 24 leaves was a corner square! Clearly, in order to do surgery while maintaining the net assumption, one must be careful that the transplanted leaves do not add any more corner squares.

More importantly, this example shows that there are many interesting tetraflexagons that do not satisfy the net assumption. In fact, perhaps our favorite flexagon violates both the net and symmetry assumptions, as it is folded from regular pentagons. Of course, this flexagon cannot be folded flat, but we were impressed that such an object could exist at all. This flexagon, and many others, are described in McIntosh's notes [7] as well as Pook's book [**14**]. We highly recommend these references, as well as an impressive report by Conrad and Hartline [**1**] (also found in [**7**]) for the reader interested in deepening his or her background about flexagons and their kin.

We have found flexagons to be interesting mathematical objects at many levels. Since they are easily folded, they are a good addition to an introductory mathematics class, giving students an opportunity to look for patterns and explore a topic at their own pace. In a more advanced setting, such as a course in geometry or algebra, flexagons can be used to introduce notions of symmetry and transformation. They are also an excellent topic of study for a senior thesis, as most of the flexagon materials are written at an elementary level and are appropriate for a student's introduction to the reading of mathematical articles. And, of course, there are many open questions that are easily asked but difficult to answer. Despite their long history, flexagons still give an exciting twist to an otherwise boring strip of paper, and are well worth a little study.

Acknowledgments. We would like to thank Harold McIntosh for informative correspondence, and Liz McMahon, Kyra Berkove, and the anonymous referees, whose comments greatly improved the exposition of this paper.

REFERENCES

1. Anthony Conrad and Daniel Hartline, Flexagons, *RIAS Technical Report 62-11*, Baltimore, MD, 1962.
2. P. B. Chapman, Square flexagons, *Math. Gaz.*, **45**, (1961), 192–194.
3. Antonio De Queioroz, The HexaFind Program, http://www.coe.ufrj.br/~acmq/hexaflexagons.
4. Martin Gardner, *Hexaflexagons and Other Mathematical Diversions*, University of Chicago Press, Chicago, 1988.
5. ———, *The 2nd Scientific American Book of Mathematical Puzzles and Diversions*, Simon and Schuster, New York, 1961.
6. ———, Mathematical games, tetraflexagons and tetraflexigation, *Sci. Amer.*, May 1958, 122–126.
7. Harold McIntosh, Flexagons, http://delta.cs.cinvestav.mx/~mcintosh/oldweb/pflexagon.html.
8. Michael Gilpin, Symmetries of the trihexaflexagon, this MAGAZINE, **49**:4 (1976), 189–192.
9. Peter Hilton, Jean Pedersen, and Hans Walser, The faces of the trihexaflexagon, this MAGAZINE, **70**:4 (1997), 243–251.
10. Joseph Madachy, *Mathematics on Vacation*, Charles Scribner, New York, 1966.
11. C. O. Oakley and R. J. Wisner, Flexagons, *Amer. Math. Monthly*, **64**, (1957), 143–154.
12. Thomas O'Reilly, Classifying and counting hexaflexagrams, *J. Recreational Math.*, **8**:3, (1975), 182–187.
13. Jean Pedersen, Sneaking up on a group, *The Two Year College Math. J.*, **3**: 2, (1972), 9–12.
14. Les Pook, *Flexagons Inside Out*, Cambridge University Press, Cambridge, 2003.
15. Roger Wheeler, The flexagon family, *Math. Gaz.*, **42**, (1958), 1–6.

Loxodromes: A Rhumb Way to Go

JAMES ALEXANDER
Case Western Reserve University
Cleveland, OH 44106
james.alexander@cwru.edu

A rhumb is a course on the Earth of constant bearing. For example, to travel from New York to London a voyager could head at a constant bearing 73° east of north. *Loxodrome* is a Latin synonym for *rhumb*, and has come to be used more as a geometric term—the course is a rhumb, the curve is a loxodrome. On a surface of revolution, *meridians* are copies of the revolved curve; on the earth, they are north-south lines of constant longitude. A loxodrome intersects all the meridians at the same angle. A circle of constant latitude is a loxodrome (perpendicular to meridians). Any other loxodrome can be continued forwards and backwards, winding infinitely often around the poles in limiting logarithmic spirals, as in FIGURE 1.

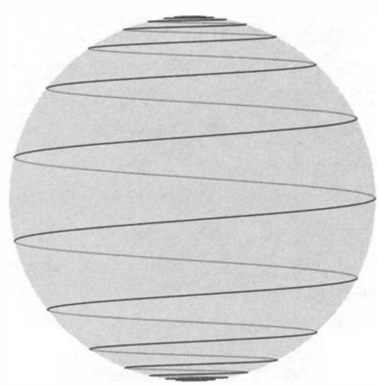


Figure 1 Loxodrome on a semi-transparent sphere. Compare also M. C. Escher's "Sphere Surface with Fish" and "Sphere Spirals" [**3**]. All loxodromes, except circles of latitude, look like this, differing only in the slope. Of course, for navigation, only a portion of the full loxodrome is relevant.

A rhumb course is not generally a great circle, and thus not the shortest route from one point on the Earth to another. However, a great-circle course requires continually resetting the bearing, an impossible task in the early days of ocean navigation. Usually a navigator would approximate a great circle by a series of rhumbs. Even today, with the availability of global positioning system (GPS), sailors and airplane pilots must know about rhumbs.

The history of loxodromes goes back to the days when voyagers first realized that the Earth is not flat and they had to take the curvature into account. They had to develop the mathematics of loxodromes. A major development was the introduction of the Mercator map projection—a rhumb is a straight line on a Mercator map. Here we consider three related mathematical problems:

I. Constructing the Mercator projection,

II. determining the rhumb heading from one location to another, and

III. computing the rhumb distance between two locations on the earth.

Some history

Economically, spice was the oil of the early Renaissance in Europe [**1**]. Economies depended on the spice trade. The Dutch and British East India Companies, incorporated for the spice trade, were the Shell and British Petroleum of their times. Spices were used for food preservation and preparation (a major consideration before refrigeration), for medicine, for personal grooming, and for conspicuous consumption. Originally, spices were transported from various points in Asia overland or along the coast of the Arabian Sea, and through the Middle East and Egypt. There were lots of middlemen (and women—in 595, Muhammad married a spice trader) along the route; each one took his (or her) cut, and prices were whatever the market would bear. A small satchel of smuggled pepper could financially set a man up for life. Small wonder that the Iberian countries, at the very end of this long route, took the lead in exploring alternative routes to the Far East.

A water route was a natural for these countries. But navigating the open Atlantic was a different ballgame than the Mediterranean. Perhaps the only certainty about the Atlantic was that it was home to sea monsters. But the economics were compelling, and navigation became a science. In the 1500s, the technology of navigation was cutting-edge for the countries of Europe. Charts, of course, could be developed only for known areas. Less than 100 years earlier, the Canaries and Azores had been *terra incognita.* There was a mystic island of Brazil somewhere out there. The Portuguese had, in a succession of voyages, sailed down the west coast of Africa, eventually losing their view of the North Star, requiring new techniques for determining latitude. Christopher Columbus had traversed the Atlantic in 1492 (but not to the spices of India), and in 1501, Amerigo Vespucci sailed down the east coast of South America. Sometime after, Portuguese sailors made south Atlantic crossings, and realized that the traverse was shorter than charts with equally-spaced meridians indicated.

Much like today, governments acted to protect and enhance technologies important to their economies. In the 1500s, Portugal was a major seafaring country. The Portuguese king prohibited the use of (newly high-tech) globes for navigation, presumably as an export control, to prevent them from falling into foreign hands. In 1534, a Chair of Mathematics was inaugurated at the University of Coimbra, underwritten with the intent of bringing mathematical methods to navigation. The polymath Pedro Nunes (1502–1578) was appointed to the new position. Nunes (pronounced noo′ nush) originally took a degree in medicine, but then held chairs in moral philosophy, logic, and metaphysics at the University of Lisbon before moving to Coimbra. He wrote treatises on mathematics, cosmography, physics, and navigation. He was appointed Royal Cosmographer in 1529 and Chief Royal Cosmographer in 1547.

Navigators came to realize in the early 1500s that a course of constant bearing is not the same as a great circle. A (probably apocryphal) tale is that Nunes was introduced to the issue by a sea captain complaining that when he set out, say, eastward, and tied the rudder to hold a straight course, he would find his ship turning towards the equator [**12**]. In 1537, Nunes published two treatises analyzing the geometry of such courses, followed by a synthesis and expansion in 1566 in Latin [**11**], where he first used the word *rumbo.* Nunes's work was controversial. Indeed, his analysis met opposition. At one point, claiming calumny on the part of his critics, he wrote, "I have decided, for this reason, to polish up some of the things I have written and set about studying philosophy and abandon mathematics, in the study of which, I have irretrievably lost my health" [**12**, translated]. The word *loxodrome* (Greek *loxos* = *oblique*, *dromos* = *bearing*) is a 1624 Latinization by Willebrord Snell (Snel) (1580–1626) [**15**] (of Snell's law in optics) of the Dutch word *kromstrijk* (curved direction), used by Simon Stevin in his description of Nunes's work.

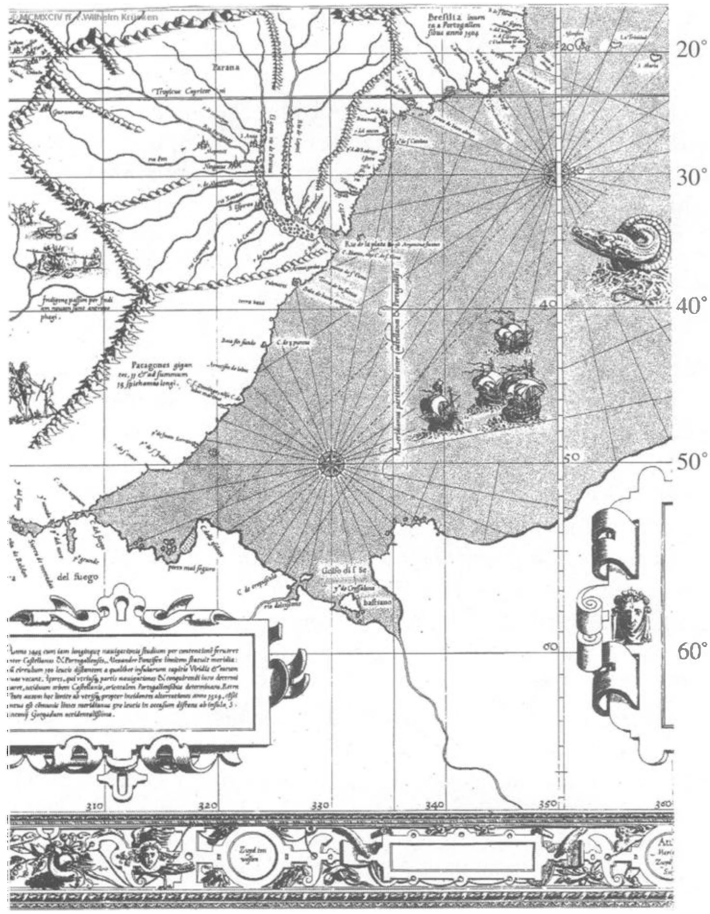


Figure 2 A detail of Mercator's 1569 map, showing a portion of South America, including latitude lines for 20 through 60 degrees South. Although the concept of "increasing latitudes," that is, the uneven spacing of the parallels of latitude, was developed early in the 1500s, Mercator figured how to space the latitudes so that a loxodrome appears as a straight line. For a complete map and other details, see [**9**]. ©F. Wilhelm Krücken, used with permission.

Another major development was the construction in 1569 by Gerhardus Mercator (Latinized from Gerhard de Cremer) of his world map [**9**]. His map, a portion of which is shown in FIGURE 2, had the great virtue that a straight line is a rhumb and (not coincidentally) angles on the map equal angles on the earth. To set a course from one location to another, a navigator drew a straight line on the map, determined the bearing, and set off. The Mercator projection became the standard for navigation for centuries, and, perhaps unfortunately given its distortions [**5**], also for atlases, wall maps, and geography books. The mathematical problem, the "true division of the nautical meridian"—the exact spacing of the meridians—remained into the next century [**8, 10**]. (Side note: The Inquisitions were in full bloom in the 1500s, and affected both Nunes and Mercator. Many Jews in Spain were converted to Roman Catholicism; such people were called *conversos*. Nunes was converted as a child. The primary targets of the later Spanish inquisition were descendants of conversos, who were persecuted under the vague charge of *judaizing*. According to one authority, this happened to Nunes's grandsons in the early 1600s [**14**, p. 96]. Mercator was (evidently falsely) accused of *lutherye* in an inquisition originated by Queen Maria of Hungary and spent seven months in 1544–45 imprisoned in the castle at Rupelmonde, narrowly escaping execution [**2**, chap. 15]. He moved to Duisburg in 1552 to escape further persecution.)

The importance of navigation, and hence of these problems, in the 16th and 17th century is hard to overestimate. Geometry was premier applied mathematics. At the time, calculus was not available. Today, these problems are straightforward applications of calculus to geometry and navigation.

Calculations

Mercator mathematics On a Mercator map, meridians are vertical lines. They do not converge at the poles. Although, as is well known, the Mercator map distorts distances, it does not distort angles. At any point, it distorts east-west directions exactly the same as north-south directions—the distortion factor depends only on position, not on direction. Such a map is called *conformal* [**4**]. Moreover, the distortion depends only on the latitude, not on longitude. A line of longitude, or meridian, is of course half a great circle, running between the north and south poles. On the other hand, latitudes are parallel circles, but shrinking in radius away from the equator. On a spherical Earth, the parallel of latitude L (in degrees north or south of the equator) is shrunk by the factor $\cos L$ compared to the equator. The equator has length $2\pi R$, where $R \approx 6371$ km ≈ 3959 mi is the mean radius of the earth. Inversely, at latitude L, distances on the Earth are stretched by the reciprocal $\sec L$ on the Mercator map.

More generally, on any surface of revolution, let $\sigma(L)$ denote the *local stretching factor* at latitude L, the reciprocal of the amount that parallels of latitude L are shrunk, compared to the equator. On a spherical Earth, $\sigma(L) = \sec L$. Some geodesists' computations are more exact, and take into account the fact that the earth is more an ellipsoid than a sphere. In this case, R is set equal to the equatorial radius 6378 km $\approx$ 3963 mi, and

$$\sigma(L) = \frac{(1 - e^2)\sec L}{1 - e^2 \sin^2 L},$$

where the eccentricity $e \approx .081$ specifies how far off from spherical the earth is. Here L is the *geodetic latitude*, the complement of the angle of the perpendicular to the surface with the axis of the earth.

Let $\Sigma(L) = \int_0^L \sigma(\ell)\, d\ell$ denote the *total stretching* from the equator to L, as in FIGURE 3. Here and below, all angles in the analysis are represented in radians. The total stretching solves problem I. The east-west scale on the equator sets the scale for the map. The parallel at latitude L is placed at north-south distance $\Sigma(L)$ from the equator. If a standard schoolhouse ruler is laid N-S on a Mercator map, the distance markings on the ruler are proportional to Σ. Thus Σ serves as a linear coordinate on a Mercator map.

Rhumb directions The Mercator mathematics of the previous section provides the basis for determining rhumb courses. Thus, if λ denotes longitude, the east-west coordinate, a straight line on a Mercator map has the form

$$\Sigma = m\,(\lambda - \lambda_1) + b. \tag{1}$$

Since the Mercator projection is conformal and meridians are parallel, such a line corresponds to a curve of constant bearing on the Earth, that is, a loxodrome. Thus, the rhumb line between two points of longitudes λ_1 and λ_2 and latitudes L_1 and L_2 has

$$m = \frac{\Sigma_2 - \Sigma_1}{\lambda_2 - \lambda_1}, \qquad b = \Sigma_1.$$

In particular, the bearing from location 1 to location 2 is (one of the two values of)

$$\theta = \operatorname{arccot} \frac{\Sigma_2 - \Sigma_1}{\lambda_2 - \lambda_1} \text{ (degrees)} \tag{2}$$

from north, solving problem II.

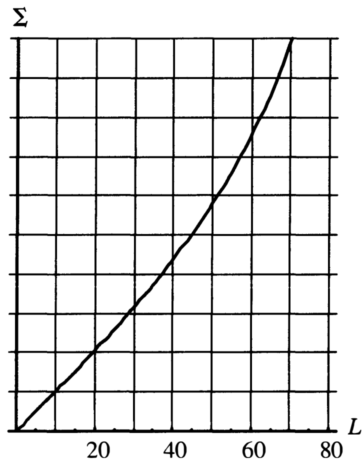


Figure 3 The graph of $\Sigma(L)$, solving problem I. The horizontal axis is measured in degrees and the vertical axis is scaled accordingly. As $L \to 90°$, $\Sigma \to \infty$ and the parallels of latitude spread apart more and more. Edward Wright [**16**] realized Σ was obtained by "perpetuall addition of the Secantes," and created a table using increments of one minute of arc, essentially amounting to numerical integration [**8, 13**]. The identification of an explicit expression for Σ (equation (6)) has a story of its own [**13**].

Rhumb distances In a Riemannian geometry, the differential of arc length, ds, is integrated over a rectifiable curve to obtain its length. By considering small (infinitesimal) right triangles, with sides aligned with latitudes and longitudes, one finds that, in terms of latitude and longitude, the square differential of arc length is

$$\left(\frac{ds}{R}\right)^2 = dL^2 + \left(\frac{d\lambda}{\sigma(L)}\right)^2.$$

Thus

$$\frac{ds}{dL} = R\sqrt{1 + \frac{1}{\sigma^2(L)}\left(\frac{d\lambda}{dL}\right)^2} = R\sqrt{1 + \frac{1}{\sigma^2(L)}\left(\frac{d\lambda}{d\Sigma}\right)^2\left(\frac{d\Sigma}{dL}\right)^2} = R\sqrt{1 + \frac{1}{m^2}},$$

so that the rhumb distance (up to sign) between the two points is

$$D_{\text{rh}} = \int_{\text{start}}^{\text{end}} ds = \int_{L_1}^{L_2} \frac{ds}{dL} dL = R\sqrt{1 + \frac{1}{m^2}} \int_{L_1}^{L_2} dL = R\sqrt{1 + \frac{1}{m^2}}\,(L_2 - L_1). \tag{3}$$

Since, from equation (1), m is the cotangent of the bearing θ, equation (3) can be written

$$D_{\text{rh}} = R\,|L_2 - L_1|\,|\sec\theta|. \tag{4}$$

An alternative form for equation (3) is

$$D_{\text{rh}} = R\frac{\sqrt{\Delta\Sigma^2 + \Delta\lambda^2}}{|\Delta\Sigma/\Delta L|}. \tag{5}$$

Any of (3), (4), or (5) solves problem III. Table 1 gives some examples. Equation (5) can be read as the "Mercator Euclidean" distance divided by the relative total stretching between the two latitudes. From formula (5), in the limiting case $L_2 = L_1 = L$, the distance $D_{\text{rh}} = R|\Delta\lambda|/\sigma(L)$, as we already knew. Formula (5) is easy to program into a calculator or computer. Note also that a loxodrome extending all the way to the poles has finite length, although it spirals infinitely.

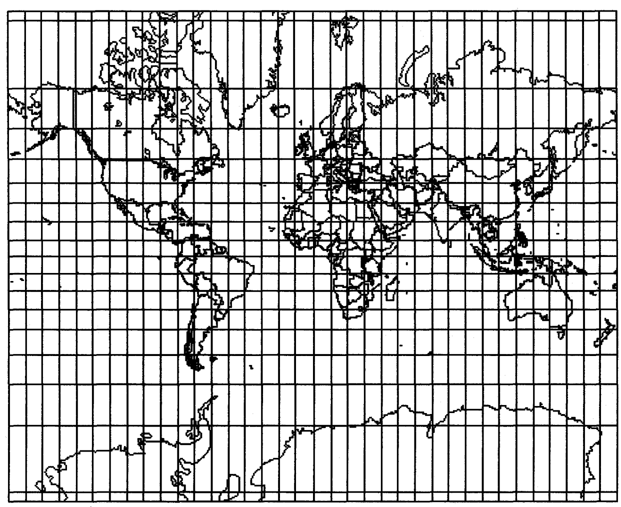


Figure 4 A modern Mercator map, vertically centered at the equator, with 10° latitude and longitude grid. The vertical spacing of the parallels of latitude is given by Σ, as graphed in FIGURE 3. The area distortion at high latitudes is evident. A rhumb course from New York to Beijing goes almost directly west across the 40th parallel of north latitude, while the great circle route comes close to the North Pole.

For a spherical earth, $\sigma(L) = \sec L$, so

$$\Sigma(L) = \Sigma_{\text{sphere}}(L) = \ln(\sec L + \tan L) = \ln\left(\tan\frac{1}{2}\left(\frac{\pi}{2} + L\right)\right), \tag{6}$$

and the rhumb distance (5) can be computed by algebra (code can be downloaded from the web for calculators or PDAs). For an ellipsoidal earth, the stretching factor can also be integrated explicitly:

$$\Sigma_{\text{ellipsoid}}(L) = \ln[\sec L + \tan L] - \frac{e}{2}\ln\left(\frac{1 + e\sin L}{1 - e\sin L}\right),$$

and more precise loxodromic calculations can be made. In fact one could push further; the earth is slightly pear-shaped (technically, the spherical-harmonic coefficient J_3 of the earth is nonzero), and one could make even more precise loxodromic calculations. However, to this level of precision, there are variations that depend on longitude and the precision becomes meaningless.

A rhumb between two points of the same longitude is an arc of a great circle (for instance, New York and Bogotá); a rhumb deviates most from a great circle when the two points have the same latitude. For comparison, the great-circle distance between two locations on a spherical earth is

$$D_{gc} = R \arccos\big(\sin L_1 \sin L_2 + \cos L_1 \cos L_2 \cos(\lambda_2 - \lambda_1)\big)$$

(as can be obtained from a dot product of position vectors), which can also be programmed into a calculator. Some examples of distances from New York (latitude/longitude = $40°45'N/73°58'W$) are given in Table 1.

TABLE 1: Great circle and rhumb distances between New York and selected cities (distances in kilometers)

City	Latitude/Longitude	Comments	D_{gc}	D_{rh}
London	$51°32'N/0°10'W$		5,564	5,802
Bogotá	$4°32'N/74°5'W$	colongitudinal	4,024	4,030
Beijing	$39°55'N/116°23'E$	colatitudinal	11,019	14,380
Canberra	$35°31'S/149°10'E$		16,230	16,408

If the rhumb is to cross the international date line, the calculation must be modified slightly. In equation (2), 2π must be added to or subtracted from the denominator. In equation (3) or (5), m or $\Delta\lambda$ must be calculated appropriately. Otherwise, one determines the rhumb that goes around the earth in the opposite direction. In fact, there are an infinite number of rhumbs between any two non-polar, non-colatitudinal, points, differing in the number of times they encircle the earth, say indexed by the number of times they cross the international date line (positively or negatively). These can all be visualized by putting an infinite number of Mercator maps side by side. The lines from one point to the infinite number of copies of the other point correspond to the infinite number of rhumbs. The headings and distances are calculated by modifying the longitudinal differences in (2) and (5) by the index times 2π. If the index is very large, the rhumb will come close to the target point, but then swing away for another cycle around the earth.

Loxodrome equations

A loxodrome can be represented parametrically. Inverting equation (6) and using equation (1) with $\lambda_1 = b = 0$ yields $L = \pi/2 - 2\arctan(\exp(-m\lambda))$. Substituting into spherical coordinates and simplifying with various trigonometric formulae yields $x = R\cos\lambda\,\operatorname{sech} m\lambda$, $y = R\sin\lambda\,\operatorname{sech} m\lambda$, $z = R\tanh m\lambda$. Thus can loxodromes be graphed, as in FIGURE 1 (where $m = .075$). The stereographic projection to the tangent plane at the pole is the logarithmic spiral $r = Re^{m\theta}$, a result found by Edmond Halley by synthetic construction [**6, 7**], but an easy exercise with analytic geometry.

A question

These results provoke the question: for what positions for locations 1 and 2 are loxodromic distances the largest compared to great circles? Briefly we consider a restricted case, leaving a larger answer to the interested reader. Suppose 1 and 2 are at the same latitude L (in radians), but 180° apart in longitude. The great circle route goes over the pole, with length $D_{gc} = R(\pi - 2L)$. The loxodromic distance is $D_{rh} = \pi R \cos L$. The ratio D_{rh}/D_{gc} increases from 1 at the equator to $\pi/2$ at the pole. On the other hand, the difference $D_{rh} - D_{gc}$ is greatest when $\sin L = 2/\pi$, so $L \approx .69 \approx 40°$ (see New York-Beijing). This seems to be a maximum for locations that are 180° apart in longitude (with no constraints on the latitudes).

Acknowledgment. The author would like to thank a number of people—the referees and others—whose comments on earlier drafts materially improved the paper.

REFERENCES

1. American Spice Trade Association (ASTA), The history of the spice trade, http://www.astaspice.org/history/frame_history.htm, undated.
2. Nicholas Crane, *Mercator: The Man Who Mapped the Planet,* Henry Holt, New York, 2002.
3. M. C. Escher Foundation, *M. C. Escher: the official website,* http://www.mcescher.com, undated.
4. Timothy G. Feeman, Conformality, the exponential function, and world map projections, *Coll. Math. Journal* **32**:5 (2001), 334–342.
5. ———, *Portraits of the Earth: A Mathematician Looks at a Map*, American Mathematical Society, Providence, RI, 2002.
6. Ronald Gowing, Halley, Cotes, and the nautical meridian, *Historica Math.* **22**:1 (1995), 19–32.
7. Edmond Halley, An easie demonstration of the analogy of the logarithmick tangents to the meridian line or sum of the tangents, with various methods for computing the same, to the utmost exactness, *Phil. Trans. Royal Soc. London* **19** (1696), 199–214.
8. Raymond D'Hollander, Historique de la loxodromie, *Mare Liberum* **1** (1990), 29–69.
9. F. Wilhelm Krücken, http://www.wilhelmkruecken.de (2002) (click on *Gerhard Mercator*, then on links under *Die Weltkarte*).
10. ———, *Wissenschaftsgeschihtliche und –theoretische berlegungen zur Entstehung der Mercator-Weltakerte 1569*, Sonderdruck aus Duisbeurger Forschungen Bd 41, 1993.
11. Pedro Nunes, *Opera,* Basel, 1566.
12. W. G. L. Randles, Pedro Nunes' discovery of the loxodromic curve (1537): how Portuguese sailors in the early sixteenth century, navigating with globes, had failed to solve the difficulties encountered with the plane chart, *Journal of Navigation* **50** (1997), 85-96, reprinted as chap. XIV in W. G. L. Randles, *Geography, Cartography and Nautical Science in the Renaissance: The Impact of the Great Discoveries*, Variorum Collected Studies, Ashgate, 2000.
13. V. Frederick Rickey and Philip M. Tuchinsky, An application of geography to mathematics: history of the integral of the secant, this MAGAZINE **53** (May 1980), 162–166.
14. Cecil Roth, *The Jewish Contribution to Civilization*, 4th ed,, Union of American Hebrew Congregations, Cincinnati, 1940.
15. Willebrord Snel van Royen, *Hypomnemata mathematica*, Lyons, 1605–08.
16. Edward Wright, *Certaine errors of Navigation,* London, 1599.

50 Years Ago in the MAGAZINE

from "On the Stability of Differential Expressions," by S. W. Ulam and D. H. Hyers, Vol. **28**, No. 2, (Nov.–Dec., 1954), 59:

> Every student of calculus knows [!] that two functions may differ uniformly by a small amount and yet their derivatives may differ widely. On the other hand, it is more or less obvious intuitively, and quite easily proved, that if a continuous function f on a finite closed interval has a proper maximum at a point $x = a$, then any continuous function g sufficiently close to f also has a maximum arbitrarily close to $x = a$.

An interview with Ulam appeared in the June 1981 issue of the *College Mathematics Journal*, known at the time as *The Two-Year College Mathematics Journal*, **12**: 3 (1981), 182–189.

Which Way Did You Say That Bicycle Went?

DAVID L. FINN
Rose-Hulman Institute of Technology
Terre Haute, IN 47803
david.finn@rose-hulman.edu

Imagine a 20-foot wide mud patch through which a bicycle has just passed, with its front and rear tires leaving tracks as illustrated in FIGURE 1. *In which direction was the bicyclist traveling?*

Figure 1 Which way did the bicycle go?

You may recognize this problem from the title of the book *Which Way Did the Bicycle Go?* [**8**]. For most pairs of tire tracks, you can determine which direction the bicycle was traveling. However, there are exceptions. For example, it is impossible to determine which direction a bicycle was traveling given a single straight track or a pair of tire tracks in the form of concentric circles.

In this paper, we provide a method for constructing such *ambiguous tire tracks*, where one cannot determine from the shape of the tracks which direction the bicycle was traveling. The pair of tracks in FIGURE 1 is in fact an example of such a pair of ambiguous tracks. Our method to produce these involves only a slight variation of the solution method for the original problem [**8**].

Our method for constructing ambiguous tire tracks starts with the construction of a special initial back-tire track segment. From this initial back-tire track segment, we can construct a front-tire track segment (which will be longer) by pushing the bicycle forward in both possible directions of motion along the back-tire track. The next step is to extend the back-tire track by pushing the bicycle backwards along the newly created front-tire track in both possible directions. This physical construction can be continued iteratively until either the front-tire track or the back-tire track develops a singularity. To see a demonstration of this mathematical construction, the interested reader can view animations at the author's website [**5**].

A problem with solving the original problem

To determine from its tire tracks which direction a bicycle went, we need to know how the positions of the front and back tires are related. We will assume for simplicity that the bicycle is ridden on a perfectly flat surface and that the plane of each tire meets the plane of the surface in a right angle. These assumptions will be used throughout this paper to simplify the mathematics, as without them we must account for more variables in the problem (the angle at which the bicycle is banked, the shape of the surface on which the bicycle is ridden, the changing point of contact on the tire between the tire and the ground, etc.). The practical impact of these assumptions is that the position of the front tire $\alpha(t)$ at time t is related to the position of the back tire $\beta(t)$ at time t by

$$\alpha(t) = \beta(t) + L\,\mathbf{T}_\beta(t), \tag{1}$$

where $\mathbf{T}_\beta(t)$ is the unit tangent vector of the back-tire track and L is a constant representing the length of the bicycle.

We obtain (1) as a simple consequence of our assumptions and two facts about the manner in which a bicycle is built. First, the back tire is aligned with the frame. This implies that the unit tangent of the back-tire track is a multiple of the displacement vector $\alpha(t) - \beta(t)$. Second, the frame of a bicycle is rigid, which means that the vector $\alpha(t) - \beta(t)$ is of constant length, L. We note that in practice the distance between $\alpha(t)$ and $\beta(t)$ is not actually a constant, but only nearly a constant: the variation depends on changing points of contact between the tires and the road; it arises mainly from the deviations of the tires from perfect circles as the bicycle is ridden.

We can now try using (1) to determine which way the bicycle went in FIGURE 1. First, we need to determine which track was created by the front tire and which was created by the back tire. One can do this by trial and error. However, it is normally possible to determine which tire is which by inspection, using the simple fact that the front-tire track should vary more than the back-tire track. This means the curve that deviates more from a straight line (has a larger amplitude) should be the front-tire track, and the curve that deviates less from a straight line (has a smaller amplitude) should be the back-tire track. Once we know which is which, all we need to do is draw tangent lines at a few points on the back-tire track and measure the distance between the point of tangency and the intersections with the front-tire track. The bicycle then went in the direction in which the measured distance does not change.

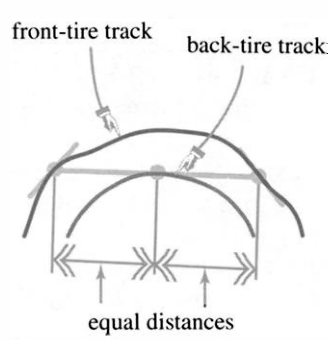


Figure 2 Example of ambiguity of direction

The tracks in FIGURE 1 present problems with this solution method, since like the tracks in FIGURE 2 above, the distances are constant in both directions; see also the animations at the author's website [**5**]. This gives us the defining property for ambiguous tire tracks: given a back-tire track β, the front-tire track of a set of ambiguous tire tracks can be described by either $\beta(t) + L\,\mathbf{T}_\beta(t)$ or $\beta(t) - L\,\mathbf{T}_\beta(t)$. This means that in order to create a pair of ambiguous tire tracks, we need a back-tire track such that for every t there is a $\tau > t$, continuously varying with t, such that

$$\beta(t) + L\,\mathbf{T}_\beta(t) = \beta(\tau) - L\,\mathbf{T}_\beta(\tau).$$

Creating such a back-tire track is thus the key part in our method for constructing ambiguous tracks. In fact, as stated earlier, we need only construct a finite segment of such a back-tire track.

To preview our construction of an initial back-tire track segment for a set of ambiguous tire tracks, we will suppose that we have a set of ambiguous tracks and describe some of the necessary conditions on a segment of the back-tire track. The condition above implies that we may take the back-tire track $\beta : [0, 1] \to \mathbf{R}^2$ to be parameterized so that

$$\beta(0) + L\,\mathbf{T}_\beta(0) = \beta(1) - L\,\mathbf{T}_\beta(1).$$

This condition implies that the forward and reverse front-tire tracks $\alpha_f(t) = \beta(t) + L\,\mathbf{T}_\beta$ and $\alpha_r(t) = \beta(t) - L\,\mathbf{T}_\beta(t)$ meet at $\alpha_f(0) = \alpha_r(0)$. This places compatibility conditions on the geometry of an initial back-tire track segment at the end points of the segment. Other more technical conditions will be required on the geometry of the endpoints of the back-tire track to ensure that the two possible front-tire tracks $\alpha_f(t)$ and $\alpha_r(t)$ join smoothly at $\alpha_f(0) = \alpha_r(1)$. These more technical conditions will be discussed in detail later in the paper.

Constructing such an initial back-tire track and the technical conditions involved in describing the geometry of the tracks are conceptually the hardest part of the process. However, once an initial back-tire track segment has been constructed, it is relatively easy to construct a front-tire segment by joining $\beta(t) + L\,\mathbf{T}_\beta(t)$ and $\beta(t) - L\,\mathbf{T}_\beta(t)$. Extending the back-tire track so that we can continue to extend the front-tire track is the most technical part of the entire discussion. Physically, the extension of the back-tire track is accomplished by pushing the bicycle backwards, steering the bicycle in such a manner as to keep the front tire on the newly created front-tire track; again see the relevant animations [**5**]. Mathematically, the extension amounts to solving a differential equation that governs the physical construction. We can extend the front-tire track and the back-tire track in this manner to produce an arbitrarily long ambiguous tire track, at least given a sufficiently nice initial back-tire track.

The geometry of bicycle tracks

In this paper, we construct ambiguous bicycle tire tracks by means of a geometric approach. This is the same approach we used to describe the creation of a unicycle track with a bicycle [**4**]. Tabachnikov [**11**] uses a similar approach to examine closed ambiguous tracks, but he describes ambiguous tracks entirely through the front-tire track. One can also use a more analytic approach using Ricatti equations [**2, 7**]. However, we prefer the geometric approach for its intuitive appeal, matching the geometric nature of the actual problem.

The geometric relations between the front-tire track and the back-tire track are derived by differentiating $\alpha(t) = \beta(t) + L\,\mathbf{T}_\beta(t)$. These relations are easiest to state and understand in terms of elementary differential geometry (curvature, Frenet frames, and the fundamental theorem of plane curves). Most of the differential geometry that we require can be found in calculus textbooks [**3, 10**]. However, for completeness, we will briefly explain the necessary facts. The interested reader can consult texts on differential geometry for more details [**1, 6, 9**].

First, we define the *Frenet frame* of a plane curve $c(t)$ to be the ordered pair of vectors $[\mathbf{T}(t), \mathbf{N}(t)]$, where $\mathbf{T}(t)$ is the unit tangent vector $c'(t)/\|c'(t)\|$ of c at time t, and $\mathbf{N}(t)$ is the principal unit normal vector of c at time t. Here, we will define the vector $\mathbf{N}(t)$ by rotating $\mathbf{T}(t)$ counterclockwise by $\pi/2$ radians, that is, if

$$\mathbf{T} = \cos(\phi)\,\mathbf{i} + \sin(\phi)\,\mathbf{j}, \quad \text{then} \quad \mathbf{N} = -\sin(\phi)\,\mathbf{i} + \cos(\phi)\,\mathbf{j},$$

since $\cos(\phi + \pi/2) = -\sin(\phi)$ and $\sin(\phi + \pi/2) = \cos(\phi)$. For each t the pair of vectors $[\mathbf{T}(t), \mathbf{N}(t)]$ is a set of orthogonal unit vectors in the planes. (Note that we tacitly assume that the curve is *regular*, meaning that $c'(t) \neq 0$ at every point.)

The importance of the Frenet frame is that the geometry of the curve is revealed in how the frame changes. Specifically, the signed curvature κ of the curve c is given by the Frenet frame equations:

$$\frac{d\mathbf{T}}{ds} = \kappa\,\mathbf{N} \quad \text{and} \quad \frac{d\mathbf{N}}{ds} = -\kappa\,\mathbf{T},$$

where s is the arc-length parameter of the curve. It is worth noting that functions $\mathbf{T}$, $\mathbf{N}$, and κ depend only on the point on the curve and not the parameterization of the curve; thus, they are naturally defined as functions of the arc-length parameter. However, as the arc-length parameter is not convenient for calculations, we often write the Frenet frame equations, using the chain rule, as

$$\frac{d\mathbf{T}}{dt} = \kappa\,\mathbf{N}\,\frac{ds}{dt} \quad \text{and} \quad \frac{d\mathbf{N}}{dt} = -\kappa\,\mathbf{T}\,\frac{ds}{dt},$$

where we view t as time and ds/dt is the speed of the particle creating the curve.

The signed curvature κ as defined above is different from the curvature defined in some standard calculus texts **[3, 10]**. Such texts often limit the discussion to the unsigned curvature, $|\kappa| = \|d\mathbf{T}/ds\|$, which can be shown to be the reciprocal of the radius of the circle that best fits the curve at a point. For us, the importance of the signed curvature κ is that it determines the curve up to a translation and a rotation. This important fact is known as the *fundamental theorem of plane curves*.

It is not too difficult to see why the signed curvature determines the curve up to a translation and a rotation: Differentiating $\mathbf{T} = \cos(\phi)\mathbf{i} + \sin(\phi)\mathbf{j}$ with respect to s, we find $\kappa = d\phi/ds$. Therefore, knowing κ as function of s, we can determine the angle ϕ defining T, at least up to a constant angle of rotation by integration. Once we know $\mathbf{T}$, we can determine $\mathbf{c}$, at least up to a translation by integration, as $\mathbf{T} = d\mathbf{c}/ds$.

The fundamental theorem of plane curves allows us to construct a tire track from its curvature. This means that we can construct the front-tire track from the back-tire track, provided we can relate the two curvatures. We thus distill the system of differential equations $\alpha(t) = \beta(t) + L\,\mathbf{T}_\beta(t)$ into a single differential equation, which we can use to construct ambiguous tire tracks.

In the remainder of this section, we relate the curvatures of the front- and back-tire tracks, α and β. To distinguish between the features associated with the two curves, we will use subscripts. For instance, we will use $[\mathbf{T}_\alpha, \mathbf{N}_\alpha]$ for the Frenet frame of α and $[\mathbf{T}_\beta, \mathbf{N}_\beta]$ for the Frenet frame of β.

The curvatures κ_α and κ_β are related by differentiating the fundamental relation $\alpha(t) = \beta(t) + L\,\mathbf{T}_\beta(t)$, using the Frenet frame equations to substitute for the derivatives of the unit tangent and normal vectors. Differentiating $\alpha(t) = \beta(t) + L\,\mathbf{T}_\beta(t)$ with respect to t gives

$$\mathbf{T}_\alpha\,\frac{ds_\alpha}{dt} = \mathbf{T}_\beta\,\frac{ds_\beta}{dt} + L\,\kappa_\beta\,\mathbf{N}_\beta\,\frac{ds_\beta}{dt},$$

where s_α and s_β are the arc-length parameters of the front- and back-tire tracks, α and β. Since $\mathbf{T}_\beta$ and $\mathbf{N}_\beta$ are orthogonal, we have

$$\left(\frac{ds_\alpha}{dt}\right)^2 = \left(1 + (L\kappa_\beta)^2\right)\left(\frac{ds_\beta}{dt}\right)^2.$$

Orienting α and β so that ds_α/dt and ds_β/dt have the same sign, we find

$$\begin{cases} \mathbf{T}_\alpha = \dfrac{1}{\sqrt{1+(L\,\kappa_\beta)^2}}\,\mathbf{T}_\beta + \dfrac{L\,\kappa_\beta}{\sqrt{1+(L\,\kappa_\beta)^2}}\,\mathbf{N}_\beta \\ \mathbf{N}_\alpha = \dfrac{-L\,\kappa_\beta}{\sqrt{1+(L\,\kappa_\beta)^2}}\,\mathbf{T}_\beta + \dfrac{1}{\sqrt{1+(L\,\kappa_\beta)^2}}\,\mathbf{N}_\beta. \end{cases} \tag{2}$$

This relates the Frenet frame of α to the Frenet frame of β through the curvature κ_β of β.

Using a standard method in differential geometry, we differentiate again, this time starting with the equation for $\mathbf{T}_\alpha$ in (2). Differentiate with respect to t, using the Frenet frame equations. Use the relation (2) to express $\mathbf{N}_\alpha$ in terms of $\mathbf{T}_\beta$ and $\mathbf{N}_\beta$. Then equate the coefficients of $\mathbf{T}_\beta$ in both sides of the resulting equation. When all is done, we have

$$\kappa_\alpha \frac{ds_\alpha}{dt} = \left[\frac{L\,\frac{d\kappa_\beta}{ds_\beta}}{1+(L\,\kappa_\beta)^2} + \kappa_\beta \right] \frac{ds_\beta}{dt}. \tag{3}$$

Furthermore, using the relation between ds_α/dt and ds_β/dt, we find

$$\kappa_\alpha = \frac{L\,\frac{d\kappa_\beta}{ds_\beta}}{(1+(L\,\kappa_\beta)^2)^{3/2}} + \frac{\kappa_\beta}{(1+(L\,\kappa_\beta)^2)^{1/2}}. \tag{4}$$

Notice that (3) and (4) imply that the curvature κ_α of the front-tire track α depends only on the curvature κ_β of the back-tire track.

If we know the curvature of the front tire, then either (3) or (4) can be viewed as a differential equation that we may solve for the curvature of the back tire. However, as differential equations to solve for κ_β, they present some problems. They are highly nonlinear and κ_α is not naturally given in terms of s_β, as the arc-length parameters of the two curves are related by the unknown κ_β. We can overcome these difficulties by introducing alternative variables that are naturally associated to riding bicycles.

Those who have ridden bicycles may realize that they have direct control over the speed of the back tire, ds_β/dt, and the turning angle, Θ, between the frame and the front tire, which is also the angle between the vectors $\mathbf{T}_\alpha$ and $\mathbf{T}_\beta$. For our purposes, the angle Θ should be signed and satisfy $-\pi/2 < \Theta < \pi/2$. The sign requirement is so that we can distinguish between right-hand and left-hand turns. Let us declare that $\Theta > 0$ for a left-hand turn and $\Theta < 0$ for a right-hand turn. The requirement that $-\pi/2 < \Theta < \pi/2$ implies that ds_α/dt and ds_β/dt have the same sign, and will hold until a singularity develops in the tire tracks.

We can now write (3) and (4) in terms of these natural control parameters for a bicycle, thus creating a bicycle track naturally, by specifying how the bicycle is turned and how fast it goes. The principal relation we need uses the angle Θ to relate the Frenet frames of the front and back tires:

$$\begin{cases} \mathbf{T}_\alpha = \cos\Theta\,\mathbf{T}_\beta + \sin\Theta\,\mathbf{N}_\beta \\ \mathbf{N}_\alpha = -\sin\Theta\,\mathbf{T}_\beta + \cos\Theta\,\mathbf{N}_\beta, \end{cases} \tag{5}$$

which allows us to compute the relations (3) and (4) in terms of Θ.

Combining (2) with the formula for $\mathbf{T}_\alpha$ in (5), we see that

$$\cos\Theta = \frac{1}{\sqrt{1+(L\,\kappa_\beta)^2}} \quad \text{and} \quad \sin\Theta = \frac{L\,\kappa_\beta}{\sqrt{1+(L\,\kappa_\beta)^2}},$$

and thus

$$\kappa_\beta = \frac{\tan\Theta}{L} \quad \text{and} \quad \frac{ds_\alpha}{dt} = \sec\Theta\,\frac{ds_\beta}{dt}. \tag{6}$$

We can now rewrite (4) as

$$\kappa_\alpha = \cos\Theta\,\frac{d\Theta}{ds_\beta} + \frac{\sin\Theta}{L}, \tag{7}$$

where we interpret

$$\frac{d\Theta}{ds_\beta} \quad \text{as} \quad \frac{d\Theta}{dt}\Big/\frac{ds_\beta}{dt}.$$

Using (6), we can also write (7) as

$$\kappa_\alpha = \frac{d\Theta}{ds_\alpha} + \frac{\sin\Theta}{L}. \tag{8}$$

Equations (6) to (8) allow us to create a pair of tire tracks knowing how the bicycle is turned and the speed of the back tire; we just solve the relevant Frenet frame equations to recover the curves from their curvature. Equations (7) and (8) also allow us to construct the back-tire track knowing the front-tire track: If we know κ_α, we can solve for the angle Θ and then solve for the back-tire track either by the appropriate Frenet frame equations or directly from (1) and (5):

$$\beta = \alpha - L(\cos\Theta\,\mathbf{T}_\alpha - \sin\Theta\,\mathbf{N}_\alpha).$$

In the remainder of this paper, we will use t as the parameter for the bicycle tracks α and β. However, we will occasionally use the arc-length parameters. We view the time parameter t as the natural one for constructing bicycle tracks, while the arc-length parameters are more useful in geometric calculations.

Creating the initial back-tire track segment

Given the relations between the front and back tires, we can now derive conditions that will allow us to create an initial back-tire track segment for ambiguous tracks using an interpolation scheme. Recall that the defining property of ambiguous bicycle tracks is that the front-tire track α can be described as either $\beta(t) + L\,T_\beta(t)$ or $\beta(t) - L\,T_\beta(t)$ in terms of the back-tire track β. This implies that for each t_1 there is a $t_2 > t_1$ such that

$$\beta(t_1) + L\mathbf{T}_\beta(t_1) = \beta(t_2) - L\,\mathbf{T}_\beta(t_2). \tag{9}$$

This then implies that the geometry (Frenet frames and curvatures) of the front-tire track can be computed from either description and we must obtain the same values at the corresponding t-values t_1 and t_2. Therefore, we must have

$$\begin{aligned} &\left[\frac{1}{\sqrt{1+(L\kappa_\beta)^2}}\mathbf{T}_\beta + \frac{L\kappa_\beta}{\sqrt{1+(L\kappa_\beta)^2}}\mathbf{N}_\beta\right]\Bigg|_{t=t_1} \\ &\quad = \left[\frac{1}{\sqrt{1+(L\kappa_\beta)^2}}\mathbf{T}_\beta - \frac{L\kappa_\beta}{\sqrt{1+(L\kappa_\beta)^2}}\mathbf{N}_\beta\right]\Bigg|_{t=t_2} \end{aligned} \tag{10}$$

and

$$\left[\frac{L\frac{d\kappa_\beta}{ds_\beta}}{(\sqrt{1+(L\kappa_\beta)^2})^3}+\frac{\kappa_\beta}{\sqrt{1+(L\kappa_\beta)^2}}\right]\Bigg|_{t=t_1} = \left[-\frac{L\frac{d\kappa_\beta}{ds_\beta}}{(\sqrt{1+(L\kappa_\beta)^2})^3}+\frac{\kappa_\beta}{\sqrt{1+(L\kappa_\beta)^2}}\right]\Bigg|_{t=t_2}. \tag{11}$$

These properties are enough to generate a pair of ambiguous tire tracks that are thrice differentiable. It is possible to enforce additional smoothness conditions by differentiating the curvature of the front-tire track with respect to arc length, but these are not needed in our method.

From (9), (10), and (11), we can easily construct an initial back-tire track segment $\beta : [0, 1] \to \mathbb{R}^2$ that will be thrice differentiable. First, we choose an isosceles triangle PQR with $|PQ| = |QR| = L$. Then, we set $\beta(0) = P$, $\beta(1) = R$,

$$\mathbf{T}_\beta(0) = \overrightarrow{PQ}/|\overrightarrow{PQ}| \quad \text{and} \quad \mathbf{T}_\beta(1) = \overrightarrow{QR}/|\overrightarrow{QR}|. \tag{12}$$

This ensures that

$$Q = \beta(0) + L\,\mathbf{T}_\beta(0) = \beta(1) - L\,\mathbf{T}_\beta(1).$$

Next, we choose a tangent vector $\mathbf{T}_\alpha$ and a curvature κ_α for the front-tire track at Q. These are used to define the curvature of β at P and R from (2), and the derivative of the curvature $d\kappa_\beta/ds_\beta$ at P and R from (3). The curvature κ_β at P and R are given respectively by

$$\kappa_\beta(0) = \frac{(\mathbf{T}_\alpha\cdot\mathbf{N}_\beta)}{L\,(\mathbf{T}_\alpha\cdot\mathbf{T}_\beta)}\bigg|_{t=0} \quad \text{and} \quad \kappa_\beta(1) = -\frac{(\mathbf{T}_\alpha\cdot\mathbf{N}_\beta)}{L\,(\mathbf{T}_\alpha\cdot\mathbf{T}_\beta)}\bigg|_{t=1}. \tag{13}$$

The derivative $d\kappa_\beta/ds_\beta$ of the curvature with respect to arc length at P and R are then given by solving

$$\kappa_\alpha = \left[L\frac{d\kappa_\beta}{ds_\beta}(\mathbf{T}_\alpha\cdot\mathbf{T}_\beta)^3 + \frac{(\mathbf{T}_\alpha\cdot\mathbf{N}_\beta)}{L}\right]\bigg|_{t=0} \tag{14}$$

and

$$\kappa_\alpha = \left[L\frac{d\kappa_\beta}{ds_\beta}(\mathbf{T}_\alpha\cdot\mathbf{T}_\beta)^3 - \frac{(\mathbf{T}_\alpha\cdot\mathbf{N}_\beta)}{L}\right]\bigg|_{t=1} \tag{15}$$

for $\frac{d\kappa_\beta}{ds_\beta}\big|_{t=0}$ and $\frac{d\kappa_\beta}{ds_\beta}\big|_{t=1}$.

It is now a standard exercise in interpolation to create a curve $\beta : [0, 1] \to \mathbb{R}^2$ that has these properties: prescribed values for β, T_β, κ_β, and $d\kappa_\beta/ds_\beta$ when $t = 0$ and $t = 1$. One method is to look for a seventh-order polynomial curve of the form

$$\beta(t) = (1-t)^4\left(\mathbf{a}_0 + \mathbf{a}_1 t + \frac{1}{2}\mathbf{a}_2 t^2 + \frac{1}{6}\mathbf{a}_3 t^3\right) + t^4\left(\mathbf{b}_0 + \mathbf{b}_1(t-1) + \frac{1}{2}\mathbf{b}_2(t-1)^2 + \frac{1}{6}\mathbf{b}_3(t-1)^3\right). \tag{16}$$

Standard geometric relations arise from differentiationg β and using the Frenet frame equations:

$$\begin{cases} \beta' = \dfrac{ds_\beta}{dt}\,\mathbf{T}_\beta \\ \beta'' = \dfrac{d^2s_\beta}{dt^2}\,\mathbf{T}_\beta + \kappa_\beta \left(\dfrac{ds_\beta}{dt}\right)^2 \mathbf{N}_\beta \\ \beta''' = \left(\dfrac{d^3s_\beta}{dt}^3 - \kappa_\beta^2 \left(\dfrac{ds_\beta}{dt}\right)^3\right)\mathbf{T}_\beta \\ \qquad + \left(3\kappa_\beta \left(\dfrac{ds_\beta}{dt}\right)\left(\dfrac{d^2s_\beta}{dt^2}\right) + \dfrac{d\kappa_\beta}{ds_\beta}\left(\dfrac{ds_\beta}{dt}\right)^3\right)\mathbf{N}_\beta. \end{cases}$$

By choosing

$$\frac{ds_\beta}{dt} = \lambda_0, \quad \frac{d^2s_\beta}{dt^2} = 0, \quad \frac{d^3s_\beta}{dt^3} = 0$$

when $t = 0$, and we find

$$\begin{cases} \mathbf{a}_0 = P \\ \mathbf{a}_1 - 4\mathbf{a}_0 = \lambda_0\,\mathbf{T}_\beta \\ \mathbf{a}_2 - 8\mathbf{a}_1 + 12\mathbf{a}_0 = \kappa_\beta\,\lambda_0^2\,\mathbf{N}_\beta \\ \mathbf{a}_3 - 12\mathbf{a}_2 + 36\mathbf{a}_1 - 24\mathbf{a}_0 = -\kappa_\beta^2\,\lambda_0^3\,\mathbf{T}_\beta + \dfrac{d\kappa_\beta}{ds_\beta}\,\lambda_0^3\,\mathbf{N}_\beta. \end{cases} \tag{17}$$

Choosing

$$\frac{ds_\beta}{dt} = \lambda_1, \quad \frac{d^2s_\beta}{dt^2} = 0, \quad \frac{d^3s_\beta}{dt^3} = 0$$

when $t = 1$ gives

$$\begin{cases} \mathbf{b}_0 = R \\ \mathbf{b}_1 + 4\mathbf{b}_0 = \lambda_1\,\mathbf{T}_\beta \\ \mathbf{b}_2 + 8\mathbf{b}_1 + 12\mathbf{b}_0 = \kappa_\beta\,\lambda_1^2\,\mathbf{N}_\beta \\ \mathbf{b}_3 + 12\mathbf{b}_2 + 36\mathbf{b}_1 + 24\mathbf{b}_0 = -\kappa_\beta^2\,\lambda_1^3\,\mathbf{T}_\beta + \dfrac{d\kappa_\beta}{ds_\beta}\,\lambda_1^3\,\mathbf{N}_\beta. \end{cases} \tag{18}$$

Given an isosceles triangle PQR, it is now possible to construct an initial back-tire track of form (16) from a choice of T_α, κ_α, λ_0, and λ_1 by using (12) to define $\mathbf{T}_\beta$, using (13) to define $\kappa_\beta(0)$ and $\kappa_\beta(1)$, and then (14) and (15) to define $d\kappa_\beta/ds_\beta$ when $t = 0$ and when $t = 1$. These quantities are then used to define $\mathbf{a}_i$ in (17) and $\mathbf{b}_i$ in (18).

As an example, consider the above procedure with $P = [-2, 0]$, $Q = [0, 1]$, $R = [0, 2]$, $T_\alpha = [12/13, 5/13]$, $\kappa_\alpha = -1$. We then find, solving for κ_β and $d\kappa_\beta/ds_\beta$ when $t = 0$ and $t = 1$, that

$$\kappa_\beta(0) = -\frac{2\sqrt{5}}{145} \approx -0.031, \quad \kappa_\beta(1) = -\frac{22\sqrt{5}}{95} \approx -0.518$$

and

$$\left.\frac{d\kappa_\beta}{ds_\beta}\right|_{t=0} = -\frac{10647}{24389} \approx -0.437, \quad \left.\frac{d\kappa_\beta}{ds_\beta}\right|_{t=1} = -\frac{7267}{6859} \approx -1.059.$$

Using $\lambda_0 = 5$ and $\lambda_1 = 3$, we get the initial back-tire track

$$\beta(t) \approx \begin{bmatrix} -2.00 + 4.47t + 0.17t^2 - 4.05t^3 - 9.61t^4 + 2.60t^5 + 5.07t^6 - 2.75t^7 \\ 2.24t - 0.34t^2 - 8.14t^3 + 4.73t^4 + 10.55t^5 - 13.06t^6 + 4.03t^7 \end{bmatrix}$$

displayed in the diagram below with the front-tire track it generates.

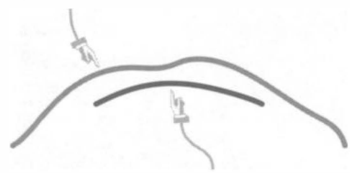


Figure 3 An initial back-tire track

The method described above gives one way to create such an initial back-tire track using polynomial curves with some free parameters. It is useful to note that very general forms arise from perturbing the simple ambiguous tracks constructed as straight lines and concentric circles. For instance, a perturbation of a straight line is given by any curve of the form

$$\beta(t) = \big[-L + 2Lt,\, \nu(t)\big],$$

where $\nu(t) = t^4(1-t)^4\varphi(t)$ with φ any smooth function on $0 \le t \le 1$. It is a straightforward exercise to verify that such a curve satisfies (13), (14), and (15) with $\mathbf{T}_\alpha = [1, 0]$ and $\kappa_\alpha = 0$. A perturbation of a circle is given by any curve of the form

$$\beta(t) = \big[\big(r + \nu(t)\big)\sin(-\psi + 2\psi t),\, \big(r + \nu(t)\big)\cos(-\psi + 2\psi t) - r\cos(\psi)\big],$$

where $r = L\cot(\psi)$ with ψ an arbitrary chosen angle satisfying $0 < \psi < \pi/2$ and $\nu(t)$ is a smooth function of the form $\nu(t) = t^4(1-t)^4\varphi(t)$. Again, it is a straightforward exercise to verify that such a curve satisfies (13), (14), and (15) with $T_\alpha = [1, 0]$ and $\kappa_\alpha = \sin(\psi)/L$. In fact, FIGURE 1 was generated as a perturbation of a straight line using an appropriate scalar multiple of $(4t^2 - 1)$ as $\varphi(t)$. FIGURE 2 was created by perturbing a circle using a multiple of $\sin(2\pi t)$ as $\varphi(t)$. Additional examples of perturbations of straight lines and concentric circles appear at the author's website **[5]**.

Creating ambiguous tire tracks

Suppose we have constructed an initial back-tire track segment $\beta : [0, 1] \to \mathbb{R}^2$ by one of the methods in the previous section. Let $\alpha : [0, 1] \to \mathbb{R}^2$ be the front-tire track obtained by pushing the bicycle in both possible directions on the back-tire track, that is,

$$\alpha(t) = \begin{cases} \beta(1+t) - L\,\mathbf{T}_\beta(1+t) & \text{for } -1 \le t \le 0, \\ \beta(t) + L\,\mathbf{T}_\beta(t) & \text{for } 0 \le t \le 1. \end{cases} \tag{19}$$

The construction of β ensures that this α is twice differentiable with respect to arc length. In this section, we use the differential equations describing the geometry of bicycle tracks to extend the back-tire track β to $-1 \le t \le 0$. We will only describe the method for extending the back-tire track to this domain, as once we have the new back-tire track we can repeat this procedure to extend the track to $t < -1$. The extension for $t > 1$ is then established by applying the same method to the initial back-tire track $\beta_1(t) = \beta(1 - t)$ that has the opposite orientation as β.

Let Θ be the turning angle between $\mathbf{T}_\alpha$ and $\mathbf{T}_\beta$, defined for $0 \le t \le 1$. Our construction method is based upon extending Θ to $-1 \le t \le 0$, and then using Θ to define the curvature of the back-tire track for $-1 \le t \le 0$; finally, we extend the back-tire track, by the fundamental theorem of plane curves. To extend the turning angle Θ, we solve the differential equation

$$\kappa_\alpha = \frac{d\Theta}{ds_\alpha} + \frac{\sin(\Theta)}{L}, \tag{20}$$

using the definition of α in (19) to compute the curvature κ_α for $-1 \le t \le 0$. The existence of a solution to (20) for a short time $-\epsilon < t < 0$ follows from the existence and uniqueness of solutions to differential equations. Using comparison theorems, it is not hard to show that there is a solution to (20) for $-1 \le t \le 0$. However, it is not possible to show that a generic solution must satisfy $-\pi/2 < \Theta < \pi/2$. In fact, one can easily construct examples of initial segments for which the turning angle leaves this domain.

Rather than give an exhaustive analysis of the existence of solutions to (20) satisfying $-\pi/2 < \Theta < \pi/2$, we supply an informal perturbation argument that one can extend some initial segments as long as one pleases. We first note that if Θ is constant on $0 \le t \le 1$, then the extension of Θ is constant on $-1 \le t \le 0$. Thus, an initial segment of constant Θ (an arc of a circle or a line segment) will generate a circle or a straight line (curves of constant curvature). If we start with a small perturbation of a segment of constant curvature, then it seems reasonable that with a small enough perturbation one may extend the track indefinitely. Of course, we will not specify what we mean by small.

Our argument relies on the use of comparison theorems to show the existence of a solution Θ for $-1 \le t \le 0$, and the continuous dependence of the solution on the parameters of the differential equation that is the curvature of the front tire κ_α and the initial condition $\Theta(0)$. The existence of a solution for $-1 \le t \le 0$ is obtained by comparing the solution Θ of (20) to solutions of the linear equations

$$\frac{d\Theta}{ds} + \frac{\Theta}{L} = \kappa_\alpha \quad \text{and} \quad \frac{d\Theta}{ds} - \frac{\Theta}{L} = \kappa_\alpha.$$

From the continuous dependence on the parameters, it follows that if κ_α is close to a constant, then $\Theta(t)$ is close to $\Theta(0)$ for $-1 \le t \le 0$; therefore, $-\pi/2 < \Theta(T) < \pi/2$. Furthermore, $\Theta(t)$ is close to $\Theta(0)$ for $-1 \le t \le 0$ and κ_α is close to a constant for $-1 \le t \le 0$. The iterative nature of our construction then implies that we should be able to apply the same argument repeatedly as long as we start with a sufficiently small perturbation of a curve of constant curvature, which can be extended indefinitely. The catch in this heuristic argument is that the technical meaning of the word *close* may change as we iterate along. However, if we start with a very small perturbation we should be able to extend the curve as far as we want.

REFERENCES

1. M. DoCarmo, *Differential Geometry of Curves and Surfaces*, Prentice Hall, 1976.
2. S. R. Dunbar, R. J. Bosman, and S. E. M. Nooij, The track of a bicycle back tire, this MAGAZINE **74**:4 (2001), 273–287.

3. C. H. Edwards and D. E. Penney, *Calculus, with Analytic Geometry and Early Transcendentals*, 5th ed., Prentice Hall, 1998, Section 12.6, pp. 759–773.
4. D. L. Finn, Can a bicycle create a unicycle track?, *Coll. Math. J.* **33**:4 (2002), 283–292.
5. ———, http://www.rose-hulman.edu/~finn/research/ambiguous.htm.
6. A. Gray, *Modern Differential Geometry of Curves and Surfaces*, CRC Press, 1993.
7. H. I. Freedman and S. D. Riemenschneider, Determining the path of the rear wheels of a bus, *SIAM Review* **25** (1983), 561–567.
8. J. Konhauser, D. Velleman, and S. Wagon, *Which Way Did the Bicycle Go?*, Dolciani Mathematical Expositions #18, MAA, Washington, D.C., 1996.
9. J. Oprea, *Differential Geometry and its Applications*, Prentice Hall, 1997.
10. J. Stewart, *Calculus: Concepts and Contexts*, 2nd ed., Brooks/Cole, 2001, Section 10.3, pp. 717–725.
11. S. Tabachnikov, Tire track geometry: variations on a theme, preprint.

50 Years Ago in the MAGAZINE: Mathematics and Style

from "The Influence of Newtonian Mathematics on Literature and Aesthetics," by Morris Kline, Vol. **28**, No. 2, (Nov.–Dec., 1954), 95:

> Dryden went so far as to declare: "A man should be learned in several sciences, and should have a reasonable, philosophical and in some measure, a mathematical head to be a complete and excellent poet ...."
>
> Young America also fell under the new influences.
>
> > We do not listen with the best regard to the verses of a man who is only a poet, nor to his problems if he is only an algebraist; but if a man is at once acquainted with the geometric foundation of things and with their festal splendor, his poetry is exact and his arithmetic musical.
>
> This from Emerson.
>
> The works of the outstanding mathematicians were set up as literary models by the eighteenth century. Descartes' style was extolled for its clarity, neatness, readability, and perspicuity, and Cartesianism became a style as well as a philosophy. The elegance and rationality of Pascal's manner, especially in his *Lettres Provinciales*, were hailed as superb elements of literary style. Writers in almost all fields began to ape as closely as their subject matter permitted the works of Descartes, Pascal, Huygens, Galileo, and Newton.

This issue also includes a review of Kline's book, *Mathematics in Western Culture*, "Oxford University Press, 1953, 472 pp., $7.50," written by Kline himself.

NOTES

The Egg-Drop Numbers

MICHAEL BOARDMAN
Pacific University
Forest Grove, OR 97116
boardman@pacificu.edu

Each month, the mathematics students at Pacific University are given a challenge problem in what we call the "Pizza Problem of the Month." One of the most rewarding problems came from *Which Way Did the Bicycle Go?* [**3**, p. 53]:

> **An egg-drop experiment** *We wish to know which windows in a thirty-six-story building are safe to drop eggs from, and which are high enough to cause the eggs to break on landing .... Suppose two eggs are available. What is the* least *number of egg-droppings that is guaranteed to work in all cases?*

To make this problem mathematical, the authors make some simplifying assumptions including

- Eggs that survive can be used again and are not weakened. Eggs that break are history.
- Eggs that break at a particular floor would break from higher floors as well.
- Eggs that survive from a particular floor would survive from lower floors as well.

This problem piqued my interest. Unaware of the solution in the appendix, I found my own solution and stayed up late working out a nice presentation. The next morning, I shared this with a colleague who responded "Interesting. I wonder if this works with more eggs?" This Note answers the question. In working on it, I found that the egg-drop problem is a fruitful setting in which to introduce students to recurrence relations, generating functions, and other counting methods.

Two eggs, thirty-six floors To find a solution to the original problem, imagine that we have one egg rather than two. In this case, we must start at the first floor and work our way up one floor at a time until we have discovered where eggs begin breaking. When we have two eggs, we can use a more efficient strategy until the first egg breaks. Then, we must resort to the original one-egg strategy.

Table 1 illustrates our solution, which begins with a drop from the eighth floor. If this egg breaks, we move back down to the first floor and work our way up until the second egg breaks, unless it survives the drop from the seventh floor, in which case there is no need to drop it from the eighth. If the first egg does not break, we can move up to the fifteenth floor and repeat the process. In any case, we will have made no more than 8 drops. It is easy to see that this solution is optimal; if we started with a drop from the seventh floor, we would not be able to reach the thirty-sixth in 8 drops, whereas if we started on the ninth floor, we might need 9 drops to learn that the egg would break on the eighth.

TABLE 1: Where to drop the eggs

First Egg	Second Egg if First Egg Breaks												
8	1	→	2	→	3	→	4	→	5	→	6	→	7
15	9	→	10	→	11	→	12	→	13	→	14		
21	16	→	17	→	18	→	19	→	20				
26	22	→	23	→	24	→	25						
30	27	→	28	→	29								
33	31	→	32										
35	34												
36													

More eggs, more drops Our natural inclination is to generalize this algorithm for taller buildings. As we do so, we note a tricky feature of the original problem: It asked us to determine that 8 drops would be optimal, given 2 eggs and thirty-six floors, but it is actually simpler to determine that thirty-six floors can be reached, given 2 eggs and 8 drops.

FIGURE 1 shows which floors we can reach for various numbers of drops of 2 eggs and leads us to state the problem in a different way:

A generalized egg-drop problem *Given k eggs, how many floors can we reach if we have at most n drops?*

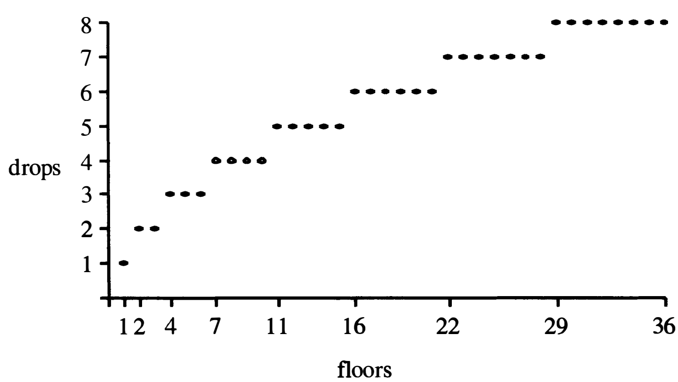


Figure 1 Solutions to two-egg problem

We label the answer to this question $\left\langle {n \atop k} \right\rangle$ (read "n drop k") and refer to the collection of all such numbers as the *egg-drop numbers*. Our goal is to characterize and compute these numbers, directly if possible.

The heart of the solution to our original problem lies in the recurrence behavior. When the first egg broke, we had only one egg left and needed to resort to the one-egg solution. We can use this idea to construct a recurrence relation for $\left\langle {n \atop k} \right\rangle$. For when we begin with k eggs and n available drops, after the first drop, we either still have k eggs (the egg survived) or we have $k-1$ eggs (the egg broke). In either case, we have $n-1$ drops available. This reasoning, as illustrated in FIGURE 2, provides our recurrence relation:

$$\left\langle {n \atop k} \right\rangle = \left\langle {n-1 \atop k} \right\rangle + \left\langle {n-1 \atop k-1} \right\rangle + 1 \quad \text{for all } n \geq 1, k \geq 1,$$

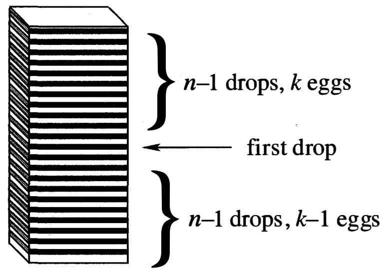


Figure 2 Motivating the recurrence relation

with

$$\left\langle {0 \atop k} \right\rangle = 0 \quad \text{for all } k \geq 0, \qquad \left\langle {n \atop 0} \right\rangle = 0 \quad \text{for all } n \geq 0.$$

The boundary conditions in the recurrence relation come from the fact that we cannot do anything without drops or eggs.

We now can recursively calculate the egg-drop numbers. Table 2 gives some of their values. You may recognize some of these numbers. For instance, the $k = 2$ column is made up of the triangular numbers (as can be seen from the solution to the two-egg problem). We would like be able to directly calculate egg-drop numbers. We illustrate two approaches to this problem.

TABLE 2: Some egg-drop numbers $\langle {n \atop k} \rangle$

$n \backslash k$	0	1	2	3	4	5	6	7	8
0	0	0	0	0	0	0	0	0	0
1	0	1	1	1	1	1	1	1	1
2	0	2	3	3	3	3	3	3	3
3	0	3	6	7	7	7	7	7	7
4	0	4	10	14	15	15	15	15	15
5	0	5	15	25	30	31	31	31	31
6	0	6	21	41	56	62	63	63	63
7	0	7	28	63	98	119	126	127	127
8	0	8	36	92	162	218	246	254	255

Generating Functions Given an infinite sequence $\{a_k\}_{k=0}^{\infty}$, the *ordinary power series generating function* (*ogf*) of the sequence is the symbolic power series

$$g(x) = \sum_{k=0}^{\infty} a_k x^k = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k + \cdots$$

*ogf*s are used extensively in combinatorics to find or solve recurrence relations and to discover relations between sequences. They are especially useful because we can manipulate them symbolically, in the ring of formal power series. Thus, we do not need to concern ourselves with issues of convergence. Wilf [**2**] gives a comprehensive study of the use of *ogf*s.

Because the egg-drop numbers are doubly-indexed, with n and k, we can construct a one-parameter sequence of generating functions, $\{g_n\}$. For each $n \geq 1$,

$$g_n(x) = \sum_{k=0}^{\infty} \left\langle {n \atop k} \right\rangle x^k = \left\langle {n \atop 0} \right\rangle + \left\langle {n \atop 1} \right\rangle x + \left\langle {n \atop 2} \right\rangle x^2 + \cdots + \left\langle {n \atop j} \right\rangle x^j + \cdots$$

Note that the constant term $\langle {n \atop 0} \rangle$ drops off as it is 0.

We begin by showing that our generating functions satisfy a recurrence relation related to the recurrence relation that defines the egg-drop numbers. We offer two proofs of this lemma to illustrate various ways to work with *ogf*s.

LEMMA 1. *For each* $n \geq 1$, $g_n(x) = (1+x)g_{n-1}(x) + g_1(x)$

Proof.

$$\begin{aligned} g_n(x) - g_{n-1}(x) &= \sum_{k=1}^{\infty} \left\langle {n \atop k} \right\rangle x^k - \sum_{k=1}^{\infty} \left\langle {n-1 \atop k} \right\rangle x^k = \sum_{k=1}^{\infty} \left(\left\langle {n \atop k} \right\rangle - \left\langle {n-1 \atop k} \right\rangle \right) x^k \\ &= \sum_{k=1}^{\infty} \left(\left\langle {n-1 \atop k-1} \right\rangle + 1 \right) x^k = \sum_{k=1}^{\infty} \left\langle {n-1 \atop k-1} \right\rangle x^k + \sum_{k=1}^{\infty} x^k \\ &= x \cdot g_{n-1}(x) + g_1(x) \end{aligned}$$

■

Alternate Proof. This more general method, discussed by Wilf [**2**] takes us from the recurrence relation for $\langle {n \atop k} \rangle$ to a recurrence relation for the sequence of generating functions. We multiply each term of the recurrence relation by x^k and sum over all k for which this recurrence is valid. In our case, for fixed $n \geq 1$,

$$\left\langle {n \atop k} \right\rangle = \left\langle {n-1 \atop k} \right\rangle + \left\langle {n-1 \atop k-1} \right\rangle + 1, \quad \text{for } k \geq 1.$$

Thus,

$$\sum_{k=1}^{\infty} \left\langle {n \atop k} \right\rangle x^k = \sum_{k=1}^{\infty} \left\langle {n-1 \atop k} \right\rangle x^k + \sum_{k=1}^{\infty} \left\langle {n-1 \atop k-1} \right\rangle x^k + \sum_{k=1}^{\infty} x^k$$

And hence

$$g_n(x) = g_{n-1}(x) + x g_{n-1}(x) + g_1(x) = (1+x)g_{n-1}(x) + g_1(x)$$

■

We can use this recurrence relation to find a closed form for each generating function. The proof of the following lemma, left to the reader, is a good exercise in induction.

LEMMA 2. *For each* $n \geq 1$,

$$g_n(x) = \frac{(1+x)^n - 1}{1-x}$$

Now that we know the generating functions, $g_n(x)$, we obtain our main result.

THEOREM. $\left\langle {n \atop k} \right\rangle = \sum_{j=1}^{k} \binom{n}{j}$

Proof. On the one hand,

$$g_n(x) = \left\langle {n \atop 1} \right\rangle x + \left\langle {n \atop 2} \right\rangle x^2 + \cdots + \left\langle {n \atop n} \right\rangle x^n + \left\langle {n \atop n+1} \right\rangle x^{n+1} + \cdots$$

On the other hand,

$$
\begin{aligned}
g_n(x) &= \frac{1}{1-x}\cdot\big((1+x)^n - 1\big)\\
&= (1+x+x^2+x^3+\cdots)\left(\binom{n}{1}x+\binom{n}{2}x^2+\cdots+\binom{n}{n}x^n\right)\\
&= \binom{n}{1}x+\left[\binom{n}{1}+\binom{n}{2}\right]x^2+\cdots+\left[\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}\right]x^n\\
&\quad+\left[\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}\right]x^{n+1}\\
&\quad+\left[\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}\right]x^{n+2}+\cdots
\end{aligned}
$$

Thus, g_n is also the generating function of $\sum_{j=1}^{k}\binom{n}{j}$. But two sequences with the same generating function are equal. ■

Direct counting approach Given the simplicity of the formula for the egg-drop numbers, one might suspect that there is a direct counting technique we could use. Indeed, there is. Consider a specific sequence of at most n drops with k eggs. Each of the drops has two possible outcomes: either the egg breaks or the egg does not break. Let 0 represent a drop without a break and 1 a drop with a break. Our sequence of drops thus yields a binary word of length at most n and having between zero and k 1s.

With this representation of drops, each word corresponds to a unique floor. For example, in the case of 8 drops and 2 eggs, the word 01001 corresponds to floor 11. In the general case, we make words of length $m \leq n$ have length n by adding $n - m$ trailing zeros. Note that we only need do this for words that have exactly k 1s, since the efficiency of our procedure ensures that words with fewer than k broken eggs are guaranteed to use all n drops. Also note that the trailing zeros are merely placeholders and do not represent drops.

Thus, there is a one-to-one correspondence between the floors in our building and the number of binary words of length n with at least one and no more than k 1s. The latter is easily shown to be $\sum_{j=1}^{k}\binom{n}{j}$ while the former is $\left\langle {n \atop k} \right\rangle$.

Conclusion Our main theorem is both exciting and disheartening. For we have found a beautiful characterization of the egg-drop numbers. But, it is well known that there is no closed form (that is, direct formula) for the partial sum of binomial coefficients **[1]**. Alas, we cannot calculate the egg-drop numbers without a tedious recursive calculation. Nevertheless, students may find the egg-drop problem a fun way to learn the power of generating functions and other basic combinatorial tools.

Acknowledgment. Sincere thank you to the referees for their suggestions for improvement.

REFERENCES

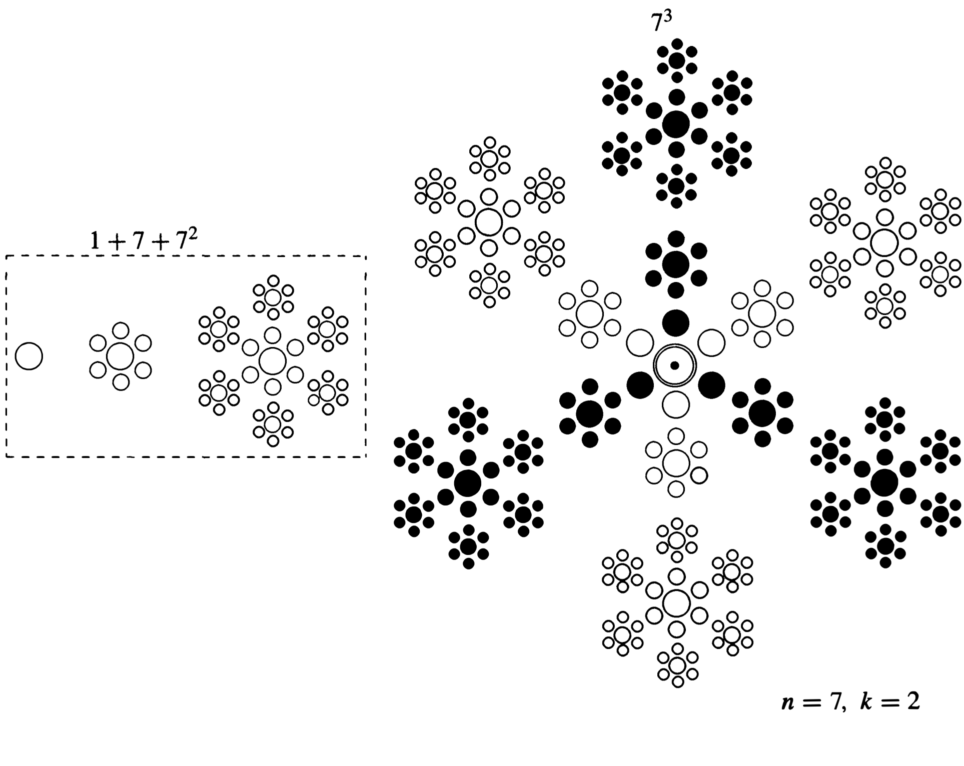
1. M. Petkovsek, H. Wilf, and D. Zeilberger, $A = B$, A. K. Peters, Wellesley, MA, 1996.
2. H. Wilf, *Generatingfunctionology*, Academic Press, Philadelphia, PA, 1994.
3. J. Konhauser, D. Velleman, and S. Wagon, *Which Way Did the Bicycle Go?*, Dolciani Mathematical Expositions #18, MAA, Washington, D.C., 1996.

Proof Without Words: Sums of Consecutive Powers of n via Self-Similarity

MINGJANG CHEN
Center for General Education
National Chiao Tung University
Hsinchu, Taiwan
mjchen@mail.nctu.edu.tw

For any nonnegative integers $n \geq 4$ and $k \geq 0$,

$$1 + n + n^2 + \cdots + n^k = \frac{n^{k+1} - 1}{n - 1}.$$



REFERENCES

1. David B. Sher, Proof without words: $3^n - 1 = \sum_{k=0}^{n-1} 2 \times 3^k$, *Math. Comput. Ed.* **31**:2 (1997), 190.
2. ———, Proof without words: $1 + \sum_{x=0}^{k}(n-1)n^x = n^{k+1}$, *Math. Comput. Ed.* **32**:2 (1998), 51.
3. ———, Sums of powers of three, *Proofs Without Words II*, MAA (2000), p. 108.
4. Rick Mabry, Mathematics without words, *College Math. J.*, **32**:1, (2001), 19.

When Are Two Subgroups of the Rationals Isomorphic?

FRIEDRICH L. KLUEMPEN
ETS
Princeton, NJ 08541
fkluempen@ets.org

DENISE M. REBOLI
King's College
Wilkes-Barre, PA 18711
dmreboli@kings.edu

To Reinhold Baer on his hundredth birthday, July 22, 2002.

Since most of us have studied fractions in various ways since elementary school, it seems that almost any question about them could be answered with little effort. Consider the following question: "Given two subgroups of the rationals under addition, can we decide whether these subgroups are isomorphic?" Given arbitrary groups, it is typically easier to determine that they are *not* isomorphic than to show that they are. Some of the features at our disposal to show that two groups cannot be isomorphic are element orders, group orders, and certain characteristic subgroups—the center of the group, the commutator subgroup, etc.—which must match up precisely in isomorphic groups.

When we consider additive subgroups of the rationals, we lose all of these typical approaches: Every nonzero element has infinite order (a group with this property is said to be *torsion-free*); since the rationals are commutative under addition, the center of a subgroup is the subgroup itself and commutator subgroups are trivial. Consider, for example, G_1, the subgroup consisting of the rationals with square-free denominators, and G_2, the subgroup consisting of the rationals with square-free odd denominators. The subgroups G_1 and G_2 are generated by reciprocals of primes and odd primes, respectively. Each element in these groups is a finite linear combination of the said reciprocals with integer coefficients. We write $G_1 = \langle 1/2, 1/3, 1/5, \ldots \rangle$, and similarly, $G_2 = \langle 1/3, 1/5, 1/7, \ldots \rangle$. Since $1/2 \notin G_2$, G_2 is a proper subgroup of G_1 (written $G_2 < G_1$). To show that G_1 is isomorphic to G_2, we will find a mapping between the groups and then prove that the mapping is an isomorphism. To establish this mapping, we look at the solvability of equations in the two groups. For example, the equation $3x = 1$ is solvable in both groups, since there is a unique element x in each group that satisfies the equation. (The uniqueness of solutions to equations in torsion-free abelian groups is important and we make repeated use of it throughout the discussion.)

This avenue will lead us to the goal of deciding whether subgroups of the rationals are isomorphic, but there is work to be done en route. In particular, for our examples G_1 and G_2, the equation $2x = 1$ is solvable in G_1 but not in G_2, an observable difference between these subgroups. To see that this does not necessarily mean that the subgroups are nonisomorphic, consider the following: The groups $\langle 1/2 \rangle$ and $\langle 1/3 \rangle$ are two subgroups of $\mathbb{Q}$ that are both infinite cyclic and therefore isomorphic. However, the equation $2x = 1$ has a solution in $\langle 1/2 \rangle$, but not in $\langle 1/3 \rangle$. We conclude that the solvability or nonsolvability of particular equations in isolated instances is not enough to prove that two groups are not isomorphic. In the groups G_1 and G_2, the only equations of the form $mx = 1$ that are solvable in G_1 and not in G_2 are those with $m = 2^i \cdot j$, where i and j are positive integers. We will need better tools to fully understand the structure of subgroups of the rationals.

Reinhold Baer **[1]** gave an elementary and elegant solution to our problem. Unfortunately, the result is part of a more general discussion (see Fuchs **[3, 4]**) that is presented in language better suited for a more sophisticated audience. We discuss the portions of Baer's result that are pertinent here so that this result might be examined in any undergraduate course in abstract algebra. Baer solved the isomorphism problem for subgroups of the rationals by finding a complete set of invariants. An invariant of groups is a function that takes on the same value for isomorphic groups (see Macdonald **[5]**), and a set of invariants is complete if two groups having the same invariants are isomorphic. The set of invariants used by Baer includes the concept of *height*, which will be developed in the next section. Before turning to the discussion of height, we refine our given problem by observing the structure of some of the subgroups of the rationals.

THEOREM 1. *Every finitely generated subgroup of the rationals is cyclic. (We call such groups locally cyclic.)*

Proof. Consider $S = \langle r/s, t/u \rangle$. We wish to show that S is cyclic. Consider the cyclic group $\langle 1/su \rangle$. Since $r/s = ru/su$ and $t/u = ts/su$, $S \leq \langle 1/su \rangle$. Since S is a subgroup of a cyclic group, it must be cyclic. By induction it follows that $\langle a_1, \ldots, a_k \rangle$ is cyclic for rationals $a_1, a_2, \ldots, a_k$. We ask the reader to supply the needed justification for the inductive step. (This is one of several problems we offer the reader at the end of the paper.) ∎

This result answers our question for all finitely generated subgroups of the rationals. To answer the question for infinitely generated subgroups, we must develop some additional ideas.

The height of an element The solvability of particular equations will help us answer our question, particularly in the case of infinitely generated subgroups. To work with the solvability question, we develop the concept of *height*. Since this concept is rarely discussed at the undergraduate level, we introduce it here by example.

EXAMPLE. *Let* $A_{2^\infty} = \langle u/2^k \mid u \in \mathbb{Z},\ k = 0, 1, 2, \ldots \rangle$ *and consider* $a = 45/8$.

The *2-height* of a in the group A_{2^∞} is the largest integer k for which $2^k x = 45/8$ has a solution in A. So to find the 2-height of a in A_{2^∞}, we must find an element x of A_{2^∞} for which $2^k x = 45/8$ has a solution. Since $x = 45/2^{k+3}$ will give a solution for any nonnegative integer k, we say that the 2-height of a is infinite. Similarly, to find the 3-height, we want to find the largest integer k for which $3^k x = 45/8$ has a solution in A_{2^∞}. The reader can check that $9x = 45/8$ has a solution in A_{2^∞}, while $27x = 45/8$ does not. We conclude that the 3-height of the element is 2. Similar calculations will show that the 5-height of a is 1 and that the p-height for any prime other than 2, 3, or 5 is 0. This leads us to the formal definition of p-height.

DEFINITION 2. *Let A be an additive subgroup of $\mathbb{Q}$ (written $A \leq \mathbb{Q}$), $a \in A$, k a nonnegative integer, and p a prime number. Then the p-height of a in A is k if $p^k x = a$ is solvable in A and $p^{k+1} x = a$ is not. If $p^k x = a$ has a solution for every k, then we say that the p-height of a is infinite, where we note that the p-height of 0, the identity element, is defined to be ∞.*

We note that the p-height of an element depends on the group in which it is considered. For example, the 3-height of the element $a = 45/8$ in $\mathbb{Q}$ is ∞, since $x = 45/(8 \cdot 3^k)$ will give a solution for any k.

Rather than focus on individual primes, we wish to keep track of a list that gives the p-height of an element for each prime p. Some notation will help with the bookkeeping. If $A \leq \mathbb{Q}$, $a \in A$, and $p_1 < p_2 < p_3 < \cdots$ is the list of all the primes, define $H_i(a)$ to be the p_i-height of a in A. Furthermore, define the *height* of a in A, denoted $H_A(a)$ or $H(a)$ if the subgroup A is clear from context, to be the infinite tuple $H(a) = (H_1(a), H_2(a), \ldots, H_i(a), \ldots)$. Then $H(a) = H(b)$ if and only if $H_i(a) = H_i(b)$ for every $i \geq 1$. Also, we say $H(a) \leq H(b)$ if $H_i(a) \leq H_i(b)$ for all i.

Let us apply this notation to the example given at the beginning of this section: Since $p_1 = 2$, $p_2 = 3$, and $p_3 = 5$, our calculations for the example show that for $a = 45/8$, $H_1(a) = \infty$, $H_2(a) = 2$, and $H_3(a) = 1$, while $H_i(a) = 0$ for $i \geq 4$. As a result, $H(a) = (\infty, 2, 1, 0, 0, 0, \ldots)$.

Now let's find out what happens to height under the group operation. Given a prime q, consider qa, the sum of q copies of the element a. Since the solvability of equations involving primes not equal to q is unchanged, the p-height of qa is identical to the p-height of a for those primes. If the q-height of a is k, where $k < \infty$, then by our definition, $q^k x = a$ has a solution in A, but $q^{k+1}x = a$ does not. Applying this definition to the element qa, we see that $q^{k+1}x = qa$ is solvable in A, so that the q-height of qa is at least $k+1$. Assuming the height is greater than $k+1$, we have $q^{k+2}x = qa$ is solvable in A, so that $q(q^{k+1}x - a) = 0$ for some x. Since nonzero elements of $\mathbb{Q}$ have infinite order, this means that $q^{k+1}x - a = 0$. Thus, $q^{k+1}x = a$ has a solution in A. But this means that the q-height of a is greater than k, contradicting our initial assumption; we conclude that the q-height of qa is $k+1$. If the q-height of a is infinite, then the q-height of qa is infinite as well. This argument can be applied repeatedly, either for powers of q, or for a product of primes. We conclude the following for ma, the sum of m copies of a. If $m = p_{i_1}^{\alpha_1} p_{i_2}^{\alpha_2} \ldots p_{i_r}^{\alpha_r}$, for α_i positive integers and $H(a) = (k_1, k_2, \ldots)$, then $H(ma) = (k_1, k_2, \ldots, k_{i_1} + \alpha_1, \ldots, k_{i_r} + \alpha_r, k_{i_r+1}, \ldots)$, where $k_i + \alpha = \infty$ if $k_i = \infty$.

In making the observation regarding qa above, we applied the definition of height to reach the desired contradiction. Examining the statements carefully leads to the observation that for $k_{i_1}, k_{i_2}, \ldots, k_{i_l}$ positive integers, then $p_{i_1}^{k_{i_1}} p_{i_2}^{k_{i_2}} \ldots p_{i_l}^{k_{i_l}} x = a$ is solvable in A if and only if $k_{i_1} \leq H_{i_1}(a), k_{i_2} \leq H_{i_2}(a), \ldots,$ and $k_{i_l} \leq H_{i_l}(a)$. This observation is stated as a lemma in the problem section (Problem 2) and we encourage the reader to supply the proof. Armed with these facts about height, we define an equivalence relation as follows.

DEFINITION 3. *Given*

$$A \leq \mathbb{Q}, \quad a, b \in A, \quad H(a) = (k_1, k_2, \ldots) \quad \textit{and} \quad H(b) = (l_1, l_2, \ldots),$$

we say that $H(a)$ is equivalent to $H(b)$ and write $H(a) \sim H(b)$ if $k_i = l_i$ for all but finitely many indices i and in case $k_i \neq l_i$, then both k_i and l_i are finite.

The proof that this is an equivalence relation is left as another of the exercises. To clarify the definition, observe, for example,

$$(\infty, \infty, 0, 0, 0, \ldots) \sim (\infty, \infty, 9, 2, 0, 0, \ldots),$$

since the coordinates that differ are finite and appear in a finite number of positions. On the other hand, $(\infty, \infty, 0, 0, 0, \ldots) \not\sim (\infty, 1, 1, 1, \ldots)$ for two reasons: the heights differ for infinitely many indices and they differ at the second coordinate, where the first height is infinite and the second is finite.

Given any equivalence relation, the associated equivalence classes deserve some attention. For $A \leq \mathbb{Q}$ and $a \in A$, define $T(a) = \{H(b) \mid H(a) \sim H(b), b \in A\}$ to be the equivalence class of $H(a)$ under $\sim$, which we refer to as the *type* of a. We have already observed that multiplying an element a by a positive integer m only affects a finite number of finite indices in $H(a)$, so that $H(ma) \sim H(a)$. Given any nonzero elements a, b of A, we write $a = k/l$ and $b = m/n$. Simple algebra shows that $(lm)a = (nk)b$ and it follows that $H(a) \sim H(lma) = H(nkb) \sim H(b)$. Hence any two nonzero elements have the same type and accordingly there is only one equivalence class for a given subgroup A of $\mathbb{Q}$. We may write $T(A)$ without any ambiguity and refer to the *type of* A. From these observations, we note that the type of a subgroup is determined by finding the height of an arbitrary nonzero element. For example, consider G_1 and G_2 as defined above. We have $H_{G_1}(1) = (1, 1, \ldots)$ and $H_{G_2}(1) = (0, 1, 1, \ldots)$. Thus $H_{G_1}(1) \sim H_{G_2}(1)$ and $T(G_1) = T(G_2)$. We also note here that $T(\mathbb{Q}) = \{(\infty, \infty, \ldots)\}$, while $T(\mathbb{Z}) = \{(0, 0, \ldots)\}$. We will show type to be an invariant, but is it a complete invariant? If we find that type is a complete invariant, we will be able to decide whether two subgroups of the rationals are isomorphic by determining their type.

Subgroups of $\mathbb{Q}$ Having developed the concept of type, we will apply it to subgroups of the rational numbers. But first, what does it mean for two subgroups to have the same type? Interpreting type in terms of solvability of equations, we can say that if two subgroups of the rationals have the same type, then an equation that is solvable in one subgroup, but not in the other does not happen too often (that is, at most finitely many times). We now have the tools to prove that type determines the isomorphism class of a subgroup.

THEOREM 4. *If A, $B \leq \mathbb{Q}$, then $T(A) = T(B)$ if and only if $A \cong B$.*

Proof. Suppose first that $A \cong B$. Denoting this isomorphism with ϕ, let $\phi(a) = b$ for an arbitrary nonzero element $a \in A$. If $p_i^k x = a$ is solvable in A, then $\phi(p_i^k x) = \phi(a)$ is solvable in B. However, the previous equation may be rewritten as $p_i^k \phi(x) = b$, and it follows that $H_i(a) \leq H_i(b)$. Since ϕ is an isomorphism, the above argument may be reversed and we conclude that $H_i(a) = H_i(b)$ for all i. Equivalently, we can say that $H(a) = H(b)$, or that $T(A) = T(B)$, as desired.

To prove the converse, consider any arbitrary nonzero elements $a' \in A$ and $b' \in B$. Since $T(A) = T(B)$ by assumption, we have that $H(a') \sim H(b')$. Letting $H(a') = (k'_1, k'_2, \ldots)$ and $H(b') = (l'_1, l'_2, \ldots)$, we obtain $k'_i = l'_i$ for all but finitely many entries and that the entries that differ are finite. Suppose then, that $k'_{i_1}, \ldots, k'_{i_s}$ and $l'_{i_1}, \ldots, l'_{i_s}$ are the entries that differ.

By the definition of height, $p_{i_1}^{k'_{i_1}} \cdots p_{i_s}^{k'_{i_s}} x = a'$ has a solution in A. Let $a \in A$ be that solution and suppose $H(a) = (k_1, k_2, \ldots)$ is the height of a in A. Using the definition of height, it can be shown that $k_j = 0$ for $j \in \{i_1, \ldots, i_s\}$ and that $k_j = k'_j$ otherwise. In the same manner, we can find $b \in B$, which is a solution to the equation $p_{i_1}^{l'_{i_1}} \cdots p_{i_s}^{l'_{i_s}} x = b'$. If $H(b) = (l_1, l_2, \ldots)$, we can show that $l_j = 0$ for $j \in \{i_1, \ldots, i_s\}$ and that $l_j = l'_j$ otherwise. We conclude that $H(a) = H(b)$.

Using the nonzero elements $a \in A$ and $b \in B$, we will obtain an isomorphism from A to B. To do this, consider the following observations. Given any nonzero element x of A, we can find nonzero integers m and t depending on x, such that $mx = ta$. Utilizing that $H(a) = H(b)$, and subsequently $H(ta) = H(tb)$, it follows that the equation $my = tb$ is solvable in B if and only if the equation $mx = ta$ is solvable in A. The justification also requires the use of the lemma stated in Problem 2 at the end. This sets up a correspondence between elements x of A and y of B, which we will denote by $\phi(x) = y$.

We will show that this correspondence is onto, one-to-one, and addition-preserving. That ϕ is a function from A to B follows from the fact that a y corresponding to a given x is unique. For suppose y and y' correspond to the same x, then we have that $my = tb$ and $my' = tb$. This implies that $m = 0$ or that $y = y'$. Our assumption that $x \neq 0$ precludes the case that $m = 0$, allowing us to conclude that $y = y'$. The fact that $my = tb$ is solvable in B if and only if the equation $mx = ta$ is solvable in A establishes that the domain of ϕ is A, and that the correspondence of the elements is onto. To establish that ϕ is injective, suppose that for a nonzero element $x \in A$, we have that $\phi(x) = 0$. Given m and t for which $mx = ta$, we observe that both m and t are nonzero. Since $\phi(x) = 0$, we have that $m \cdot 0 = tb$. Since $b \neq 0$, we conclude that $t = 0$, which is a contradiction.

To show that ϕ preserves addition, consider $x, x' \in A$, along with their corresponding elements y and y' in B and the corresponding integers m and t, m' and t', respectively. In other words, given integers m'' and t'' such that $m''(x + x') = t''a$, we must demonstrate that $m''(y + y') = t''b$. Simple algebra will show that setting $m'' = mm'$ and $t'' = m't + mt'$ will yield the desired result. ∎

To illustrate the construction outlined in the proof of the previous theorem, we examine the groups G_1 and G_2. From previous calculations, we know that $H_{G_1}(1) = (1, 1, \ldots)$ and $H_{G_2}(1) = (0, 1, 1, \ldots)$. The first step in the process is to find two elements, one in each group, that have identical heights. Since the heights of the two groups only differ in the first coordinate, these elements are found as the solutions to the equations $2^1 x = 1$ in G_1 and $2^0 y = 1$ in G_2. The respective solutions are $x = 1/2$ and $y = 1$. The reader may check that indeed $H_{G_1}(1/2) = (0, 1, 1, \ldots)$, as desired. Given an arbitrary element r/s in G_1, we must find integers m and t for which $m \cdot r/s = (1/2) \cdot t$. Choosing $m = s$ and $t = 2r$ satisfies the equation. Using these integers m and t, we now look for a corresponding element y' in G_2 that satisfies the equation $m \cdot y' = 1 \cdot t$. Substituting the values for m and t, we see that $y' = 2r/s$ satisfies the equation. So for this pair of groups, the mapping ϕ constructed in the proof of THEOREM 4 is $\phi(x) = 2x$, which is easily seen to be an isomorphism from G_1 to G_2.

Examples of subgroups of $\mathbb{Q}$ As previously observed, $T(G_1) = T(G_2)$ so THEOREM 4 allows us to conclude that $G_1 \cong G_2$, even though G_2 is a proper subgroup of G_1. In THEOREM 1, we proved that any finitely generated subgroup of the rationals is cyclic. We now can give a separate proof of the result using THEOREM 4. In particular, if $S \leq \mathbb{Q}$ is a finitely generated subgroup and $x \in S$, then $H_S(x) = (k_1, k_2, \ldots, k_n, 0, 0, \ldots)$, where each k_i is finite. But then, $H_S(x) \sim (0, 0, 0, \ldots)$ and it follows that $T(S) = T(\mathbb{Z})$. By THEOREM 4 we conclude that $S \cong \mathbb{Z}$.

To complete our discussion, we give an affirmative answer to the following question. Given a type, can we find a subgroup A of $\mathbb{Q}$ for which this type can be realized?

Suppose that the given type is $(k_1, k_2, \ldots)$ and consider $A = \langle 1/p_i^{l_i} \mid l_i \leq k_i \text{ for } i = 1, 2, \ldots \rangle$. Now $1 \in A$, and using the definition of height, we can show that $H_A(1) = (k_1, k_2, \ldots)$. Since we have observed previously that all heights represent the same equivalence class, we conclude that $T(A) = \{(k_1, k_2, \ldots)\}$, as desired.

Problems for the Reader

1. Supply the induction in THEOREM 1.
2. LEMMA. *If $A \leq \mathbb{Q}$, $a \in A$, then for $k_{i_1}, k_{i_2}, \ldots, k_{i_l}$ positive integers,*

$$p_{i_1}^{k_{i_1}} p_{i_2}^{k_{i_2}} \cdots p_{i_l}^{k_{i_l}} x = a$$

is solvable in A if and only if $k_{i_1} \leq H_{i_1}(a)$, $k_{i_2} \leq H_{i_2}(a)$, …, and $k_{i_l} \leq H_{i_l}(a)$.

3. Prove that "$\sim$," as given in DEFINITION 3, is an equivalence relation.
4. Show that $T(\mathbb{Q}) = \{(\infty, \infty, \ldots)\}$ and show that $T(\mathbb{Z}) = \{(0, 0, \ldots)\}$.

5. A nonzero torsion-free abelian group A is said to be of rank 1, if for every pair of elements $x, y \in A$, there are integers r and s, not both zero, such that $rx = sy$. Prove that if A is a group of rank 1, then A is isomorphic to a subgroup of $\mathbb{Q}$ under addition.

Acknowledgment. The authors would like to thank Professor Luise-Charlotte Kappe for her suggestions and her input, and Craig Wright for his insightful comments.

REFERENCES

1. R. Baer, Abelian groups without elements of finite order, *Duke Math. J.* **3** (1937), 68–122.
2. R. A. Beaumont and R. S. Pierce, Torsion-free groups of rank two, *Mem. Amer. Math. Soc* **38** (1961), 41 pp.
3. L. Fuchs, *Abelian Groups*, Publishing House of the Hungarian Acad. Sciences, Budapest, 1958.
4. ———, *Infinite Abelian Groups II*, Academic Press, New York, 1973.
5. I. Macdonald, *The Theory of Groups*, Robert E. Krieger Publishing Company, Florida, 1988.
6. O. Mutzbauer, Klassifizierung torsionsfreier abelscher Gruppen des Ranges 2, *Rend. Sem. Mat. Univ. Padova* **55** (1976), 195–208.

Surfaces of Revolution in Four Dimensions

BENNETT EISENBERG
Lehigh University
Bethlehem, PA 18015
be01@lehigh.edu

Most calculus students have seen curves revolved around the x- and y-axes in three dimensions to form 2-dimensional surfaces of revolution. With a little imagination, we can also revolve curves around axes in four dimensions or *hyperspace* to form 3-dimensional *hypersurfaces of revolution.* Extending the formulas for surface area of surfaces of revolution to formulas for surface measure of hypersurfaces of revolution is a challenging exercise and can serve as an interesting introduction to the fourth dimension. It requires us to consider thought provoking questions such as: "What do such hypersurfaces look like?" and "What is meant by the surface area of a hypersurface?"

In this paper we describe two ways to generate a hypersurface of revolution. We then use a variety of heuristic arguments to find formulas for the 3-dimensional areas of the hypersurfaces of revolution and the 4-dimensional volumes of the corresponding solids of revolution. The formulas are generalizations of the formulas for surface area and volume in three dimensions, where the expressions corresponding to the circumference and area of a circle are replaced with expressions for the surface area and volume of a sphere. The formulas allow us to find volumes and lateral surface areas of hyperspheres, hypercylinders, and hypercones as well as some figures that do not come directly from geometry.

The validity of the heuristic arguments is confirmed by analytic proofs based on Fubini's theorem, the use of spherical coordinates, and standard formulas for the surface measure of a parametrically described 3-dimensional surface in four dimensions. The heuristic arguments show why the formulas are true; the analytic arguments prove that the formulas are correct. Each type of argument has its value.

Hypersurfaces of revolution about the x-axis All of our surfaces will be generated from the graph of a smooth nonnegative function $y = f(x)$ for $a \le x \le b$. We view the graph as sitting in the x-y plane embedded in $\mathbb{R}^3$ or $\mathbb{R}^4$. When we revolve the graph around an axis, we spawn new coordinate axes perpendicular to all previous ones. We then label the axes so as to distinguish the axis of revolution. For example, when the revolution is about the x-axis in $\mathbb{R}^3$, we label the axes x-y_1-y_2, whereas when the revolution is about the y-axis in $\mathbb{R}^3$, we label the axes x_1-x_2-y.

In calculus we revolve the graph about the x axis to form a surface of revolution defined implicitly by the equation $y_1^2 + y_2^2 = f^2(x)$. The cross section of this surface with the plane $x = t$ is a circle of radius $f(t)$. The corresponding solid of revolution is given by $y_1^2 + y_2^2 \leq f^2(x)$ and the corresponding cross section is a *disc* of radius $f(t)$.

If we move up one dimension to hyperspace, then by analogy, revolving the graph about the x-axis gives the *hypersurface of revolution about the x-axis,* defined implicitly by the equation $y_1^2 + y_2^2 + y_3^2 = f^2(x)$. The cross section of this hypersurface with a 3-dimensional plane perpendicular to the x-axis at $x = t$ is a sphere of radius $f(t)$. The corresponding solid (or hypersolid) is given by $y_1^2 + y_2^2 + y_3^2 \leq f^2(x)$ and the corresponding cross section is a *solid sphere* or *ball* of radius $f(t)$.

Returning to three dimensions, the standard disc method shows that the volume of a solid of revolution about the x-axis is the integral of its infinitesimal volume element $dV = \pi f^2(x)\,dx$. The heuristic behind this is that the infinitesimal solid is the Cartesian product of the disc of area $\pi f^2(x)$ and a line segment of length dx. The volume of this infinitesimal region is the product of its cross sectional area and its thickness. There are pictures of these discs in most calculus texts, where they look like coins having radius $f(x)$ and thickness dx.

More generally, the volume or measure of a Cartesian product of sets A and B is the product of the measures of the sets A and B. If A is a set in $\mathbb{R}^2$ and B is a set in $\mathbb{R}^1$, then the volume of $A \times B$ is the area of A times the length of B. The same principle applies for volumes in four dimensions. If A is a set in $\mathbb{R}^3$ and B is a set in $\mathbb{R}^1$, then the 4-dimensional volume of $A \times B$ is the 3-dimensional volume of A times the length of B.

It follows that the 4-dimensional volume element for the corresponding solid of revolution is given by $dV_4 = \frac{4}{3}\pi f^3(x)\,dx$, which is found by replacing the area of a circle with the volume of a solid sphere. The point to remember is that each of these solid spheres lies in a 3-dimensional hyperplane perpendicular to the x-axis, so for different values of x these spheres are parallel and do not overlap. All of these spheres are disjoint, just as the cross-sectional discs are disjoint in three dimensions (see FIGURE 1). In this figure the planes drawn perpendicular to the x-axis represent 3-dimensional planes and the closed curves represent ordinary spheres.

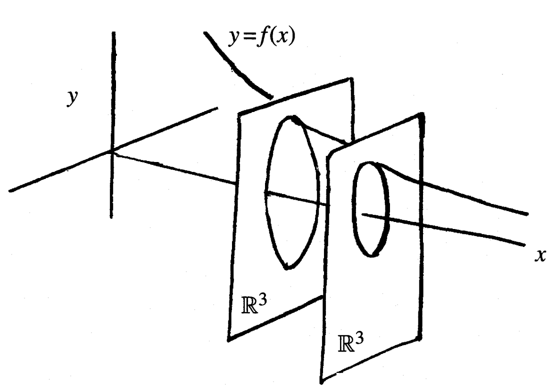


Figure 1 Spheres perpendicular to the x-axis

Integrating the differentials we find

$$V_4 = \int_a^b \frac{4}{3}\pi f^3(x)\,dx, \tag{1}$$

which generalizes the formula $V_3 = \int_a^b \pi f^2(x)\,dx$ from three dimensions.

Surface area Surface area is not as intuitive as volume. In $\mathbb{R}^3$, the area of a surface of revolution is found by breaking the surface into infinitesimal bands of surfaces of truncated cones. The cross section of the band is a circle of radius $f(x)$. The area element of the band in three dimensions is $dA = 2\pi f(x)\,ds$, where $2\pi f(x)$ is the circumference of the circle of radius $f(x)$ and $ds = \sqrt{1+(f'(x))^2}\,dx$ is an element of arclength.

The heuristic argument for this formula comes from viewing the infinitesimal band as made up of infinitesimal rectangles, each with one side of length $f(x)\,d\theta$ along the circular cross section and the other side of length ds and perpendicular to the circle, where the circle of radius $f(x)$ is parameterized by the angle θ. Viewing these regions as infinitesimal rectangles is similar to viewing infinitesimal polar rectangles as ordinary rectangles with sides $r\,d\theta$ and dr. The total area of the infinitesimal rectangles making up the infinitesimal band shown in FIGURE 2 is the circumference, $2\pi f(x)$, times ds.

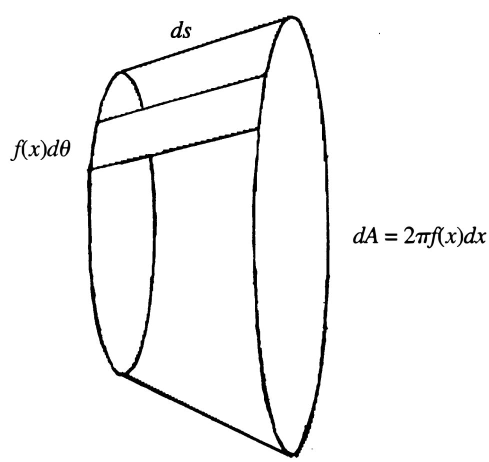


Figure 2 Infinitesimal rectangles in bands

One might be skeptical of these infinitesimal calculations. After all, the tiny regions are not really rectangles. Furthermore, we are using these infinitesimal rectangles to approximate another infinitesimal area. Nevertheless, these heuristics work; the conclusions can be verified by standard methods.

Applying a similar heuristic to the corresponding 3-dimensional area element of a hypersurface of revolution around the x-axis yields $dA_3 = 4\pi f^2(x)\,ds$, which is found by replacing the circumference of the cross section of the band with the surface area of a sphere of radius $f(x)$. The 3-dimensional infinitesimal bands of the hypersurface are made up of infinitesimal rectangular solids with one side on the sphere (as shown in FIGURE 1) and with an edge perpendicular to the sphere along the graph of the function and of length ds. The total volume of these rectangular solids is thus the area of the sphere times ds.

This gives the formula for the lateral surface area as

$$A_3 = \int_a^b 4\pi f^2(x)\sqrt{1+(f'(x))^2}\,dx. \tag{2}$$

This generalizes the formula $A_2 = \int_a^b 2\pi f(x)\sqrt{1+(f'(x))^2}\,dx$ in three dimensions. We use the word *lateral* because the areas of the sides of the corresponding solid are not included.

Hypersurfaces of revolution about the *y*-axis If we revolve the graph of $y = f(x)$ for $a \le x \le b$ about the y-axis in $\mathbb{R}^3$ we get a surface whose equation is

$$y = f\left(\sqrt{x_1^2+x_2^2}\right), \quad \text{for} \quad a \le \sqrt{x_1^2+x_2^2} \le b,$$

the graph of a function of two variables over an annulus. The corresponding solid of revolution is given by

$$y \le f\left(\sqrt{x_1^2+x_2^2}\right) \quad \text{for} \quad a \le \sqrt{x_1^2+x_2^2} \le b.$$

The volume of this 3-dimensional solid is found by breaking it up into infinitesimal shells of volume $dV = 2\pi x f(x)\,dx$, where $2\pi x\,dx$ is the area of an infinitesimal circular annulus at the base of the shell and $f(x)$ is the height of the shell.

If we move up a dimension, then by analogy revolving the graph about the y-axis gives a *hypersurface of revolution about the y-axis*, whose equation is

$$y = f\left(\sqrt{x_1^2+x_2^2+x_3^2}\right).$$

The corresponding solid (or hypersolid, for the purists) of revolution is given by

$$y \le f\left(\sqrt{x_1^2+x_2^2+x_3^2}\right) \quad \text{for} \quad a \le \sqrt{x_1^2+x_2^2+x_3^2} \le b.$$

To find the corresponding volume of this 4-dimensional solid, we break it up into infinitesimal regions that lie under the hypersurface and over the shells of spheres of radius x and thickness dx. In such a case the surface area of the sphere is $4\pi x^2$ and the thickness is dx. The volume element below the revolved function is then $dV_4 = 4\pi x^2 f(x)\,dx$. The volume is given by

$$V_4 = \int_a^b 4\pi x^2 f(x)\,dx. \tag{3}$$

This corresponds to the formula $V_3 = \int_a^b 2\pi x f(x)\,dx$ in three dimensions.

The surface area is found in a fashion similar to that for hypersurfaces about the x-axis, but this time the radius is x since the revolution is around the y-axis. This leads to the formula for the lateral surface area

$$A_3 = \int_a^b 4\pi x^2\sqrt{1+(f'(x))^2}\,dx. \tag{4}$$

This corresponds to the formula $A_2 = \int_a^b 2\pi x\sqrt{1+(f'(x))^2}\,dx$ in three dimensions.

Trying out the formulas The first three examples apply the formulas to functions that generate analogs of simple geometric figures.

Example 1: Hyperspheres. Revolve the graph of $f(x) = \sqrt{r^2 - x^2}$ for $-r \le x \le r$ about the x-axis in four dimensions to generate the surface $y_1^2 + y_2^2 + y_3^2 = r^2 - x^2$, a hypersphere of radius r. Using the formulas (1) and (2), we obtain the well-known [**4**, p. 488] results $V_4 = \pi^2 r^4/2$ and $A_3 = 2\pi^2 r^3$.

Developing an intuitive understanding of the hypersphere is not easy. It is a bounded region that is locally 3-dimensional and is called a *3-sphere* by topologists. Some have suggested that for any fixed time our 3-dimensional universe might really be a hypersphere [**2**, p. 167]. The expanding universe would then be described by hyperspheres with increasing radii. Popular science books sometimes oversimplify this by describing the expanding universe as having galaxies on the surface of a balloon that is being blown up.

In one important way our intuition doesn't fail us. Notice that the derivative of the volume of the hypersphere with respect to its radius r equals its surface area. This agrees with what happens in three dimensions for spheres and in two dimensions for circles and is the way we would expect volume and surface area to relate for spheres in any dimension.

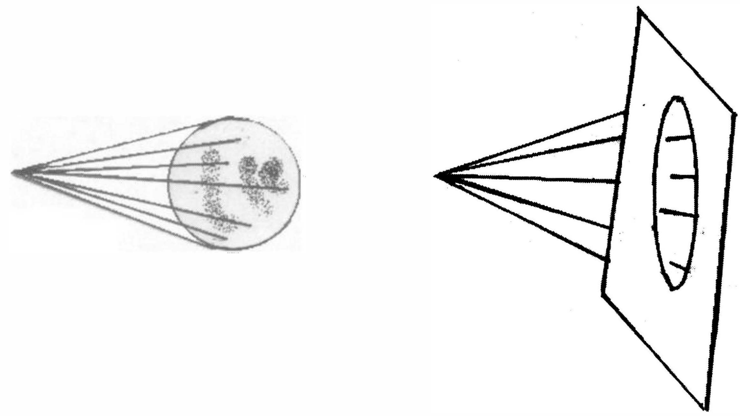
Example 2: Hypercylinders. Revolving the graph of $f(x) = r$ for $0 \le x \le h$ about the x-axis in four dimensions generates a 4-dimensional cylinder $y_1^2 + y_2^2 + y_3^2 = r^2$, for $0 \le x \le h$. The cross sections with the planes $x = t$ are ordinary 2-spheres of radius r.

The 2-spheres in the planes $x = 0$ and $x = h$ do not form the sides of the cylinder because they do not enclose it, just as circles do not form the sides of an ordinary cylinder. The sides of an ordinary cylinder are discs, not circles. Recall that the spheres are in 3-dimensional planes perpendicular to the x-axis. The axis itself is given by the equations $y_1 = y_2 = y_3 = 0$, and thus passes through the spheres without touching them just as the axis of a 3-dimensional cylinder passes right through the circles at the ends without touching them.

This is similar to the situation described in the the 19th-century book *Flatland* [**1**, pp. 89–90]. Davis [**2**, pp. 124–126] gives a different account, where an invader from the fourth dimension can enter an apparently closed region in three dimensions.

Using (1) and (2), we find the volume and lateral surface area to be $V_4 = \frac{4}{3}\pi r^3 h$ and $A_3 = 4\pi r^2 h$. These correspond to the formulas $\pi r^2 h$ and $2\pi r h$ for the volumes and lateral areas of cylinders in three dimensions of radius r and height h. The formulas are not surprising since the solid hypercylinder is the Cartesian product of a *ball* of radius r and an interval of length h. Its lateral surface is the Cartesian product of a *sphere* of radius r and an interval of length h. Thus its volume should be h times the volume of the ball and its surface area should be h times the surface area of the sphere.

Example 3: Hypercones. Revolving the graph of $f(x) = rx/h$ for $0 \le x \le h$ about the x-axis in four dimensions generates a hypercone $h^2(y_1^2 + y_2^2 + y_3^2) = r^2 x^2$. The cross section with the plane $x = h$ is a sphere of radius r. The hypercone consists of rays from the origin to points on the sphere at $x = h$. Rays from the origin to the points on a sphere in three dimensions would form a solid spherical cone in the shape of an ice cream cone, but in four dimensions it is just a hypersurface. The points on the surface are given by $\{t(h, y_1, y_2, y_3) : 0 \le t \le 1,\ y_1^2 + y_2^2 + y_3^2 = r^2\}$. A point (h, u_1, u_2, u_3) with $u_1^2 + u_2^2 + u_3^2 < 1$ (in the interior of the base sphere) is not on this surface. As with the cylinder, the x-axis, where $y_1 = y_2 = y_3 = 0$, passes right through the base sphere without touching it, as in FIGURE 3.



Rays hit the interior of the sphere. Rays miss the interior of the sphere.

Figure 3 Spherical cones in 3 and 4 dimensions

The volume of a cone in three dimensions is $\frac{1}{3}\pi r^2 h$, or $1/3$ of the area of the base circle times the altitude. The surface area is $\pi r L$ or $1/2$ of the circumference of the base circle times the slant height. Using (1) and (2), we find the volume and surface area of the hypercone to be $V_4 = \frac{1}{4}(\frac{4}{3}\pi r^3)h$ and $A_3 = \frac{1}{3}(4\pi r^2)L$, where $L = \sqrt{h^2 + r^2}$ is the slant height of the cone. In other words, the volume is $1/4$ of the volume of the base sphere times the altitude and the lateral surface area is $1/3$ of the surface area of the base sphere times the slant height.

The next example generalizes a standard example in calculus courses.

Example 4. Gabriel's Horn in the Fourth Dimension. If we revolve the graph of $y = x^{-1}$ for $1 \leq x < \infty$ about the x-axis in $\mathbb{R}^3$, we get what is called *Gabriel's Horn*. It has the paradoxical property that the corresponding solid has finite volume $\int_1^\infty \pi x^{-2}\,dx$, while the lateral surface has infinite area $\int_1^\infty 2\pi x^{-1}\sqrt{1 + x^{-4}}\,dx$. It is as if the horn could be filled with a finite amount of paint, but would require an infinite amount of paint to give its surface a uniform coating. The same paradox holds for $y = x^{-p}$ for $1/2 < p \leq 1$.

If we revolve the graph of $y = x^{-p}$ for $1 \leq x < \infty$ about the x-axis in four dimensions we form a hyperhorn with volume $\int_1^\infty \frac{4}{3}\pi x^{-3p}\,dx$ and lateral surface area $\int_1^\infty 4\pi x^{-2p}\sqrt{1 + p^2 x^{-2p-2}}\,dx$. This time the phenomenon of finite volume and infinite surface area occurs if $1/3 < p \leq 1/2$.

The last example involves revolution about the y-axis; in this case, the integrals have closed-form solutions.

Example 5. Revolve the graph of $y = \ln x$, for $1 \leq x \leq t$ about the y-axis in hyperspace. Using formulas (3) and (4), the reader may verify that $V_4 = \int_1^t 4\pi x^2 \ln x\,dx = \frac{4\pi}{9}(3t^3 \ln t - t^3 + 1)$ and $A_3 = \int_1^t 4\pi x^2\sqrt{1 + x^{-2}}\,dx = \frac{4\pi}{3}((1 + t^2)^{3/2} - 2^{3/2})$.

Proofs of the formulas for volume and surface area We have argued by analogy to arrive at our formulas for the volume and surface area differentials. It is now time to justify them rigorously.

The triple integral of a positive function may be viewed as a 4-dimensional volume. The volume for the solid of revolution of $y = f(x)$ for $a \leq x \leq b$ about the y-axis is then given by

$$\int_D f\left(\sqrt{x_1^2+x_2^2+x_3^2}\right) dx_1\, dx_2\, dx_3,$$

where D is the region

$$\left\{a \le \sqrt{x_1^2+x_2^2+x_3^2} \le b\right\}.$$

Changing to spherical coordinates in the x_1-x_2-x_3 space shows that

$$V_4 = \int_0^{2\pi}\int_0^{\pi}\int_a^b f(\rho)\rho^2 \sin\phi\, d\rho\, d\phi\, d\theta = \int_a^b 4\pi\rho^2 f(\rho)\, d\rho.$$

This gives (3).

In the case of revolution about the x-axis the formula for dV follows from Fubini's theorem. Let S be the solid whose volume is being computed and let 1_S be the indicator function of S. That is, $1_S(x, y_1, y_2, y_3) = 1$ if $(x, y_1, y_2, y_3) \in S$ and is 0 otherwise. Then

$$V_4(S) = \int_a^b \left(\int_{(y_1,y_2,y_3)} 1_S(x, y_1, y_2, y_3)\, dy_1\, dy_2\, dy_3\right) dx = \int_a^b V_x\, dx,$$

where V_t is the 3-dimensional volume of the cross section of S with the hyperplane $x = t$. Equation (1) follows since $V_x = (4/3)\pi f^3(x)$ for $a \le x \le b$.

The formulas for surface area require that the function $f(x)$ be continuously differentiable. The results then follow from the general result for the 3-dimensional surface area or surface measure of a parameterized surface in $\mathbb{R}^4$. Let $(\theta_1, \theta_2, \theta_3) \in E$, an open set in $\mathbb{R}^3$, and let $T : (\theta_1, \theta_2, \theta_3) \to (z_1, z_2, z_3, z_4)$ in $\mathbb{R}^4$ be a differentiable transformation. Then the surface area of the image $T(E)$ is given by

$$A_3\big(T(E)\big) = \int_E \sqrt{\det(\mathbf{B}'\mathbf{B})}\, d\theta_1\, d\theta_2\, d\theta_3, \tag{5}$$

where $\mathbf{B}$ is a matrix with $\mathbf{B}_{ij} = \partial z_i/\partial\theta_j$ for i =1 to 4, j=1 to 3 [**3**, p. 489].

The heuristic argument for this formula is similar to that for a parameterized surface in $\mathbb{R}^3$. The transformation from the parameter space to $\mathbb{R}^4$ takes the infinitesimal rectangular solid $d\theta_1\, d\theta_2\, d\theta_3$ into an infinitesimal parallelepiped in $\mathbb{R}^4$ formed by vectors $\mathbf{V}_1\, d\theta_1, \mathbf{V}_2\, d\theta_2, \mathbf{V}_3\, d\theta_3$, where $\mathbf{V}_j(i) = (\partial z_i/\partial\theta_j)$. Let $\mathbf{B}$ be the matrix of the form $[\mathbf{V}_1\ \mathbf{V}_2\ \mathbf{V}_3]$. The 3-dimensional volume of the parallelepiped is then $\sqrt{\det(\mathbf{B}'\mathbf{B})}\, d\theta_1\, d\theta_2\, d\theta_3$ [**3**, p. 472; **4**]. The surface $T(E)$ is approximated by these parallelepipeds in the same way that a surface in three dimensions is approximated by parallelograms. Thus the surface area of $T(E)$ should be given by (5).

In calculus courses, surfaces are parameterized by two parameters and $\sqrt{\det(\mathbf{B}'\mathbf{B})}$ is written as the magnitude of a cross product $|\mathbf{V}_1 \times \mathbf{V}_2|$. It is easily checked that this magnitude can also be written as the square root of a determinant as above with $\mathbf{B} = [\mathbf{V}_1\ \mathbf{V}_2]$. For curves parameterized by a single parameter θ_1, we have $\sqrt{\det(\mathbf{B}'\mathbf{B})} = |\mathbf{V}_1|$, so that the arc-length formula is given by a 1-dimensional form of (5).

If the surface of revolution is the graph of a function $F(x_1, x_2, x_3)$ of three variables, then from (5) the area is given by

$$\int_E \sqrt{1 + F_1^2 + F_2^2 + F_3^2}\, dx_1\, dx_2\, dx_3.$$

In the revolution of a graph $y = f(x_1)$ about the y-axis, we let

$$F(x_1, x_2, x_3) = f\left(\sqrt{x_1^2 + x_2^2 + x_3^2}\right).$$

Using spherical coordinates the integral becomes

$$\int_0^{2\pi}\int_0^{\pi}\int_a^b \sqrt{1 + (f'(\rho))^2}\,\rho^2 \sin\phi \, d\rho \, d\phi \, d\theta = \int_a^b 4\pi\rho^2\sqrt{1 + (f'(\rho))^2}\, d\rho.$$

Formula (4) follows.

For revolution about the x-axis, the surface $y_1^2 + y_2^2 + y_3^2 = f^2(x)$ can be parameterized using spherical coordinates as

$$x = t, \quad y_1 = f(t)\sin\phi\cos\theta, \quad y_2 = f(t)\sin\phi\sin\theta, \quad y_3 = f(t)\cos\phi$$

for $a \le t \le b$, $0 \le \phi \le \pi$, $0 \le \theta \le 2\pi$. Formula (4) follows after a straightforward calculation of the determinant in (5).

Epilogue Now that we have extended the results to four dimensions, we might consider extending them further to n dimensions. In fact, in higher dimensions there are other ways to revolve graphs about axes. For example, in four dimensions we could consider the double rotation, where the graph is first revolved about the y-axis in three dimensions and then the resulting surface is revolved about the x-axis in four dimensions.

REFERENCES

1. E. A. Abbott, *Flatland*, Harper and Row Publishers, NY, 1983.
2. D. Davis, *The Nature and Power of Mathematics*, Princeton University Press, Princeton, NJ, 1993.
3. J. H. Hubbard and B. B. Hubbard, *Vector Calculus, Linear Algebra, and Differential Forms*, Prentice Hall, Upper Saddle River, NJ, 1999.
4. G. J. Porter, k-volume in $\mathbb{R}^n$ and the Generalized Pythagorean Theorem, *Amer. Math. Monthly* **103** (1996), 252–256.

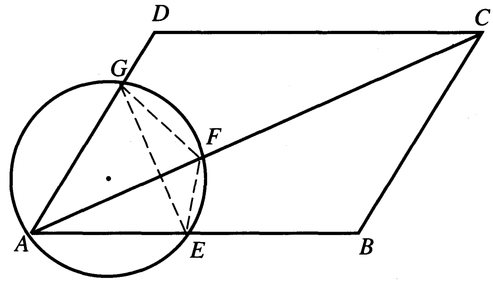
A Vector Approach to Ptolemy's Theorem

ERWIN JUST
NORMAN SCHAUMBERGER
Bronx Community College of the City University of New York
Bronx, NY 10453
erwinjust@verizon.net

Ptolemy's Theorem [**1**] states that the product of the diagonals of a cyclic quadrilateral (a quadrilateral that can be inscribed in a circle) is equal to the sum of the products of its opposite sides. Our purpose is to prove Ptolemy's Theorem by incorporating the use of vectors, an approach which we have never before seen. The notion that we might succeed in this effort occurred to us after observing that Ptolemy's Theorem may be used to prove the following result [**2**]:

THEOREM 1. *Suppose a circle contains point A of parallelogram ABCD and intersects side $\overline{AB}$, side $\overline{AD}$, and diagonal $\overline{AC}$ in points E, G, and F, respectively. Then*

$$|AF||AC| = |AE||AB| + |AG||AD|.$$



Proof. We readily deduce that $\triangle ABC \sim \triangle GFE$, from which it follows that

$$\frac{|AC|}{|GE|} = \frac{|AB|}{|FG|} = \frac{|AD|}{|EF|}.$$

We apply Ptolemy's Theorem to quadrilateral *AEFG* to obtain

$$|AF||GE| = |AE||FG| + |AG||EF|.$$

Multiplying through gives

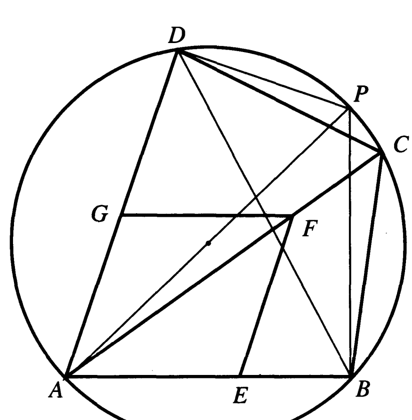
$$|AF||GE|\frac{|AC|}{|GE|} = |AE||FG|\frac{|AB|}{|FG|} + |AG||EF|\frac{|AD|}{|EF|},$$

which yields the required conclusion. ■

It was our hope that if we could avoid using Ptolemy's Theorem in the proof of Theorem 1, then perhaps we could use Theorem 1 to deduce Ptolemy's Theorem. By incorporating a vector approach, Theorem 1 can indeed be proved independently of Ptolemy's Theorem. This is described in the body of the proof of Theorem 2. (Subsequently, we found another proof of Theorem 1 that does not use Ptolemy's Theorem [**3**]). It turns out that, unlike in Theorem 1, none of the points of the parallelogram used in the proof of Theorem 2 need be exterior to the circle.

THEOREM 2. (PTOLEMY'S THEOREM) *Let ABCD be a cyclic quadrilateral. Then*

$$|AC||BD| = |AB||CD| + |AD||BC|.$$



Let F be a point on chord $\overline{AC}$. Let E and G be points on chords $\overline{AB}$ and $\overline{AD}$ (extended if necessary) such that quadrilateral $AEFG$ is a parallelogram. Let P be the point on the circle for which $\overline{AP}$ is a diameter. Then $\angle ABP$, $\angle ACP$, and $\angle ADP$ are right angles or, in the case that diameter $\overline{AP}$ coincides with one of the chords $\overline{AB}$, $\overline{AC}$, or $\overline{AD}$, two of these angles are right angles. In either circumstance, it follows that

$$\begin{aligned}|AF||AC| &= \overrightarrow{AP}\cdot\overrightarrow{AF} = \overrightarrow{AP}\cdot\left(\overrightarrow{AE}+\overrightarrow{AG}\right)\\ &= \overrightarrow{AP}\cdot\overrightarrow{AE}+\overrightarrow{AP}\cdot\overrightarrow{AG} = |AE||AB|+|AG||AD|.\end{aligned}$$

Since $|AG| = |EF|$, this can be rearranged as

$$|AC||BD|\frac{|AF|}{|BD|} = |AB||CD|\frac{|AE|}{|CD|} + |AD||BC|\frac{|EF|}{|BC|}. \tag{1}$$

Since $\angle FEA \equiv \angle BCD$ and $\angle AFE \equiv \angle DAC \equiv \angle DBC$, we have $\triangle AEF \sim \triangle BCD$, from which we obtain

$$\frac{|AF|}{|BD|} = \frac{|AE|}{|CD|} = \frac{|EF|}{|BC|}. \tag{2}$$

We can now conclude from (1) and (2) that

$$|AC||BD| = |AB||CD| + |AD||BC|,$$

which completes the proof.

REFERENCES

1. Nathan Altshiller Court, *College Geometry*, Barnes & Noble, 1952, p. 128.
2. H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited*, Random House, 1967, p. 43.
3. M. N. Aref and William Wernick, *Problems and solutions in Euclidean Geometry*, Dover Publications, 1968, p. 134, Problem 13.

Projected Rotating Polygons

DIMITRIOS KODOKOSTAS
Department of Computer Science
Technological Education Institute of Larissa
Larissa, 41110 Greece
dkodokostas@alumni.nd.edu
dkodokostas@teilar.gr

Take any number of lines $l_1, l_2, \ldots, l_n$ through the center of a circle; pick a point M of the circle and project it perpendicularly onto each line, creating points M_1, M_2, $\ldots$, M_n. Connect the points by segments to form a polygon $M_1M_2\ldots M_n$ as in FIGURE 1). Its shape depends on M, right? The surprising answer is that it does not!

The reader can have fun verifying this fact with the use of dynamic mathematical software like the *Geometer's Sketchpad*. Seeing the polygon dance around the circle can be a real joy.

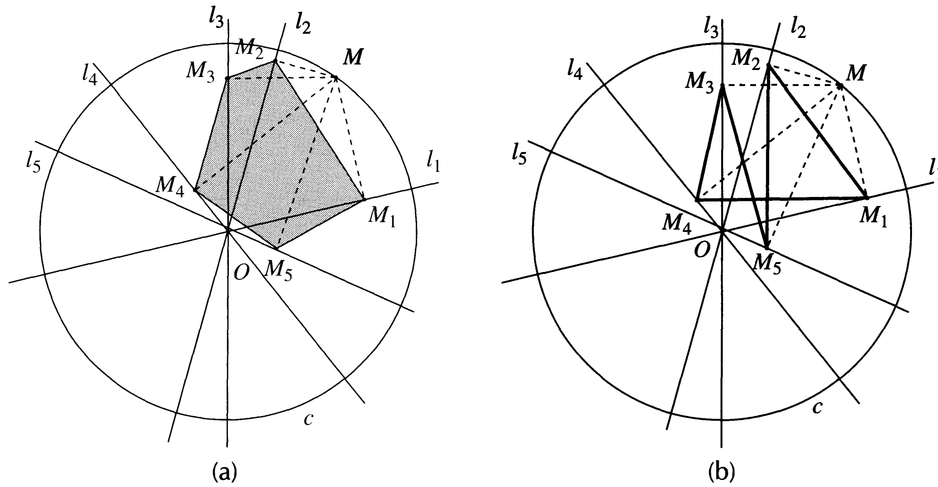


Figure 1 Polygons $M_1M_2M_3M_4M_5$ and $M_1M_2M_5M_3M_4$ retain their shape and size as M spins around c

To verify this fact rigorously, one can start with only two lines l_1 and l_2 through the center O of the circle c, as in FIGURE 2a, and show that the length of M_1M_2 does not change as M spins around c.

Indeed, the quadrilateral MM_1OM_2 is cyclic since $\angle OM_1M + \angle OM_2M = 180°$. So $M_1M_2 = 2R\sin(\angle M_1OM_2)$ where R is the radius of the circle inscribed about MM_1OM_2. But observe that OM is a diameter of this circle. So $2R = r$, thus $M_1M_2 = r/2\sin(\angle M_1OM_2)$, which is constant since $\angle M_1OM_2$ is one of the two supplementary angles of l_1, l_2 and the function sin has the same value on either of them.

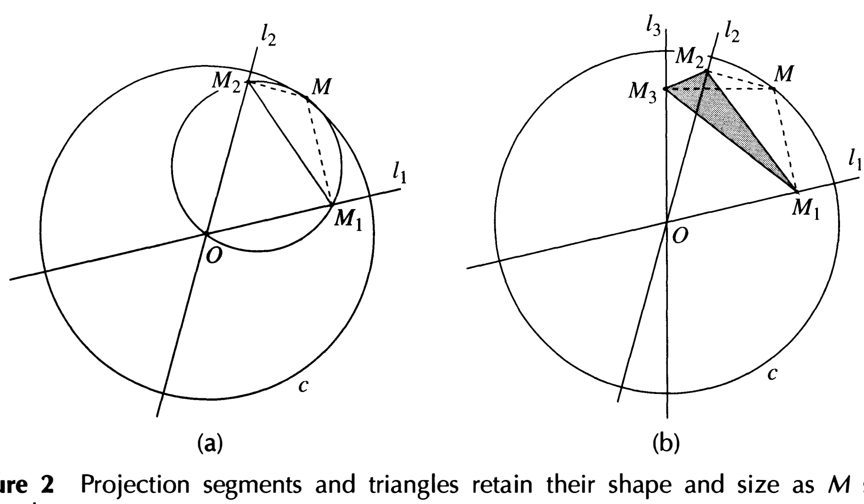


Figure 2 Projection segments and triangles retain their shape and size as M spins around c

Now if l_1, l_2, and l_3 are three lines through O, as in FIGURE 2b, then each one of the projection segments M_1M_2, M_2M_3, M_3M_1 retains its length as M moves around c, for the same reason as above. But then of course the triangle $M_1M_2M_3$ retains its shape and size as M moves around c!

Finally, for more than three lines $l_1, l_2, \ldots, l_n$ through O, consider the closed polygon $M_1M_2\ldots M_n$ formed by the projection points. It can either be convex (FIGURE 1a) or a self-intersecting (FIGURE 1b), but there is always a way to re-order the vertices to make it convex, because it is cyclic (just notice that all points M_i lie on the circle of diameter OM). In any case, this polygon always retains its shape and size as M moves around c, as follows:

If $M_1M_2 \ldots M_n$ is convex, the result follows immediately by using diagonals to dissect the polygon into triangles with diagonals and recalling the truth of the result for triangles.

If $M_1M_2 \ldots M_n$ is self-intersecting, we choose an ordering $M_{i_1}, M_{i_2}, \ldots, M_{i_n}$ of the points $M_1, M_2, \ldots, M_n$ so that the polygon $M_{i_1}M_{i_2} \ldots M_{i_n}$ is convex. Then the convex polygon $M_{i_1}M_{i_2} \ldots M_{i_n}$ retains its shape and size as M spins around c. Surely any linear figure formed by sides and diagonals of $M_{i_1}M_{i_2} \ldots M_{i_n}$ (for example, $M_1M_2 \ldots M_n$) also retains its size and shape as M spins around c and we are done.

With only slightly more extra care, similar results can be proved whenever the projections of the point M onto the lines through the center of the circle are not perpendicular but rather of a fixed signed angle.

Acknowledgment. I would like to thank the referees for their valuable comments.

The Probability That a Randomly Generated Quadratic Is Factorable

MICHAEL J. BOSSÉ
Department of Mathematics and Science Education
East Carolina University
Greenville, NC 27858
bossem@mail.ecu.edu

N. R. NANDAKUMAR
Delaware State University
Dover, DE 19901
nnandaku@desu.edu

The question is quite simple: Given a quadratic $Ax^2 + Bx + C$ with integer coefficients A, B, and C less than or equal to N in magnitude and such that $A \neq 0$, what is the probability, Pr_N, that the quadratic is factorable, assuming that A, B, and C are chosen randomly? Intuition leads one to assume that as N approaches ∞, the probability of factorability approaches zero. The proof of the following theorem will formally confirm this supposition.

MAIN THEOREM. *Let $A \neq 0$, B, and C be random integers such that $-N \leq A, B, C \leq N$ where N is a positive integer. Then*

$$\lim_{N\to\infty} \mathrm{Pr}_N = 0.$$

Setting up the problem What do we mean by a "randomly generated quadratic"? For the quadratic expression $Ax^2 + Bx + C$, let us randomly choose integers $A \neq 0$, B, and C, such that $-N \leq A, B, C \leq N$ where N is a positive integer. If we form ordered triplets with these integers, (A, B, C) and let $\mathcal{S}_N = \{(A, B, C) \mid -N \leq A, B, C \leq N \text{ and } A \neq 0\}$, then $\mathcal{S}_N$ is the sample space of all possible quadratics with integer coefficients less than or equal to N in magnitude. We assume that all of these are equally likely. Then let $\mathcal{F}_N = \{(A, B, C) \in \mathcal{S}_N | Ax^2 + Bx + C \text{ is factorable}\}$ be the set of all ordered triplets in $\mathcal{S}_N$ with the additional condition that $Ax^2 + Bx + C$ is factorable.

As before, we let $\Pr_N$ represent the probability that the quadratic $Ax^2 + Bx + C$ is factorable where $A \neq 0$, B, and C are randomly chosen integers between $-N$ and N. Then our definitions imply that

$$\Pr_N = \text{card}(\mathcal{F}_N)/\text{card}(\mathcal{S}_N).$$

Evaluating the denominator of this expression is elementary: Since $-N \leq A, B, C \leq N$ with $A \neq 0$, we have $2N$ choices for A, and $2N + 1$ choices for each of B and C; thus $\text{card}(\mathcal{S}_N) = (2N)(2N + 1)(2N + 1)$ and we have

$$\Pr_N = \frac{\text{card}(\mathcal{F}_N)}{(2N)(2N + 1)(2N + 1)}.$$

In order to demonstrate that $\Pr_N$ approaches 0 as N approaches infinity, we must find an upper bound for $\text{card}(\mathcal{F}_N)$.

Solving the problem By the quadratic formula, an integer quadratic $Ax^2 + Bx + C$ factors over the integers if and only if $B^2 - 4AC = D^2$ for some nonnegative integer D. Our overall plan is this: For each of the $2N + 1$ possible choices of B, we will find a bound on the number of A, C pairs for which $B^2 - 4AC$ is a perfect square. Our fundamental equation thus becomes $AC = (B^2 - D^2)/4$.

Since A and C are selected from the interval $[-N, N]$, all solutions must have $(B^2 - D^2)/4 \geq -N^2$. Thus, we must have $D^2 \leq B^2 + 4N^2 \leq 5N^2$, and consequently, $D \leq \sqrt{5}N < 3N$. This means that D can be chosen in fewer than $3N$ ways. (We could reduce this further by observing that B and D must also have the same parity, but this upper bound will do the trick.)

Next, letting $n = (B^2 - D^2)/4$, we need to count the number of ways to choose A and C from $[-N, N]$ so that $AC = n$. If $D = \pm B$, then $n = 0$, and thus A can be chosen $2N$ ways and C must be zero. Otherwise, $n \neq 0$ and A can be chosen $2d(n)$ ways, where $d(n)$ is the number of positive divisors of n. Once A is chosen, then the selection of C is forced to be n/A.

LEMMA. *If n has prime factorization $n = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$, then $d(n) = (1 + n_1) \cdot (1 + n_2) \cdots (1 + n_k)$ and $d(n) \leq 729n^{1/4}$.*

Proof of Lemma Observe that q is a divisor of n if and only if $q = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$, where $0 \leq m_k \leq n_k$ for all k. Since each m_k can be chosen $1 + n_k$ ways as an integer in the interval $[0, n_k]$, the first part of the lemma follows.

Next we can show by induction on m that for all $m \geq 0$,

$$(1 + m) \leq 3 \cdot 2^{m/4}.$$

This is clearly true for $m = 0, \ldots, 6$, and for $m \geq 6$, we can see that $(2 + m)^4 = ((2 + m)/(1 + m))^4 (1 + m)^4 \leq (7/6)^4 (1 + m)^4 < 2(1 + m)^4 < 3^4 \cdot 2^{m+1}$. Likewise, it is obvious that for $m \geq 0$,

$$(1 + m) \leq 2^m = 16^{m/4}.$$

Now let $n = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$, where we assume that $2 \leq p_1 < p_2 < \cdots < p_k$. (Consequently, for all $i \geq 7$, $p_i \geq 17$.) For $i \leq 6$ we have

$$(1 + n_i) \leq 3 \cdot 2^{n_i/4} \leq 3 \cdot p_i^{n_i/4}$$

and for $i \geq 7$,

$$(1 + n_i) \leq 16^{n_i/4} < p_i^{n_i/4}.$$

Thus, $d(n) = (1 + n_1)(1 + n_2) \cdots (1 + n_k) < 3^6 \prod_{i=1}^{k} p_i^{n_i/4} = 729n^{1/4}$, as asserted.

Proof of Main Theorem Counting the number of solutions to $AC = (B^2 - D^2)/4$, we have $2N + 1$ choices for B and at most $3N$ choices for D. Letting $n = (B^2 - D^2)/4$ we have, when $n \neq 0$, the number of ways to pick A and C is at most $2d(n) < 1458n^{1/4} < 729\sqrt{N}$, since $n \leq B^2/4 \leq N^2/4$. Hence, there are at most

$$(2N + 1)(3N)729\sqrt{N}$$

solutions when $n \neq 0$. For large N, this behaves like a constant times $N^{2.5}$, so we refer to it as $O(N^{2.5})$, indicating that it has *order* $N^{2.5}$. The case where $n = 0$ adds only $(2N + 1)(2)(2N)$ more solutions, a quantity of lower order than $N^{2.5}$.

Thus, card(F_N) is on the order of $N^{2.5}$. And since card(S_N) is $O(N^3)$, it follows that

$$\lim_{N \to \infty} \Pr_N = 0.$$

Remark. It has been shown that $d(n) = O(n^{\delta})$ for all positive δ where $d(n)$ is the number of divisors of an integer n [**1**, pp. 260–261]. One can use this estimate to arrive at the above result. However, we have presented a more elementary proof using combinatorial arguments.

Acknowledgment. The authors wish to thank the referees for their helpful suggestions in the presentation of this paper.

REFERENCE

1. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., Clarendon Press, Oxford, 1979.

Our Searchable Database

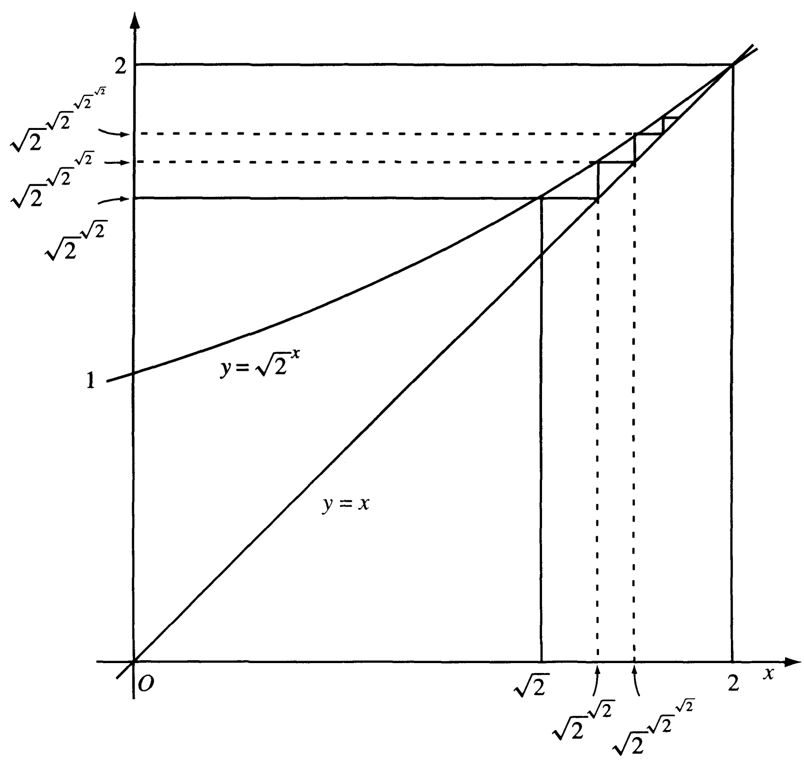
Readers are reminded that an online index of the entire backlog of the MAGAZINE and the *College Mathematics Journal* can be searched on the web at `www.math.hmc.edu/journalsearch/`. This is a valuable resource for teaching and research.

Some of this information is available elsewhere: Many articles are indexed in *Mathematical Reviews* and can be found via *MathSciNet*; material more than five years old appears in the *Arts and Sciences II* portion of *JSTOR*. However, only in this database, maintained by Harvey Mudd College, can you find an abstract (or first paragraph) of every piece that has appeared in our pages.

This online index turns your backlog of the *Magazine* into a treasure chest of information. We hope that readers are developing their own colorful shelves of back issues.

Proof Without Words: Convergence of a Hyperpower Sequence

$$\sqrt{2}^{\sqrt{2}^{\sqrt{2}^{\cdots}}} = 2$$



—F. AZARPANAH
CHAMRAN UNIVERSITY
AHVAZ, IRAN
azarpanah@ipm.ir

Constructive Geometry

When I think of the circle, I am filled with awe.
Not the circle of the sewer lid, or the Hula Hoop,
But the one that spins in Euclid's mind:
The collection of all points in a plane
Equidistant from the center, also a point,
That thing "which hath no part." So beautiful!
But think of the work involved. An infinite number of laborers
Each pushing a wheelbarrow filled with tiny black points
Placing them precisely, just so many centimeters from the center.

When they finally complete the job, they sit down,
Get their sandwiches and thermos of coffee out of their lunch boxes
And with a sigh, begin to eat.
But then, I imagine, the boss drives by in his Cadillac,
And calls their attention to some gaps: here, here, and over there—an infinite number of gaps.
The laborers sigh, roll their eyes, put away their sandwiches, get up
And begin wheeling their barrows full of points again.
This happens an infinite number of times, all done at the speed of light.

Euclid's straight line would be just as much work:
A collection of points that is the shortest distance between two points.
The men would be working with their tape measures, carefully checking,
And placing their points,
When the boss comes by again. He tells them, "Hey! Look at that kink!
That makes it a nanometer too long!"

But of all Euclid's definitions the one I like best,
Is the bittersweet one for parallel lines:
Two straight lines in a plane, that never intersect.
I think of two people, a man and a woman,
In adjoining seats on a train that goes on forever
Down a perfectly straight track.
They become friends, fall in love;
He yearns to take her in his arms, she longs to kiss him,
But they can't even touch their little fingers together.
So they sit there, looking at each other,
As the train thunders through tunnels, rumbles across bridges,
And whistles at crossings,
Hoping that the old man will turn his back, or nod off for just a minute.

—BEN VINEYARD
1417 NORTH 24TH STREET
ST. JOSEPH, MO 64506

Proof Without Words: A Triangular Identity

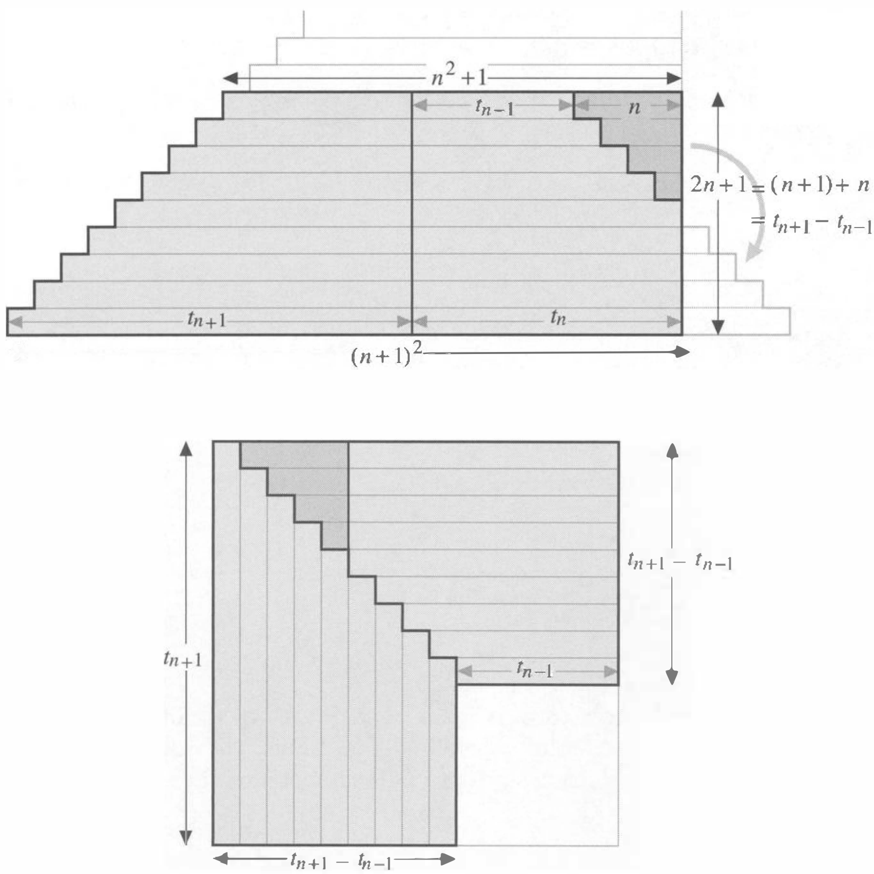
$$2 + 3 + 4 = 9 = 3^2 - 0^2$$

$$5 + 6 + 7 + 8 + 9 = 35 = 6^2 - 1^2$$

$$10 + 11 + 12 + 13 + 14 + 15 + 16 = 91 = 10^2 - 3^2$$

$$\vdots$$

$$t_n = 1 + 2 + \cdots + n \Rightarrow t_{(n+1)^2} - t_{n^2} = t_{n+1}^2 - t_{n-1}^2$$



—ROGER B. NELSEN
LEWIS & CLARK COLLEGE
PORTLAND OR 97219

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

Proposals

To be considered for publication, solutions should be received by May 1, 2005.

1706. *Proposed by Steve Edwards and James Whitenton, Southern Polytechnic State University, GA.*

Let $0 < a, b < 1$. Evaluate

$$\prod_{n=-\infty}^{\infty} \frac{1+b^{2^n}}{1+a^{2^n}}.$$

1707. *Proposed by Barthel Wayne Huff, Salt Lake City, Utah.*

Let k and n be positive integers. Evaluate

$$\sum_{j=1}^{n} \left(\frac{jk}{n-j+k} \prod_{r=1}^{j-1} \frac{n-r}{n-r+k} \right),$$

where the empty product is equal to 1.

1708. *Proposed by Stephen J. Herschkorn, Highland Park, NJ.*

It is well known that the area of a square is half the square of the length of its diagonal. Show that if the area of a parallelogram is half the square of one of its diagonals, and if the area and each side have rational measure, then the parallelogram is a square.

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.

Solutions and new proposals should be mailed to Elgin Johnston, Problems Editor, Department of Mathematics, Iowa State University, Ames IA 50011, or mailed electronically (ideally as a LaTeX file) to ehjohnst@iastate.edu. All communications should include the readers name, full address, and an e-mail address and/or FAX number.

1709. *Proposed by Mihály Bencze, Săcele-Négyfalu, Romania.*

Let $x_1,\ x_2, \dots, x_{3n} \geq 0$. Prove that

$$2^n \prod_{k=1}^{3n} \frac{1+x_k^2}{1+x_k} \geq \left(1 + \prod_{k=1}^{3n} x_k^{1/n}\right)^n .$$

1710. *Proposed by William D. Weakley, Indiana-Purdue University at Fort Wayne, Fort Wayne, IN.*

Let $n > 1$ be an integer and let $[a_1\ a_2 \cdots a_n]$ and $[b_1\ b_2 \cdots b_n]$ be $1 \times n$ matrices with integer entries. Show that $\gcd(a_1,\ a_2, \dots, a_n) = \gcd(b_1,\ b_2, \dots, b_n)$ if and only if there is an $n \times n$ matrix M with integer entries and $\det(M) = 1$ such that

$$[a_1\ a_2 \cdots a_n]M = [b_1\ b_2 \cdots b_n].$$

Quickies

Answers to the Quickies are on page 403.

Q945. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, AB, Canada.*

Determine the maximum value of

$$\prod_{k=1}^{n+1}(1 + \tanh x_k) \Big/ \prod_{k=1}^{n+1}(1 - \tanh x_k),$$

for real numbers $x_1, x_2, \dots, x_{n+1}$ with $\sum_{k=1}^{n+1} x_k = 0$.

Q946. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, AB, Canada.*

Let $ABCD$ be a tetrahedron. Let ℓ_A denote the line through the centroid of face BCD and perpendicular to the face, and let ℓ_B, ℓ_C, and ℓ_D be defined in a similar way. Prove that ℓ_A, ℓ_B, ℓ_C, and ℓ_D are concurrent if and only if the four altitudes of the tetrahedron are concurrent.

Solutions

An Irreducible Polynomial **December 2003**

1681. *Proposed by Mihai Manea, Princeton University, Princeton, NJ.*

Let p be a prime number. Prove that the polynomial

$$x^{p-1} + 2x^{p-2} + 3x^{p-3} + \cdots + (p-1)x + p$$

is irreducible in $\mathbb{Z}[x]$.

Solution by Li Zhou, Polk Community College, Winter Haven, FL.

Let f denote the polynomial. Because $f(\pm 1) > 0$ and $f(\pm p) > 0$, it follows from the rational root theorem that $f(x)$ has no linear factor in $\mathbb{Z}[x]$. Since

$$f(x) = \sum_{k=1}^{p} \frac{x^k - 1}{x-1} = \frac{x(x^p - 1) - p(x-1)}{(x-1)^2},$$

we have

$$f(x+1)=\frac{(x+1)\left((x+1)^p-1\right)-px}{x^2}=x^{p-1}+(p+1)x^{p-2}+\sum_{k=0}^{p-3}a_kx^k,$$

where $a_k=\binom{p+1}{k+2}$, $0\le k\le p-3$. Because $p\nmid(p+1)$, $p\mid a_k$, $0\le k\le p-3$, and $p^2\nmid a_0$, it follows from a modification of Eisenstein's criterion that $f(x+1)$ has an irreducible factor of degree at least $p-2$ over $\mathbb{Z}[x]$. However, $f(x+1)$ has no factor of degree $p-2$ because if it did, the other factor would be linear. If follows that $f(x)$ is irreducible in $\mathbb{Z}[x]$.

Also solved by Michel Bataille (France), Minh Can, John Christopher, Knut Dale (Norway), Daniele Donini (Italy), Lorraine L. Foster, G.R.A.20 Problems Group (Italy), Leslie D. Johnson, Kee-Wai Lau (China), Heinz-Jürgen Seiffert (Germany), Achilleas Sinefakopoulos, Nicholas C. Singer, and the proposer. There were two incorrect submissions.

Summing Inverted Binomial Coefficients **December 2003**

1682. *Proposed by Paul Bracken, University of Texas, Edinburg, TX.*

Let n be a positive integer. Prove that

$$\sum_{j=1}^{n}1\Big/\binom{n}{j}=\frac{n+1}{2^n}\sum_{j=0}^{n-1}\frac{2^j}{j+1}.$$

I. *Many readers submitted a solution similar to the following.*

It is easy to check that the result is true for $n=1$. Assume that the identity holds for a positive integer n and observe that

$$\binom{n+1}{j}^{-1}+\binom{n+1}{j+1}^{-1}=\frac{n+2}{n+1}\binom{n}{j}^{-1}.$$

We then have

$$\begin{aligned}\sum_{j=1}^{n+1}\binom{n+1}{j}^{-1}&=\frac{1}{2}\left(\sum_{j=1}^{n+1}\binom{n+1}{j}^{-1}+\sum_{j=0}^{n}\binom{n+1}{j+1}^{-1}\right)\\&=\frac{1}{2}\left(\binom{n+1}{1}^{-1}+\frac{n+2}{n+1}\sum_{j=1}^{n}\binom{n}{j}^{-1}+\binom{n+1}{n+1}^{-1}\right)\\&=\frac{1}{2}\left(\frac{n+2}{n+1}+\frac{n+2}{2^n}\sum_{j=0}^{n-1}\frac{2^j}{j+1}\right)=\frac{n+2}{2^{n+1}}\sum_{j=0}^{n}\frac{2^j}{j+1}.\end{aligned}$$

Thus, by induction, the desired identity holds.

II. *Solution by Chu Wenchang and Di Claudio Leontina Veliana, Università degli Studi di Lecce, Lecce, Italy.*

First note that

$$\begin{aligned}\binom{n}{k}^{-1}&=\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+1)}=(n+1)\beta(k+1,\,n-k+1)\\&=(n+1)\int_0^1x^k(1-x)^{n-k}\,dx.\end{aligned}$$

Hence

$$\sum_{k=0}^{n}\binom{n}{k}^{-1} = (n+1)\int_0^1\left(\sum_{k=0}^{n}x^k(1-x)^{n-k}\right)dx$$
$$= (n+1)\int_0^1 \frac{(1-x)^{n+1}-x^{n+1}}{1-2x}\,dx.$$

Make the change of variables $x = (1-y)/2$ and note that the resulting integrand is even to get

$$\sum_{k=0}^{n}\binom{n}{k}^{-1} = \frac{n+1}{2^{n+1}}\int_0^1 \frac{(1+y)^{n+1}-(1-y)^{n+1}}{y}\,dy.$$

Using

$$(1+y)^{n+1} = \big(2y+(1-y)\big)^{n+1} = (1-y)^{n+1} + \sum_{k=0}^{n}\binom{n+1}{k+1}(1-y)^{n-k}(2y)^{k+1}$$

in this last expression, we obtain

$$\sum_{k=0}^{n}\binom{n}{k}^{-1} = \frac{n+1}{2^n}\sum_{k=0}^{n}2^k\binom{n+1}{k+1}\int_0^1 y^k(1-y)^{n-k}\,dy$$
$$= \frac{n+1}{2^n}\sum_{k=0}^{n}2^k\binom{n+1}{k+1}\beta(k+1,\, n-k+1) = \frac{n+1}{2^n}\sum_{k=0}^{n}\frac{2^k}{k+1}.$$

This is equivalent to the identity in the problem statement.

Note: Several readers pointed out that this identity has appeared in other places. Among the sources cited were *Concrete Mathematics*, by Ronald Graham, Donald Knuth, and Oren Patashnick, Second Edition, Addison Wesley, 1994 (page 254, Exercise 100); and *The Green Book of Mathematical Problems*, by K. Hardy and K. S. Williams, Dover, 1997 (page 97, Problem 97).

Also solved by Michael Andreolli, Michel Bataille (France), K. S. Bhanu and M. N. Deshpande (India), Jany C. Binz (Switzerland), Brian Bradie, Samuel J. Buelk, Stan Byrd, Minh Can, Jeremy Case, Con Amore Problem Group (Denmark), Knut Dale (Norway), Livia De Donno (Italy), José Luis Díaz-Barrero (Spain), Daniele Donini (Italy), Marty Getz and Dixon Jones, Hurlee Gonchigdanzan, G.R.A. 20 Problems Group (Italy), Enkel Hysnelaj (Australia), Stephen Kaczkowski, Brian Krummel, Lau Sai Luk (China), Eric Malm, Juniad N. Mansuri, Robert Poodiack, Rob Pratt, Nelisa Roach and Yisa Rumula, Rolf Richberg (Germany), Ossama A. Saleh and Terry J. Walters, Heinz-Jürgen Seiffert, Andrei Simion, Achilleas Sinefakopoulos, Nicholas C. Singer, Satyanand Singh, John W. Spellman and Ricardo M. Torrejón, Mike Spivey, H. T. Tang, Michael Vowe (Switzerland), Li Zhou, and the proposer.

An Infrequent Inequality **December 2003**

1683. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, AB, Canada.*

For integer $n \geq 2$ and nonnegative real numbers $x_1, x_2, \ldots, x_n$, define

$$A_n = (x_1^2+x_2^2)(x_2^2+x_3^2)\cdots(x_n^2+x_1^2)/2^n,$$
$$B_n = (x_1x_2+x_2x_3+\cdots+x_nx_1)^n/n^n.$$

a. Determine all n, if any, such that $A_n \geq B_n$ for all choices of x_ks.

b. Determine all n, if any, such that $B_n \geq A_n$ for all choices of x_ks.

Solution by Roy Barbara, American University of Beirut, Beirut, Lebanon.

a. We show that $A_n \geq B_n$ always holds for $n = 2, 3$, but need not hold for $n \geq 4$. For $n = 2$ the inequality follows immediately from $(x_1 - x_2)^2 \geq 0$. For the case $n = 3$, let $S_1 = x_1 + x_2 + x_3$, $S_2 = x_1x_2 + x_2x_3 + x_3x_1$, and $S_3 = x_1x_2x_3$. We may assume that $S_2 = 1$. Then the $n = 3$ case of the inequality becomes

$$\frac{1}{8}(x_1^2 + x_2^2)(x_2^2 + x_3^2)(x_3^2 + x_1^2) \geq \frac{1}{27}.$$

To prove this, first observe that

$$S_1^2 = x_1^2 + x_2^2 + x_3^2 + 2S_2 \geq 3S_2 = 3, \quad \text{so} \quad S_1 \geq \sqrt{3},$$

and by the arithmetic-geometric mean inequality

$$\frac{1}{3} = \frac{S_2}{3} \geq S_3^{2/3}, \quad \text{so} \quad S_3 \leq \frac{\sqrt{3}}{9}.$$

Thus

$$\begin{aligned}\frac{1}{8}(x_1^2 + x_2^2)(x_2^2 + x_3^2)(x_3^2 + x_1^2) &\geq \left(\frac{x_1 + x_2}{2}\right)^2 \left(\frac{x_2 + x_3}{2}\right)^2 \left(\frac{x_3 + x_2}{2}\right)^2 \\ &= \frac{1}{64}(S_1 - x_3)^2(S_1 - x_1)^2(S_1 - x_2)^2 \\ &= \frac{1}{64}(S_1S_2 - S_3)^2 \geq \frac{1}{64}\left(\sqrt{3} - \frac{\sqrt{3}}{9}\right)^2 = \frac{1}{27},\end{aligned}$$

as desired.

Now let $n \geq 4$. Setting $x_1 = x_2 = 0$ and $x_3 = x_4 = \cdots = x_n = 1$, we get $A_n = 0$ and $B_n > 0$, showing that $A_n < B_n$ is possible.

b. For $n \geq 2$, the inequality $B_n \geq A_n$ does not hold for all real x_ks. Set $x_k = 1$ for k odd and $x_k = 0$ for k even. Then for even $n \geq 2$,

$$B_n = 0 < \frac{1}{2^n} = A_n,$$

and for odd $n \geq 3$,

$$B_n = \frac{1}{n^n} < \frac{1}{2^{n-1}} = A_n.$$

Also solved by Daniele Donini (Italy), Eugene Curtin and John W. Spellman and Ricardo M. Torrejón, T. L. McCoy, Rolf Richberg (Germany), Nora Thornber, Chu Wenchang (Italy), Li Zhou, and the proposer. There was one incomplete submission.

Palindromes Need not Apply **December 2003**

1684. *Proposed by Ethan S. Brown, Massachusetts Institute of Technology, Cambridge, MA, and Christopher J. Hillar, University of California, Berkeley, CA.*

Let S be the set of all n letter words in two letters, say a and b. Define an equivalence relation on S as follows: given a word W, the reverse of W, the complement of W (that is, change all as to bs and all bs to as) and the reverse of the complement are all equivalent to W. Find the number of equivalence classes of S that do not contain any palindromes.

Solution by Darren D. Wick, Ashland University, Ashland, OH.

For n even, there are 2^{n-2} such classes, and for $n > 1$ odd, there are $2^{n-2} - 2^{(n-3)/2}$ classes.

For $W \in S$, denote the reverse of W by W^r and the complement of W by W^c. For any $W \in S$, $W^{cc} = W = W^{rr}$ and $W^{rc} = W^{cr}$. Thus, the equivalence class of W, denoted by $[W]$, is a subset of $\{W, W^c, W^r, W^{cr}\}$. If W is a palindrome, then $[W] = \{W, W^c\}$ and W^c is also a palindrome.

Let n be even. Then there are $2^{n/2}$ palindromes in S. If $W^c = W^r$, then $|[W]| = 2$. There are $2^{n/2}$ such W, and hence $2^{n/2}/2$ equivalence classes of order 2 that do not contain a palindrome. The number of equivalence classes of order 4 (not containing a palindrome) is $(2^n - 2^{n/2} - 2^{n/2})/4 = 2^{n-2} - 2^{n/2}/2$. Thus the number of equivalence classes that do not contain a palindrome is $2^{n/2}/2 + (2^{n-2} - 2^{n/2}/2) = 2^{n-2}$.

Now let n be odd. There are $2^{(n+1)/2}$ palindromes in S and there are no $W \in S$ with $W^r = W^c$. Thus there are $2^n - 2^{(n+1)/2}$ words that are not palindromes and all are in equivalence classes of order 4. Thus the number of equivalence classes that do not contain a palindrome is $(2^n - 2^{(n+1)/2})/4 = 2^{n-2} - 2^{(n-3)/2}$.

Also solved by JPV Abad, Michael Andreolli, Roy Barbara (Lebanon), Michel Bataille (France), K. S. Bhanu and M. N. Deshpande (India), Jany C. Binz (Switzerland), Mark Bowron, Marc Brodie, B. J. Falkowski, Dimitry Fleishman, Marty Getz and Dixon Jones, G.R.A.20 Problems Group (Italy), Jerrold W. Grossman, Kathleen E. Lewis, J. Andrew Long, Arturo Magidin, Rob Pratt, Manuel Reyes, R. P. Sealy (Canada), Nicholas C. Singer, Mike Spivey, Li Zhou, Paul J. Zwier, and the proposer. There was one incorrect submission.

Sums Equal Products — December 2003

1685. *Proposed by Michel Bataille, Rouen, France.*

Find all continuous functions $f : (0, \infty) \longrightarrow (0, \infty)$ such that

$$f(a) + f(b) + f(c) = f(a)f(b)f(c)$$

whenever $ab + bc + ca = 1$.

Solution by Chip Curtis, Missouri Southern State University, Joplin, MO.

The solutions are those functions f defined by

$$f(s) = f_k(s) = \tan\left(\frac{\pi}{6}(k+2) - k \arctan s\right),$$

for $-1/2 \le k \le 1$.

Let f be a solution to the functional equation. For $a, b, c \in (0, \infty)$, define $x, y, z, u, v, w \in (0, \pi/2)$ by

$$x = \arctan a, \qquad y = \arctan b, \qquad z = \arctan c$$

and

$$u = \arctan f(a), \qquad v = \arctan f(b), \qquad w = \arctan f(c).$$

Then

$$ab + bc + ca = 1 \Longleftrightarrow c = \frac{1 - ab}{a + b}$$

$$\Longleftrightarrow \tan z = \frac{1 - \tan x \tan y}{\tan x + \tan y} \Longleftrightarrow \tan z = \cot(x + y) \Longleftrightarrow x + y + z = \frac{\pi}{2}.$$

Next observe that if $f(a)+f(b)+f(c)=f(a)f(b)f(c)$, then $f(a)f(b)\neq 1$ because this would imply that $f(a)+1/f(a)=0$, which is impossible. Thus

$$\begin{aligned} f(a)+f(b)+f(c)=f(a)f(b)f(c) &\iff f(c)=-\frac{f(a)+f(b)}{1-f(a)f(b)} \\ \iff \tan w=-\frac{\tan u+\tan v}{1-\tan u\tan v} &\iff \tan w=-\tan(u+v) \\ &\iff u+v+w=\pi. \end{aligned}$$

Define $h:(0,\frac{\pi}{2})\to(0,\frac{\pi}{2})$ by $h(t)=\arctan(f(\tan t))$. Because f is continuous on $(0,\infty)$, it follows that h is continuous on $(0,\frac{\pi}{2})$. The functional equation for f is equivalent to the the requirement that

$$h(x)+h(y)+h(z)=\pi \quad \text{whenever} \quad x+y+z=\frac{\pi}{2},$$

and hence that

$$h\left(\frac{\pi}{2}-x-y\right)=\pi-h(x)-h(y) \quad \text{if} \quad x,y,x+y\in\left(0,\frac{\pi}{2}\right).$$

Define H for $-\frac{\pi}{3}<t<\frac{\pi}{6}$ by

$$H(t)=h\left(\frac{\pi}{6}-t\right)-\left(\frac{\pi}{3}+t\right), \quad \text{so that} \quad h(t)=H\left(\frac{\pi}{6}-t\right)+\left(\frac{\pi}{2}-t\right).$$

Then

$$\begin{aligned} H(x+y) &= h\left(\frac{\pi}{6}-x-y\right)-\left(\frac{\pi}{3}+x+y\right) \\ &= h\left(\frac{\pi}{2}-\left(x+\frac{\pi}{6}\right)-\left(y-\frac{\pi}{6}\right)\right)-\left(\frac{\pi}{3}+x+y\right) \\ &= \pi-h\left(x+\frac{\pi}{6}\right)-h\left(y+\frac{\pi}{6}\right)-\left(\frac{\pi}{3}+x+y\right) \\ &= \pi-\left(H(-x)+\left(\frac{\pi}{3}-x\right)\right)-\left(H(-y)+\left(\frac{\pi}{3}-y\right)\right)-\left(\frac{\pi}{3}+x+y\right) \\ &= -H(-x)-H(-y). \end{aligned}$$

Setting $x=y=0$ we find that $H(0)=0$ and then that $H(x)=-H(-x)$ so that $H(x+y)=H(x)+H(y)$. For H defined on an interval containing 0, it is well known that the only continuous solutions to this functional equation are those of the form $H(t)=mt$, where m is a constant. We then find

$$h(t)=m\left(\frac{\pi}{6}-t\right)+\left(\frac{\pi}{2}-t\right)=\frac{\pi}{6}(k+2)-kt,$$

where $k=m+1$. To ensure that $h:(0,\frac{\pi}{2})\to(0,\frac{\pi}{2})$, we must have

$$0\leq\frac{\pi}{6}(k+2)\leq\frac{\pi}{2} \quad \text{and} \quad 0\leq\frac{\pi}{6}(k+2)-\frac{k\pi}{2}\leq\frac{\pi}{2},$$

from which $-1/2\leq k\leq 1$. If follows from our definition of h that

$$f(s)=f_k(s)=\tan\left(\frac{\pi}{6}(k+2)-k\arctan s\right), \quad \text{for some} \quad -\frac{1}{2}\leq k\leq 1.$$

For some values of k the expression for f_k can be simplified. For example,

$$f_0(s) \equiv \sqrt{3}, \quad f_1(s) = \frac{1}{s}, \quad \text{and} \quad f_{-1/2}(s) = s + \sqrt{1+s^2}.$$

Also solved by John A. Baker (Canada), Christopher Hill, Northwestern University Math Problem Solving Group, Rolf Richberg (Germany), Nicholas C. Singer, Li Zhou, and the proposer.

Answers

Solutions to the Quickies from page 397.

A945. Because

$$\tanh(a+b) = \frac{\tanh a + \tanh b}{1 + \tanh a \tanh b},$$

it follows by induction that

$$\tanh(x_1 + x_2 + \cdots + x_n) = \frac{T_1 + T_3 + T_5 + \cdots}{1 + T_2 + T_4 + \cdots},$$

where

$$T_k = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \tanh x_{i_1} \tanh x_{i_2} \cdots \tanh x_{i_k}$$

is the symmetric sum of the products of the $\tanh x_j$ taken k at a time. Hence

$$1 + \tanh x_{n+1} = 1 - \tanh(x_1 + x_2 + \cdots + x_n) = \frac{1 - T_1 + T_2 - T_3 + \cdots}{1 + T_2 + T_4 + T_6 + \cdots},$$

and

$$1 - \tanh x_{n+1} = \frac{1 + T_1 + T_2 + T_3 + \cdots}{1 + T_2 + T_4 + T_6 + \cdots}.$$

We then have

$$\frac{\prod_{k=1}^{n+1}(1 + \tanh x_k)}{\prod_{k=1}^{n+1}(1 - \tanh x_k)} = \frac{1 + T_1 + T_2 + T_3 + \cdots}{1 - T_1 + T_2 - T_3 + \cdots} \cdot \frac{1 + \tanh x_{n+1}}{1 - \tanh x_{n+1}} = 1,$$

showing that the expression is identically 1.

A946. Let $\mathbf{X}$ denote the vector from the origin to the point X. First assume that ℓ_A, ℓ_B, ℓ_C, and ℓ_D are concurrent at P. Then ℓ_A is parallel to the vector $\frac{1}{3}(\mathbf{B} + \mathbf{C} + \mathbf{D}) - \mathbf{P}$. The altitude from A is parallel to ℓ_A so has vector equation

$$\mathbf{R}_A(t) = \mathbf{A} + t\left(\frac{1}{3}(\mathbf{B} + \mathbf{C} + \mathbf{D}) - \mathbf{P}\right).$$

Setting $t = 3$ we obtain $\mathbf{R}_A(3) = 4\mathbf{G} - 3\mathbf{P}$, where G is the centroid of $ABCD$. By a similar argument, each of the other three altitudes also contains this point. Hence the four altitudes are concurrent.

Next assume that the altitudes are concurrent at H. Then the vector equation of ℓ_A is

$$\mathbf{L}_A(t) = \frac{1}{3}(\mathbf{B} + \mathbf{C} + \mathbf{D}) + t(\mathbf{A} - \mathbf{H}).$$

Setting $t = \frac{1}{3}$ we obtain $\mathbf{L}(\frac{1}{3}) = \frac{4}{3}\mathbf{G} - \frac{1}{3}\mathbf{H}$. By a similar argument, ℓ_B, ℓ_C, and ℓ_D also contain this point, so the four lines are concurrent.

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

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

Watson, Christopher, Maths millions still safe, *The Australian* (8 September 2004), `http://www.theaustralian.news.com.au/common/story_page/0,5744,10706836%255E30417,00.html` . Sabbagh, Karl, The strange case of Louis de Branges, *London Review of Books* 26 (14) (22 July 2004), `http://www.lrb.co.uk/v26/n14/sabb01_.html` . Shipman, Jerome, A $459,279 question (letter), *London Review of Books* (19 August 2004), `http://www.lrb.co.uk/v26/n16/letters.html#4`. Radford, Tim, Maths holy grail could bring disaster for internet, *The Guardian* (7 September 2004), `http://www.guardian.co.uk/online/news/0,12597,1299014,00.html` .

Marcus du Sautoy (author of a popular book about the Riemann hypothesis) is still dubious that Louis de Branges (Purdue University) has proved the Riemann hypothesis. On the other hand, Karl Sabbagh (not a mathematician but author of a "competing" popular book) claims that "there is no evidence that ... any mathematician has read [de Branges's paper]; de Branges and his proof appear to have been ostracised by the profession .... It may be that a possible solution of one of the most important problems in mathematics is never investigated because no one likes the solution's author." However, Jerome Shipman notes that de Branges's work was supported by the National Science Foundation and that Sabbagh "shouldn't mistake scepticism for ostracism." Sabbagh explains that de Branges is not a "crank" but that he is "cranky," with "disastrous" (de Branges's word) relationships with colleagues, who have been skeptical of his work ever since a claim of his in 1964 that he could not substantiate. Meanwhile, du Sautoy recently asserted that proving the Riemann hypothesis would give mathematicians a "prime spectrometer, like the machine chemists use to tell them what things are made of. If we had something like that it would bring the world of e-commerce to its knees overnight." That sort of claim gets headlines but won't endear mathematics or mathematicians to business people.

Robinson, Sara, Tweaking the math to make happier medical marriages, *New York Times* (24 August 2004) F2; `http://www.nytimes.com/2004/08/24/science/24matc.html` . Crawford, Vincent P., The flexible-salary match: A proposal to increase the salary flexibility of the National Resident Matching Program `http://weber.ucsd.edu/~vcrawfor/FlexibleSalaryMatch.pdf` . Bulow, Jeremy I., and Jonathan D. Levin, Matching and price competition (July 2003), Stanford University Research Paper No. 1818, `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=441006` .

In the 1950s, hospitals instituted the Residency Match for assigning medical students to residencies in hospitals. The procedure used was what mathematicians later named the "marriage algorithm"; it produces a stable solution that favors the hospitals. A recent suit by residents complained that the Match is unfair because it does not allow hospitals to compete by making higher offers to residents considered more desirable. Although a federal district course dismissed this antitrust case in August, economists have taken over where mathematicians left off, trying to devise an algorithm incorporating salary considerations.

Ross, Ken, *A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans*, Pi Press (Pearson Education), 2004; xv + 190 pp, $19.95. ISBN 0–13–147990–3.

Ken Ross (University of Oregon) taught for 35 years and was President of the Mathematical Association of America. His book reflects the keenness and longing of a baseball fan who grew up (and still lives) far from a major-league team, plus the sensibility of a mathematician and the pleasant prose of a master writer and educator. The book treats the averages used in baseball statistics, probability and odds, expected value, betting strategies, conditional probability, repeated Bernoulli trials (e.g., the World Series), and streaks, with an appendix on the binomial theorem. Algebraic notation is used but does not overwhelm; the book should do well, if it can get into one of the few distribution channels.

Netz, Reviel, *The Works of Archimedes: Translation and Commentary*. Vol. 1: *The Two Books* On the Sphere and the Cylinder, Cambridge University Press, 2004; x + 375 pp, $125. ISBN 0–521–66160–9. Klarreich, Erica, Glimpses of genius: Mathematicians and historians piece together a puzzle that Archimedes pondered, *Science News* (15 May 2004) 314–315.

Reviel Netz (Stanford) has begun the definitive translation/commentary of Archimedes' works with the two books of *On the Sphere and the Cylinder* and Eutocius's commentaries, "largely unaffected by the ... [rediscovered] Palimpsest." Meanwhile, the *Stomachion* fragment in the Palimpsest, now thought to be about a geometric tiling, has generated new discoveries about the number of tilings with the pieces and the tree structure of the tilings, by Ron Graham and Fan Chung (University of California–San Diego).

Grossman, Wendy, New tack wins Prisoner's Dilemma, *Wired* (13 October 2004), `http://www.wired.com/news/culture/0,1284,65317,00.html` .

In the Prisoner's Dilemma game, two accomplices are interrogated separately; either can "defect" (implicate the other) or "cooperate" (remain silent). If both cooperate, both get a minimal sentence; if both defect, both get long sentences; if one cooperates and one defects, the defector goes free and the cooperator gets a very long sentence. The long-standing champion among strategies for repeatedly playing the game has been Tit-for-Tat, which first cooperates and then echoes the other player. How to get cooperation to emerge from selfish agents? Well, how else, but by collusion? In this year's competition, teams from Southampton University executed sequences of "moves" to recognize each other; then one team thereafter always defected (thereby "winning") and the other always cooperate (thereby "losing"); against unrecognized teams, Southampton's always defected. So some teams from Southampton won the contest, while others were at the very bottom.

Pikul, Corrie, The calculus of coitus, *Salon* (13 September 2004), `http://www.salon.com/mwt/feature/2004/09/13/math/index_np.html` . Baxter, Mark, Mathematics, sex, actuarial studies, *Actuary* (Australia) (April 2004) 8–9. Cresswell, Clio, *Mathematics and Sex*, Allen & Unwin Pty., Limited (Australia), U.S. $14.95 (P). ISBN 1–74114159–1.

Clio Cresswell is a celebrity in Australia—perhaps not mainly because she is a mathematician (University of New South Wales), since she also is a radio host, promotes a brand of clothing, and writes a column on relationships for a women's magazine. She would have become a celebrity anyway after her book *Mathematics and Sex*. I hoped to review the book here, but the publisher did not respond, perhaps judging that with its title the book doesn't need any hyping in this MAGAZINE. So I can only titillate you with snippets about it: "The book presents current mathematical research that can be used to answer questions like: How will we know when we've found 'the one'? How much should ... partners compromise ... ? Who has better orgasms, men or women?" The Secretary Problem from optimal stopping theory comes into play to answer the first question, in Cresswell's juicy "Rule of 12 Bonks" and its corollary that "most people will end up happier if they actively proposition as many desirable partners as possible." I won't try to guess what mathematics addresses the other questions. The main selling point may be that there is much more sex, and much less mathematics, than you might anticipate from the title.

NEWS AND LETTERS

Acknowledgments

In addition to our Associate Editors, the following referees have assisted the MAGAZINE *during the past year. We thank them for their time and care.*

Allman, Elizabeth S., *University of Southern Maine, Falmouth, ME*

Bauer, Craig P., *York College, York, PA*

Beals, Richard W., *Yale University, New Haven, CT*

Beauregard, Raymond A., *University of Rhode Island, Kingston, RI*

Beezer, Robert A., *University of Puget Sound, Tacoma, WA*

Borwein, Jonathan M., *Dalhousie University, Halifax, Nova Scotia, Canada*

Bressoud, David M., *Macalester College, St. Paul, MN*

Brunson, Barry W., *Western Kentucky University, Bowling Green, KY*

Burton, David M., *University of New Hampshire, Durham, NH*

Callan, C. David, *University of Wisconsin, Madison, WI*

Case, Jeremy S., *Taylor University, Upland, IN*

Chakerian, G. Don, *University of California, Davis, Davis, CA*

Chinn, Phyllis Z., *Humboldt State University, Arcata, CA*

DeTemple, Duane W., *Washington State University, Pullman, WA*

Dodge, Clayton W., *University of Maine, Orono, ME*

Dunne, Edward G., *American Mathematical Society, Providence, RI*

Eisenberg, Bennett, *Lehigh University, Bethelem, PA*

Elderkin, Richard H., *Pomona College, Claremont, CA*

Fisher, Evan, *Lafayette College, Easton, PA*

Fisher, J. Chris, *University of Regina, Regina, Saskatchewan, Canada*

Frantz, Marc, *Indiana University, Bloomington, IN*

Fung, Maria G., *Western Oregon University, Monmouth, OR*

Gordon, Gary, *Lafayette College, Easton, PA*

Gordon, Russell A., *Whitman College, Walla Walla, WA*

Green, Euline I., *Abilene Christian University, Abilene, TX*

Grosshans, Frank D., *West Chester University of Pennsylvania, Baltimore, MD*

Guichard, David R., *Whitman College, Walla Walla, WA*

Hayashi, Elmer K., *Wake Forest University, Winston-Salem, NC*

Horn, Roger, *University of Utah, Salt Lake City, UT*

Hull, Thomas, *Merrimack College, North Andover, MA*

Isaac, Richard, *Herbert H. Lehman College CUNY, Bronx, NY*

Janke, Steven J., *Colorado College, Colorado Springs, CO*

Jepsen, Charles H., *Grinnell College, Grinnell, IA*

Johnson, Brenda, *Union College, Schenectady, NY*

Johnson, Diane L., *Humboldt State University, Arcata, CA*

Johnson, Warren P., *Connecticut College, New London, CT*

Jones, Michael A., *Montclair State University, Upper Montclair, NJ*

Jordan, James H., *Washington State University, Pullman, WA*

Kimberling, Clark, *University of Evansville, Evansville, IN*

Kuzmanovich, James J., *Wake Forest University, Winston-Salem, NC*

Leise, Tanya L., *Amherst College, Amherst, MA*

Libeskind-Hadas, Ran, *Harvey Mudd College, Claremont, CA*

Lo Bello, Anthony J., *Allegheny College, Meadville, PA*

Lotto, Benjamin A., *Vassar College, Poughkeepsie, NY*

Loud, Warren S., , *Minneapolis, MN*

Lynch, Mark, *Millsaps College, Jackson, MS*

Manvel, Bennet, *Colorado State University, Ft. Collins, CO*

Mazur, Mark S., *Duquesne University, Pittsburgh, PA*

McCleary, John H., *Vassar College, Poughkeepsie, NY*

McDermot, Richard F., *Allegheny College, Meadville, PA*

Merrill, Kathy D., *Colorado College, Colorado Springs, CO*

Michaels, John G., *State University of New York At Brockport, Brockport, NY*

Molinek, Donna Kay, *Davidson College, Davidson, NC*

Monson, Barry, *University of New Brunswick, Fredericton, NB, Canada*

Morics, Steven W., *University of Redlands, Redlands, CA*

Mulcahy, Colm K., *Spelman College, Atlanta, GA*

Neidinger, Richard D., *Davidson College, Davidson, NC*

Nelsen, Roger B., *Lewis & Clark College, Portland, OR*

Neudauer, Nancy Ann, *Pacific University, Forest Grove, OR*

Ostrov, Daniel N., *Santa Clara University, Santa Clara, CA*

Pedersen, Jean J., *Santa Clara University, Santa Clara, CA*

Pfaff, Thomas J., *Ithaca College, Ithaca, NY*

Pfiefer, Richard E., *San Jose State University, San Jose, CA*

Rodger, Chris, *Auburn University, Montgomery, AL*

Ross, Kenneth A., *University of Oregon, Eugene, OR*

Ross, Peter, *Santa Clara University, Santa Clara, CA*

Salwach, Chester J., *Lafayette College, Easton, PA*

Schumer, Peter D., *Middlebury College, Middlebury, VT*

Sellers, James A., *Pennsylvania State University, University Park, PA*

Sherman, Gary, *Rose-Hulman Institute of Technology, Terre Haute, IN*

Shifrin, Theodore, *University of Georgia, Athens, GA*

Simoson, Andrew J., *King College, Bristol, TN*

Sloughter, Daniel C., *Furman University, Greenville, SC*

Solheid, Ernie S., *California State University, Fullerton, Fullerton, CA*

Szabó, Tamás, *Weber State University, Ogden, UT*

Teets, Donald, *South Dakota School of Mines and Technology, Rapid City, SD*

Towse, Christopher W., *Scripps College, Claremont, CA*

Villalobos, Ph. D., Cristina, *The University of Texas-Pan American, Edinburg, TX*

Walden, Byron, *Santa Clara University, Santa Clara, CA*

Whitehead, Tamsen, *Santa Clara University, Santa Clara, CA*

Zorn, Paul, *St. Olaf College, Northfield, MN*

Index to Volume 77

AUTHORS

TITLES

PROBLEMS

The letters P, Q, and S refer to Proposals, Quickies, and Solutions, respectively; page numbers appear in parentheses. For example, P1690 (70) refers to Proposal 1690, which appears on page 70.

I would also like to acknowledge those who have refereed proposals for the Problem Column: Steven Dunbar, Richard Gibbs, Sunder Sethuraman, Ananda Weerasinghe, James Wilson, Gary Lieberman, and A. M. Fink.

New Editor-Elect of MATHEMATICS MAGAZINE

As of January 1, 2005, new submissions should be sent to the new Editor-Elect:

Allen Schwenk
Department of Mathematics
Western Michigan University
Kalamazoo, MI 49008

United States Postal Service

Statement of Ownership, Management, and Circulation

1 Publication Title	2 Publication Number	3 Filing Date
MATHEMATICS MAGAZINE	0025-570X	9/8/04
4 Issue Frequency	**5 Number of Issues Published Annually**	**6 Annual Subscription Price**
Bimonthly except July and August	5	$20

7 Complete Mailing Address of Known Office of Publication (Not printer) (Street, city, county, state, and ZIP+4)
MAA
1529 18th St. N.W.
Washington, D.C. 20036

Contact Person: H. Waldman
Telephone: 202-387-5200

8 Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)
same

9 Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank)

Publisher (Name and complete mailing address)
MAA
1529 18th St. N.W.
Washington, D.C. 20036

Editor (Name and complete mailing address)
FRANK FARRIS
SANTA CLARA UNIVERSITY
SANTA CLARA, CA 95053

Managing Editor (Name and complete mailing address)
HARRY WALDMAN

10 Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)

Full Name	Complete Mailing Address
MATH ASSN OF AMERICA	1529 18th St. N.W., Washington, DC 20036

11 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address

12 Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes.
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

PS Form 3526, October 1999 (See Instructions on Reverse)

13 Publication Title		14 Issue Date for Circulation Data Below	
MATHEMATICS MAGAZINE		June, 2004	
15 Extent and Nature of Circulation		**Average No. Copies Each Issue During Preceding 12 Months**	**No. Copies of Single Issue Published Nearest to Filing Date**
a Total Number of Copies (Net press run)		14,000	15,000
b Paid and/or Requested Circulation	(1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541 (Include advertiser's proof and exchange copies)	12,500	13,500
	(2) Paid In-County Subscriptions Stated on Form 3541 (Include advertiser's proof and exchange copies)	0	0
	(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	0	0
	(4) Other Classes Mailed Through the USPS	0	0
c Total Paid and/or Requested Circulation [Sum of 15b (1), (2), (3), and (4)] ▶		12,500	13,500
d Free Distribution by Mail (Samples, complimentary, and other free)	(1) Outside-County as Stated on Form 3541	0	0
	(2) In-County as Stated on Form 3541	0	0
	(3) Other Classes Mailed Through the USPS	1,000	1,000
e Free Distribution Outside the Mail (Carriers or other means)		0	0
f Total Free Distribution (Sum of 15d and 15e) ▶		1,000	1,000
g Total Distribution (Sum of 15c and 15f) ▶		13,500	14,500
h Copies not Distributed		500	500
i Total (Sum of 15g. and h) ▶		14,000	15,000
j Percent Paid and/or Requested Circulation (15c. divided by 15g. times 100)		90%	90%

16 Publication of Statement of Ownership
☒ Publication required. Will be printed in the December 2004 issue of this publication. ☐ Publication not required

17 Signature and Title of Editor, Publisher, Business Manager, or Owner: Donald J. Albers — Date: 7 Sept. 2004

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

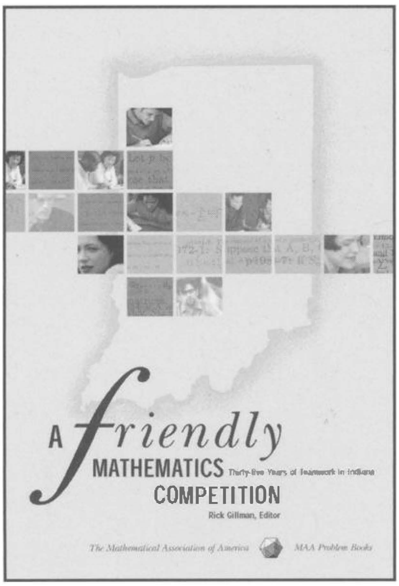
Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.
5. If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
6. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
7. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form 3526, October 1999 (Reverse)

from the Mathematical Association of America



A Friendly Mathematics Competition:
Thirty-five Years of Teamwork in Indiana

Rick Gillman, Editor

Series: Problem Books

A Friendly Mathematics Competition tells the story of the Indiana College Mathematics Competition (ICMC) by presenting the problems, solutions, and results of the first thirty-five years of the ICMC. The ICMC was organized in reaction to the Putnam Exam—its problems were to be more representative of the undergraduate curriculum, and students could work on them in teams. Originally participation was restricted to the small, private colleges and universities of the state, but it was later opened up to students from all of the schools in Indiana. The competition was quickly nicknamed the "Friendly" Competition because of its focus on solving mathematical problems, which brought faculty and students together, rather than on the competitive nature of winning. Organized by year, the problems and solutions in this volume present an excellent archive of information about what has been expected of an undergraduate mathematics major over the past thirty-five years. With more than 245 problems and solutions, the book is also a must buy for faculty and students interested in problem-solving.

Catalog Code: FMC/JR • 169 pp., Paperbound, 2003 • 0-88385-808-8
List: $29.95 • Member: $24.95

Order your copy today!
www.maa.org or 1-800-331-1622

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, DC 20036