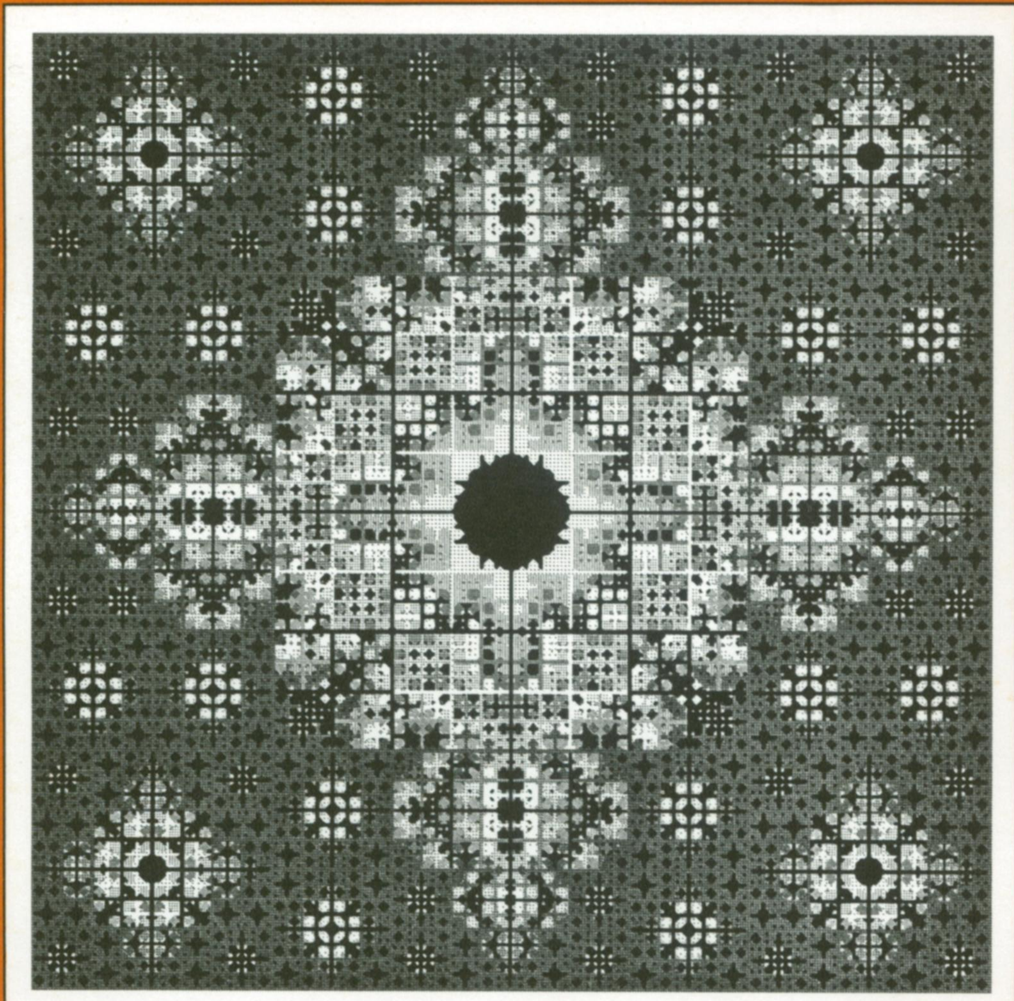
Vol. 70, No. 3 June 1997

MATHEMATICS MAGAZINE



"Persian" Recursion (see pp. 196–199)

- Descartes' Curve-Drawing Devices
- Knight's Tours on a Torus
- Thales Meets Poincaré

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

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

Send new manuscripts to Paul Zorn, Editor, Department of Mathematics, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

Cover illustration: A recursively-defined "Persian Rug" design, by Anne M. Burns. For details, see pp. 196–199.

AUTHORS

David Dennis teaches in the Department of Mathematical Sciences at the University of Texas at El Paso. His research interests focus on the relations among history of mathematics, education, geometry, number theory, and technology. At Cornell University he received an M.S. in mathematical logic in 1977 under Anil Nerode, who also sparked in him a great interest in the history of mathematics. After teaching at Queens College (CUNY) and Ithaca College, he returned to Cornell University and combined his various interests into a research project in mathematics education leading to a Ph.D., under Jere Confrey, in 1995. He is now part of a National Science Foundation project in El Paso known as the Partnership for Excellence in Teacher Education (PETE).

David E. Dobbs received his bachelor's and master's degrees from the University of Manitoba, in 1964 and 1965. His master's thesis concerned coordinatization problems in affine geometry. He received his Ph.D. from Cornell University in 1969, with a dissertation on algebraic geometry. After a postdoctoral year at UCLA and five years of teaching at Rutgers University, he moved to the University of Tennessee in 1975. His main research interest is algebra, especially commutative algebra, on which he has published three books and more than 130 research articles. His work on mathematics education has included three textbooks and more than 50 articles. *Thales Meets Poincaré* was developed for a course that he taught in the summer of 1995—his first contact with non-Euclidean geometry since graduate student days.

Rebecca Hoenigman graduated with Distinction in Chemistry in 1995 from Colorado College, where she also pursued interests in mathematics, music, and lacrosse. Now working on a doctorate in chemistry at the University of Washington, she is interested in theoretical organic chemistry and hopes to synthesize compounds that are theoretically interesting, and to use *ab initio* calculations to understand the experimental results.

John Watkins graduated from Oberlin College; he then taught mathematics in Ghana, as a Peace Corps volunteer, before receiving his doctorate from the University of Kansas in 1980. He has been at Colorado College since 1977 and is an author, with Robin Wilson, of the book *Graphs: An Introductory Approach*. John plans to return to Africa during his next sabbatical to direct a program for American undergraduates in Zimbabwe.

Vol. 70, No. 3, June 1997

MATHEMATICS MAGAZINE

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

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

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

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

ARTICLES

René Descartes' Curve-Drawing Devices: Experiments in the Relations Between Mechanical Motion and Symbolic Language

DAVID DENNIS
University of Texas at El Paso
El Paso, TX 79968-0514

Introduction

By the beginning of the seventeenth century it had become possible to represent a wide variety of arithmetic concepts and relationships in the newly evolved language of symbolic algebra **[19]**. Geometry, however, held a preeminent position as an older and far more trusted form of mathematics. Throughout the scientific revolution geometry continued to be thought of as the primary and most reliable form of mathematics, but a continuing series of investigations took place that examined the extent to which algebra and geometry might be compatible. These experiments in compatibility were quite opposite from most of the ancient classics. Euclid, for example, describes in Book 8–10 of the *Elements* a number of important theorems of number theory cloaked awkwardly in a geometrical representation **[16]**.[1] The experiments of the seventeenth century, conversely, probed the possibilities of representing geometrical concepts and constructions in the language of symbolic algebra. To what extent could it be done? Would contradictions emerge if one moved freely back and forth between geometric and algebraic representations?

Questions of appropriate forms of representation dominated the intellectual activities of seventeenth century Europe, not just in science and mathematics but perhaps even more pervasively in religious, political, legal, and philosophical discussions **[13, 24, 25]**. Seen in the context of this social history it is not surprising that mathematicians like René Descartes and G. W. von Leibniz would have seen their new symbolic mathematical representations in the context of their extensive philosophical works. Descartes' *Geometry* **[11]** was originally published as an appendix to his larger philosophical work, the *Discourse on Method.* Conversely, political thinkers like Thomas Hobbes commented extensively on the latest developments in physics and mathematics **[25, 4]**. Questions of the appropriate forms of scientific symbolism and discourse were seen as closely connected to questions about the construction of the new apparatuses of the modern state. This is particularly evident, for example, in the work of the physicist Robert Boyle **[25]**.

[1] See, for example, Book 10, Lemma 1 before Prop. 29, where Euclid generates all Pythagorean triples geometrically even though he violates the dimensional integrity of his argument. Areas, in the form of "similar plane numbers," are multiplied by areas to yield areas. There seems to be no way to reconcile dimension and still obtain the result.

This paper will investigate in detail two of the curve-drawing constructions from the *Geometry* of Descartes in such a way as to highlight the issue of the coordination of multiple representations (see, e.g., [**6**]). The profound impact of Descartes' mathematics was rooted in the bold and fluid ways in which he shifted between geometrical and algebraic forms of representation, demonstrating the compatibility of these seemingly separate forms of expression. Descartes is touted to students today as the originator of analytic geometry, but nowhere in the *Geometry* did he ever graph an equation. Curves were constructed from geometrical actions, many of which were pictured as mechanical apparatuses. After curves had been drawn Descartes introduced coordinates and then analyzed the curve-drawing actions in order to arrive at an equation that represented the curve. Equations did not create curves; curves gave rise to equations.[2] Descartes used equations to create a taxonomy of curves [**20**].

It can be difficult for a person well schooled in modern mathematics to enter into and appreciate the philosophical and linguistic issues involved in seventeenth century mathematics and science. We have all been thoroughly trained in algebra and calculus and have come to rely on this language and grammar as a dominant form of mathematical representation. We inherently trust that these symbolic manipulations will give results that are compatible with geometry; a trust that did not fully emerge in mathematics until the early works of Euler more than a century after Descartes. Such trust became possible because of an extensive set of representational experiments conducted throughout the seventeenth century which tested the ability of symbolic algebraic language to represent geometry faithfully [**5**, **7**]. Descartes' *Geometry* is one of the earliest and most notable of these linguistic experiments. Because of our cultural trust in the reliability of symbolic languages applied to geometry, many of those schooled in mathematics today have learned comparatively little about geometry in its own right.

Descartes wrote for an audience with opposite predispositions. He assumed that his readers were thoroughly acquainted with geometry, in particular the works of Apollonius (ca. 200 BC) on conic sections [**1**, **15**]. In order to appreciate the accomplishments of Descartes one must be able to check back and forth between representations and see that the results of symbolic algebraic manipulations are consistent with independently established geometrical results. The seventeenth century witnessed an increasingly subtle and persuasive series of such linguistic experiments in the work of Roberval, Cavalieri, Pascal, Wallis, and Newton [**8**, **9**]. These led eventually to Leibniz's creation of a general symbolic language capable of fully representing all known geometry of his day, that being his "calculus" [**5**, **7**].

Because many of the most simple and beautiful results of Apollonius are scarcely known to modern mathematicians, it can be difficult to recreate one essential element of the linguistic achievements of Descartes—checking algebraic manipulations against independently established geometrical results. In this article I will ask the reader to become a kind of intellectual Merlin and live history backwards. After we explore one of Descartes' curve-drawing devices, we will use the resulting bridge between geometry and algebra to regain a compelling result from Apollonius concerning hyperbolic tangents. The reader can choose to regard the investigation either as a philosophical demonstration of the consistency between algebra and geometry or as a simple analytical demonstration of a powerful ancient result of Apollonius. By adopt-

[2] Descartes' contemporary, Fermat, did begin graphing equations but his work did not have nearly the philosophical or scientific impact of Descartes'. Fermat's original problematic contexts came from financial work rather than engineering and mechanics.

ing *both* views one gains a fully flexible cognitive feedback loop of the sort that my students and I have found most enlightening **[6]**.

I was recently discussing my work on curve-drawing devices and their possible educational implications with a friend. His initial reaction was surprise: "Surely you don't advocate the revival of geometrical methods; progress in mathematics has been made only to the extent to which geometry has been eliminated." This claim has historical validity, especially since the eighteenth century, but my response was that such progress was possible only after mathematicians had achieved a basic faith in the ability of algebraic language to represent and model geometry accurately. I argued that one cannot appreciate the profundity of calculus unless one is aware of the issue of coordination of independent representations. Many students seem to learn and even master the manipulations of calculus without ever having questioned or tested the language's ability to model geometry precisely. Even Leibniz, no lover of geometry, would feel that such a student had missed the main point of his symbolic achievement **[5]**. On this point my friend and I agreed.

Descartes' curve-drawing devices poignantly raise the issue of technology and its relation to mathematical investigation. During the seventeenth century there was a distinct turning away from the classical Greek orientation that had been popular during the Renaissance in favor of pragmatic and stoic Roman philosophy. During much of the seventeenth century a class in "Geometry" would concern itself mainly with the design of fortifications, siege engines, canals, water systems, and hoisting devices—what we would call civil and mechanical engineering. Descartes' *Geometry* was not about static constructions and axiomatic proofs, but concerned itself instead with mechanical motions and their possible representation by algebraic equations. Classical problems were addressed, but they were all transformed into locus problems, through the use of a wide variety of motions and devices that went far beyond the classical restriction to straight-edge and compass. Descartes sought to build a geometry that included all curves whose construction he considered "clear and distinct" **[11, 20]**. An examination of his work shows that what he meant by this was any curve that could be drawn with a "linkage," i.e., a device made of hinged rigid rods. Descartes' work indicates that he was well aware that this class of curves is exactly the class of all algebraic curves, although he gave no formal proof of this. This theorem is scarcely known among modern mathematicians, although it can be proved straightforwardly by looking at linkages that add, subtract, multiply, divide, and generate integer powers **[3]**. Descartes' linkage for generating any integer power was used repeatedly in the *Geometry* and has many interesting possibilities **[10]**.

This transformation of geometry from classical static constructions to problems involving motions and their resultant loci has once again raised itself in light of modern computer technology, specifically the advent of dynamic geometry software such as *Cabri* and *Geometer's Sketchpad*. Many new educational and research possibilities have emerged recently in response to these technological developments **[26]**. It seems, indeed, that seventeenth century mechanical geometry may yet rise from the ashes of history and regain a new electronic life in our mathematics classrooms. (It has always had a life in our schools of engineering, where the finding of equations that model motion has always been a fundamental concern.) My own explorations of seventeenth century dynamic geometry have been conducted with a combination of physical models and devices along with computer animations made using *Geometer's Sketchpad* **[18]**. The first figure in this paper is taken directly from Descartes, but all the others were made using *Geometer's Sketchpad*. This software allows a more authentic historical exploration since curves are generated from geometrical actions rather than as the graphs of equations. Static figures cannot vividly convey the sense of motion that is necessary for a complete understanding of these devices.[3] In the generation of the figures in this paper no equations were typed into the computer.

FIGURE 1 is reproduced from the (original) 1637 edition of Descartes' *Geometry* [**11**, p. 50]. Descartes described the device as follows:

> *Suppose the curve EC to be described by the intersection of the ruler GL and the rectilinear plane figure NKL, whose side KN is produced indefinitely in the direction of C, and which, being moved in the same plane in such a way that its diameter KL always coincides with some part of the line BA (produced in both directions), imparts to the ruler GL a rotary motion about G (the ruler being hinged to the figure NKL at L). If I wish to find out to what class this curve belongs, I choose a straight line, as AB, to which to refer all its points, and on AB I choose a point A at which to begin the investigation. I say "choose this and that," because we are free to choose what we will, for, while it is necessary to use care in the choice, in order to make the equation as short and simple as possible, yet no matter what line I should take instead of AB the curve would always prove to be of the same class, a fact easily demonstrated.*

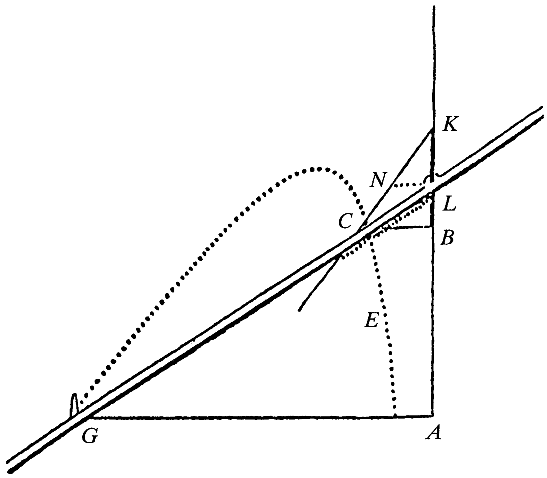


FIGURE 1
Descartes' Hyperbolic Device

Descartes addressed here several of his main points concerning the relations between geometrical actions and their symbolic representations. His "classes of curves" refer to the use of algebraic degrees to create a taxonomy of curves. He is asserting that the algebraic degree of an equation representing a curve is independent of how one chooses to impose a coordinate system. Scale, starting point, and even the angle between axes will not change the degree of the equation, although this "fact easily demonstrated" is never given anything like a formal proof in the *Geometry*. Descartes also mentioned here the issue of a judicious choice of coordinates, an important scientific issue that goes largely unaddressed in modern mathematics curricula until an advanced level, at which point geometry is scarcely mentioned.

Descartes went on to find the equation of the curve in FIGURE 1 as follows. Introduce the variables (Descartes used the term "unknown and indeterminate

[3] Animated figures made in *Geometer's Sketchpad* are available from the author by e-mail (dennis@math.utep.edu).

quantities") $AB = y$, $BC = x$ (in modern notation, $C = (x, y)$), and then the constants ("known quantities") $GA = a$, $KL = b$, and $NL = c$. Descartes routinely used the lower case letters x, y, and z as variables, and a, b, and c as constants; our modern convention stems from his usage. Descartes, however, had no convention about which variable was used horizontally, or in which direction (right or left) a variable was measured (here, x is measured to the left). There was, in general, no demand that x and y be measured at right angles to each other. The variables were tailored to the geometric situation. There was a very hesitant use of negative values (often called "false roots"), and in most geometric situations they were avoided.

Continuing with the derivation, since the triangles KLN and KBC are similar, we have $\frac{c}{b} = \frac{x}{BK}$, hence $BK = \frac{b}{c}x$, hence $BL = \frac{b}{c}x - b$. From this it follows that $AL = y + BL = y + \frac{b}{c}x - b$. Since triangles LBC and LAG are similar, we have $\frac{BC}{BL} = \frac{AG}{AL}$. This implies the following chain of equations:

$$\frac{x}{\frac{b}{c}x - b} = \frac{a}{y + \frac{b}{c}x - b} \Leftrightarrow x\left(y + \frac{b}{c}x - b\right) = a\left(\frac{b}{c}x - b\right)$$

$$\Leftrightarrow xy + \frac{b}{c}x^2 - bx = \frac{ab}{c}x - ab$$

$$\Leftrightarrow x^2 = cx - \frac{c}{b}xy + ax - ac. \qquad (1)$$

Descartes left the equation in this form because he wished to emphasize its second degree. He concluded that the curve is a hyperbola. How does this follow? As we said before Descartes assumed that his readers were well acquainted with Apollonius. We will return to this issue shortly.

If one continues to let the triangle NLK rise along the vertical line, and keeps tracing the locus of the intersection of GL with NK, the lines will eventually become parallel (see FIGURE 2), and after that the other branch of the hyperbola will appear (see FIGURE 3).

These figures were made with *Geometer's Sketchpad*, although I have altered slightly the values of the constants a, b, and c from those in FIGURE 1. In FIGURE 2, the line KN is in the asymptotic position, i.e., parallel to GL. I will hereafter refer to this particular position of the point K, as point O. In this position triangles NLK and GAL are similar, so $AK = AO = \frac{ab}{c} + b$ (the y-intercept of the asymptote). The slope of the asymptote is the same as the fixed slope of KN, i.e., b/c. (Recall that $KL = b$, $NL = c$, and $GA = a$.)

To rewrite Equation 1 using A as the origin in the usual modern sense, with x measured positively to the right, we can substitute $-x$ for x. With this substitution, solving Equation 1 for y yields

$$y = ab\frac{1}{x} + \frac{b}{c}x + \left(\frac{ab}{c} + b\right). \qquad (2)$$

In Equation 2, the linear equation of the asymptote appears as the last two terms. In FIGURE 3, I have shown, to the right, the lengths that represent the values of the three terms in Equation 2, for the point P. (The labels 1, 2, and 3 represent, respectively, the inverse term, the linear term, and the constant term.) Term 3 accounts for the rise from the x-axis to the level of point O (the intercept of the asymptote). Adding term 2 raises one to the level of the asymptote, and term 1 completes the ordinate to the curve.

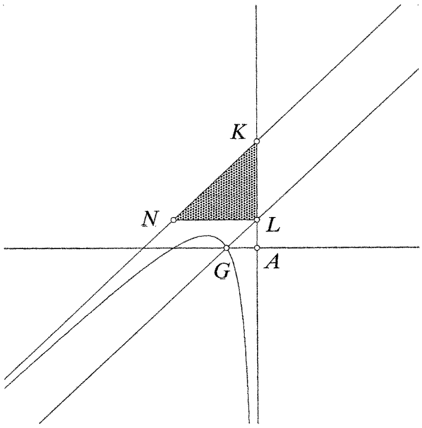


FIGURE 2
Descartes' Device in the Asymptotic Position

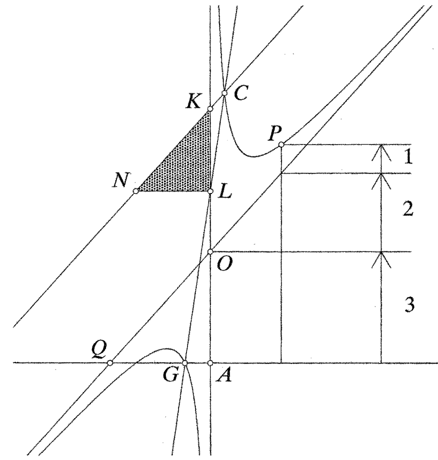


FIGURE 3
Geometrical Display of the Terms in the Hyperbolic Equation

As a geometric construction, the hyperbola is drawn from parameters that specify the angle between the asymptotes ($\angle NKL$), and a point on the curve (G). If one changes the position of the point N without changing the angle $\angle NKL$, the curve is unaffected, as in FIGURE 4. The derivation of the equation depends only on similarity, and not on having perpendicular coordinates. As long as GA (which determines the coordinate system) is parallel to NL, the derivation of the equation is the same except for the values of the constants $NL = c$, and $GA = a$ (both have become larger in FIGURE 4). Of course this equation is in the oblique coordinate system of the lines GA (x-axis) and AK (y-axis). It is the same curve geometrically, with the same form of equation, but with new constant values that refer to an oblique coordinate system. As long as angle $\angle NKL$ remains the same, and G is taken at the same distance from the line KL, the device will draw the same curve. This form of a hyperbolic equation, as an inverse term plus linear terms, depends only on using at least one of the asymptotes as an axis.

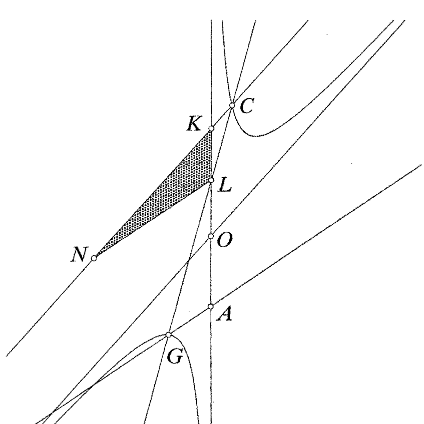


FIGURE 4
Hyperbola in Skewed Coordinates

I have encountered many students who are well acquainted with the function $y = \frac{1}{x}$, and yet have no idea that its graph is an hyperbola. Descartes' construction can be adjusted to draw right hyperbolas. Consider the special case in which the line KN is parallel to the x-axis (see FIGURE 5). The point G is on the negative x-axis. Let $KC = x$, and $AK = y$ (i.e., $C = (x, y)$), $AG = a$, and $KL = b$. Now $AL = y - b$, and since triangles LKC, and LAG are similar, we have $\frac{KC}{KL} = \frac{AG}{AL}$, or, equivalently, $\frac{x}{b} = \frac{a}{y-b}$. Hence the curve has equation

$$y = ab\frac{1}{x} + b. \tag{3}$$

A vertical translation by b would move the origin to the point O, and letting $a = b = 1$, would put G at the vertex $(-1, -1)$, yielding the curve with equation $y = 1/x$.

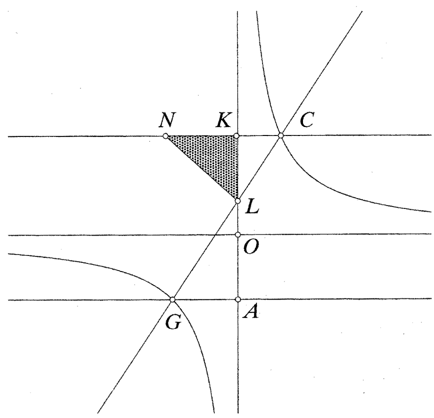


FIGURE 5
Device Adjusted to Draw Right Hyperbolas

Equation 3 can be seen as a special case of Equation 2, obtained by substituting ∞ for c, where c is thought of as the horizontal distance from L to the line KN. All translations and rescalings of the multiplicative inverse function can be directly seen as special members of the family of hyperbolas, using this construction.

Appollonius Regained

How do we know that these curves are, in fact, hyperbolas? Descartes said that this is implied by Equation 1. In his commentaries on Descartes, van Schooten gives us more detail [**11**, p. 55, note 86]. Once again these mathematicians assumed that their readers were familiar with a variety of ratio properties from Book 2 of the *Conics* of Apollonius [**1**, **15**] that are equivalent to Equation 1. I will not give a full set of formal proofs, but will instead suggest means for exploring these relations.

Several beautiful theorems of Apollonius concerning the relations between tangents and asymptotes are easily explored in this setting. Using the asymptotes of the curve in FIGURE 5 as edges to define rectangles, one sees that the points on the curve define a family of rectangles, all with the same area (see FIGURE 6). Indeed, if M and N are any two points on the curve, Equation 3 implies that $OPMS$ and $OQNR$ both have area equal to $a \cdot b$, the product of the constants used in drawing the curve. Another interesting geometric property is that the triangles TSM and NQU are always congruent. This congruence provides one way to dissect and compare these rectangles in a geometric manner [**17**].

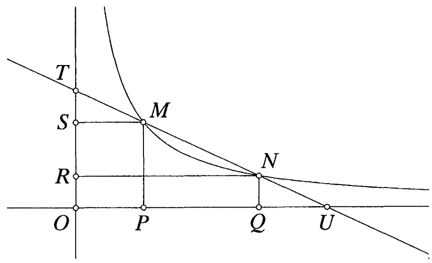


FIGURE 6
Hyperbola as a Family of Equal Area Rectangles

Approaching these equations analytically, assume that the curve in FIGURE 6 has the equation $x \cdot y = k$ (using O as the origin). Let $M = (m, k/m)$ and $N = (n, k/m)$, i.e., $OP = m$ and $OQ = n$. The line through M and N has equation $y = \frac{-k}{mn}x + \left(\frac{k}{m} + \frac{k}{n}\right)$. Hence $TO = \frac{k}{m} + \frac{k}{n}$, and, since $SO = \frac{k}{m}$, this implies that $TS = \frac{k}{n} = NQ$. Since triangles TSM and NQU are clearly similar, $TS = NQ$ implies that they are congruent and that $TM = NU$. Now let the points M and N get close to each other; then the line MN gets close to a tangent line, and one can perceive a theorem of Apollonius:

> *Given any tangent line to a hyperbola, the segment of the tangent contained between the two asymptotes is always bisected by the point of tangency to the curve* [**1**, **15**].

This property is a defining characteristic of hyperbolas. This simple and beautiful theorem immediately implies, among other things, that the derivative of $\frac{1}{x}$ is $\frac{-1}{x^2}$. (Look at the congruent triangles and compute the rise over run for the tangent.) This gives a student an independent geometrical check on the validity of the calculus derivation.

This bisection property of hyperbolic tangents is not restricted to the right hyperbola. Looking back at FIGURE 3 and Equation 2, one sees that any hyperbola coordinatized along both its asymptotes will always have an equation of the form $x \cdot y = k$ for some constant k. To see this, subtract off the linear and constant terms from the y-coordinate, and then rescale the x-coordinates by a constant factor that projects them in the asymptotic direction (in FIGURE 7 the new x-coordinate in this skew system is OQ). In the general case the curve can be seen as the set of corners of a family of equi-angular parallelograms, all with the same area. In FIGURE 7, for any two points M and N on the curve, the parallelograms $OQNR$ and $OPMS$ have equal

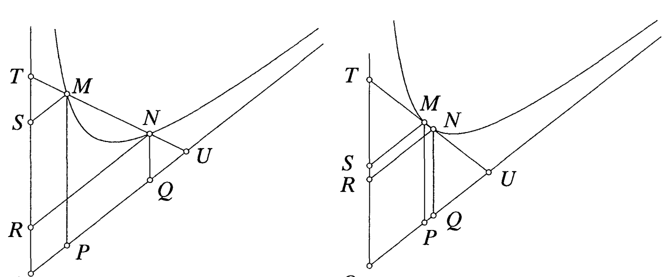


FIGURE 7
Bisection Property of Hyperbolic Tangents

areas. Since the triangles TSM and NQU are congruent, by letting M and N get close together one sees that any tangent segment TU is bisected by the point of tangency (M or N).

An alternative view of the situations just described is to imagine any line parallel to TU meeting the asymptotes and the curve in corresponding points T', M', and U'. Then the product $T'M' \cdot M'U' = TM \cdot MU$. That is to say, parallel chords between the asymptotes of a hyperbola are divided by the curve into pieces with a constant product. This follows from our discussion, because the pieces are constant projections of the sides of the parallelograms just discussed. This form of the statement was most often used by van Schooten, Newton, Euler and others in the seventeenth century. This statement (from Book 2 of Apollonius [**1**, **15**]) was traditionally used as an identifying property of hyperbolas. This constant product was given as a proof by van Schooten that the curve drawn by Descartes' device was indeed a hyperbola [**11**, p. 55]. Apollonius derived these properties directly from sections of a general cone.

In this way it is possible to investigate hyperbolas, using both geometric and algebraic representations, to create a complete cognitive feedback loop. Neither representation is used as a foundation for proof; instead, one is led to a belief in a relative consistency between certain aspects of geometry and algebra through checking back and forth between alternative representations. A calculus derivation of the derivative of $y = 1/x$ becomes, in this setting, a limited special case of the bisection property of hyperbolic tangents. It can be very satisfying to see symbolic algebra arrange itself into answers that are consistent with physical and geometric experience. Students of calculus can then experience the elation of Leibniz, as they build up a vocabulary of viable notation, capable of being checked against independently verifiable physical and geometric experience. Mathematical language is then seen as a powerful code for aspects of experience, rather than as the sole dictator of truth.

Conchoids Generalized from Hyperbolas

The hyperbolic device is only the beginning of what appears in Descartes' *Geometry*. He discussed several cases where curve-drawing constructions can be progressively iterated to produce curves of higher and higher algebraic degree [**11**, **10**]. It is usually mentioned in histories of mathematics that Descartes was the first to classify curves according to the algebraic degree of their equations. This is not quite accurate. Descartes classified curves according to *pairs* of algebraic degrees; i.e., lines and conics form his first class (he used the term *genre*), curves with third or fourth degree equations form his second class, etc. [**11**, p. 48]. This classification is quite natural if one is working with mechanical linkages and loci. With most examples of iterated linkage, each iteration raises the degree of the curve's equation by two, with some special cases that collapse back to an odd algebraic degree [**7**].[4] What follows is an example of such an interation based on the hyperbolic device.

---

[4] Descartes' linkages led directly to Newton's universal method for drawing conics, which is essentially a projective method [**7**, **23**]. This same classification by pairs of degrees is used in modern topology in the definition of "genus." The "genus" of a non-singular algebraic plane curve can be thought of topologically as the number of "handles" on the curve when defined in complex projective space. In complex projective space, linear and quadratic non-singular curves have genus 0, and are topologically sphere-like. Similarly, curves of degrees 3 and 4 are topologically torus-like, and have genus 1. Curves of degrees 5 and 6 are topologically double-holed and have genus 2, etc. In the real model, (i.e., when considering only real solutions of one real equation in 2 variables) the genus 0 curves consist of at most one oval when you join up the asymptotes. The genus 1 curves will have two ovals, which is what you'd expect when cutting through a toric by a plane, etc. (This comment was made to me by Paul Pedersen.)

Descartes generalized the previous hyperbola construction method by replacing the triangle *KLN* with any previously constructed curve. For example, let a circle with center L be moved along one axis and let the points C and C' be the intersections of the circle with the line LG, where G is any fixed point in the plane and LG is a ruler hinged at point L just as in the hyperbolic device (see FIGURE 8). Then C traces out a curve of degree four, known in ancient times as a *conchoid* [**11**, p. 55]. The two geometric parameters involved in the device are the radius of the circle (r), and the distance (a) between the point G and the axis along which L moves.

FIGURE 8 shows three examples of conchoids for $a > r$, $a = r$, and $a < r$. If the curve is coordinatized along the path of L, and a perpendicular line through G (OG), then its equation can be found by looking at the similar triangles GOL and CXL (top of FIGURE 8). Since $GO = a$, $LC = r$, $CX = y$, $OX = x$, and $XL = \sqrt{r^2 - y^2}$, one obtains the ratios of the legs of the triangles as follows: $\frac{\sqrt{r^2 - y^2}}{y} = \frac{\sqrt{r^2 - y^2} + x}{a}$. This is equivalent to $x^2y^2 = (r^2 - y^2)(a - y)^2$, an equation of fourth degree, or of Descartes' second class. (The squared form of the equation has both branches of the curve, above and below the axis, as solutions.)

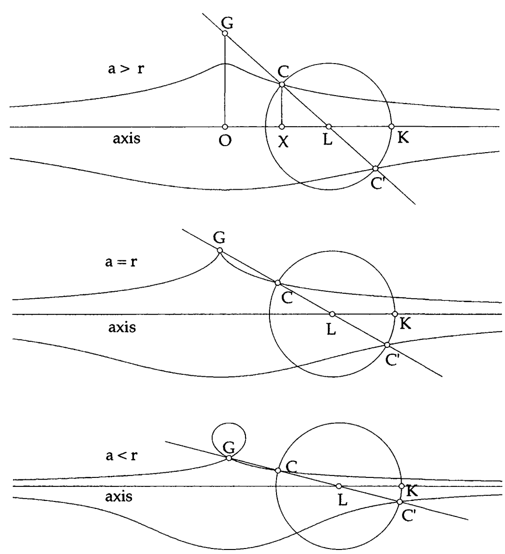


FIGURE 8
Conchoids Drawn by Dragging a Circle Along a Line

This example demonstrates Descartes' claim that, as one uses previously constructed curves to draw new curves, one gets chains of constructed curves that go up by pairs of algebraic degrees. Descartes called the conchoid a curve of the second class, i.e., of degree three or four. Dragging any rigid conic-sectioned shape along the axis, and drawing a curve in this manner will produce curves in the second class. Dragging curves of the second class will produce curves of the third class (i.e., degree five or six), etc. Descartes demonstrated this general principle through many examples [**11**, **7**, **10**], but he offered nothing like a formal proof, either geometric or algebraic. His definition of curve classes was justified by his geometric experience.

Notice that when $a \leq r$, the point G becomes a cusp or a crossover point. When singularities like cusps or crossover points occur, these tend to occur at important parts of the apparatus, like a pivot point (such as G) or a point on an axis of motion.

Other important examples of this phenomena can be found in Newton's notebooks [**22**, **23**]. I am not asserting any particular or explicit mathematical theorem here. This general observation is based upon my own historical research and empirical experience with curve-drawing devices. There are probably several ways to make this observation into an explicit mathematical statement, subject to proof (Newton attempted several [**23**]). There are many open questions concerning these forms of curve iteration and the relations between the parts of the physical devices and the singularities of the curves [**7**]. Students might benefit from such empirical experience—regardless of the extent to which they eventually formalize that experience in strictly algebraic or logical language. An instinctual sense of where curve singularities might occur is fundamentally useful in many sciences [**2**]. Modern computer software makes such investigations routinely possible with a minimum of technical expertise.

Conclusion

Descartes wrote his *Rules for the Direction of the Mind* [**12**] in 1625, twelve years before he would publish his famous *Geometry*. In this earlier work he emphasized the importance of making strong connections between physical actions and their possible representations in diagrams and language. Here are a few quotes:

> *Rule 13: If we understand a problem perfectly, it should be considered apart from all superfluous concepts, reduced to its simplest form, and divided by enumeration into the smallest possible parts.*
>
> *Rule 14: The same problem should be understood as relating to the actual extension of bodies and at the same time should be completely represented by diagrams to the imagination, for thus will it be much more distinctly perceived by the intellect.*
>
> *Rule 15: It is usually helpful, also, to draw these diagrams and observe them through the external senses, so that by this means our thought can more easily remain attentive.*

These lines from Descartes sound much like parts of the hands-on, problem-solving educational philosophy of mathematics put forth by the National Council of Teachers of Mathematics [**21**]. Descartes' entire approach to mathematics had problem solving as its foundation [**14**], but we must not allow ourselves to read into him too modern a perspective. He was constructing a new method of mathematical representation that responded to both the new symbolic language of his time (algebra) and to the new technology of his time (mechanical engineering). He was not seeking the broad educational goals of the NCTM. In fact, his *Geometry* was not widely read in the seventeenth century until it was republished, in 1657, with extensive commentaries by Franz van Schooten.

Nonetheless, Descartes' approach to geometry through curve-drawing devices and locus problems has important implications for education. His work connects important classical and Arabic traditions with modern algebraic formalisms [**7**]. It provides the missing linkages (pun intended). These linkage and loci problems, combined with the new dynamic geometry software, allow a new kind of exploration of curves that could go far towards ending the isolation of geometry in our mathematics curriculum. One can use geometrical curve generation to recreate calculus concepts such as tangents and areas in a much more elementary and physical setting [**7**, **8**, **10**], as well as to explore complicated questions about algebraic curves left open since the seventeenth century **[7, 23]**. Computer graphic techniques have already led to new branches of mathematics, such as fractals. Perhaps a new phase of computer-assisted empirical geometrical investigation of curves and surfaces has already begun. If this new beginning proves as revolutionary as the century that began with Descartes' *Geometry*, then we are in for some very exciting times.

Acknowledgment. This research was funded by a grant from the National Science Foundation.

REFERENCES

1. Apollonius of Perga, *On Conic Sections*, in Vol. 11 of *The Great Books of the Western World*, Encyclopedia Britannica, Chicago, IL, 1952.
2. V. I. Arnol'd, *Huygens & Barrow, Newton & Hooke*, Birkhaüser Verlag, Boston, MA, 1990.
3. I. I. Artobolevskii, *Mechanisms for the Generation of Plane Curves*, Macmillan, New York, NY, 1964.
4. F. Cajori, Controversies on mathematics between Wallis, Hobbes, and Barrow, *The Mathematics Teacher* 22 (1929), 146–151.
5. J. M. Child, *The Early Mathematical Manuscripts of Leibniz*, Open Court, Chicago, IL, 1920.
6. J. Confrey, A theory of intellectual development, *For the Learning of Mathematics* (in three consecutive issues) 14 (1994), 2–8; 15 (1994), 38–48; 15 (1995).
7. D. Dennis, *Historical Perspectives for the Reform of Mathematics Curriculum: Geometric Curve Drawing Devices and their Role in the Transition to an Algebraic Description of Functions*, unpublished doctoral dissertation, Cornell University, Ithaca, NY, 1995.
8. D. Dennis and J. Confrey, Functions of a curve: Leibniz's original notion of functions and its meaning for the parabola, *The College Mathematics Journal* 26 (1995), 124–130.
9. D. Dennis and J. Confrey, The Creation of Continuous Exponents: A Study of the Methods and Epistemology of Alhazen and Wallis. In J. Kaput & E. Dubinsky (Eds.) *Research in Collegiate Mathematics II*. CBMS Vol. 6, pp. 33–60, American Mathematical Society, Providence, RI, 1996.
10. D. Dennis and J. Confrey, Drawing Logarithmic Curves with *Geometer's Sketchpad*: A Method Inspired by Historical Sources. In J. King and D. Schattschneider (Ed.), *Geometry Turned On: Dynamic Software in Learning, Teaching, and Research*, MAA, Washington, DC, in press.
11. R. Descartes, *The Geometry*, Dover, New York, NY, 1954.
12. R. Descartes, *Rules For the Direction of the Mind*, Bobbs-Merrill, New York, NY, 1961.
13. M. Foucault, *The Order Of Things: An Archeology Of The Human Sciences*, Pantheon Books, New York, NY, 1970.
14. J. Grabiner, Descartes and problem solving, this MAGAZINE 86 (1995), 83–97.
15. T. L. Heath, *Apollonius of Perga: Treatise on Conic Sections*, Barnes & Noble, New York, NY, 1961.
16. T. L. Heath, *The Thirteen Books of Euclid's Elements*, Dover, New York, NY, 1956.
17. D. Henderson, *Experiencing Geometry on Plane and Sphere*, Prentice-Hall, Englewood Cliffs, NJ, 1996.
18. N. Jackiw, *Geometer's Sketchpad* (computer program), Key Curriculum Press, Berkeley, CA, 1996.
19. J. Klein, *Greek Mathematical Thought and the Origin of Algebra*, M.I.T. Press, Cambridge, MA, 1968.
20. T. Lenoir, Descartes and the geometrization of thought: The methodological background of Descartes' geometry, *Historia Mathematica* 6 (1979), 355–379.
21. National Council of Teachers of Mathematics, *Professional Standards for Teaching Mathematics*, NCTM, Reston, VA, 1991.
22. I. Newton, *The Mathematical Papers of Isaac Newton*, Vol. 1 (1664–1666), Cambridge University Press, Cambridge, UK, 1967.
23. I. Newton, *The Mathematical Papers of Isaac Newton*, Vol. 2 (1667–1670), Cambridge University Press, Cambridge, UK, 1968.
24. B. Rotman, *Signifying Nothing: The Semiotics of Zero*, Stanford University Press, Stanford, CA, 1987.
25. S. Shapin and S. Schaffer, *Leviathan and the Air Pump: Hobbes, Boyle, and the Experimental Life*, Princeton University Press, Princeton, NJ, 1985.
26. D. Schattschneider and J. King (Eds.), *Geometry Turned On: Dynamic Software in Learning, Teaching, and Research*, MAA, Washington, DC, in press.

Knight's Tours on a Torus

JOHN J. WATKINS
Colorado College
Colorado Springs, CO 80903

REBECCA L. HOENIGMAN
Department of Chemistry
University of Washington
Seattle, WA 98195

Introduction

The knight is the only piece in chess that does not move in a straight line. Instead, it moves in an "L"—two squares in either a vertical or horizontal direction and then one square in a perpendicular direction. It is the strangeness of this move that has made the *Knight's Tour Problem* one of the most intriguing in all of recreational mathematics: *Can a knight visit each square of a chessboard by a sequence of knight's moves, and finish on the same square as it began*? Since a chessboard can be represented as a graph in which each vertex corresponds to a square, and edges correspond to those pairs of squares connected by a knight's move (FIGURE 1 illustrates this for a 4×4 board), finding a knight's tour amounts to finding a Hamiltonian cycle in the corresponding graph, a notoriously difficult general problem in graph theory (see [5]). However, we can easily see that there is *no* knight's tour for a 4×4 board since any Hamiltonian cycle would have to include the four edges incident to the two corner vertices indicated in FIGURE 1; this is impossible since these four edges already form a cycle that includes only four vertices.

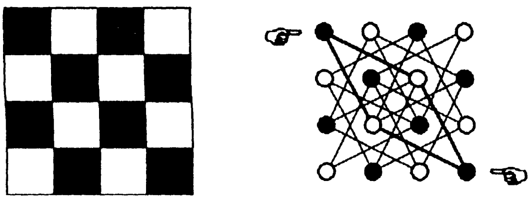


FIGURE 1
Representing a chessboard as a graph

We can also notice that the vertices in a knight's graph can be colored black and white so that *every* edge joins a black vertex and a white vertex. Such a graph is called *bipartite*. Since any cycle in a bipartite graph must have an *even* number of edges, we conclude that an $m \times n$ board with m and n odd cannot have a knight's tour, because the corresponding Hamiltonian cycle would have an odd number of edges.

There are several excellent sources for the history of this problem. We particularly recommend the discussion by W. W. Rouse Ball ([1]), which includes contributions by Euler as well as an ingenious method by the German mathematician H. C. Warnsdorff, dating from 1823, in which the knight is always moved to one of the squares from which it will have the fewest open moves. Combining this rule with Euler's techniques provides a remarkably efficient way to find knight's tours on various boards. Martin Gardner ([3]) presents several other problems involving knights, as well as giving S. W. Golomb's elegant proof that no $4 \times n$ board has a knight's tour. In 1991, Schwenk ([4]) answered the obvious question: *Which rectangular chessboards have a knight's tour*?

THEOREM. *An $m \times n$ chessboard with $m \leq n$ has a knight's tour unless one or more of these conditions holds:*

(1) *m and n are both odd;*
(2) *$m = 1$, 2, or 4; or*
(3) *$m = 3$ and $n = 4$, 6, or 8.*

But what if we allow the knight to move off the side of the board and then return to the board on the opposite side, as in some video games? (Such moves were used in [**2**] to find Hamiltonian tours for checkers.) For example, with this change it is now possible to find a knight's tour of a 5×5 board—in fact, Warnsdorff's method can be used here—since in FIGURE 2 a knight at square 25 *can* return to square 1 in a legal move by going off the bottom edge.

1	14	9	20	3
24	19	2	15	10
13	8	23	4	21
18	25	6	11	16
7	12	17	22	5

FIGURE 2
Knight's tour of a 5×5 board on a torus

This is equivalent to changing the flat chessboard into a torus (i.e., a doughnut) by gluing the top edge to the bottom edge (which creates a cylinder) and then gluing the left and right edges (which brings the two ends of the cylinder together). So we now pose the question: *Which rectangular chessboards have a knight's tour on a torus?*

Knight's Tours on a Torus

In this section we will prove that, on a torus, *every* rectangular board has a tour. First, we establish some useful notation.

A knight has eight possible moves as shown in FIGURE 3. Each move has an arithmetic description (x, y) where x indicates how many squares the knight moves to the right and y indicates how many squares down. Notice the symmetry between moves a, b, c, d and α, β, γ, δ, respectively; this will become important later.

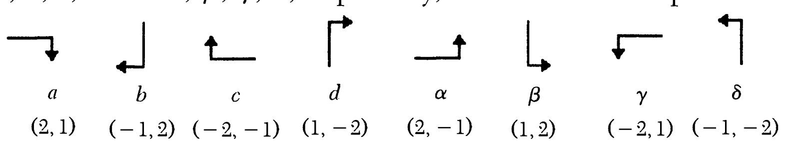


FIGURE 3
The eight knight moves

Our strategy will be to provide explicit tours for boards with a small number of rows (but any number of columns) and then to show how to 'stack' these boards together to form tours for arbitrary boards.

$1 \times n$ and $2 \times n$ boards You can easily tour any $1 \times n$ board by starting at any square and making move $\beta = (1, 2)$ n consecutive times. This is illustrated in FIGURE 4 for a 1×6 board. Similarly, you can tour any $2 \times n$ board by making move $(1, 2)$ until

1	2	3	4	5	6

1	2	3	4	5	6	7
9	8	14	13	12	11	10

FIGURE 4
Tours for 1×6 and 2×7 boards

you get stuck half way through, at which point you make move $(2,1)$ once, and then continue with move $(-1,2)$ until every square has been visited and you can take move $(-2,-1)$ back to the starting point. This is illustrated in FIGURE 4 for a 2×7 board.

3 × *n* boards You can tour any $3 \times n$ board, as long as n is not a multiple of 5, by repeating the three moves $(2,1)$, $(2,1)$, and $(1,-2)$ over and over again. If n is a multiple of 5 you can repeat the moves $(2,1)$, $(2,1)$, and $(-1,-2)$ over and over instead. These two cases are illustrated in FIGURE 5 for a 3×8 and a 3×10 board. Notice that in neither case do you ever go off the top or bottom edge.

1	16	7	22	13	4	19	10
20	11	2	17	8	23	14	5
15	6	21	12	3	18	9	24

1	22	13	4	25	16	7	28	19	10
20	11	2	23	14	5	26	17	8	29
9	30	21	12	3	24	15	6	27	18

FIGURE 5

Tours for 3×8 and 3×10 boards

4 × *n* boards There are two cases. If n is *odd* you can alternate moves $(1,2)$ and $(1,-2)$ until you get stuck half way through (and the squares in the first and third rows have all been visited), at which point you make move $(2,-1)$, from 18 to 19 in FIGURE 6, and then continue alternating with moves $(-1,2)$ and $(-1,-2)$ until every square has been visited (at 36) and you can take move $(-2,1)$ back to the starting point. If n is *even* you again alternate moves $(1,2)$ and $(1,-2)$, but this time you get stuck a quarter of the way through, at 10 in FIGURE 6, at which point you make move $(2,1)$; then alternate $(-1,-2)$ and $(-1,2)$ until you get stuck (at 20) and make move $(-2,1)$; next, alternate $(1,-2)$ and $(1,2)$ until you get stuck (at 30) and make move $(2,1)$; finally, alternate $(-1,2)$ and $(-1,-2)$ until every square has been visited (at 40) and you can take move $(-2,1)$ back to the starting point. These two cases are illustrated in FIGURE 6 for a 4×9 and a 4×10 board. Notice in each case that *only* the last move goes off the top or bottom edge.

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

1	22	3	24	5	26	7	28	9	**30**
12	31	**20**	39	18	37	16	35	14	33
21	2	23	4	25	6	27	8	29	**10**
32	11	**40**	19	38	17	36	15	34	13

FIGURE 6

Tours for 4×9 and 4×10 boards

Notice that at this point we have already taken care of exceptions (2) and (3) of Schwenk's theorem. Strictly speaking, all that remains to do is the case of an odd by odd board on a torus. However, in part for the sake of completeness and in part because we like the constructions involved, we will instead consider *all* remaining boards.

Even × odd boards In order to show that any board with an even number of rows and an odd number of columns can be toured, we will simply stack together an appropriate number of boards each having two rows. However, a difficulty arises since the tour of a $2 \times n$ board shown in FIGURE 4 uses the top and bottom edge of the board many times. Fortunately, if a $2 \times n$ board has an *odd* number of columns, there is a tour that does not use the top and bottom edge: you simply alternate moves $(2,1)$ and $(2,-1)$. This is illustrated in FIGURE 7 for a 2×7 board.

1	5	9	13	3	7	11
8	12	2	6	10	14	4

FIGURE 7

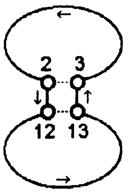
Alternate tour for a 2×7 board

It is easy to stack any number of these boards on top of one another. We illustrate this by creating a tour for a 4×7 board from the tours of two 2×7 boards. It is perhaps best to think in terms of the corresponding graphs. The idea—used by Euler—is to remove two edges, one from each Hamiltonian cycle, and then add two edges that join the two pieces into a single cycle. The only trick is to make sure the edges you add correspond to legal knight moves.

In FIGURE 8 we remove edge 2–3 from the top board and edge 12–13 from the bottom board, and then add edge 2–12 and edge 3–13, *both of which correspond to legal knight moves*. Still thinking in terms of the graph, it is now routine to do a knight's tour by beginning at square 1 on the top board, going to square 2, then to the bottom board at square 12, at which point we travel *backwards* on the bottom board until we reach square 13 from which we return to the top board at square 3 and finish the tour on the top board by taking the squares in order. The result, with the appropriate renumbering, is shown in FIGURE 8.

1	5	9	13	3	7	11
8	12	2	6	10	14	4

1	5	9	13	3	7	11
8	12	2	6	10	14	4



1	19	23	27	17	21	25
22	26	2	20	24	28	18
14	10	6	16	12	8	4
7	3	13	9	5	15	11

FIGURE 8
Stacking two 2×7 boards

It is clear that this process can be continued indefinitely; for example, we can stack another 2×7 board on top of the 4×7 board by again removing edge 2–3 from the top board and removing edge 26–27 from the bottom board. In this way we can construct a knight's tour for any board with an even number of rows and an odd number of columns (and, by symmetry, any board with an odd number of rows and an even number of columns).

Odd $\times$ odd boards We can now take care of exception (1) in Schwenk's theorem. In order to do a board with an odd number of rows (and an odd number of columns) we simply stack a board with 3 rows on top of a board with an even number of rows as done above. We illustrate this for a 7×7 board in FIGURE 9 using edge 5–6 from the top board and edge 19–20 from the bottom board.

1	10	19	7	16	4	13
5	14	2	11	20	8	17
9	18	6	15	3	12	21

1	19	23	27	17	21	25
22	26	2	20	24	28	18
14	10	6	16	12	8	4
7	3	13	9	5	15	11

1	38	47	35	44	4	41
5	42	2	39	48	36	45
37	46	34	43	3	40	49
24	6	30	26	8	32	28
31	27	23	33	29	25	7
11	15	19	9	13	17	21
18	22	12	16	20	10	14

FIGURE 9
A 7×7 board

Even $\times$ even boards We consider two cases. First, if the *number of rows is divisible by* 4, then we can stack multiple copies of $4 \times n$ boards. We illustrate this in FIGURE 10 for an 8×6 board.

1	14	3	16	5	18
8	19	12	23	10	21
13	2	15	4	17	6
20	7	24	11	22	9

1	14	3	16	5	18
8	19	12	23	10	21
13	2	15	4	17	6
20	7	24	11	22	9

1	14	3	16	5	18
8	19	12	23	10	21
13	2	15	4	17	6
20	7	24	11	22	9
25	38	27	40	29	42
32	43	36	47	34	45
37	26	39	28	41	30
44	31	48	35	46	33

FIGURE 10

An 8×6 board

We remove edge 1–24 from each board—remember this was the only move that used the top and bottom edge of the board—and join the 24 at the top to the 1 at the bottom and the 24 at the bottom to the 1 at the top. By noticing the position of the 1 and the 48 in the 8×6 board, we see that we can repeat this procedure as many times as we like, simply adding four rows at a time.

The second case—where the *number of rows is even but not divisible by* 4—is a good bit harder. This will be done by showing how to stack a board with 6 rows on a board with $4k$ rows. First we show how to join two $3 \times n$ boards to get the $6 \times n$ board which we need. From the top $3 \times n$ board remove the edge that joins the next to last square in the second row to the first square in the last row—for example, edge 14–15 in Figure 11. From the bottom $3 \times n$ board remove the edge that joins the next to last square in the first row to the second from last square in the last row—that is, edge 19–18 in Figure 11. We can now add two edges in the obvious way to create a Hamiltonian cycle—namely, edges 15–19 and 14–18 in Figure 11. Notice that in the resulting tour of the board with 6 rows, the next to the last square in the second row is connected to the second from last square in the last row—that is, 14–15 in Figure 11. It is this edge that we will remove in the next step.

1	16	7	22	13	4	19	10
20	11	2	17	8	23	14	5
15	6	21	12	3	18	9	24

1	16	7	22	13	4	19	10
20	11	2	17	8	23	14	5
15	6	21	12	3	18	9	24

1	40	7	46	13	4	43	10
44	11	2	41	8	47	14	5
39	6	45	12	3	42	9	48
32	17	26	35	20	29	38	23
37	22	31	16	25	34	19	28
18	27	36	21	30	15	24	33

FIGURE 11

A 6×8 board

Next, we get a tour for the board with $4k$ rows exactly as we did previously *except* that we begin the tour in the top row *five* squares from the right-hand edge of the board rather than in the upper left-hand corner as we usually do—notice the placement of the 1 in the 4×8 board in Figure 12. This is so that we will end up in the bottom row *three* squares from the right-hand edge of the board (at 32 in Figure 12), and we can then join the two boards with two legal knight moves—namely, 14–32 and 1–15 in Figure 12.

1	40	7	46	13	4	43	10
44	11	2	41	8	47	14	5
39	6	45	12	3	42	9	48
32	17	26	35	20	29	38	23
37	22	31	16	25	34	19	28
18	27	36	21	30	15	24	33

22	7	24	1	18	3	20	5
29	12	27	10	25	16	31	14
6	23	8	17	2	19	4	21
13	28	11	26	9	32	15	30

1	72	7	78	13	4	75	10
76	11	2	73	8	79	14	5
71	6	77	12	3	74	9	80
64	49	58	67	52	61	70	55
69	54	63	48	57	66	51	60
50	59	68	53	62	47	56	65
25	40	23	46	29	44	27	42
18	35	20	37	22	31	16	33
41	24	39	30	45	28	43	26
34	19	36	21	38	15	32	17

FIGURE 12
A 10×8 board

This completes the proof of the following theorem.

THEOREM. *On a torus, every rectangular chessboard has a knight's tour.*

Tours on Square Boards

In particular, all square boards have knight's tours on a torus. In this section we shall see that tours on square boards can have patterns that are far nicer than those offered by the foregoing inductive procedure. Moreover, we shall see that the attractiveness of these patterns is due to an underlying algebraic structure. Interestingly, a Fulani astronomer and mathematician, Muhammad Ibn Muhammad, used similar knight's patterns in his native northern Nigeria to produce magic squares at just about the same time that Euler was working on knight's tours in Europe (see **[7]**, 137–151). We will deal with $n \times n$ boards in three cases.

Case 1: $n \neq 5k$ Simply repeat the move $(2, 1)$ $n-1$ times—we call this a *stroll.* Then use the move $(1, -2)$—a *shift*—once, and resume the stroll, shifting every n moves, until you return to the starting point. FIGURE 13 shows the resulting tour for a 7×7 board. (Notice that the result is a magic square; in fact, this procedure produces a magic square for all n not a multiple of 2, 3, or 5; see **[6]**.)

1	24	47	21	37	11	34
12	35	2	25	48	15	38
16	39	13	29	3	26	49
27	43	17	40	14	30	4
31	5	28	44	18	41	8
42	9	32	6	22	45	19
46	20	36	10	33	7	23

1	12	23	9	20
10	16	2	13	24
14	25	6	17	3
18	4	15	21	7
22	8	19	5	11

FIGURE 13
Tours for 7×7 and 5×5 boards

Case 2: $n \neq 3k$ The reason the previous pattern does not work when n is a multiple of 5 is that eventually the shift can't be made. Obviously, the thing to do is to make a different shift. So we use the shift $(-1, -2)$ instead. This works for all n not a multiple of 3. We illustrate this in FIGURE 13 for a 5×5 board. (Notice that the result is a semi-magic square; in fact, this procedure produces a semi-magic square for all n not a multiple of 2 or 3; moreover, it is only the main diagonal whose sum fails to be correct in each case; see [6].)

Case 3: $n = 15k$ Unfortunately, this still leaves us having to deal with square boards where n is a multiple of 15. Our approach in this case will be very similar to the previous two cases, but the actual details turn out to be considerably more involved. Therefore, we delay our discussion of this case until the Appendix, and turn now to an alternate approach.

An Algebraic Approach

Anyone familiar with the concept of a group will have sensed that there is an underlying algebraic structure for these tours. For example, it is clear that if $\gcd(m, n) = 1$ where, without loss, we take n to be odd, then there is a tour of the $m \times n$ board using only the move (2, 1). This, of course, is because the element (2, 1) generates the group $\mathbb{Z}_n \times \mathbb{Z}_m$.

Similarly, we see that in the tour of the 7×7 board in FIGURE 13, the first stroll, which uses (2, 1) six times, yields the subgroup of $\mathbb{Z}_7 \times \mathbb{Z}_7$ generated by the element (2, 1)—namely, $\{(0,0),(2,1),(4,2),(6,3),(1,4),(3,5),(5,6)\}$. The shift $(1, -2)$ then moves us to a different coset of this subgroup, where the stroll now takes us through this new coset. In this way, we tour the entire group, one coset at a time.

In order to see how this works in general for a square board, we label the $n \times n$ board by the elements of $\mathbb{Z}_n \times \mathbb{Z}_n$ viewed as vectors (a, b), $a, b \in \mathbb{Z}_n$. In particular, the upper left-hand corner is labeled (0, 0). We can then make a *change of coordinates*, such as

$$(a,b) = c(2,1) + d(1,-2).$$

So, for example, a knight at position $(a, b) = (2, 1)$ in the original coordinates would be at $(c, d) = (1, 0)$ under the change of coordinates, or a knight at (4, 2) would now be at (2, 0). In this way, the 7×7 knight's tour in FIGURE 13, under the change of coordinates, becomes the 1-step rook's tour in FIGURE 14. (Such tours are discussed in [2].)

1	2	3	4	5	6	7
9	10	11	12	13	14	8
17	18	19	20	21	15	16
25	26	27	28	22	23	24
33	34	35	29	30	31	32
41	42	36	37	38	39	40
49	43	44	45	46	47	48

FIGURE 14
A 1-step rook's tour

Thus, we can turn the knight's tour problem into an obviously simpler rook's tour problem, a process that worked in this case because (2, 1) and $(1, -2)$ form a basis for $\mathbb{Z}_7 \times \mathbb{Z}_7$. This happened, in turn, because $\det\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$ is a unit in the ring $\mathbb{Z}_7$. This particular change of coordinates, therefore, will work as long as 5 does not divide n.

On the other hand, since $\det\begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} = 4$, the change of coordinates given by $(a, b) = c(2, 1) + d(-2, 1)$ will work as long as n is odd. It is worth noting, however, that the two tour problems are not equivalent. For example, the knight's tour for the 5×5 board in FIGURE 13 does not become a rook's tour under this particular change of variables. This is not at all surprising since a knight has more moves than a 1-step rook. Similarly, the change of variables given by $(a, b) = c(2,1) + d(1,2)$ works as long as 3 does not divide n. There are three additional changes of variables that are possible, but they are equivalent to the three already mentioned. This method, therefore, handles any $n \times n$ board where n is not divisible by 30.

Open Questions

There are several directions for further study. Since our proof for the torus only rarely makes use of *both* the top to bottom and the left to right identifications, the most obvious question is to ask which rectangular boards have tours on a cylinder. In addition, there are always projective planes and Klein bottles on which to put chessboards. Finally, the algebraic approach could be applied to rectangular boards.

REFERENCES

1. W. W. Rouse Ball, *Mathematical Recreations and Problems of Past and Present Times*, Macmillan, London, UK, 1982.
2. J. A. Gallian and D. Witte, Hamiltonian chessboards, this MAGAZINE 57 (1984), 291–294.
3. M. Gardner, Mathematical games: problems that are built on the knight's move in chess, *Scientific American* 217 (1967), 128–132.
4. A. J. Schwenk, Which rectangular chessboards have a knight's tour?, this MAGAZINE 64 (1991), 325–332.
5. R. J. Wilson and J. J. Watkins, *Graphs, an Introductory Approach*, John Wiley and Sons, Inc., New York, NY, 1990, pp. 143–153.
6. J. J. Watkins and B. A. Balof, Knight's tours and magic squares, *Congressus Numerantium* (to appear).
7. C. Zaslavsky, *Africa Counts*, Prindle, Weber and Schmidt, Inc., Boston, MA, 1973.

Appendix

We now return to our discussion of Case 3 for square boards where n is a multiple of 15. In order to better understand our solution for the general $15k \times 15k$ board, it is worth looking at the 15×15 board in some detail. Let us begin with the move $(2,1)$ as a stroll. After 14 moves we make a shift using $(1, -2)$. All goes well in this fashion until exactly $1/5$ of the squares have been visited and we are unable to use our shift at square 45, as we see in FIGURE 15.

What we notice, however, is that the 45 squares that have already been visited form a perfectly arranged lattice on the board. Furthermore, from any of these squares, any of the knight moves a, b, c, d—see FIGURE 3—takes you to another of these squares; whereas, any of the knight moves $\alpha, \beta, \gamma, \delta$ takes you to a *previously unvisited* square. Now, it is clear what to do: use move a as a stroll and use move d as a shift (every 15 moves) until you reach 45, then use, say γ, once before resuming the strolling and shifting with a and d, using γ at 45, 90, 135, 180, and 225.

Since it is far less confusing if one uses colored pens when doing this by hand—red for squares 1–45, green for 46–90, and so on—we call a move such as γ a *color change*. In this way, γ acts as a translation of a lattice of one color to an identical lattice of another color. The five disjoint, but identical, lattices comprise the board.

This strategy certainly allows a knight to visit every square on a $15k \times 15k$ board, but does not always produce a closed tour. In fact, using a, d, and γ in this same way on a 30×30 board leaves an exhausted knight stranded after 900 moves in the 16th row and 16th column. In order to find a *closed* tour we use a little algebra.

1					19					37				
		2					20					38		
				3					21					39
	40					4					22			
			41					5					23	
24					42					6				
		25					43					7		
				26					44					8
	9					27					45			
			10					28					31	
32					11					29				
		33					12					30		
				34					13					16
	17					35					14			
			18					36					15	

FIGURE 15
Start of a 15×15 tour

Let us examine the case $n = 15$ more closely. We consider three variables, s, t, and ω, representing the stroll, the shift, and the color change, respectively. In a tour of a 15×15 board, we stroll for 14 moves and then shift, and repeat for a total of 3 strolls and 2 shifts for each of 5 colors—that is, we repeat the sequence $3(14s) + 2t + \omega$ five times, once for each color, and end up back where we started. Substituting a, d, and γ for s, t, and ω, and multiplying by 5, we get $15(14s) + 10t + 5\omega = 210 \cdot (2,1) + 10 \cdot (1,-2) + 5 \cdot (-2,1) = (420, 195) \equiv (0,0) \pmod{15}$ which explains precisely why this pattern returns us to the starting point. (A similar computation for the case $n = 30$ also shows why the knight ends up stuck in the 16th row and 16th column.)

Let us now turn to the general case $n = 15k$. It is necessary to allow the stroll and shift to vary from color to color, and to use different color changes as well. We thus have *fifteen* variables s_i, t_i, and ω_i, for $i = 1, \ldots, 5$. Since we stroll for $15k - 1$ moves and then shift, and repeat for a total of $3k$ strolls and $3k - 1$ shifts for each of 5 colors, the result of all the moves is given by

$$\sum_{i=1}^{5} 3k(15k-1)s_i + (3k-1)t_i + \omega_i \equiv \sum_{i=1}^{5} (3k-1)t_i - 3ks_i + \omega_i \pmod{15k}.$$

Thus, we are looking to solve the following congruence

$$(**) \quad (3k-1)(t_1 + \cdots + t_5) - 3k(s_1 + \cdots + s_5) + (\omega_1 + \cdots + \omega_5) \equiv 0 \pmod{15k},$$

where $s_i, t_i \in \{a, b, c, d\}$ and $\omega_i \in \{\alpha, \beta, \gamma, \delta\}$ for each i. Moreover, we obviously require that $s_i \neq \pm t_i$ for any i.

One further restriction applies to the color changes, since not every sequence of 5 colors changes will cycle you through all 5 colors. It is easy to find appropriate sequences by constructing a directed graph with 5 vertices, one for each color, and joining each ordered pair of vertices with an arc labeled with the color change α, β, γ, or δ which takes you between the corresponding colors. We can thus see that there are 24 allowable sequences. Since we are only concerned with the arithmetic at present, these can be grouped into the following 8 classes where, in each case, we give

1	132	211	161	89	44	108	202	175	46	30	99	193	147	75
146	74	2	91	185	160	88	45	124	218	174	60	16	115	209
131	225	145	73	3	107	201	159	87	31	98	192	173	59	17
58	18	105	184	144	72	4	123	217	158	86	32	114	208	172
224	171	57	19	106	200	143	71	5	97	191	157	85	33	130
34	104	183	170	56	20	122	216	142	70	6	113	207	156	84
155	83	35	120	199	169	55	21	96	190	141	69	7	129	223
103	182	154	82	36	121	215	168	54	22	112	206	140	68	8
67	9	119	198	153	81	37	95	189	167	53	23	128	222	139
181	138	66	10	135	214	152	80	38	111	205	166	52	24	102
25	118	197	137	65	11	94	188	151	79	39	127	221	180	51
179	50	26	134	213	136	64	12	110	204	165	78	40	101	195
117	196	178	49	27	93	187	150	63	13	126	220	164	77	41
76	42	133	212	177	48	28	109	203	149	62	14	100	194	163
210	162	90	43	92	186	176	47	29	125	219	148	61	15	116

FIGURE 16
Knight's tour of a 15×15 board

the total effect of the five moves:

$$\alpha^2\beta^2\delta = (5,0) \quad \alpha^2\gamma\delta^2 = (0,-5) \quad \beta\gamma^2\delta^2 = (-5,0) \quad \alpha\beta^2\gamma^2 = (0,5)$$
$$\alpha^5 = (10,-5) \quad \beta^5 = (5,10) \quad \gamma^5 = (-10,5) \quad \delta^5 = (-5,-10)$$

We are now ready to present a solution of the Knight's Tour Problem for a $15k \times 15k$ chessboard! In fact, the following moves provide a solution that works for all $15k \times 15k$ boards.

$$\begin{array}{lll} s_1 = a = (2,1) & t_1 = b = (-1,2) & \omega_1 = \alpha = (2,-1) \\ s_2 = c = (-2,-1) & t_2 = d = (1,-2) & \omega_2 = \beta = (1,2) \\ s_3 = d = (1,-2) & t_3 = a = (2,1) & \omega_3 = \beta = (1,2) \\ s_4 = c = (-2,-1) & t_4 = d = (1,-2) & \omega_4 = \alpha = (2,-1) \\ s_5 = d = (1,-2) & t_5 = a = (2,1) & \omega_5 = \delta = (-1,-2) \end{array}$$

In order to see that this does yield a knight's tour, note that

$$s_1 + s_2 + s_3 + s_4 + s_5 = (0,-5)$$
$$t_1 + t_2 + t_3 + t_4 + t_5 = (5,0)$$
$$\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 = (5,0)$$

so that congruence $(**)$ becomes

$$(3k-1)\cdot(5,0) - 3k\cdot(0,-5) + (5,0) = (15k,15k) \equiv (0,0) \pmod{15k}$$

which shows that our wandering knight does indeed return to the original square. You might notice that the key was to make the sum of the five shifts and the sum of the five color changes equal; thus, other solutions are possible. FIGURE 15 shows the tour produced by this particular solution for a 15×15 board. We encourage you to grab five colored pens, a 30×30 grid, and have at it!

Thales Meets Poincaré

DAVID E. DOBBS
University of Tennessee
Knoxville, TN 37996-1300

1. Introduction

According to [**2**, Theorem 39, p. 100], Euclidean geometry and hyperbolic (Bolyai-Lobachevskian) geometry are the only types of absolute (also known as "neutral") plane geometry. These geometries are distinguished by the number of lines that pass through a given point and are parallel to a given line. This number may be 1 (the Euclidean case) or greater than 1 (the hyperbolic case): see axioms **E** and **BL** in [**2**, p. 197] or postulates (I) and (II) in [**3**, p. 317]. Euclidean and hyperbolic geometry may also be distinguished by considering the sum of the (radian) measures of the (interior) angles in any triangle. Indeed, in Euclidean geometry, this sum is π, while in hyperbolic geometry, this sum is less than π (and may vary from triangle to triangle): see [**2**, Theorem 2, p. 264 and Theorem 2, p. 278], [**3**, p. 118], [**4**, Theorems 10.1 and 10.3]. The main purpose of this paper is to see whether these two geometries may also be distinguished by Thales' Theorem [**2**, Theorem 13, p. 269], a classical result on triangles in Euclidean geometry whose proof depends on the behavior of parallel lines and similar triangles in the Euclidean setting.

To study the possible validity of Thales' Theorem in hyperbolic geometry, we shall work inside the Poincaré half-plane model, whose salient features are reviewed in Section 2. There is no loss of generality in using this model, as hyperbolic geometry is "categorical" [**2**, Proposition 7, p. 345], in the sense that all its models are isomorphic. One benefit of using the Poincaré model is that the question of a possible "hyperbolic Thales' Theorem" comes down to asking whether two specific numbers are equal; the calculations in the Example in Section 2 give a negative answer. This accomplishes our main purpose: Thales' Theorem *does* distinguish Euclidean geometry from hyperbolic geometry.

But we can say more. In Section 3, continuing to work in the Poincaré model, we further analyze the two numbers that need to be calculated and compared in testing for a "hyperbolic Thales' Theorem." By subjecting the triangular data to a limiting process that is designed to, so to speak, minimize the difference between the hyperbolic and Euclidean metrics, we show (see the Theorem in Section 3) that the ratio of the two numbers in question has limit 1. Thus, although Thales' Theorem is false in hyperbolic geometry, we can say that it holds "in the limit," for a suitable Euclidean-seeking limit process.

We hope that this work finds use as enrichment material in model-oriented courses on absolute geometry. However, the technical details in Section 3 depend on a more central part of the curriculum, namely real-analytic functions, as studied in advanced calculus. The material being reinforced includes the Binomial Theorem for power series, the calculations involved in multiplying, dividing, or composing functions defined by power series, and the Maclaurin series for $\ln(1+x)$. A suitable reference for this material is [**1**].

2. Poincaré's Half-Plane Violates Thales' Theorem

The form of Thales' Theorem that we focus on is the consequence identified in [**2**, Theorem 14, p. 271] and illustrated in FIGURE 1: If ΔABC is a triangle, with D an interior point of the segment $\overline{AB}$ and E an interior point of $\overline{AC}$ such that the lines DE and BC are parallel, then $d(A,B)/d(A,D) = d(B,C)/d(D,E)$. (As usual, $d(P,Q)$ denotes the distance between points P and Q.) In fact, we focus further on the special case in which D and E are the midpoints of $\overline{AB}$ and $\overline{AC}$, respectively (in which case, DE and BC are automatically parallel). Thus, for our purposes, Thales' Theorem is the statement that if D and E are the midpoints of $\overline{AB}$ and $\overline{AC}$, respectively, in ΔABC, then $d(D,E)/d(B,C) = 1/2$, or equivalently, $d(B,C) = 2d(D,E)$.

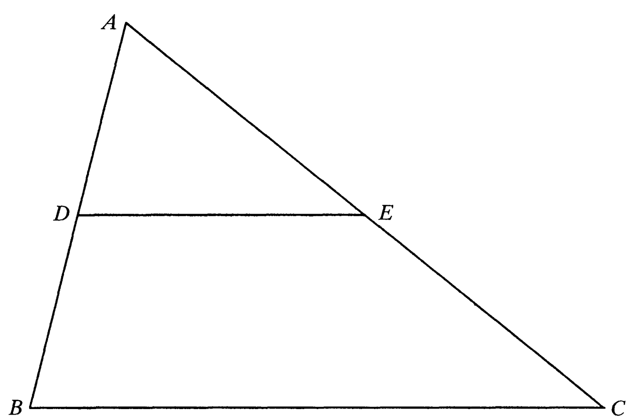


FIGURE 1
Thales' Theorem: If $DE//BC$, then $\dfrac{d(A,B)}{d(A,D)} = \dfrac{d(B,C)}{d(D,E)}$

In order to interpret the statement of Thales' Theorem in hyperbolic geometry, it is convenient to review the following four features of the Poincaré half-plane model for hyperbolic geometry.

1. *Points*. The "points" in the model are the points of the Euclidean upper half-plane, equipped with Cartesian coordinates in the usual way.
2. *Lines*. The "lines" in the model arise from the "geodesic segments," which are two types: (a) *bowed geodesics*, arcs of Euclidean semicircles that are centered on the x-axis; and (b) *straight geodesics*, segments of Euclidean lines that are perpendicular to the x-axis [**4**, Theorem 4.4].
3. *Distance*. The *hyperbolic distance along a curve* Γ is given by the integral $\int_\Gamma \frac{\sqrt{(dx)^2 + (dy)^2}}{y}$. If Γ is a geodesic segment as in (2), then hyperbolic distance is termed *hyperbolic length*. In particular (cf. [**4**, Proposition 4.1]), if q is a circle with center K on the x-axis and P and Q are points of q such that the radii $\overline{KP}$ and $\overline{KQ}$ make angles α and β, respectively, with the positive x-axis, where $0 < \alpha < \beta < \pi$, then the hyperbolic length from P to Q, measured necessarily along a bowed geodesic, is $\ln \frac{\csc\beta - \cot\beta}{\csc\alpha - \cot\alpha}$. Similarly [**4**, Proposition 4.3], if

$1 \leq k_1 < k_2$, then the hyperbolic length from $(0, k_1)$ to $(0, k_2)$, measured necessarily along a straight geodesic, is $\ln \frac{k_2}{k_1}$ [4, pp. 64–65].

4. *Angles*. "Angles" in the model may have curvilinear sides. The (*hyperbolic*) *measure of an angle* is the same as its Euclidean measure, namely the usual measure of the angle formed by the two tangent lines to the sides of the angle at its vertex [4, pp. 70–71].

We are ready to test the validity of a "hyperbolic Thales' Theorem." The triangular data are summarized in FIGURE 2, and the analysis in the Example below proceeds using the above features (1)–(4). These methods will be needed again for the more general analysis in Section 3. In interpreting the "lines" in FIGURE 2, view $\overline{AB}$, $\overline{AD}$ and $\overline{DB}$ as straight geodesics (as they appear), but view all the other segments as bowed geodesics.

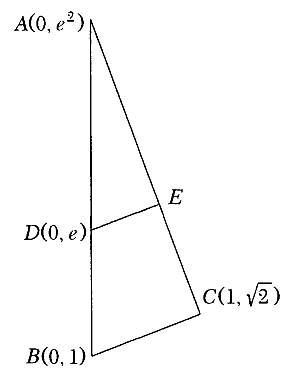


FIGURE 2
Data analyzed in Example

Example. For the data in FIGURE 2, with D and E the (hyperbolic) midpoints of $\overline{AB}$ and $\overline{AC}$, respectively, we find the hyperbolic lengths $d(B,C) = \ln(\sqrt{2} + 1) \approx 0.881373587$ and $d(D, E) \approx 0.333182944$. In particular, $d(B,C) \neq 2d(D, E)$, and so Thales' Theorem is not valid in hyperbolic geometry.

Proof. By the second fact in (3), $d(D, B) = d(B, D) = \ln \frac{e}{1} = 1$ and $d(A, B) = d(B, A) = \ln \frac{e^2}{1} = 2$, and so D is indeed the (hyperbolic) midpoint of $\overline{AB}$. We proceed next to show that $d(B,C) = \ln(\sqrt{2} + 1)$. To do this using the first fact in (3), consider C_1, the circle which is centered on the x-axis and passes through B and C. This circle has a Cartesian equation of the form $x^2 + y^2 + \lambda x + \mu = 0$ for suitable real numbers λ and μ. By substituting the coordinates of B and C, we can solve for $\lambda = -2$, $\mu = -1$. Thus, C_1 is given by $x^2 + y^2 - 2x - 1 = 0$, or $(x - 1)^2 + y^2 = 2$. In particular, the center of C_1 is $K_1(1, 0)$. The radii $\overline{K_1C}$ and $\overline{K_1B}$ make angles $\frac{\pi}{2}$ and $\frac{3\pi}{4}$, respectively, with the positive x-axis. By (3),

$$d(B,C) = \ln \frac{\csc \frac{3\pi}{4} - \cot \frac{3\pi}{4}}{\csc \frac{\pi}{2} - \cot \frac{\pi}{2}} = \ln \frac{\sqrt{2} - (-1)}{1 - 0} = \ln(\sqrt{2} + 1),$$

as asserted.

Finding $d(D, E)$ is more involved than finding $d(B,C)$ because, in addition to finding an equation of the circle C_3 which is centered on the x-axis and passes through D and E, we must first find E! And, for this, we need to find an equation for the circle C_2 which is centered on the x-axis and passes through A and C. Fortunately, some of the process is as in the preceding paragraph, and so we may leave many details to the reader.

As above, one can produce an equation for C_2: $x^2 + y^2 + (e^4 - 3)x - e^4 = 0$, or "approximately" $x^2 + y^2 + 51.59815003x - 54.59815003 = 0$. The center of C_2 is $K_2\left(\frac{-e^4+3}{2}, 0\right)$, or approximately $(-25.79907502, 0)$. We now turn to finding the coordinates of E, certainly the most tedious part of the analysis. For this, FIGURE 3 (not drawn to scale) will be helpful.

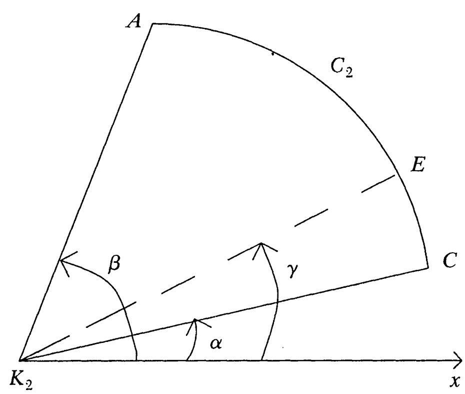


FIGURE 3
Data used to find E in Example

Since E is the midpoint of $\overline{AC}$, we have $d(C, E) = d(E, A)$. Rewriting using (3) and applying some trigonometric identities, we find

$$\left[\frac{1-\cos\gamma}{\sin\gamma}\right]\Bigg/\left[\frac{1-\cos\alpha}{\sin\alpha}\right] = \left[\frac{1-\cos\beta}{\sin\beta}\right]\Bigg/\left[\frac{1-\cos\gamma}{\sin\gamma}\right]$$

and so

$$\frac{1-\cos\gamma}{\sin\gamma} = \sqrt{\frac{(1-\cos\beta)(1-\cos\alpha)}{(\sin\beta)(\sin\alpha)}}.$$

The right-hand side of the preceding expression can be evaluated using the definition of cosine and the Pythagorean identity $\sin\theta = \sqrt{1-\cos^2\theta}$ for $0 \le \theta \le \pi$. It follows that

$$\cos\beta \approx \frac{25.79907502}{26.83636379} \approx 0.961347641, \qquad \sin\beta \approx 0.275337454,$$

$$\cos\alpha \approx \frac{25.79907502 + 1}{26.83636379} \approx 0.998610513, \qquad \sin\alpha \approx 0.052697654.$$

Hence, $\frac{1-\cos\gamma}{\sin\gamma} \approx 0.060839685$, and so $t := \cos\gamma$ satisfies

$$\frac{\sqrt{1-t}}{\sqrt{1+t}} = \frac{1-t}{\sqrt{1-t^2}} \approx 0.060839685.$$

Squaring both sides and rewriting, we obtain the linear "equation"

$$1.003701467t - 0.9962985327 \approx 0.$$

Thus, $\cos\gamma = t \approx 0.992624366$, whence $\gamma = \cos^{-1} t \approx 0.121529573$. (Readers using other computational devices may find a value of t which differs from ours after "several" decimal places, with resulting "small" deviations from our subsequent calculations.)

Now, the slope of the (Euclidean) line K_2E is $\tan\gamma \approx 0.122131437$. Thus, the point-slope form for the equation of K_2E is "approximately" $y = 0.122131437(x + 25.79907502)$. Solving this equation simultaneously with the above equation for C_2, we find the coordinates of E. Bearing in mind that the x-coordinate of E is positive, the reader can resolve a sign ambiguity arising from the quadratic formula and thus determine that E is approximately $(0.839353580, 3.25338956)$.

Now that we have the coordinates of E, we can find that an equation for C_3 is $x^2 + y^2 + \delta x - e^2 = 0$, where $\delta \approx -4.646435132$. Thus, the center of C_3 is approximately $K_3(2.323217566, 0)$. By analyzing the analogue of FIGURE 3 for C_3 and reasoning as above, the reader can verify, using (3), that

$$d(D,E) \approx \ln\frac{1.315465894 + 0.854663980}{1.099102020 + 0.456097851} \approx \ln(1.395402555) \approx 0.333182944.$$

All the assertions have now been verified. This completes the proof.

It is tempting to seek a facile explanation for the discrepancy between $2d(D,E) \approx 0.666365888$ and $d(B,C) \approx 0.881373587$ in the above Example. The answer does not lie exactly in analyzing only the y-coordinates. In fact, the arithmetic mean of the y-coordinates of D and E (resp., of B and C) is $y_1 \approx 2.985835694$ (resp., $y_2 \approx 1.207106781$). Heuristic arguments using (3) might lead one to suspect the relevance of $\dfrac{1/y_1}{1/y_2} \approx 0.404277698$ or $\dfrac{\ln y_1}{\ln y_2} \approx 5.81151016$, but neither is "approximately" $\dfrac{d(D,E)}{d(B,C)} \approx 0.378026922$. (Geometric means fare no better in such heuristics.) Rather than try further to "explain" the Example, we turn next to examining Thales' Theorem within a part of the Poincaré half-plane model in which the bowed geodesics look "more Euclidean" or "less bowed" than in the above Example.

3. Thales' Theorem Pushes Poincaré's Half-Plane to the Limit

The Example in Section 2 showed that Thales' Theorem is not valid in hyperbolic geometry. Nevertheless, the Theorem in this section will establish that when such an example is subjected to a suitable limiting process, then a hyperbolic Thales' Theorem becomes valid "in the limit." To make matters precise, let's consider the data in FIGURE 4. Here, $\delta > 0$ and $\varepsilon > 1$. During the limit process, $\delta \to 0^+$ and $\varepsilon \to 1^+$. For simplicity, we assume early in the analysis that $\delta = \sqrt{\varepsilon^4 - 1}$ (equivalently, $\varepsilon = \sqrt[4]{1 + \delta^2}$).

The above expressions for δ and ε suggest (correctly) that the power series form of the Binomial Theorem will be useful. Moreover, we shall need some results from advanced calculus stating that polynomial-like calculations are valid on the interior of intervals of convergence for functions which (as our functions always will be) are defined by power series with positive radii of convergence. Specifically, we assume

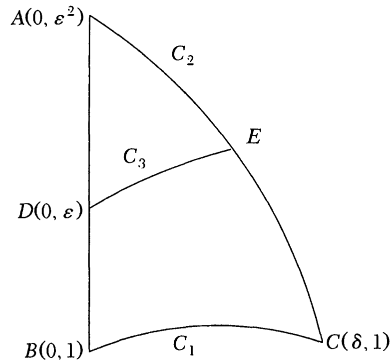


FIGURE 4
Data analyzed in Theorem

familiarity with a corollary of Mertens' Theorem [**1**, Theorem 8.46, p. 204], which permits taking products of analytic functions by using Cauchy multiplication [**1**, Theorem 9.24, p. 237]; as well as the computational ways to carry out composition [**1**, Theorem 9.25, p. 238] and reciprocation/division [**1**, Theorem 9.26, p. 239] for analytic functions.

With our context and prerequisites in place, we can now proceed to our main result.

THEOREM. *Consider the data in* FIGURE 4, *with* $\delta > 0$ *and* $\varepsilon = \sqrt[4]{1+\delta^2} > 1$, *and with* D *and* E *the (hyperbolic) midpoints of* $\overline{AB}$ *and* $\overline{AC}$, *respectively. Then, as analytic functions of* δ, *the hyperbolic lengths*

$$d(B,C) = \delta + 0\delta^2 - \frac{1}{24}\delta^3 + \cdots \quad (\textit{higher degree terms})$$

$$\textit{and} \quad d(D,E) = \frac{1}{2}\delta + 0\delta^2 + \cdots.$$

In particular, $\lim_{\delta \to 0^+} \dfrac{d(D,E)}{d(B,C)} = \dfrac{1}{2}$.

Proof. Insofar as possible, we argue as in the earlier Example, omitting analogous details. In particular, C_1, C_2, and C_3 denote the bowed geodesics through B and C, A and C, and D and E, respectively; and K_i denotes the center of C_i $(i = 1,2,3)$. Notice that C_1 has Cartesian equation $x^2 - \delta x + y^2 - 1 = 0$ and center $K_1\left(\frac{\delta}{2}, 0\right)$. In particular, the maximum (Euclidean) height on C_1 above its "equilibrium" y-value of 1 is $\sqrt{1 + \frac{\delta^2}{4}} - 1$, which tends to 0 as $\delta \to 0^+$. In fact, one can show by l'Hôpital's Rule that more is true, namely $\lim_{\delta \to 0^+} \dfrac{\sqrt{1+\frac{\delta^2}{4}} - 1}{\delta} = 0$. Thus, the "line" C_1 looks increasingly "flatter," or "more Euclidean," as $\delta \to 0^+$. We will eventually be able to show the same for C_3.

Using fact (3) from Section 2 and simplifying, we find that

$$d(B,C)=\ln\frac{\dfrac{\sqrt{4+\delta^2}}{2}-\left(-\dfrac{\delta}{2}\right)}{\dfrac{\sqrt{4+\delta^2}}{2}-\dfrac{\delta}{2}}=\ln\frac{\sqrt{4+\delta^2}+\delta}{\sqrt{4+\delta^2}-\delta}$$

$$=\ln\left(1+\frac{2\delta}{\sqrt{4+\delta^2}-\delta}\right)=\ln\left(1+\frac{\delta}{\sqrt{1+\dfrac{\delta^2}{4}}-\dfrac{\delta}{2}}\right)$$

$$=\ln\left(1+\frac{\delta}{\left[1+\dfrac{1}{2}\left(\dfrac{\delta^2}{4}\right)-\dfrac{1}{8}\left(\dfrac{\delta^2}{4}\right)^2+\cdots\right]-\dfrac{\delta}{2}}\right).$$

The preceding application of the Binomial Theorem is valid if $\left|\dfrac{\delta^2}{4}\right|<1$ and thus holds for all sufficiently small $\delta>0$. The same proviso holds without explicit mention for all the subsequent machinations.

The denominator in the last-displayed fraction is $1-\dfrac{\delta}{2}+\dfrac{\delta^2}{8}-\dfrac{\delta^4}{128}+\cdots$. We need to find its reciprocal, and we will need to find the reciprocal of many such series in the work below. One familiar way to proceed is to solve recursively for the coefficients when the reciprocal is expressed as a power series. For our purposes, it may be faster to use a less familiar "long division" process. We carry out one calculation in detail and then leave all similar calculations to the reader.

$$\begin{array}{r|l}
 & 1+\dfrac{\delta}{2}+\dfrac{\delta^2}{8}-\dfrac{\delta^4}{128}+\cdots \\ \hline
1-\dfrac{\delta}{2}+\dfrac{\delta^2}{8}-\dfrac{\delta^4}{128}+\cdots & 1 \\
 & 1-\dfrac{\delta}{2}+\dfrac{\delta^2}{8}-\dfrac{\delta^4}{128}+\cdots \\ \cline{2-2}
 & \dfrac{\delta}{2}-\dfrac{\delta^2}{8}+\dfrac{\delta^4}{128}+\cdots \\
 & \dfrac{\delta}{2}-\dfrac{\delta^2}{4}+\dfrac{\delta^3}{16}+0\delta^4+\cdots \\ \cline{2-2}
 & \dfrac{\delta^2}{8}-\dfrac{\delta^3}{16}+\dfrac{\delta^4}{128}+\cdots \\
 & \dfrac{\delta^2}{8}-\dfrac{\delta^3}{16}+\dfrac{\delta^4}{64}+\cdots \\ \cline{2-2}
 & -\dfrac{\delta^4}{128}+\cdots
\end{array}$$

Thus,

$$d(B,C)=\ln\left(1+\delta\left[1+\frac{\delta}{2}+\frac{\delta^2}{8}-\frac{\delta^4}{128}+\cdots\right]\right)$$

$$=\ln(1+\Delta),\text{ with }\Delta=\delta+\frac{\delta^2}{2}+\frac{\delta^3}{8}-\frac{\delta^5}{128}+\cdots.$$

Using the Maclaurin series for $\ln(1+x)$ and then simplifying via Cauchy multiplication, we find

$$d(B,C) = \Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \cdots$$
$$= \left(\delta + \frac{\delta^2}{2} + \frac{\delta^3}{8} + \cdots\right) - \frac{(\delta^2 + \delta^3 + \cdots)}{2} + \frac{\delta^3 + \cdots}{3}$$

$= \delta + 0\delta^2 - \dfrac{\delta^3}{24} + \cdots$, as asserted.

The above work did not need the condition $\varepsilon = \sqrt[4]{1+\delta^2}$, but we will use this hypothesis from now on. Its first contribution is a simple form for an equation for C_2, namely $x^2 + y^2 - \varepsilon^4 = 0$, whence C_2 has center $K_2(0,0)$ and radius $\varepsilon^2 = \sqrt{1+\delta^2}$.

We turn to finding the coordinates of E, using FIGURE 3 of the Example. In the present context, $\cos\alpha = \dfrac{\delta}{\varepsilon^2}$, $\sin\alpha = \dfrac{1}{\varepsilon^2}$, and $\beta = \dfrac{\pi}{2}$. It follows, as in the Example, that

$$\frac{1-\cos\gamma}{\sin\gamma} = \nu, \quad \text{where } \nu := \sqrt{\sqrt{1+\delta^2} - \delta}\,.$$

As $0 < \gamma < \frac{\pi}{2}$, $\sin\gamma = \sqrt{1-\cos^2\gamma}$, and so $\sqrt{1-\cos\gamma}\,/\sqrt{1+\cos\gamma} = \nu$. As in the Example, squaring both sides leads easily to $\cos\gamma = \dfrac{1-\nu^2}{1+\nu^2}$, and then it follows from the Pythagorean identity that $\sin\gamma = \dfrac{2\nu}{1+\nu^2}$. Hence, the coordinates of E are $\left(\varepsilon^2\left(\dfrac{1-\nu^2}{1+\nu^2}\right), \dfrac{2\varepsilon^2\nu}{1+\nu^2}\right)$.

With the coordinates for E in hand, the reader can now check that an equation for C_3 is $x^2 + y^2 + Fx - \varepsilon^2 = 0$, where

$$F := \frac{(1-\varepsilon^2)(1+\nu^2)}{1-\nu^2} < 0.$$

Hence, the center of C_3 is $K_3\left(-\frac{F}{2}, 0\right)$, and the radius of C_3 is

$$R := \frac{\sqrt{4\varepsilon^2 + F^2}}{2}.$$

Notice that the maximum (Euclidean) height on C_3 is R. To show that the portion of C_3 between D and E becomes "flatter" or "more Euclidean" during the limit process, just as was the case for C_1, it suffices to prove that

$$\lim_{\delta\to 0^+} \max\left(R - \varepsilon, R - \frac{2\varepsilon^2\nu}{1+\nu^2}\right) = 0.$$

In fact,

$$\lim_{\delta\to 0^+} \frac{R-\varepsilon}{\delta} = 0 = \lim_{\delta\to 0^+} \frac{R - \dfrac{2\varepsilon^2\nu}{1+\nu^2}}{\delta^2}.$$

This may be easily shown using the following expansions, established below:

$$R = 1 + \frac{3}{8}\delta^2 + \cdots; \qquad \varepsilon = 1 + \frac{\delta^2}{4} + \cdots; \qquad \varepsilon^2 = 1 + \frac{\delta^2}{2} + \cdots;$$

$$\nu = 1 - \frac{\delta}{2} + \frac{\delta^2}{8} + \cdots; \qquad \nu^2 = 1 - \delta + \frac{\delta^2}{2} + \cdots.$$

For this reason, the present situation was described in the Introduction as a "Euclidean-seeking limit process," as C_1 and C_3 are approaching in appearance the Euclidean segments $\overline{BC}$ and $\overline{DE}$ in FIGURE 1.

Before using the above information to find $d(D, E)$, we need to show that E is to the right of K_3, as such information is needed in determining the quadrants (and thus the sign of the trigonometric functions) of the angles involved in applying (3). Thus, we proceed to show that (for all sufficiently small $\delta > 0$) the x-coordinate of E is greater than the x-coordinate of K_3; that is,

$$\frac{\varepsilon^2(1-\nu^2)}{1+\nu^2} > \frac{(\varepsilon^2-1)(1+\nu^2)}{2(1-\nu^2)},$$

or, equivalently,

$$\left(\frac{1-\nu^2}{1+\nu^2}\right)^2 > \frac{\varepsilon^2-1}{2\varepsilon^2}, \text{ or } \left(\frac{1-\sqrt{1+\delta^2}+\delta}{1+\sqrt{1+\delta^2}-\delta}\right)^2 > \frac{\sqrt{1+\delta^2}-1}{2\sqrt{1+\delta^2}}.$$

Since the Binomial Theorem gives $\sqrt{1+\delta^2} = 1 + \frac{\delta^2}{2} - \frac{\delta^4}{8} + \frac{\delta^6}{16} + \cdots$, we have

$$\left(\frac{1-\sqrt{1+\delta^2}+\delta}{1+\sqrt{1+\delta^2}-\delta}\right)^2 = \left(\frac{\delta - \frac{\delta^2}{2} + \frac{\delta^4}{8} - \frac{\delta^6}{16} + \cdots}{2 - \delta + \frac{\delta^2}{2} - \frac{\delta^4}{8} + \frac{\delta^6}{16} + \cdots}\right)^2,$$

$$= \left(\frac{\delta}{2} - \frac{\delta^3}{8} + \frac{\delta^5}{16} + 0\delta^6 + \cdots\right)^2 = \frac{\delta^2}{4} - \frac{\delta^4}{8} + \frac{5}{64}\delta^6 + \cdots.$$

Similarly,

$$\frac{\sqrt{1-\delta^2}-1}{2\sqrt{1+\delta^2}} = \frac{\frac{\delta^2}{2} - \frac{\delta^4}{8} + \frac{\delta^6}{16} + \cdots}{2\left(1 + \frac{\delta^2}{2} - \frac{\delta^4}{8} + \frac{\delta^6}{16} + \cdots\right)} = \frac{\delta^2}{4} - \frac{3}{16}\delta^4 + \cdots.$$

For all sufficiently small $\delta > 0$, $\frac{\delta^2}{4} - \frac{\delta^4}{8} + \cdots > \frac{\delta^2}{4} - \frac{3}{16}\delta^4 + \cdots$, that is, $\frac{\delta^4}{16} + \cdots = \delta^4(\frac{1}{16} + \cdots) > 0$, since the continuity of analytic functions ensures that $\lim_{\delta \to 0^+}(\frac{1}{16} + \cdots) = \frac{1}{16} > 0$. This completes the verification that E is to the right of K_3.

The above information and (3) show that $d(D, E)$ is the natural logarithm of

$$\left[\frac{1-\dfrac{F}{2R}}{\dfrac{\varepsilon}{R}}\right]\Bigg/\left[\frac{1-\dfrac{\dfrac{F}{2}+\varepsilon^2\left(\dfrac{1-\nu^2}{1+\nu^2}\right)}{R}}{\dfrac{2\varepsilon^2\nu}{(1+\nu^2)R}}\right]=\left[\frac{2\varepsilon\nu}{1+\nu^2}\right]\Bigg/\left[1-\frac{2\varepsilon^2(1-\nu^2)}{(2R-F)(1+\nu^2)}\right].$$

We next proceed to express the key ingredients in the preceding formula as analytic functions of δ.

First,

$$\nu=\sqrt{\sqrt{1+\delta^2}-\delta}=\sqrt{1-\delta+\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots}$$

$$=1+\frac{1}{2}\left(-\delta+\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)-\frac{1}{8}\left(-\delta+\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)^2+\cdots$$

$$=1-\frac{\delta}{2}+\frac{\delta^2}{8}+\cdots;$$

thus $\nu^2=1-\delta+\dfrac{\delta^2}{2}-\dfrac{\delta^4}{8}+\dfrac{\delta^6}{16}+\cdots$. Similarly, $\varepsilon^2=\sqrt{1+\delta^2}=1+\dfrac{\delta^2}{2}-\dfrac{\delta^4}{8}+\dfrac{\delta^6}{16}+\cdots$, and so

$$\varepsilon=\sqrt{1+\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots}$$

$$=1+\frac{1}{2}\left(\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)-\frac{1}{8}\left(\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)^2+\cdots$$

$$=1+\frac{\delta^2}{4}-\frac{3}{32}\delta^4+\cdots.$$

Next,

$$F=\frac{(1-\varepsilon^2)(1+\nu^2)}{1-\nu^2}=\frac{\left(-\dfrac{\delta^2}{2}+\dfrac{\delta^4}{8}-\dfrac{\delta^6}{16}+\cdots\right)\left(2-\delta+\dfrac{\delta^2}{2}-\dfrac{\delta^4}{8}+\dfrac{\delta^6}{16}+\cdots\right)}{\delta-\dfrac{\delta^2}{2}+\dfrac{\delta^4}{8}-\dfrac{\delta^6}{16}+\cdots}$$

$$=\frac{-\delta^2+\dfrac{\delta^3}{2}-\dfrac{\delta^5}{8}+0\delta^6+\cdots}{\delta-\dfrac{\delta^2}{2}+\dfrac{\delta^4}{8}-\dfrac{\delta^6}{16}+\cdots}=-\delta+0\delta^5+\cdots,$$

and so $F^2=\delta^2+0\delta^6+\cdots$. Finally,

$$R=\frac{\sqrt{4\varepsilon^2+F^2}}{2}$$

$$=\frac{\sqrt{4\left(1+\dfrac{\delta^2}{2}-\dfrac{\delta^4}{8}+\dfrac{\delta^6}{16}+\cdots\right)+(\delta^2+0\delta^6+\cdots)}}{2}$$

$$= \frac{\sqrt{4+3\delta^2-\frac{\delta^4}{2}+\frac{\delta^6}{4}+\cdots}}{2} = \sqrt{1+\frac{3}{4}\delta^2-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots}$$

$$= 1+\frac{1}{2}\left(\frac{3}{4}\delta^2-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)-\frac{1}{8}\left(\frac{3}{4}\delta^2-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)^2$$

$$+\frac{1}{16}\left(\frac{3}{4}\delta^2-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)^3+\cdots$$

$$= 1+\frac{3}{8}\delta^2-\frac{17}{128}\delta^4+\frac{83}{1024}\delta^6+\cdots.$$

Substituting the above expressions, we have that $d(D,E)$ is the natural logarithm of

$$\frac{\dfrac{2\left(1+\frac{\delta^2}{4}-\frac{3}{32}\delta^4+\cdots\right)\left(1-\frac{\delta}{2}+\frac{\delta^2}{8}+\cdots\right)}{2-\delta+\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots}}{1-\dfrac{2\left(1+\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)\left(\delta-\frac{\delta^2}{2}+\frac{\delta^4}{8}-\frac{\delta^6}{16}+\cdots\right)}{\left(2+\delta+\frac{3}{4}\delta^2-\frac{17}{64}\delta^4+\cdots\right)\left(2-\delta+\frac{\delta^2}{2}-\frac{\delta^4}{8}+\frac{\delta^6}{16}+\cdots\right)}}$$

$$= \frac{1+\frac{1}{8}\delta^2+\cdots}{1-\frac{\delta}{2}+\frac{\delta^2}{4}-\frac{\delta^3}{16}-\frac{\delta^4}{16}+\frac{13}{128}\delta^5+\cdots}$$

$$= 1+\frac{\delta}{2}+\frac{\delta^2}{8}+\cdots.$$

Hence,

$$d(D,E) = \left(\frac{\delta}{2}+\frac{\delta^2}{8}+\cdots\right)-\frac{\left(\frac{\delta}{2}+\frac{\delta^2}{8}+\cdots\right)^2}{2}+\cdots = \frac{\delta}{2}+0\delta^2+\cdots,$$

as asserted. Therefore,

$$\lim_{\delta\to 0^+}\frac{d(D,E)}{d(B,C)} = \lim_{\delta\to 0^+}\frac{\frac{\delta}{2}+0\delta^2+\cdots}{\delta+0\delta^2-\frac{\delta^3}{24}+\cdots} = \lim_{\delta\to 0^+}\frac{\frac{\delta}{2}+0\delta^2+\cdots}{\delta-\frac{\delta^3}{24}+\cdots} = \frac{1}{2},$$

to complete the proof.

REFERENCES

1. T. M. Apostol, *Mathematical Analysis*, 2nd edition, Addison-Wesley Publ. Co., Reading, MA, 1974.
2. K. Borsuk and W. Szmielew, *Foundations of Geometry*, North-Holland Publ. Co., Amsterdam, 1960.
3. E. E. Moise, *Elementary Geometry from an Advanced Standpoint*, Addison-Wesley Publ. Co., Reading, MA, 1974.
4. S. Stahl, *The Poincaré Half-Plane: A Gateway to Modern Geometry*, Jones and Bartlett Publ., Boston, MA, 1993.

NOTES

"Persian" Recursion

ANNE M. BURNS
Long Island University
C.W. Post Campus
Brookville, NY 11548

In this note we present a very simple recursive procedure that produces a variety of attractive patterns resembling Persian rugs. Recursion, by its very nature, can lead to surprisingly complicated and self-similar patterns. The idea behind the "Persian" recursion is to start with a large square, subdivide into four equal squares, and continue the process until we cannot go any further.

We begin with a $2^n + 1$ by $2^n + 1$ matrix of cells. Our objective is to "color" each cell. We shall do this by assigning it a number from 0 to $m - 1$, where m is the number of colors available. The values $n = 8$ and $m = 16$ work well on most computers.

The first step is to "color" the outermost cells arbitrarily to produce a "bordered square." The simplest way to do this is to color them all the same color. Then the following scheme is applied recursively: (a) Use the four corner cells and any convenient function of four variables to determine a new color. (b) Assign this new color to all interior cells in the middle row and middle column. (c) Apply the same procedure to each of the four new "bordered squares." The process terminates when all cells have been "colored."

A simple function of four variables for determining the next color is adding a predefined constant a to the sum of the four corner colors and reducing the result modulo m:

$$f(c_1, c_2, c_3, c_4) = (c_1 + c_2 + c_3 + c_4 + a) \bmod m \tag{1}$$

It is instructive to trace through a few stages by hand to see how the colors are generated. FIGURE 1 shows how the colors are assigned using (1) with $m = 16$, $n = 3$, and $a = 0$, and initializing the border to color 1.

The simple formula (1) produces an amazing variety of patterns. FIGURE 2 illustrates (1) with $m = 16, n = 8, a = 0$ and the border color 1.

A simple variation on (1) is adding a to the truncated average of the four corner colors and reducing the result modulo m:

$$f(c_1, c_2, c_3, c_4) = \big(\mathrm{trunc}((c_1 + c_2 + c_3 + c_4)/4) + a\big) \bmod m \tag{2}$$

The appendix contains a simple Basic program based on (2).

It is remarkable, too, that the same number patterns with different color assignments will appear dissimilar, as different color assignments emphasize different structures hiding in the number patterns. In FIGURE 3 we used the color function (2) with $m = 16, n = 8, a = 2$, a border color of 14 and two different palettes. In color the differences are even more striking. As is often the case with recursion, if the process is carried out for large values of n we observe that the patterns repeat on different scales.

Even more variety can be obtained by experimenting with other functions that determine the new color from the four corner colors. Quadratic functions, cubic

1	1	1	1	1	1	1	1	1
1	10	7	13	4	13	7	10	1
1	7	7	7	4	7	7	7	1
1	13	7	3	4	3	7	13	1
1	4	4	4	4	4	4	4	1
1	13	7	3	4	3	7	13	1
1	7	7	7	4	7	7	7	1
1	10	7	13	4	13	7	10	1
1	1	1	1	1	1	1	1	1

FIGURE 1
$M = 16$, $n = 3$, border color = 1

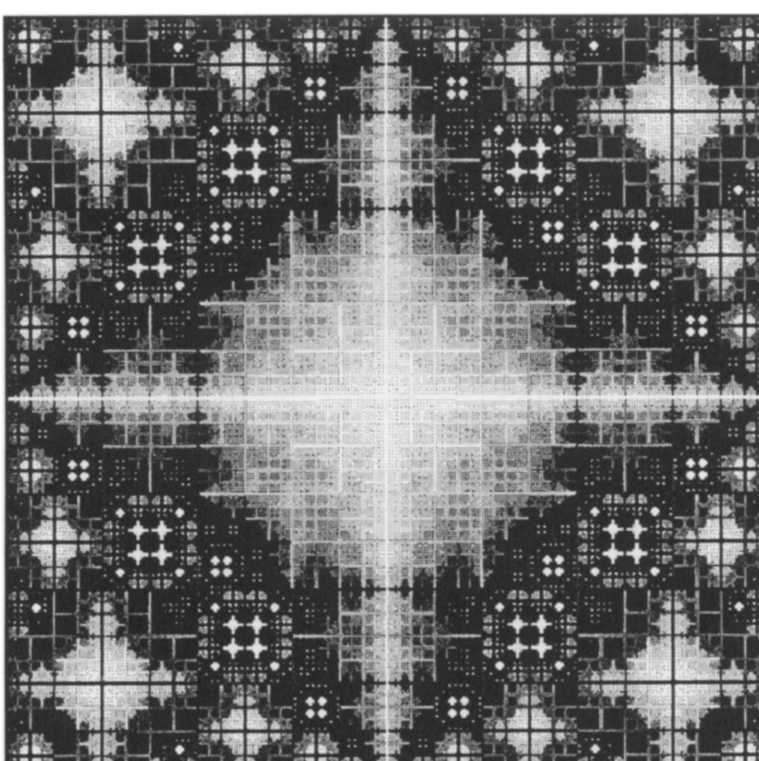


FIGURE 2
$m = 16$, $n = 8$, $a = 0$, border color = 1

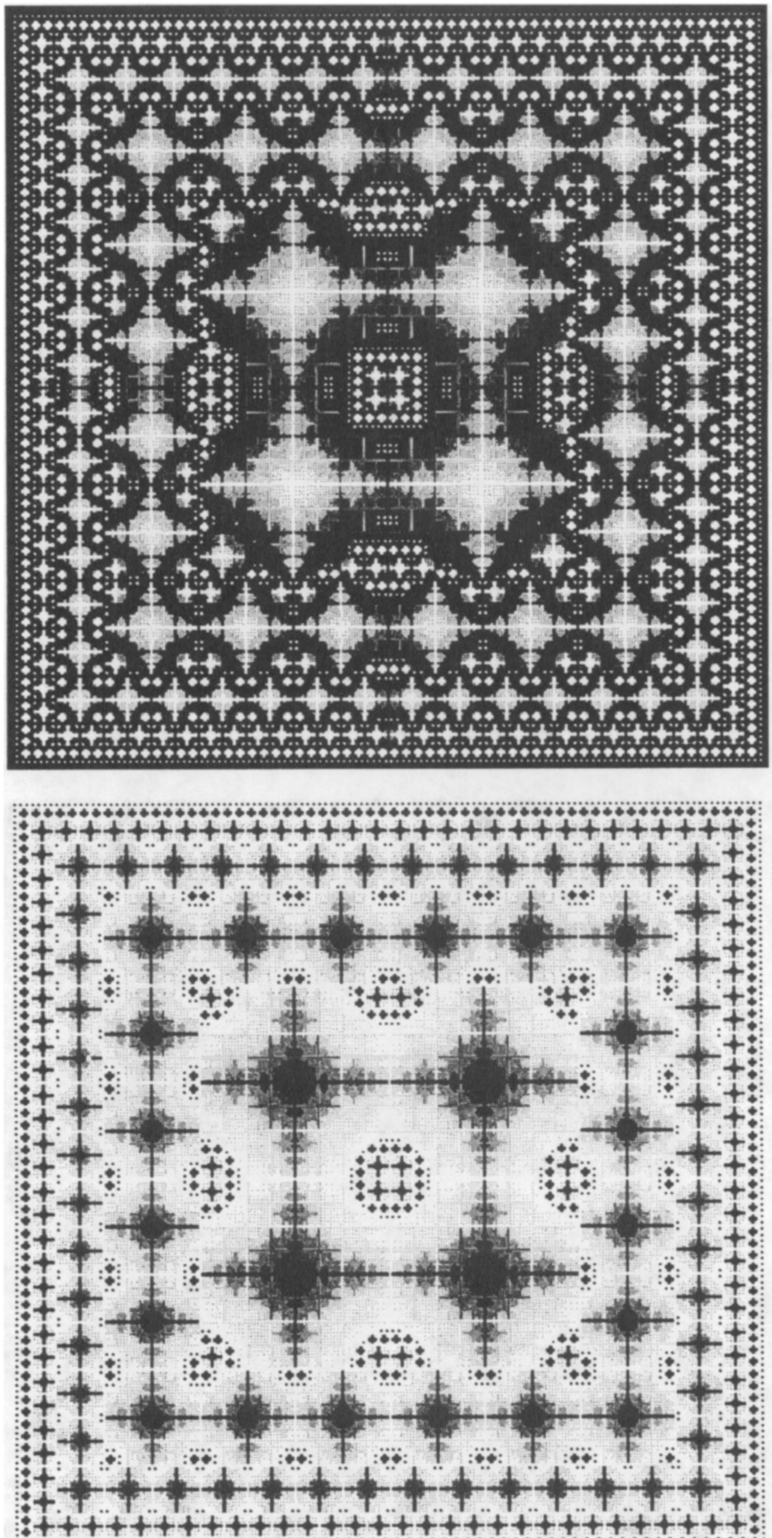


FIGURE 3

Two rugs with the same number patterns but different palettes
$m = 16$, $n = 8$, $a = 2$, border color $= 14$

functions and combinations of sine and cosine functions are just a few examples. Another way to achieve variety is to color the individual cells on the border in a symmetric pattern, such as dark to light from the beginning of the edge to the center, then light to dark from the center to the end of the edge. For a student project we might investigate making patterns from other figures such as triangles or hexagons.

There are many other recursive and iterative methods for coloring a grid. The idea of basing the color of a cell on the colors of its nearest neighbors is explored in depth in **[2]**. Many other ideas for coloring grids can be found in the chapter "Wallpaper for the Mind" in **[1]**.

Appendix

The following Basic program is based on (2) with $a = 3$. A recursive function draws two lines dividing the current square into four new squares and then calls itself four times, once for each of the new squares.

```
DECLARE FUNCTION ColorGrid! (left! right!, top!, bottom!)
DECLARE FUNCTION f! (left!, right!, top!, bottom!)
INPUT "Enter the border color (1 through 15) ", bordercolor
SCREEN 12
CLS
left = 0
right = 256
top = 0
bottom = 256
LINE (left, top)-(right, top), bordercolor
LINE (left, bottom)-(right, bottom), bordercolor
LINE (left, top)-(left, bottom), bordercolor
LINE (right, top)-(right, bottom), bordercolor
k = ColorGrid(left, right, top, bottom)
END
FUNCTION ColorGrid (left, right, top, bottom)
IF left < right - 1 THEN
  c = f(left, right, top, bottom)
  middlecol = (left + right) / 2
  middlerow = (top + bottom) / 2
  LINE (left + 1, middlerow)-(right - 1, middlerow), c
  LINE (middlecol, top +1)-(middlecol, bottom - 1), c
  ColorGrid = ColorGrid(left, middlecol, top, middlerow)
  ColorGrid = ColorGrid(middlecol, right, top, middlerow)
  ColorGrid = ColorGrid(left, middlecol, middlerow, bottom)
  ColorGrid = ColorGrid(middlecol, right, middlerow, bottom)
END IF
END FUNCTION
FUNCTION  f (left, right, top, bottom)
p = POINT(left,top)+POINT(right,top)+POINT(left,bottom)
    + POINT(right,bottom)
f = (p / 4 + 3) MOD 16
END FUNCTION
```

REFERENCES

1. A. K. Dewdney, *The Armchair Universe*, W. H. Freeman and Co., New York, NY, 1988.
2. Tommaso Toffoli and Norman Margolus, *Cellular Automata Machines: A New Environment for Modeling*, The MIT Press, Cambridge, MA, 1987.

Another Proof of Pick's Area Theorem

CHRISTIAN BLATTER
Department of Mathematics
ETH Zürich
Switzerland

Let P be a *simple lattice polygon*, i.e., a polygonal Jordan domain in the plane whose vertices have integer coordinates. *Pick's theorem* says that the area $\mu(P)$ is given by

$$\mu(P) = i + \frac{b}{2} - 1,$$

where i and b denote the number of lattice points in the interior and on the boundary of P, respectively. Several proofs of this theorem can be found in the literature (see **[1]** for a recent list of references); most of them use a dissection argument combined with an analysis of certain triangles. Here I offer a proof of a more conceptual nature; it has the form of a *Gedankenexperiment* (thought-experiment).

Assume that at time 0 a unit of heat is concentrated at each lattice point. This heat will be distributed over the whole plane by heat conduction, and at time ∞ it is equally distributed on the plane with density 1. In particular, the amount of heat contained in P will be $\mu(P)$. Where does this amount of heat come from? Consider an edge e of P with midpoint m. The lattice points not on e come in pairs lying symmetrically with respect to m, and two such lattice points will send the same amount of heat across e, but in opposite directions. This implies that the total heat flux across e is 0, so that the final amount of heat within P comes from the interior lattice points and from the lattice points lying on the boundary ∂P. To account for the latter, orient ∂P so that the interior is to the left of ∂P. A lattice point on ∂P which is not a vertex sends half its heat into the interior. The amount of heat going from a vertex into the interior is again a half, minus the turning angle of ∂P at that vertex, measured in units of 2π. Since the sum of all turning angles for a simple polygon is known to be one full turn, we arrive at the stated formula.

REFERENCE

1. B. Grünbaum and G. C. Shephard: Pick's theorem, *Amer. Math. Monthly* 100 (1993), 150–160.

A New Proof of the Formula for the Number of 3 × 3 Magic Squares

MIKLÓS BÓNA
Massachusetts Institute of Technology
Cambridge, MA 02139

Introduction For our purposes, a *magic square* is a square matrix with nonnegative integer entries in which all row sums and column sums are equal. (We note that other notions of magic squares exist. For example, one might require that the diagonal sums be the same, too, or one might forbid repetitions of elements in any line.) Let $H_n(r)$ denote the number of $n \times n$ magic squares of line sum r. It has long been known (see **[1]** and **[4]**) that

$$H_3(r) = \binom{r+4}{4} + \binom{r+3}{4} + \binom{r+2}{4}. \tag{1}$$

However, we haven't seen any proof that actually showed *why* the result is the sum of three binomial coefficients. This has suggested that a more natural proof could be found. In this paper we are going to give such a proof. In fact, we divide the 3×3 magic squares into three different classes, each of which will provide one of the binomial coefficients in (1) in a natural way.

Our proof Take any magic square of line sum r and side length 3. It is clear that the four elements shown in the figure determine all the rest of the square.

a	d	
	b	
		c

Indeed, the next figure shows our only possible choice for each remaining entry. Thus we need only compute the number of ways we can choose a, b, c and d so that we indeed *have* that one choice, i.e., so that all the entries of the magic square are nonnegative.

a	d	$r-a-d$
$r+c-$ $(a+d+b)$	b	$a+d-c$
$b+d-c$	$r-b-d$	c

The preceding figure shows that all the entries of our matrix will be nonnegative if and only if the following inequalities hold:

$$a + d \leq r \tag{2}$$

$$b + d \leq r \tag{3}$$

$$c \leq a + d \tag{4}$$

$$c \leq b + d \tag{5}$$

$$a + d + b - c \leq r. \tag{6}$$

To prove (1), we will consider three different cases, according to the position of the smallest element on the main diagonal. In each case, at least three of the five conditions above will become redundant, and we will only need to deal with the remaining one or two.

1. Suppose $0 \leq a \leq b$ and $0 \leq a \leq c$. In this case conditions (2), (5), and (6) are clearly redundant, because they are implied by (3) and (4). The crucial observation is that in all the three cases we can collect all our conditions into one chain of inequalities. In this case, we do it as follows:

$$a \leq 2a + d - c \leq a + b + d - c \leq b + d \leq r. \tag{7}$$

   Indeed, the first inequality is equivalent to (4), the second one is equivalent to our assumption that $a \leq b$, the third one is equivalent to our assumption that $a \leq c$, and the last one is equivalent to (3).

   Moreover, note that once we know the terms of this chain (i.e., $a, 2a + d - c$, $a + b + d - c, b + d$), then we know a, b, c, and d, too, so we have determined the magic square. Thus we need only count how many ways there are to choose these four terms. Inequality (7) shows that these terms are nondecreasing, therefore the number of ways to choose them is simply the number of 4-combinations of $r + 1$ elements with repetitions allowed, which is $\binom{r+4}{4}$. (Recall that 0 is allowed to be an entry.)

2. Now suppose $a > b$ and $c \geq b$. Then (3), (5), and (6) are redundant. Consider the chain of inequalities

$$b \leq 2b + d - c \leq a + b + d - c - 1 \leq a + d - 1 \leq r - 1. \tag{8}$$

   We can use the argument of the previous case to prove that (8) is equivalent to (2), (6), and our assumptions, as the roles of a and b are completely symmetric. The only change is that here we don't count those magic squares in which $a = b$—this explains the -1 in the last three terms. Thus here we have to choose 4 elements in nondecreasing order out of the set $\{0, 1, \cdots, r - 1\}$; this can be done in $\binom{r+3}{4}$ ways.

3. Finally, suppose $a > c$ and $b > c$. Then (2), (3), (4), and (5) are redundant. Condition (6) and our assumptions can be collected into the following chain:

$$c \leq b - 1 \leq b + d - 1 \leq a + b + d - c - 2 \leq r - 2. \tag{9}$$

   Here the first inequality is equivalent to our assumption $c < b$, the second one says that d is nonnegative, the third one is equivalent to our assumption $c < a$, and the last one is equivalent to (6). The four terms of (9) determine, a, b, c, and d, and they can be chosen in $\binom{r+2}{4}$ ways, which completes the proof.

Thus the number of 3×3 magic squares of line sum r is indeed $\binom{r+4}{4} + \binom{r+3}{4} + \binom{r+2}{4}$. Furthermore, the three terms in this sum count the magic squares in which the minimal element of the main diagonal first occurs in the first, second, or third position.

For results on larger magic squares, see **[6]** or **[4]**.

REFERENCES

1. H. Anand, V. C. Dumir and H. Gupta, A combinatorial distribution problem, *Duke Math. J.* 33 (1966), 757–769.
2. M. Bóna, There are a lot of magic squares, *Studies in Applied Mathematics* 94 (1995), 415–421.
3. C. J. Henrich, Magic squares and linear algebra, *Amer. Math. Monthly* 98 (1991), 481–488.
4. P. A. MacMahon, *Combinatorial Analysis*, vols. 1–2, Cambridge Univ. Pr., Cambridge, UK, 1916 (reprinted by Chelsea, New York, NY, 1960).
5. D. B. Shapiro, A replication property for magic squares, this MAGAZINE 65 (1992), 155–160.
6. R. P. Stanley, *Combinatorics and Commutative Algebra*, Progress in Mathematics 41, Birkhäuser, Boston, MA, 1983.

Bell's Conjecture*

For math, the Oscar envelope
(Assured by Price and Waterhouse)
Would list a three-way tie, I'd hope:
Archimedes, Newton, Gauss.

fine

Archimedes' *modern* mind
(Narrowly he bounded pi),
Impelled to seek and swift to find,
Defined the Hellenistic high.

Newton's fluxions formed the frame
That fit the Universal Law.
Even Leibniz spread his fame:
"We know the Lion by his claw."

Many Magi graced the scene
But Gauss was greater than all since.
If Number Theory is the Queen,
Carl Friedrich is its freshest Prince.

D.C.

*E. T. Bell, *Men of Mathematics*, Simon and Schuster, New York, 1961.

—J. D. MEMORY
THE UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NC 27515

Thus the number of 3×3 magic squares of line sum r is indeed $\binom{r+4}{4} + \binom{r+3}{4} + \binom{r+2}{4}$. Furthermore, the three terms in this sum count the magic squares in which the minimal element of the main diagonal first occurs in the first, second, or third position.

For results on larger magic squares, see **[6]** or **[4]**.

REFERENCES

1. H. Anand, V. C. Dumir and H. Gupta, A combinatorial distribution problem, *Duke Math. J.* 33 (1966), 757–769.
2. M. Bóna, There are a lot of magic squares, *Studies in Applied Mathematics* 94 (1995), 415–421.
3. C. J. Henrich, Magic squares and linear algebra, *Amer. Math. Monthly* 98 (1991), 481–488.
4. P. A. MacMahon, *Combinatorial Analysis*, vols. 1–2, Cambridge Univ. Pr., Cambridge, UK, 1916 (reprinted by Chelsea, New York, NY, 1960).
5. D. B. Shapiro, A replication property for magic squares, this MAGAZINE 65 (1992), 155–160.
6. R. P. Stanley, *Combinatorics and Commutative Algebra*, Progress in Mathematics 41, Birkhäuser, Boston, MA, 1983.

Bell's Conjecture*

For math, the Oscar envelope
(Assured by Price and Waterhouse)
Would list a three-way tie, I'd hope:
Archimedes, Newton, Gauss.

fine

Archimedes' *modern* mind
(Narrowly he bounded pi),
Impelled to seek and swift to find,
Defined the Hellenistic high.

Newton's fluxions formed the frame
That fit the Universal Law.
Even Leibniz spread his fame:
"We know the Lion by his claw."

Many Magi graced the scene
But Gauss was greater than all since.
If Number Theory is the Queen,
Carl Friedrich is its freshest Prince.

D.C.

*E. T. Bell, *Men of Mathematics*, Simon and Schuster, New York, 1961.

—J. D. MEMORY
THE UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NC 27515

Can One Load a Set of Dice So That the Sum Is Uniformly Distributed?

GUANTAO CHEN
Georgia State University
Atlanta, GA 30303

M. BHASKARA RAO
North Dakota State University
Fargo, ND 58105

WARREN E. SHREVE
North Dakota State University
Fargo, ND 58105

Introduction Suppose we have two ordinary six-faced dice, and S is the sum of the numbers shown when the dice are rolled once. Then the random variable S can take any of the eleven integer values from 2 to 12. Can one load the dice so that S is uniformly distributed, i.e., $\Pr(S=i)=1/11$ for $i=2,3,\dots,12$?

The answer is no, as can be shown by an elegant probabilistic argument using only the simple inequality that $x+1/x\geq 2$ for all $x>0$. (See Problem 52 in [**4**, p. 130].) Another (analytical) proof uses a polynomial factorization, as follows. Let p_i and r_i be the probabilities of the number i appearing on the first and second dice, respectively, for $i=1,2,3,4,5,6$. Suppose that the distribution of S is uniform. Then the probability generating function of S (for an exposition on probability generating functions, see [**3**, p. 177]) satisfies the following identity:

$$\frac{1}{11}(x^2+x^3+\cdots+x^{12})=(p_1x+p_2x^2+\cdots+p_6x^6)(r_1x+r_2x^2+\cdots+r_6x^6),$$

or, equivalently,

$$1+x+\cdots+x^{10}=11(p_1+p_2x+\cdots+p_6x^5)(r_1+r_2x+\cdots+r_6x^5).$$

Now observe that the complex roots of the polynomial $1+x+x^2+\cdots+x^{10}$ are symmetrically spaced on the unit circle in the complex plane, and that none is real. Since

$$\frac{1}{11}=\Pr(S=2)=p_1r_1=\Pr(S=12)=p_6r_6,$$

it follows that none of p_1, p_6, r_1, and r_6 vanishes. Thus each factor on the right side of (1) is actually a fifth-degree polynomial; this implies the contradiction that each factor has at least one real root.

A generalization We now generalize the above problem to the case of n dice, each m-faced with faces numbered from 1 to m. We will show that the answer to the question posed in the title is negative.

More precisely, suppose we have n independent random variables $X_1, X_2,\dots,X_n$, such that each X_i takes values $1,2,\dots,m$, with respective probabilities

$$\Pr(X_i=1)=p_{i1},\Pr(X_i=2)=p_{i2},\dots,\Pr(X_i=m)=p_{im}.$$

Let $S=X_1+X_2+\cdots+X_n$. Is it possible to choose the probabilities p_{ij} such that the distribution of S is uniform, i.e., such that

$$q_j=\Pr(S=j)=\frac{1}{nm-n+1}\text{ for } j=n,n+1,\dots nm?$$

If m is even and $n>1$, one can show that the answer is negative by a factorization argument similar to that in the introduction. However, we prove the following more general result, using probabilistic arguments.

THEOREM. *If $n>1$, then for any choice of probabilities p_{ij}, the distribution of S satisfies the inequality*

$$\max_{n\le i,j\le nm}|q_i-q_j|\ge\frac{n-1}{n^2m}.$$

The desired result follows immediately (otherwise, $|q_i-q_j|=0$ for all i,j):

COROLLARY. *If $n>1$, there do not exist p_{ij}'s such that S is uniformly distributed.*

Proof of theorem. Let $X_1', X_2',\dots,X_n'$, be independent random variables, such that each X_i' takes values $1,2,\dots,m$, with respectively "reversed" probabilities, as follows:

$$\begin{array}{ccccccc} X_i' & : & 1 & 2 & \cdots & m-1 & m \\ \Pr & : & p_{im} & p_{i(m-1)} & \cdots & p_{i2} & p_{i1} \end{array}.$$

Let $S'=X_1'+X_2'+\cdots+X_n'$, and $q_i'=\Pr(S'=i)$, $i=n,\ n+1,\dots,nm$. Note that each q_i' is equal to some q_j, and conversely. This implies that

$$\max_{n\le i,j\le nm}|q_i-q_j|=\max_{n\le i,j\le nm}|q_i'-q_j'|.$$

We can assume, without loss of generality, that $q_n\le q_{mn}$. To see why, note that for any choice of integers $1\le a_1,a_2,\dots,a_n\le m$,

$$\begin{aligned}&\Pr((X_1',X_2',\dots,X_n')=(a_1,a_2,\dots,a_n))\\&\quad=\Pr((X_1,X_2,\dots,X_n)=(m+1-a_1,m+1-a_2,\dots,m+1-a_n)).\end{aligned}$$

It follows that for all $i=n,\ n+1,\dots,nm$,

$$q_i'=\Pr(S'=i)=\Pr(S=n(m+1)-i)=q_{n(m+1)-i}.$$

In particular,

$$q_n'=q_{nm}\text{ and }q_{nm}'=q_n.$$

From this we conclude that at least one of the inequalities

$$q_n'\le q_{nm}'\text{ and }q_n\le q_{nm}$$

must hold. Thus, if $q_n\le q_{nm}$ fails, we may argue in terms of the variables X_i', which satisfy $q_n'\le q_{nm}'$. In the following, we assume that $q_n\le q_{nm}$, and consider two cases.

Case 1: $q_n\le 1/n^2m$.

Since $\sum_{i=n}^{nm}q_i=1$, it follows that

$$\max_{n\le i,j\le nm}q_i\ge\frac{1}{nm-n+1}\ge\frac{1}{nm}.$$

Therefore,

$$\begin{aligned}\max_{n\le i,j\le nm}|q_i-q_j|&\ge\max_{n\le i\le nm}q_i-q_n\\&\ge\frac{1}{nm}-\frac{1}{n^2m}=\frac{n-1}{n^2m}.\end{aligned}$$

This establishes the desired inequality.

Case 2: $q_n > 1/n^2 m$.

Since

$$q_n = \Pr(S = n) = p_{11} p_{21} \cdots p_{n1},$$
$$q_{nm} = \Pr(S = nm) = p_{1m} p_{2m} \cdots p_{nm}, \quad \text{and} \quad q_n \le q_{nm},$$

it follows that each p_{i1} and p_{im} is positive. Therefore

$$\begin{aligned}
q_{n+m-1} &= \Pr(S = n + m - 1) \\
&\ge \sum_{i=1}^{n} \Pr(X_1 = 1, X_2 = 1, \dots, X_{i-1} = 1, X_i = m, X_{i+1} = 1, \dots, X_n = 1) \\
&= \sum_{i=1}^{n} p_{11} p_{21} \cdots p_{(i-1)1} p_{im} p_{(i+1)1} \cdots p_{n1} \\
&= p_{11} p_{21} \cdots p_{n1} \left(\frac{p_{1m}}{p_{11}} + \frac{p_{2m}}{p_{21}} + \cdots + \frac{p_{nm}}{p_{n1}} \right) \\
&\ge q_n n (p_{1m} p_{2m} \cdots p_{nm} / p_{11} p_{21} \cdots p_{n1})^{1/n} \\
&\qquad \text{(by the Arithmetic Mean-Geometric Mean Inequality)} \\
&= q_n n (q_{nm}/q_n)^{1/n} \\
&\ge n q_n \quad (\text{since } q_n \le q_{nm}).
\end{aligned}$$

This establishes the desired inequality, and completes the proof.

Two comments A basic problem in the theory of polynomials is how to factor them. If a reducible polynomial has integer coefficients, it is interesting to find factors with integer coefficients. (See, e.g., Barbeau [**1**, p. 84].) In a similar vein, given a polynomial with nonnegative coefficients it is interesting to find factors with nonnegative coefficients. The probabilistic problem considered here throws some light on this algebraic problem. Consider the polynomial $1 + x + x^2 + \cdots + x^{nm-n}$, of degree $nm - n$. All of its coefficients are nonnegative. If $n > 1$ and $m > 1$, then it follows from the Corollary above that it is impossible to factor the polynomial into n polynomials, each of degree $m - 1$, such that each factor has nonnegative coefficients.

We remark, finally, that if one may choose the numbers to be imprinted on the dice, then a uniform distribution for S is possible, as Barnard [**2**] demonstrates as follows: Take two fair, six-faced dice. On one die, label the faces $1, 2, 3, 4, 5, 6$, as usual. On the other die, print 0 on three faces and 6 on the other three. Then S is uniformly distributed, taking the values 1 through 12, each with probability $1/12$.

REFERENCES

1. E. J. Barbeau, *Polynomials*, Springer-Verlag, New York, NY, 1989.
2. D. St. P. Barnard, *One Hundred Brain Twisters*, Van Nostrand, Princeton, NJ, 1966.
3. K. L. Chung, *Elementary Probability Theory with Stochastic Processes*, Springer-Verlag, New York, NY, 1979.
4. R. Honsberger, *Mathematical Morsels*, The Mathematical Association of America, Washington, DC, 1978.

Students Ask the Darnedest Things: A Result in Elementary Group Theory

STEVE BENSON
University of New Hampshire
Durham, NH 03824

Introduction In most elementary group theory courses, students are introduced to the Fundamental Theorem of Finite Abelian Groups (FT), which states that every finite abelian group is isomorphic to exactly one direct product of cyclic groups of prime power order. (These cyclic groups are called *invariant factors*.) As every cyclic group of order k is isomorphic to $\mathbb{Z}_k$, the additive group of integers mod k, FT asserts that every abelian group of order 12 is isomorphic either to $\mathbb{Z}_4 \times \mathbb{Z}_3$ or to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$. (I'll use the common abbreviation $\mathbb{Z}_k$ to denote $\mathbb{Z}/k\mathbb{Z}$, with apologies to my fellow number theorists, who usually use this notation to denote the k-adic integers.)

A standard exercise is to determine whether a given abelian group G of order 12 is isomorphic to $\mathbb{Z}_4 \times \mathbb{Z}_3$ or $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$ by computing the orders of the elements of G. For instance, if G contains an element of order 12 (or 4, for that matter), it must be isomorphic to the cyclic group $\mathbb{Z}_4 \times \mathbb{Z}_3$, and if G has more than 1 element of order 2, then it is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$. Notice, however that every abelian group of order 12 contains exactly two elements of order 3. An important concept reinforced through such an exercise is that isomorphisms preserve order. Specifically, if $f\colon G \to G'$ is a group isomorphism, then g and $f(g)$ have the same order in their respective groups. Thus if two finite groups G and G' are isomorphic, they must have *identical order structure* (the same number of elements of each order).

This article examines the converse question, posed by a student in a group theory course:

> *Could two finite abelian groups have the same number of elements of each order, but* not *be isomorphic?*

The answer "no" seems to be a folk theorem (everyone *knows* it's been proven, but nobody can cite a reference). The more I thought about it, the less obvious the result became. After all, both $\mathbb{Z}_2 \times \mathbb{Z}_8$ and $\mathbb{Z}_4 \times \mathbb{Z}_4$ have 3 elements of order 2, yet they're certainly not isomorphic. To satisfy my own curiosity (and to practice what I preach when I tell my students that they don't really know whether a conjecture is true until they've either understood a proof or proved it themselves), I decided to work on the problem. I succeeded in proving the expected result: if two finite abelian groups have the same order structure, then they are isomorphic. The key idea—and the reason I hope this article will interest the reader—is an algorithm that counts the elements of any given order in a finite abelian group G, in terms of the invariant factors of G. I'll illustrate the formula and proof with several examples and show what happens when either hypothesis on G and G' (finite, abelian) is violated. Finally, I'll derive a "converse" to the formula mentioned above, making it possible to determine the invariant factors of a group G given the order structure of G.

The order structure of finite abelian *p*-groups First, some preliminaries. As FT shows, the first step in studying finite abelian groups is to investigate cyclic groups of prime power order. A group G is called a *p-group* if $|G| = p^n$, for some positive integer n and prime p. As the direct product of cyclic p-groups is an abelian p-group, FT also says that every finite abelian group is isomorphic to a direct product of abelian p-groups for various primes p. Of course, if G is itself a cyclic group of order p^k for some prime p, then G is isomorphic to $\mathbb{Z}_{p^k}$. Thus FT implies that every finite abelian p-group of order p^n is isomorphic to exactly one group of the form

$$\mathbb{Z}_p^{k_1} \times \mathbb{Z}_{p^2}^{k_2} \times \cdots \times \mathbb{Z}_{p^n}^{k_n},$$

where each k_i is a non-negative integer. (Here $\mathbb{Z}_m^n$ denotes the direct product

$$\underbrace{\mathbb{Z}_m \times \mathbb{Z}_m \times \cdots \mathbb{Z}_m}_{n \text{ times}}$$

when $n > 0$ and $\mathbb{Z}_m^0$ denotes the trivial group.) For short, we'll say that G has *p-factor type* $\langle k_1, k_2, \ldots, k_n \rangle$. For example, the Klein 4-group is isomorphic to $\mathbb{Z}_2^2 \times \mathbb{Z}_4^0$ and has 2-factor type $\langle 2, 0 \rangle$; the cyclic group of order 4 is isomorphic to $\mathbb{Z}_2^0 \times \mathbb{Z}_4^1$ and so has 2-factor type $\langle 0, 1 \rangle$.

In order to derive a formula that counts the elements of order p^i in a finite abelian p-group, it is natural to *first* investigate cyclic, *then* the direct product of cyclic, p-groups. To understand the cyclic case, we consider an example. Let $G = \mathbb{Z}_{27}$. The elements of order 27 in G are 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, and 26. The elements of order 9 are $3 = 1 \cdot 3$, $6 = 2 \cdot 3$, $12 = 4 \cdot 3$, $15 = 5 \cdot 3$, $21 = 7 \cdot 3$, and $24 = 8 \cdot 3$, while the elements of order 3 are $9 = 1 \cdot 9$ and $18 = 2 \cdot 9$. That is, $\mathbb{Z}_{27}$ contains 18 elements of order 27, 6 elements of order 9, and 2 elements of order 3 (and of course 1 element of order 1).

We leave to the reader the task of confirming the following formula, which generalizes the result of the previous example. For the sake of brevity, we introduce the notation $N(G; m)$ to denote the number of elements of order m in the group G.

LEMMA 1.

$$N\left(\mathbb{Z}_{p^m}; p^i\right) = \begin{cases} \varphi(p^i) & \text{if } i \le m \\ 0 & \text{if } i > m, \end{cases} \quad \textit{where } \varphi \textit{ is the Euler } \varphi\textit{-function}.$$

Now consider the group $G = \mathbb{Z}_{27} \times \mathbb{Z}_{81}$. By Lemma 1, $\mathbb{Z}_{27}$ and $\mathbb{Z}_{81}$ both have $\varphi(9) = 6$ elements of order 9. The order of an element (a, b) in G is the least common multiple of the order of a and the order of b. Thus, (a, b) has order 9 if and only if *one* of a or b has order 9 and the other has order that *divides* 9, making the task of counting the elements of order 9 in G slightly more complicated. Notice that the order of (a, b) in G is a *divisor* of 9 if and only if the orders of a and b are both divisors of 9. But the number of elements in $\mathbb{Z}_{27}$ of an order which divides 9 is $N(\mathbb{Z}_{27}; 1) + N(\mathbb{Z}_{27}; 3) + N(\mathbb{Z}_{27}; 9) = 9$. Similarly, we see that $\mathbb{Z}_{81}$ also has 9 elements having order which divides 9. Again for the sake of brevity, we introduce $[G; m]$ to denote the number of elements in G having order that divides m. The formula $\sum_{j=0}^{i} \varphi(p^j) = p^i$, along with Lemma 1, allows us to confirm our next preliminary lemma.

LEMMA 2.

$$\left[\mathbb{Z}_{p^m}; p^i\right] = \begin{cases} p^i & \text{if } i \le m, \\ p^m & \text{if } i > m \end{cases}.$$

Back to $G = \mathbb{Z}_{27} \times \mathbb{Z}_{81}$. Since the order of (a, b) divides 9 if and only if the orders of a and b both divide 9, we see that $[G; 9] = [\mathbb{Z}_{27}; 9] \cdot [\mathbb{Z}_{81}; 9] = 81$. An analogous argument confirms our final lemma.

LEMMA 3. *If* $G_1, G_2, \ldots, G_n$ *are finite groups, then* $[G_1 \times G_2 \times \cdots \times G_n; m] = \prod_{j=1}^{n} [G_j; m]$.

Notice that if we are interested in computing $N(G; p^i)$ it is very useful to know that $N(G, p^i) = [G; p^i] - [G; p^{i-1}]$. Fortunately, $[G; p^j]$ is not terribly difficult to compute when G is an abelian p-group.

To this end, suppose G is an abelian p-group having p-factor type $\langle k_1, k_2, \ldots, k_n \rangle$. Then the preliminary lemmas give us

$$[G; p^i] = \prod_{j=1}^{n} \left[\mathbb{Z}_{p^j}^{k_j}; p^i\right] = \prod_{j=1}^{n} \left[\mathbb{Z}_{p^j}; p^i\right]^{k_j} = \prod_{j=1}^{i} (p^j)^{k_j} \prod_{j=i+1}^{n} (p^i)^{k_j}.$$

When $i = 1$, we then have $[G; p] = \prod_{j=1}^{n} p^{k_n} = p^{k_1 + k_2 + \cdots + k_n}$. Let $K = \sum_{j=1}^{n} k_j$. Then when $i > 1$,

$$\begin{aligned}
[G; p^i] &= p^{k_1 + 2k_2 + \cdots + ik_i} \cdot p^{i(k_{i+1} + k_{i+2} + \cdots + k_n)} \\
&= p^{k_1 + 2k_2 + \cdots + ik_i} \cdot p^{i(K - k_1 - k_2 - \cdots - k_i)} \\
&= p^{(k_1 + 2k_2 + \cdots + ik_i + iK - ik_1 - ik_2 - \cdots - ik_i)}.
\end{aligned}$$

Applying the fact that $N(G; p^i) = [G; p^i] - [G; p^{i-1}]$ gives the promised formula for $N(G; p^i)$ when G is an abelian p-group.

THEOREM 1. *Let p be prime. Suppose that G is an abelian group of order* p^n, *with p-factor type* $\langle k_1, k_2, \ldots k_n \rangle$. *Define* $K = \sum_{j=1}^{n} k_j$, $s_0 = 0$, $s_1 = K$, *and* $s_i = iK + \sum_{j=1}^{i-1} (j - i) k_j$ *when* $2 \le i \le n$. *Then* $N(G; p^i) = p^{s_i} - p^{s_{i-1}}$.

Examples Let's use Theorem 1 to determine the order structure of several specific p-groups. We leave to the reader the (optional, tedious) exercise of listing and computing the order of the group elements in order to confirm the results.

Example A. Let $G = \mathbb{Z}_2 \times \mathbb{Z}_4$. Then G is an abelian 2-group of order 8 having 2-factor type $\langle 1, 1, 0 \rangle$ (as $G \cong \mathbb{Z}_2^1 \times \mathbb{Z}_4^1 \times \mathbb{Z}_8^0$). Then $k_1 = 1 = k_2$, $k_3 = 0$, and $K = 1 + 1 + 0 = 2$, so $s_1 = 2$ and $s_2 = 2K - k_1 = 4 - 1 = 3 = s_3$. Therefore our formula shows that G has $2^{s_1} - 2^{s_0} = 4 - 1 = 3$ elements of order 2 and $2^{s_2} - 2^{s_1} = 8 - 4 = 4$ elements of order 4.

Example B. If $G = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_8$, then G is an abelian 2-group with 2-factor type $\langle 3, 0, 1, 0, 0, 0 \rangle$. Therefore, $k_1 = 3$, $k_3 = 1$, $k_2 = k_4 = k_5 = k_6 = 0$, $K = 4$, $s_1 = 4$, $s_2 = 2K - k_1 = 5$, $s_3 = 3K - 2k_1 - k_2 = 6$, $s_4 = 4K - 3k_1 - 2k_2 - k_3 = 6$, $s_5 = 5K - 4k_1 - 3k_2 - 2k_3 - k_4 = 6$, and $s_6 = 6K - 5k_1 - 4k_2 - 3k_3 - 2k_4 - k_5 = 6$. We then use the formula to compute $N(G; 2) = 2^K - 1 = 16 - 1 = 15$, $N(G; 4) = 2^{s_2} - 2^{s_1} = 32 - 16 = 16$, $N(G; 8) = 2^{s_3} - 2^{s_2} = 64 - 32 = 32$, $N(G; 16) = 2^{s_4} - 2^{s_3} = 64 - 64 = 0$, $N(G; 32) = 2^{s_5} - 2^{s_4} = 64 - 64 = 0$, and $N(G; 64) = 2^{s_6} - 2^{s_5} = 64 - 64 = 0$.

Example C. Let $G = \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_{27}$. Then G is a finite abelian 3-group with 3-factor type $\langle 3, 0, 1, 0, 0, 0 \rangle$. It is helpful to notice that the 3-factor type of this group is identical to the 2-factor type of the group in the previous example, so Theorem 1 shows that the values for each k_i, K, and s_i are the same as in Example B. Hence, our formulas are identical except that all powers of 2 are replaced with the corresponding power of 3. Thus G has 80 elements of order 3, 162 elements of order 9, and 486 elements of order 27.

Abelian groups of any (finite) order Having dealt with p-groups, let's now consider general finite abelian groups. Lagrange's theorem implies that if A is a group with $|A|$ *not* divisible by a prime p, then A has *no* elements of order p^i for any positive integer i. That is, $N(A; p^i) = 0$. FT implies that if G is an abelian group of order $n = p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r}$, where the p_i are distinct primes, then G is isomorphic to a group of the form $A_{p_1} \times A_{p_2} \times \cdots \times A_{p_r}$, where each A_{p_i} is an abelian p_i-group of order $p_i^{m_i}$. Thus we assume, without loss of generality, that $G = A_{p_1} \times A_{p_2} \times \cdots \times A_{p_r}$. The order of any element $(g_1, g_2, \ldots, g_r)$ in G is the least common multiple of the orders of the g_i in A_i, each of which is a non-negative power of p_i. Thus if $m = p_1^{b_1} p_2^{b_2} \cdots p_r^{b_r}$, then $N(G; m)$ is the number of ordered r-tuples $(g_1, g_2, \ldots, g_r)$ in G with each g_i of order $p_i^{b_i}$, which is clearly $\prod_{i=1}^{r} N(A_{p_i}; p_i^{b_i})$. (Note that it therefore follows that two finite abelian groups have identical order structure if and only if they have the same number of elements of each prime power order.) Theorem 1 tells us how to compute $N(A_{p_i}; p_i^{b_i})$ for each i, given the p_i-factor type of each A_{p_i}. We have, therefore, a method for counting the number of elements of G having order m. Instead of writing down a messy formula, we'll illustrate the method with two examples. The first example group is small enough to permit a determined reader to confirm the results "by hand." The second example, a group of order 5832, would be hard to check without a computer.

Example D. Let $G = \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_9$. Then $G \cong A_2 \times A_3$, where $A_2 = \mathbb{Z}_2 \times \mathbb{Z}_4$ and $A_3 = \mathbb{Z}_9$ are the 2 and 3 "parts" of G, respectively. Therefore, by the above discussion, $N(G; 2^a \cdot 3^b) = N(A_2; 2^a) \cdot N(A_3; 3^b)$. As in Example A, we know that $N(A_2; 1) = 1$, $N(A_2; 2) = 3$, and $N(A_2; 4) = 4$. Also, using the same method as in Examples A–C, A_3 is a 3-group having 3-factor type $\langle 0, 1\rangle$ (with $s_0 = 0$, $s_1 = 1$, and $s_2 = 2$), so $N(A_3; 1) = 1$, $N(A_3; 3) = 3^{s_1} - 3^{s_0} = 2$, and $N(A_3; 9) = 3^{s_2} - 3^{s_1} = 6$. Hence $N(G; 1) = 1 \cdot 1 = 1$, $N(G; 2) = 3 \cdot 1 = 3$, $N(G; 3) = 1 \cdot 2 = 2$, $N(G; 4) = 4 \cdot 1 = 4$, $N(G; 6) = 3 \cdot 2 = 6$, $N(G; 9) = 1 \cdot 6 = 6$, $N(G; 12) = 4 \cdot 2 = 8$, $N(G; 18) = 3 \cdot 6 = 18$, and $N(G; 36) = 4 \cdot 6 = 24$. (Notice that $\sum_{n|72} N(G; n) = 72$, the order of G. This *proves* nothing about the veracity of our results, but it is certainly comforting to the author.)

Example E. Now, Let $G = \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_{27}$. Then $G \cong A_2 \times A_3$, where $A_2 = \mathbb{Z}_2 \times \mathbb{Z}_4$ and $A_3 = \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_{27}$, so $N(G; 12) = N(G; 4 \cdot 3) = N(A_2; 4) \cdot N(A_3; 3)$. By Examples A and C, we know that $N(A_2; 4) = 4$ and $N(A_3; 3) = 80$, so $N(G; 12) = 320$. Similarly, $N(A_2; 2) = 3$ and $N(A_3; 9) = 162$, so $N(G; 6) = 3 \cdot 80 = 240$, $N(G; 18) = 3 \cdot 162 = 486$, and $N(G; 36) = 4 \cdot 162 = 648$, for example.

The main theorem We can now prove our main theorem, confirming our original proposition that two finite abelian groups are isomorphic if and only if they have identical order structure.

THEOREM 2. *If G and G' are finite abelian groups, then G and G' are isomorphic if and only if $N(G; p^i) = N(G'; p^i)$ for all prime p and all non-negative integers i.*

Proof. Clearly, if G is isomorphic to G', we must have $N(G; p^i) = N(G'; p^i)$ for all p and i. Conversely, suppose that $N(G; p^i) = N(G'; p^i)$ for all p and i. If G is a p-group, then the proof follows from the fact that the set of $N(G; p^i)$ completely determines the p-factor type of G. In a nutshell, the $N(G; p^i)$ determine the s_i (defined in Theorem 1), which determine the k_i, which determine the p-factor type. The interested reader is encouraged to see Theorem 3 for the explicit formula.

Now suppose G is not a p-group. As in the paragraph preceding Examples D and E, we have $N(A_p; p^i) = N(A'_p; p^i)$ for all p and i, where G is isomorphic to $\prod_p A_p$ and G' is isomorphic to $\prod_p A'_p$ (where, of course, all but finitely many A_p and A'_p are trivial). Then, by the result of the first paragraph of the proof, we see that A_p is isomorphic to A'_p for all primes p, so G must be isomorphic to G'.

Having shown that two finite abelian groups having identical order structure must be isomorphic, we will now show that both of the adjectives "finite" and "abelian" are necessary hypotheses for Theorem 2.

(Counter)Example F. Let $G = \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$ (the *elementary abelian group of order* 27). Every element of G other than the identity $(0,0,0)$ has order 3. Let G' be the multiplicative group of matrices of the form $\begin{pmatrix} 1 & a & b \\ 0 & 0 & c \\ 0 & 0 & 1 \end{pmatrix}$, where a, b, and c are elements of the ring $\mathbb{Z}_3$. It's easy to see that G' is a *nonabelian* group so can't be isomorphic to G. Notice, however, that $\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}^3 = \begin{pmatrix} 1 & 3a & 3ac-3b \\ 0 & 1 & 3c \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, so every non-identity element of G' has order 3. Thus G and G' have the same order structure. This shows that Theorem 2 fails when one of G or G' is nonabelian.

(Counter)Example G. We leave it to the reader to show that if G and G' are as in (Counter)Example F, then $G \times S_3$ and $G' \times S_3$ are both nonabelian groups of order 162 with identical order structure. However, $G \times S_3$ is not isomorphic to $G' \times S_3$, which shows that Theorem 2 cannot be generalized to include the case when both groups are nonabelian.

(Counter)Example H. To see that Theorem 2 fails for *infinite* abelian groups, let $G = \mathbb{Z}$ and $G' = \mathbb{Z} \times \mathbb{Z}$. G and G' both contain one element of order 1, and no other elements of finite order. In addition, G and G' both have countably many elements of infinite order, so G and G' have identical order structure, yet G is cyclic while G' is non-cyclic. Thus G and G' are not isomorphic.

Extension A useful consequence of this work is the following "converse" of the formula for $N(G; p^i)$ derived in Theorem 1, which gives a formula for determining the p-factor type of an abelian p-group (and therefore any finite abelian group) given the number of elements of each order.

THEOREM 3. *Suppose G is an abelian p-group, $N(G; p^i)$ is the number of elements of order p^i in G, and $[G; p^i] = \Sigma_{j=0}^{i} N(G; p^j)$. Then $G \cong \mathbb{Z}_p^{k_1} \times \mathbb{Z}_{p^2}^{k_2} \times \cdots \times \mathbb{Z}_{p^n}^{k_n}$ where*

$$k_i = \begin{cases} \log_p\left(\dfrac{[G; p^i]^2}{[G; p^{i+1}][G; p^{i-1}]} \right) & \text{if } 1 \le i \le n-1 \\[2ex] \log_p\left(\dfrac{[G; p^n]}{[G; p^{n-1}]} \right) & \text{if } i = n \end{cases}$$

Proof. Recall that $p^{s_i} = [G; p^i]$. Thus $s_i = \log_p([G; p^i])$. Referring to the definition of the s_i, one can show that $s_i - s_{i-1} = \Sigma_{j=i}^{n} k_j$, which implies

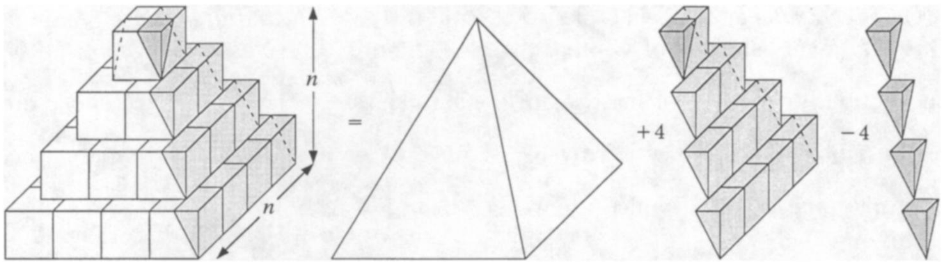
$$k_i = \begin{cases} 2s_i - s_{i+1} - s_{i-1} & \text{if } 1 \le i \le n-1 \\ s_n - s_{n-1} & \text{if } i = n, \end{cases}$$

and the theorem follows immediately.

REFERENCES

1. J. B. Fraleigh, *A First Course in Abstract Algebra*, 4th. edition, Addison-Wesley, Reading, MA, 1989.
2. J. A. Gallian, *Contemporary Abstract Algebra*, 3rd. edition, D. C. Heath, Lexington, MA, 1994.
3. I. N. Herstein, *Abstract Algebra*, 2nd. edition, Macmillan, New York, NY, 1990.

Proof Without Words: Sums of Squares



$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{3}n^2 \cdot n + 4 \cdot \frac{n(n+1)}{2} \cdot \frac{1}{4} - 4 \cdot n \cdot \frac{1}{12}$$

$$= \frac{1}{6}n(n+1)(2n+1)$$

—I. A. Sakmar
Department of Physics
University of South Florida
Tampa, FL 33620

On Groups of Order p^2

ACHILLEAS SINEFAKOPOULOS
20 Efesou Ioakim Street
Athens 17236
GREECE

Many texts on abstract algebra contain the result that a group whose order is a square of a prime must be abelian. The usual proofs make use of the class equation and occur in a section often labeled optional or placed late in the book (see, e.g., [**4**, p. 79]). Using the notation of [**4**], we shall combine well-known ideas to obtain a more elementary proof. First, two preliminary results:

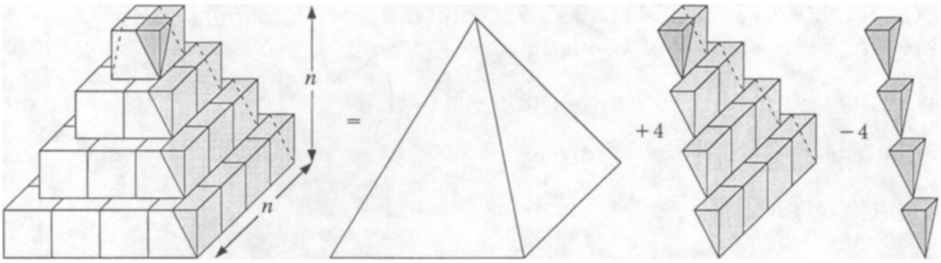
Lemma 1. *Let H and K be finite subgroups of a group G. Then*

$$|HK| = \frac{|H|\cdot|K|}{|H \cap K|}.$$

Proof. Let $J = H \cap K$ be a subgroup of K of index $n = |K|/|H \cap K|$. Also, let $K = Jk_1 \cup Jk_2 \cup \cdots \cup Jk_n$ be a decomposition of K into disjoint right cosets of J in K. It is clear that $Hk_i = Hk_j$ if and only if $Jk_i = Jk_j$; that is, if and only if $i = j$. Since $HJ = H$, these imply that HK is the disjoint union $Hk_1 \cup Hk_2 \cup \cdots \cup Hk_n$. Therefore $|HK| = |H|\cdot n = |H|\cdot|K|/|H \cap K|$, as desired.

Lemma 2. *If G is a group of order p^2, then every subgroup of G is normal.*

Proof Without Words: Sums of Squares



$$1^2+2^2+3^2+\cdots+n^2 = \frac{1}{3}n^2\cdot n + 4\cdot\frac{n(n+1)}{2}\cdot\frac{1}{4} - 4\cdot n\cdot\frac{1}{12}$$
$$= \frac{1}{6}n(n+1)(2n+1)$$

—I. A. SAKMAR
DEPARTMENT OF PHYSICS
UNIVERSITY OF SOUTH FLORIDA
TAMPA, FL 33620

On Groups of Order p^2

ACHILLEAS SINEFAKOPOULOS
20 Efesou Ioakim Street
Athens 17236
GREECE

Many texts on abstract algebra contain the result that a group whose order is a square of a prime must be abelian. The usual proofs make use of the class equation and occur in a section often labeled optional or placed late in the book (see, e.g., [**4**, p. 79]). Using the notation of [**4**], we shall combine well-known ideas to obtain a more elementary proof. First, two preliminary results:

LEMMA 1. *Let H and K be finite subgroups of a group G. Then*

$$|HK| = \frac{|H|\cdot|K|}{|H\cap K|}.$$

Proof. Let $J = H\cap K$ be a subgroup of K of index $n = |K|/|H\cap K|$. Also, let $K = Jk_1\cup Jk_2\cup\cdots\cup Jk_n$ be a decomposition of K into disjoint right cosets of J in K. It is clear that $Hk_i = Hk_j$ if and only if $Jk_i = Jk_j$; that is, if and only if $i=j$. Since $HJ = H$, these imply that HK is the disjoint union $Hk_1\cup Hk_2\cup\cdots\cup Hk_n$. Therefore $|HK| = |H|\cdot n = |H|\cdot|K|/|H\cap K|$, as desired.

LEMMA 2. *If G is a group of order p^2, then every subgroup of G is normal.*

Proof. Let H be a nontrivial proper subgroup of G. If $g \in G$ is such that $K = g^{-1}Hg \neq H$, then, by Lagrange's theorem, $H \cap K = 1$. Also, by LEMMA 1,

$$|KH| = |K| \cdot |H| / |H \cap K| = p^2 = |G|,$$

because $|K| = |H| = p$ and $|H \cap K| = 1$. Therefore $G = KH$, and so $g^{-1} = kh$ for some $k \in K$ and $h \in H$. But $k = g^{-1}h'g$ for some $h' \in H$, and so $g^{-1} = g^{-1}h'gh$. Hence $g = (h')^{-1}h^{-1} \in H$, a contradiction.

Now we can prove the main result:

THEOREM. *Every group G of order p^2 is abelian.*

Proof. If G is cyclic we have nothing to prove, so let G be a noncyclic group and use Lagrange's theorem to get two distinct subgroups H and K of G, each of order p. Then $H \cap K = 1$; by LEMMA 2, H and K are normal in G. Moreover, every element of H commutes with every element of K, because the normality of H and K in G implies that $hkh^{-1}k^{-1} \in H \cap K$ for all $h \in H$, $k \in K$. Also,

$$|HK| = \frac{|H| \cdot |K|}{|H \cap K|} = p^2 = |G|,$$

which yields $G = HK$. Since H and K are abelian, the last two results imply that G is abelian.

We conclude with two remarks:

1. The proof shows that G is either cyclic of order p^2 or isomorphic to the direct product of two cyclic groups, each of order p.
2. One could apply the following theorem instead of LEMMA 2.

THEOREM. *Suppose that G is finite and that p is the smallest prime divisor of $|G|$. If H is a subgroup of index p in G, then H is normal in G.*

There are many proofs of this theorem (see, e.g., **[1]**, **[4**, p. 75], **[5**, p. 166]); for an elementary proof, see **[2]** or **[3]**.

REFERENCES

1. H. G. Bray, A generalization of a result of Frobenius, *Amer. Math. Monthly* 76 (1969), 924–925.
2. James C. Owings, Jr., A combinatorial proof of a criterion for normality, *Amer. Math. Monthly* 92 (1985), 662–663.
3. Problem 5055, *Amer. Math. Monthly* 70 (1963), 1016. See also *Amer. Math. Monthly* 71 (1964), 328.
4. John S. Rose, *A Course on Group Theory*, Dover, New York, NY, 1994.
5. Micháel O. Seárcoid, A reordering of the Sylow theorems, *Amer. Math. Monthly* 94 (1987), 165–168.

The Relation Between the Root and Ratio Tests

DAVID CRUZ-URIBE, SFO
Trinity College
Hartford, CT 06106-3100

Two well-known tests for the convergence of a series $\Sigma_{n=0}^{\infty} a_n$, $a_n \neq 0$, are the ratio test and the root test:

Ratio Test: *If*

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

then the series Σa_n *converges.*

We call the values a_{n+1}/a_n, $n \geq 0$, the *consecutive ratios* of the series.

Root Test: *If*

$$\lim_{n\to\infty} |a_n|^{1/n} < 1 \tag{2}$$

then the series Σa_n *converges.*

(For proofs, see Krantz **[1]** or Rudin **[2]**.)

Since the limit in (1) is always greater than or equal to the limit in (2), the root test is stronger than the ratio test: there are cases in which the root test shows convergence but the ratio test does not. (In fact, the ratio test is a corollary of the root test: see Krantz **[1]**.)

We can illuminate the relationship between these two tests with a simple calculation:

$$\lim_{n\to\infty} |a_n|^{1/n} = \lim_{n\to\infty} \left| \frac{a_n}{a_{n-1}} \cdot \frac{a_{n-1}}{a_{n-2}} \cdots \frac{a_2}{a_1} \cdot \frac{a_1}{a_0} \cdot a_0 \right|^{1/n} = \lim_{n\to\infty} \left| \frac{a_n}{a_{n-1}} \cdots \frac{a_1}{a_0} \right|^{1/n}.$$

(The last equality holds since $a_0^{1/n} \to 1$ as $n \to \infty$.) The right-hand side is the limit of the geometric means of the first n consecutive ratios of the series. In other words, while the ratio test depends on the behavior (in the limit) of each consecutive ratio, the root test only considers the average behavior of these ratios. Clearly, if all the consecutive ratios get small then their average value will get small as well. The converse is false, which is why the root test is stronger. Thus, for example, the ratio test fails on the rearranged geometric series

$$1/2 + 1 + 1/8 + 1/4 + 1/32 + 1/16 + \cdots, \tag{3}$$

since the consecutive ratios alternate in value between 2 and $1/8$. However, the geometric mean of the first $2n$ consecutive ratios is

$$\left| 2^n \cdot \frac{1}{8^n} \right|^{1/2n} = 1/2,$$

so the root test shows that the series converges.

Interpreting the root test in terms of averages suggests substituting another mean for the geometric mean. By the arithmetic-geometric mean inequality, the arithmetic mean of n consecutive ratios of a series is always larger than their geometric mean. This yields a new convergence test:

Arithmetic Mean Test: *If*

$$\lim_{n\to\infty} \frac{1}{n}\left(\frac{a_n}{a_{n-1}} + \frac{a_{n-1}}{a_{n-2}} + \cdots + \frac{a_1}{a_0}\right) < 1$$

then the series Σa_n *converges.*

This test is stronger than the ratio test but weaker than the root test. However, in some cases it may be easier to compute the arithmetic mean of the consecutive ratios than it is to compute their geometric mean. For example, consider the series Σa_n, where the a_n's are defined inductively by

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

By l'Hôpital's rule,

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

so the ratio test fails. To apply the root test we would have to evaluate the limit

$$\lim_{n\to\infty}\left|\prod_{k=1}^{n}\frac{\log(1+1/(k+1))}{\log(k+1)\log(k+2)}\right|^{1/n}.$$

The arithmetic mean test, however, requires only the following calculation:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\left|\frac{\log(1+1/(k+1))}{\log(k+1)\log(k+2)}\right| &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{\log(k+1)} - \frac{1}{\log(k+2)}\right)\\ &= \lim_{n\to\infty}\frac{1}{n}\left(\frac{1}{\log 2} - \frac{1}{\log(n+2)}\right)\\ &= 0.\end{aligned}$$

Hence the series converges.

An example of a series for which the arithmetic mean test fails is series (3) above: In this case the arithmetic mean of the consecutive ratios converges to 17/16. As a exercise for the reader we leave the problem of determining the values of α, $0 < \alpha < 1$, for which the convergence of a similar rearrangement of $\Sigma \alpha^n$ can be shown using the arithmetic mean test.

REFERENCES

1. S. Krantz, *Real Analysis and Foundations*, CRC Press, Boca Raton, FL, 1991.
2. W. Rudin, *Principles of Mathematical Analysis*, 3rd Ed., McGraw Hill, New York, NY, 1976.

Another Proof of the Fundamental Theorem of Algebra

DANIEL J. VELLEMAN
Amherst College
Amherst, MA 01002

Recently, while teaching a course in Complex Analysis, I amused myself (and, I hope, my students) by having my students work out the details of several different proofs of the Fundamental Theorem of Algebra. It is interesting that the theorem follows from a number of important principles of Complex Analysis. For example, there are proofs based on Liouville's Theorem ([**6**], p. 138), Rouché's Theorem ([**8**], p. 295), the Maximum Principle ([**6**], p. 152), Picard's Theorem ([**3**]), and the Cauchy Integral Theorem ([**4**]). Even the elementary proof in [**5**], [**7**], and [**10**] (whose idea can be traced back to d'Alembert; see [**9**], pp. 195–198) is based on a fundamental property that distinguishes the complex numbers from the real numbers, namely, that every complex number has at least one nth root for every positive integer n.

When my course reached the subject of power series, it occurred to me that it might be possible to prove the Fundamental Theorem of Algebra from the fact that entire functions can always be represented by power series. Unable to find such a proof in the literature, I came up with the following proof[1]:

THEOREM. *Suppose* $P(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_n z^n$ *is a polynomial of degree* $n > 0$ *with complex coefficients. Then for some complex number* z, $P(z) = 0$.

Proof. Suppose not. Then, the function $f(z) = 1/P(z)$ is entire, and therefore can be represented by a power series throughout the complex plane; i.e., there are complex numbers $b_0, b_1, b_2, \dots$ such that for all complex numbers z,

$$f(z) = b_0 + b_1 z + b_2 z^2 + \dots .$$

The key to the proof is the following observation:

LEMMA. *There are positive real numbers* c *and* r *such that for infinitely many* k, $|b_k| > cr^k$.

To see why the theorem follows, consider substituting $1/r$ for z in the power series. For infinitely many k we have

$$|b_k z^k| = |b_k|/r^k > c > 0,$$

so the series diverges, since the terms are not approaching 0. But this is a contradiction, since the series was supposed to converge to $f(z)$ for all z.

Proof of the Lemma: Since $f(z) = 1/P(z)$, we have

$$1 = P(z)f(z) = \left(a_0 + a_1 z + a_2 z^2 + \dots + a_n z^n\right)\left(b_0 + b_1 z + b_2 z^2 + \dots\right).$$

---

[1] I have recently learned that a similar proof was discovered by Alexander Abian and James Wilson; see [**2**]. Abian has also given a proof of the Fundamental Theorem of Algebra based on Laurent series rather than Taylor series; see [**1**]. I would like to thank the referee for pointing out references [**1**], [**9**], and [**10**] to me.

Multiplying out the right-hand side and equating it to the constant function 1, we find that $a_0 b_0 = 1$, so $a_0 \neq 0$ and $b_0 \neq 0$, and from the coefficient of z^{k+n},

$$(*) \qquad a_0 b_{k+n} + a_1 b_{k+n-1} + \dots + a_n b_k = 0 \text{ for all } k \geq 0.$$

Since $b_0 \neq 0$, we can choose c so that $0 < c < |b_0|$. Since $a_n \neq 0$, we can choose r so that $r > 0$ and

$$|a_0| r^n + |a_1| r^{n-1} + \dots + |a_{n-1}| r \leq |a_n|.$$

(Note that the left-hand side of this inequality is 0 when $r = 0$, so by continuity the inequality will be true for any sufficiently small positive r. In fact, the inequality is easily seen to be true if $r \leq \min\{1, |a_n|/(|a_0| + |a_1| + \dots + |a_{n-1}|)\}$.)

We already know that $|b_0| > c = cr^0$. Now suppose $|b_k| > cr^k$. We will show that for some i between 1 and n, $|b_{k+i}| > cr^{k+i}$, which will establish the lemma.

Suppose not. Then, by equation $(*)$, we have

$$\begin{aligned}
|b_k| &= \frac{|a_0 b_{k+n} + a_1 b_{k+n-1} + \dots + a_{n-1} b_{k+1}|}{|a_n|} \\
&\leq \frac{|a_0||b_{k+n}| + |a_1||b_{k+n-1}| + \dots + |a_{n-1}||b_{k+1}|}{|a_n|} \\
&\leq \frac{|a_0| cr^{k+n} + |a_1| cr^{k+n-1} + \dots + |a_{n-1}| cr^{k+1}}{|a_n|} \\
&= cr^k \left(\frac{|a_0| r^n + |a_1| r^{n-1} + \dots + |a_{n-1}| r}{|a_n|} \right) \\
&\leq cr^k,
\end{aligned}$$

contradicting the fact that $|b_k| > cr^k$. This completes the proof of the lemma.

Note that the proof above actually gives a bound on the size of the smallest root of $P(z)$. Since the radius of convergence of the power series for $f(z)$ is at most $1/r$, where r is chosen as in the lemma, there must be some z with $|z| \leq 1/r$ such that $P(z) = 0$. It turns out that this is the same as the bound given by the proof of the Fundamental Theorem via Rouché's Theorem.

REFERENCES

1. A. Abian, A new proof of the fundamental theorem of algebra, *Caribb. J. Math.* 5 (1986), pp. 9–12.
2. Alexander Abian and James A. Wilson, A theorem equivalent to the fundamental theorem of algebra, *Nieuw Arch. Wisk.* (4) 8 (1990), pp. 47–48.
3. R. P. Boas, Jr., A proof of the fundamental theorem of algebra, *Amer. Math. Monthly* 42 (1935), pp. 501–502.
4. R. P. Boas, Jr., Yet another proof of the fundamental theorem of algebra, *Amer. Math. Monthly* 71 (1964), p. 180.
5. Charles Fefferman, An easy proof of the fundamental theorem of algebra, *Amer. Math. Monthly* 74 (1967), pp. 854–855.
6. Norman Levinson and Raymond M. Redheffer, *Complex Variables*, Holden Day, San Francisco, 1970.
7. R. M. Redheffer, What! Another note just on the fundamental theorem of algebra?, *Amer. Math. Monthly* 71 (1964), pp. 180–185.
8. E. B. Saff and A. D. Snider, *Fundamentals of Complex Analysis for Mathematics, Science, and Engineering, 2nd edition*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.
9. John Stillwell, *Mathematics and its History*, Springer-Verlag, New York, 1989.
10. Frode Terkelsen, The fundamental theorem of algebra, *Amer. Math. Monthly* 83 (1976), p. 647.

On the Differentiability of $\int_0^x \sin(1/t)\,dt$ and $\int_0^x \sin(\ln t)\,dt$

PAUL FISHBACK, NICHOLAS CEGLAREK, AND TOBIAS MOLESKI
Grand Valley State University
Allendale, MI 49401

Introduction Several recent articles ([**2**], [**3**], [**4**]) have focused on an interesting class of functions that do not satisfy all the hypotheses of the first fundamental theorem of calculus. A function f in this class has the following properties:

(a) f is defined and bounded on the interval $[0, 1]$;
(b) f is continuous on $(0, 1]$;
(c) $\lim_{x \to 0^+} f(x)$ does not exist.

It follows from (a) and (b) that the Riemann integral $\int_0^x f(t)\,dt$ exists for every $x \in [0, 1]$. However, whether this integral is right-differentiable at the origin, i.e., whether

$$\lim_{x \to 0^+} \frac{1}{x} \int_0^x f(t)\,dt$$

exists, cannot be answered from (a)–(c) alone. The articles of Ricci ([**4**]) and Klippert ([**3**]) consider, respectively, the functions

$$f_1(x) = \begin{cases} \sin(1/x) & \text{if } 0 < x \leq 1 \\ 1 & \text{if } x = 0 \end{cases} \quad \text{and} \quad f_2(x) = \begin{cases} \sin(\ln(x)) & \text{if } 0 < x \leq 1 \\ 1 & \text{if } x = 0 \end{cases}.$$

These authors show that

$$\lim_{x \to 0^+} \frac{1}{x} \int_0^x f_1(t)\,dt = 0$$

but

$$\lim_{x \to 0^+} \frac{1}{x} \int_0^x f_2(t)\,dt \text{ does not exist.}$$

What differences between f_1 and f_2 cause the different outcomes above? Klippert conjectured that the varying degrees of oscillation, as evidenced by the different spacing between the zeros of the two functions, explains the difference. He formulates a conjecture to this effect, which implies that if a function f in the above class possesses infinitely many zeros and if these zeros satisfy a certain placement condition, then $\lim_{x \to 0} \int_0^x f(t)\,dt/x$ exists. Several examples support this conjecture, including f_1, f_2, and the oscillating sawtooth functions discussed in [**2**].

It may be surprising, then, that Klippert's conjecture is false. For a function f in the above-mentioned class, *no* condition phrased solely in terms of the location of the sequence of zeros of f is sufficient to imply that $\lim_{x \to 0} \int_0^x f(t)\,dt/x$ exists.

A counterexample To see why this is so, let $\{\alpha_n\}$ be any sequence contained in $(0, 1]$ such that $\alpha_1 = 1$, $\alpha_{n+1} < \alpha_n$, and $\lim_{n\to\infty} \alpha_n = 0$. Then one may construct a subsequence $\{\alpha_{n_k}\}$ such that

$$\text{(i)}\quad \alpha_{n_1} = \alpha_1; \quad (ii)\quad \alpha_{n_{k+1}} \le \alpha_{n_k} \cdot \frac{1}{k}.$$

For each positive integer k, let ψ_k be any function satisfying the following conditions:

(1) ψ_k is continuous on $[0, 1]$;

(2) $\psi_k(x) \neq 0 \Leftrightarrow x \in (\alpha_{n_{k+1}}, \alpha_{n_k})$ and $x \notin \{\alpha_n\}$;

(3) $|\psi_k(x)| \le 1$, with $\psi_k(x) = (-1)^k$ for at least one x;

(4) $\left|\int_{\alpha_{n_{k+1}}}^{\alpha_{n_k}} (\psi_k(x) - (-1)^k)\, dx\right| < \alpha_{n_k}/k$.

Such a function can be constructed to be continuous and piecewise-linear. For example, if k is even then ψ_k might have a graph like FIGURE 1.

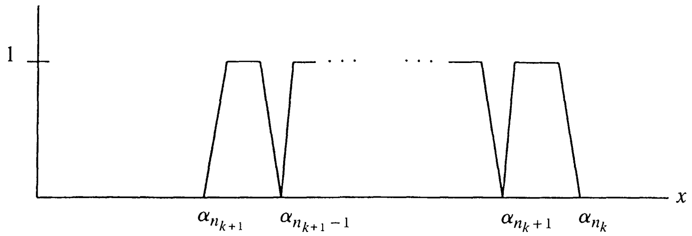


FIGURE 1
Constructing a counterexample

Now define

$$f(x) = \begin{cases} 1 & \text{if } x = 0 \\ \psi_k(x) & \text{if } \alpha_{n_{k+1}} \le x \le \alpha_{n_k}. \end{cases}$$

Note that f satisfies conditions (a)–(c) above, and has $\{\alpha_n\}$ as its set of zeros on $(0, 1]$.

The boundedness of the integrand, the properties of the subsequence $\{\alpha_{n_k}\}$, and condition (4) imply the following inequalities:

$$\left|\left(\frac{1}{\alpha_{n_k}} \int_0^{\alpha_{n_k}} f(t)\, dt\right) - (-1)^k\right|$$
$$= \left|\frac{1}{\alpha_{n_k}} \int_0^{\alpha_{n_k}} \left(f(t) - (-1)^k\right) dt\right|$$
$$\le \left|\frac{1}{\alpha_{n_k}} \int_0^{\alpha_{n_{k+1}}} \left(f(t) - (-1)^k\right) dt\right| + \left|\frac{1}{\alpha_{n_k}} \int_{\alpha_{n_{k+1}}}^{\alpha_{n_k}} \left(f(t) - (-1)^k\right) dt\right|$$
$$\le \frac{2}{k} + \frac{1}{k} = \frac{3}{k}.$$

It follows that $\lim_{k\to\infty} 1/\alpha_{n_k} \int_0^{\alpha_{n_k}} f(t)\, dt$ does not exist.

The role of the derivative The preceding argument shows that placement of zeros alone is not a determining factor. But it does not explain why $\lim_{x\to 0} \int_0^x \sin g(t)\, dt/x$ exists for $g(t) = 1/t$ and not for $g(t) = \ln t$. The difference can be explained in terms of these two functions' derivatives. Both derivatives are unbounded as $t \downarrow 0$, but one grows more quickly in magnitude than the other. The following theorem tells how quickly the magnitude of g' must grow in order for $\lim_{x\to 0} \int_0^x \sin g(t)\, dt/x$ to exist.

THEOREM 1. *Let g be a continuously differentiable function on* $(0,1)$, *such that* g' *is monotone and* $g(t)$ *either increases to* $+\infty$ *or decreases to* $-\infty$ *as* $t \downarrow 0$. *Suppose, further, that*

$$\lim_{t \to 0^+} t g'(t) = L, \quad \text{where } -\infty \le L \le \infty.$$

Then

$$\lim_{x \to 0^+} \frac{1}{x} \int_0^x \sin g(t)\, dt$$

exists if and only if $L = \pm\infty$.

In the proof we will use the following straightforward corollary to Bonnet's form of the second mean value theorem for integrals (see, e.g., [**1**, pp. 311, 328]):

PROPOSITION. *Suppose that m is a monotone function on $[y, x]$ and that p is bounded and integrable on $[y, x]$. Then there exists ζ in $[y, x]$ such that*

$$\int_y^x p(t) m(t)\, dt = m(y) \int_y^\zeta p(t)\, dt + m(x) \int_\zeta^x p(t)\, dt.$$

Proof of the theorem. Assume first that $L = \pm\infty$. For x sufficiently close to zero and for $0 < y < x$, one may write

$$\frac{1}{x} \int_y^x \sin g(t)\, dt = \frac{1}{x} \int_y^x \frac{1}{g'(t)} \sin g(t) g'(t)\, dt.$$

Because $1/g'$ is monotone on $[y, x]$ and $\sin(g) \cdot g'$ is continuous there as well, the corollary asserts that for some ζ in the interval $[y, x]$,

$$\begin{aligned}
&\frac{1}{x} \int_y^x \sin g(t)\, dt \\
&\quad = \frac{1}{x}\left[\frac{1}{g'(y)} \int_y^\zeta \sin g(t) g'(t)\, dt + \frac{1}{g'(x)} \int_\zeta^x \sin(g(t)) g'(t)\, dt\right] \\
&\quad = \frac{1}{x}\left[\frac{1}{g'(y)}(\cos g(y) - \cos g(\zeta)) - \frac{1}{g'(x)}(\cos g(x) - \cos g(\zeta))\right].
\end{aligned}$$

If we let $y \downarrow 0$ and then $x \downarrow 0$, we obtain

$$\lim_{x \to 0^+} \frac{1}{x} \int_0^x \sin g(t)\, dt = 0,$$

as desired.

To prove the converse, assume that L is finite and, without loss of generality, that $g(t)$ increases to ∞ as $t \downarrow 0$. Suppose that $0 < y < x$. Integrating by parts and rearranging terms yields

$$\begin{aligned}
\int_y^x \sin g(t)\, dt &= x \sin g(x) - y \sin g(y) \\
&\qquad + \int_y^x (L - t g'(t)) \cos g(t)\, dt - L \int_y^x \cos g(t)\, dt.
\end{aligned}$$

Integrating the last integral by parts again and rearranging terms, we obtain

$$\int_y^x \sin g(t)\,dt = x\sin g(x) - y\sin g(y)$$

$$+\int_y^x (L - tg'(t))\cos g(t)\,dt - Lx\cos g(x) + Ly\cos g(y)$$

$$+L\int_y^x (L - tg'(g))\sin g(t)\,dt - L^2\int_y^x \sin g(t)\,dt.$$

Equivalently,

$$\frac{1}{x}\int_y^x \sin g(t)\,dt - \frac{\sin g(x) - L\cos g(x)}{1+L^2}$$

$$= -\frac{y\sin g(y)}{(1+L^2)x} + \frac{Ly\cos g(y)}{(1+L^2)x}$$

$$+\frac{1}{(1+L^2)x}\int_y^x (L - tg'(t))\cos g(t)\,dt$$

$$+\frac{L}{(1+L^2)x}\int_y^x (L - tg'(t))\sin g(t)\,dt.$$

Now choose $\delta > 0$ small enough so that $|L - tg'(t)| < 1/(2(1+|L|))$ whenever $0 < t < \delta$. Then, for $x < \delta$, the sum of the last two integrals is bounded in absolute value by $1/(2(1+L^2))$. Hence

$$\left|\frac{1}{x}\int_y^x \sin g(t)\,dt - \frac{\sin g(x) - L\cos g(x)}{1+L^2}\right|$$

$$\le \frac{1}{1+L^2}\left[\left|\frac{y\sin g(y)}{x}\right| + \left|\frac{Ly\cos g(y)}{x}\right| + \frac{1}{2}\right].$$

Letting $y \downarrow 0$ yields

$$\left|\frac{1}{x}\int_0^x \sin g(t)\,dt - \frac{\sin g(x) - L\cos g(x)}{1+L^2}\right| \le \frac{1}{2}\frac{1}{1+L^2},$$

whenever $0 < x < \delta$.

Since g increases continuously to ∞ as $t \downarrow 0$, there exists a sequence $\{x_n\} \subseteq (0,1)$ that converges to zero and that satisfies $g(x_n) = (2n+1)\pi/2$ for all sufficiently large n. For such n,

$$\left|\frac{1}{x_n}\int_0^{x_n} \sin g(t)\,dt - \frac{(-1)^n}{1+L^2}\right| \le \frac{1}{2(1+L^2)}.$$

This implies that $\lim_{n\to\infty} \int_0^{x_n} \sin g(t)\,dt/x_n$ does not exist, and completes the proof.

Notes The use of the second mean value theorem for integrals in the preceding proof is inspired by van der Corput's lemma, a classical result in the theory of trigonometric series. This lemma, which estimates oscillatory integrals, is a useful tool for constructing a uniformly convergent trigonometric series that diverges absolutely (cf. [**5**, p. 197]).

Theorem 1 only scratches the surface of the class of functions described at the beginning of this paper. For example, the theorem clearly holds when the sine is replaced by the cosine, but one might ask what happens when the sine is replaced by any periodic function or, indeed, by *any* continuous, bounded function. Such questions could prompt interesting undergraduate research.

Acknowledgment. The two student authors (Ceglarek and Moleski) were supported by the Grand Valley State University Summer Undergraduate Research Program and the Council on Undergraduate Research.

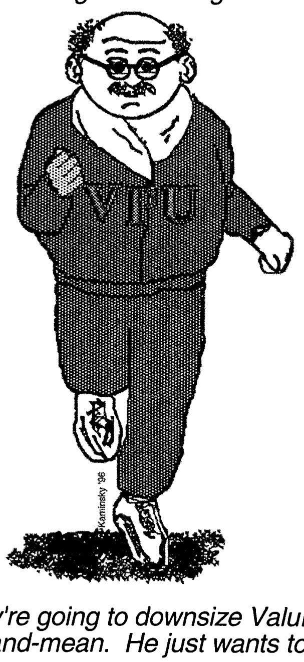
REFERENCES

1. P. Franklin, *A Treatise on Advanced Calculus*, Dover Publications, New York, NY, 1964.
2. F. Hartmann and D. Sprows, Oscillating sawtooth functions, this MAGAZINE 68 (1995), 211–212.
3. J. Klippert, On the right-hand derivative of a certain integral function, *Amer. Math. Monthly* 98 (1991), 751–752.
4. S. Ricci, Solution to proposed problem 1112, this MAGAZINE 55 (1982), 48.
5. A. Zygmund, *Trigonometric Series*, Cambridge Mathematical Society, Cambridge, UK, 1988.

Professor Fogelfroe

Professor F. Fogelfroe is Professor of Mathematics at ValuPak^{tM} University, in Margo's Forehead, Minnesota.

Why is ValuPak™ University math Prof Freddie Fogelfroe doing roadwork?



He heard they're going to downsize ValuPak™ and make it lean-and-mean. He just wants to be ready.

—KENNETH KAMINSKY
AUGSBURG COLLEGE
MINNEAPOLIS, MN 55454

Theorem 1 only scratches the surface of the class of functions described at the beginning of this paper. For example, the theorem clearly holds when the sine is replaced by the cosine, but one might ask what happens when the sine is replaced by any periodic function or, indeed, by *any* continuous, bounded function. Such questions could prompt interesting undergraduate research.

Acknowledgment. The two student authors (Ceglarek and Moleski) were supported by the Grand Valley State University Summer Undergraduate Research Program and the Council on Undergraduate Research.

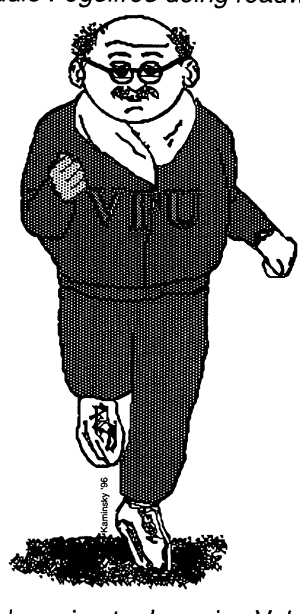
REFERENCES

1. P. Franklin, *A Treatise on Advanced Calculus*, Dover Publications, New York, NY, 1964.
2. F. Hartmann and D. Sprows, Oscillating sawtooth functions, this MAGAZINE 68 (1995), 211–212.
3. J. Klippert, On the right-hand derivative of a certain integral function, *Amer. Math. Monthly* 98 (1991), 751–752.
4. S. Ricci, Solution to proposed problem 1112, this MAGAZINE 55 (1982), 48.
5. A. Zygmund, *Trigonometric Series*, Cambridge Mathematical Society, Cambridge, UK, 1988.

Professor Fogelfroe

Professor F. Fogelfroe is Professor of Mathematics at ValuPak[tM] University, in Margo's Forehead, Minnesota.

Why is ValuPak™ University math
Prof Freddie Fogelfroe doing roadwork?



He heard they're going to downsize ValuPak™ and make it lean-and-mean. He just wants to be ready.

—KENNETH KAMINSKY
AUGSBURG COLLEGE
MINNEAPOLIS, MN 55454

PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

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

Proposals

To be considered for publication, solutions should be received by November 1, 1997.

1524. *Proposed by Ted Zerger, Kansas Wesleyan University, Salina, Kansas.*

Given ΔABC, let A', B', C' be the points on the sides BC, CA, AB, respectively, such that

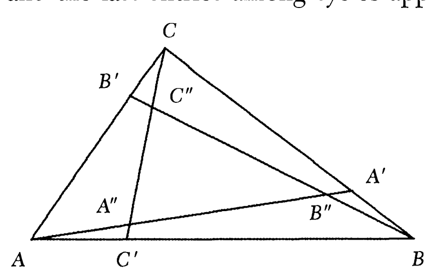
$$\frac{BA'}{BC} = \frac{CB'}{AC} = \frac{AC'}{AB} = t, \qquad 0 < t < \frac{1}{2}.$$

Let A'', B'', C'' be the points of intersection of AA' and CC', BB' and AA', CC' and BB', respectively. Prove that the ratios

$$AA' : A''B'' : B''A' = BB'' : B''C'' : C''B' = CC'' : C''A'' : A''C' = t : 1 - 2t : t^2.$$

1525. *Proposed by Emeric Deutsch, Polytechnic University, Brooklyn, New York.*

Define a mapping $f\colon S_n \to S_n$ as follows. Given a permutation π of $\{1, 2, \ldots, n\}$, express it in cycle form, including any fixed elements, such that the smallest entry of each cycle appears last, and the last entries among cycles appear in increasing order.



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

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

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

The permutation $f(\pi)$ is then defined by removing all inner parentheses and interpreting the result as the one-line representation of $f(\pi)$. In other words, the ith entry of the line is $f(\pi)(i)$. (For example, expressed in this cycle form, $f((4,6,1)(2)(5,3)) = (4,2,6,3,1)(5)$.) Characterize those π fixed by f, and determine their cardinality.

1526. *Proposed by Wu Wei Chao, He Nan Normal University, Xin Xiang City, He Nan Province, China.*

Let p be an odd prime number, and let a and b be positive integers with $1 < a < p$. Find the number of ordered pairs (x, y) of positive integers such that p divides $x + ay$ and $x + y < bp$.

1527. *Proposed by J. C. Binz, University of Bern, Bern, Switzerland.*

For n a nonnegative integer, let $A_n = (a_{i,k})_{0 \le i,k \le n}$ be the $(n+1) \times (n+1)$ matrix defined by $a_{0,k} = a_{i,0} = 1$ and

$$a_{i,k} = a_{i,k-1} + ima_{i-1,k-1} \qquad (i, k \ge 1).$$

Show that A_n is symmetric, and evaluate $a_{i,k}$.

1528. *Proposed by Florin S. Pirvănescu, Slatina, Romania.*

Let M be a point in the interior of convex polygon $A_1 A_2 \dots A_n$. If d_k is the distance from M to $A_k A_{k+1} (A_{n+1} = A_1)$, show that

$$(d_1 + d_2)(d_2 + d_3) \cdots (d_n + d_1) \le 2^n \cos^n \frac{\pi}{n} \cdot MA_1 \cdot MA_2 \cdot \cdots \cdot MA_n,$$

and determine when equality holds.

Quickies

Answers to the Quickies are on page 229.

Q865. *Proposed by Michael Andreoli, Miami Dade Community College (North Campus), Miami, Florida.*

Balls numbered 1 through n are placed in an urn and drawn out randomly without replacement. Before each draw a player is allowed to guess the number to be drawn, and is told only whether the guess is right or wrong. If the player guesses 1 until it is correct, then switches to 2 until it is correct or all balls are drawn, then switches to 3 until it is correct or all balls are drawn, and so forth until all balls are drawn, what is the expected number of correct guesses? What happens as $n \to \infty$?

Q866. *Proposed by Larry Hoehn, Austin Peay State University, Clarksville, Tennessee.*

If a, b, and c are positive real numbers, find the greatest lower bound and least upper bound of

$$\frac{a^2 + b^2 + ac}{\sqrt{(a^2 + b^2)(a^2 + b^2 + c^2 + 2ac)}}.$$

Q867. *Proposed by David Callan, University of Wisconsin, Madison, Wisconsin.*

Show that the number of $(n-1)$-element subsets of $\{1, 2, \dots, kn\}$ whose sum is divisible by n is given by $\binom{kn}{n-1}/n$.

(These are the Catalan numbers when $k = 2$. The corresponding problem for n-element subsets with n prime and $k = 2$ was Problem 6 on the 1995 IMO.)

Solutions

An Exponential Inequality June 1996

1499. *Proposed by Wu Wei Chao, He Nan Normal University, Xin Xiang City, He Nan Province, China.*

For positive numbers x and y, prove that $x^x + y^y \geq x^y + y^x$, with equality if and only if $x = y$.

Solution by Kee-Wai Lau, Hong Kong.

If $x = y$, it is clear that both sides of the proposed inequality are equal. So we may now assume without loss of generality that $x > y$.

If $y \leq 1 \leq x$, then $x^x \geq x^y$ and $y^y \geq y^x$, with at least one inequality strict, so $x^x + y^y > x^y + y^x$.

Now let $f(t) = t^x - t^y$ for $y \leq t \leq x$. We have $f'(t) = xt^{x-1} - yt^{y-1}$. We want to show that $f'(t) > 0$, which is clear in the case $1 < y < x$. Because $f'(t) > 0$ if and only if $t^{x-y} > y/x$, in the case $y < x < 1$ it suffices to show that $y^{x-y} > y/x$, or $(x - y)\ln y - \ln y + \ln x > 0$. Set $g(x) = (x-y)\ln y - \ln y + \ln x$. Now $g(y) = 0$, $g(1) = -y \ln y > 0$, and $g''(x) = -1/x^2 < 0$. If $g(x) \leq 0$ at some point in $(y, 1)$, the derivative of g would change from non-positive to positive at a pair of points on the interval, and g'' would be positive at some point in the interval. We conclude that $g(x) > 0$, hence $f'(t) > 0$. In both of these two cases, f is strictly increasing, so $f(x) > f(y)$ or $x^x + y^y > x^y + y^x$.

Comments. Several readers listed among the solvers below pointed out that this problem was proposed by Weixuan Li and Edward T. H. Wang as problem 303 in *The College Mathematics Journal* 16 (1985), p. 224, and proposed by M. Laub as problem E3116 in *The American Mathematical Monthly* 92 (1985), p. 666, with solutions appearing in *The College Mathematics Journal* 18 (1987), pp. 164–165, and *The American Mathematical Monthly* 97 (1990), pp. 65–67, respectively. Problem E3116 included the generalization to

$$x_1^{x_1} + x_2^{x_2} + \cdots + x_n^{x_n} \geq x_1^{y_1} + x_2^{y_2} + \cdots + x_n^{y_n},$$

where $(y_1, \ldots, y_n)$ is any permutation of $(x_1, \ldots, x_n)$.

Many of the incorrect solutions erred in multiplying two inequalities that could involve negative numbers.

Also solved by H. Azad and Asghar Qadir and A. B. Thaheem (Saudi Arabia), Evgenii S. Freidkin, Murray S. Klamkin (Canada), David E. Manes, Phil McCartney, Can A. Minh (student), Stephen Noltie, Heinz-Jürgen Seiffert (Germany), Volkhard Schindler (Germany), Southern Oregon State Problem Solvers Group, TAMUK Problem Solvers, Michael Vowe (Switzerland), Edward T. H. Wang (Canada), and the proposer. There were sixteen incorrect solutions.

A Limit Point of Triangles June 1996

1500. *Proposed by Saul Stahl, University of Kansas, Lawrence, Kansas.*

Let r be a positive real number and let $\Delta A_0 B_0 C_0$ be equilateral. For each $n \geq 0$ let A_{n+1} and B_{n+1} divide the sides $A_n B_n$ and $A_n C_n$, respectively, in the internal ratio $r:1$, and set $C_{n+1} = A_n$. If $P = \lim_{n \to \infty} \Delta A_n B_n C_n$, prove that the measures of $\angle B_0 P C_0$, $\angle C_0 P A_0$, and $\angle A_0 P B_0$ form an arithmetic progression.

Composite of solutions due to Volkhard Schindler, Berlin, Germany, and Ted Zerger, Kansas Wesleyan University, Salina, Kansas.

Because the measures of $\angle B_0PC_0$, $\angle C_0PA_0$, and $\angle A_0PB_0$ sum to 360°, it suffices to show that $\angle C_0PA_0 = 120°$. Representing A_0, B_0, and C_0 by complex numbers, there is no loss of generality in assuming $A_0 = 0$, $B_0 = 1$, and $C_0 = \exp(\pi i/3)$. Set $s = r/(r+1) < 1$. Then it is easy to see that, for $n > 0$,

$$A_n = A_{n-1} + s^n \exp(2\pi i(n-1)/3).$$

From the periodicity of the complex exponential function, we obtain

$$\begin{aligned} P = \lim_{n\to\infty} A_n &= (s^1 + s^4 + s^7 + \cdots) + (s^2 + s^5 + s^8 + \cdots)\exp(2\pi i/3) \\ &\quad + (s^3 + s^6 + s^9 + \cdots)\exp(4\pi i/3) \\ &= \frac{s}{1-s^3} + \frac{s^2}{1-s^3}\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) + \frac{s^3}{1-s^3}\left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) \\ &= \frac{2s+s^2}{2(1+s+s^2)} + \frac{\sqrt{3}\,s^2}{2(1+s+s^2)}i. \end{aligned}$$

We can now determine the distances $A_0P = s/\sqrt{1+s+s^2}$ and $C_0P = 1/\sqrt{1+s+s^2}$. Using the law of cosines, we find that $\cos\angle C_0PA_0 = -1/2$ or $\angle C_0PA_0 = 120°$, completing the proof.

Also solved by J. C. Binz (Switzerland), Milton P. Eisner, Michael Vowe (Switzerland), Paul J. Zwier, and the proposer.

Palindromic Numbers in Arithmetic Progressions **June 1996**

1501. *Proposed by Matúš Harminc and Roman Soták, Šafárik University, Košice, Slovakia.*

Which nonconstant arithmetic progressions of positive integers, excluding those for which every term is a multiple of 10, contain infinitely many palindromic numbers? (A palindromic number is unchanged when the order of its digits is reversed, for example 121 or 1331.)

Solution by Jack C. Abad, University of San Francisco, San Francisco, California.

We prove that all nonconstant arithmetic progressions, excluding those for which every term is a multiple of 10, contain infinitely many palindromic numbers.

Every such arithmetic progression contains an arithmetic progression of the form

$$\{a, a + b10^d, a + 2\cdot b10^d, \ldots\},$$

where 10 does not divide a, $\gcd(b, 10) = 1$, and $d > 0$. Thus, it suffices to prove the claim for such arithmetic progressions.

With this assumption, let r be the number formed by the last d digits of a, let s be the number formed by these d digits written in reverse order, and let $\phi(b)$ be the number of integers less than and relatively prime to b. For each positive integer k, the number

$$s\cdot 10^{(k-1)\phi(b)+d+1} + 10^{(k-1)\phi(b)+d} + 10^{(k-2)\phi(b)+d} + \cdots + 10^{\phi(b)+d} + 10^d + r$$

is clearly palindromic. By Euler's theorem, $10^{j\phi(b)} \equiv 1 \pmod b$, so that this number lies in the arithmetic progression

$$\{a, a + b10^d, a + 2\cdot b10^d, \ldots\}$$

if and only if

$$k \equiv \frac{a-r}{10^d} - s \cdot 10 \pmod{b}.$$

There are infinitely many such k, hence infinitely many palindromic numbers in every nonconstant arithmetic progression, excluding those for which every term is a multiple of 10.

Also solved by the proposer. There was one incomplete solution.

A Characteristic Polynomial **June 1996**

1502. *Proposed by Emeric Deutsch, Polytechnic University, Brooklyn, New York.*

Let n and k be positive integers satisfying $1 \le k \le n$. Find the characteristic polynomial of the $n \times n$ matrix

$$T_{n,k} = \begin{pmatrix} 0 & 0 & I_{n-k} \\ 0 & 1 & 0 \\ I_{k-1} & 0 & 0 \end{pmatrix},$$

where I_m denotes the $m \times m$ identity matrix.

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

More generally, suppose $T = T(a, b, c)$ is the $n \times n$ matrix

$$\begin{pmatrix} 0 & 0 & I_a \\ 0 & I_b & 0 \\ I_c & 0 & 0 \end{pmatrix},$$

with $b \ge 1$. Note that T is a permutation matrix T_π, where π is the permutation of $\{1, 2, \dots, n\}$ determined by right multiplication by T of the standard basis of row vectors. The characteristic polynomial of T_π depends only on π's cycle structure since (i) the eigenvalues of a $k \times k$ *cyclic* permutation matrix are the kth roots of unity; and (ii) if $\pi_1 \pi_2 \cdots \pi_r$ is a disjoint cycle factorization of π, then T_π is similar to a direct product of cyclic permutation matrices corresponding to the π_i's. Thus the characteristic polynomial of T_π is $\prod_{i=1}^{r}(x^{|\pi_i|} - 1)$.

To determine π's cycle structure, set $d = \gcd(a+b, b+c)$. Also let A, B, C denote the intervals of integers $[1, a], [a+1, a+b], [a+b+1, a+b+c]$, respectively, so that π permutes $A \cup B \cup C$. Observe that

$$\pi(i) = \begin{cases} i+b+c & \text{if } i \in A, \\ i+c-a & \text{if } i \in B, \\ i-a-b & \text{if } i \in C. \end{cases}$$

In particular, since d divides $b+c$, $c-a$, and $a+b$, π preserves congruence classes mod d. For any interval of integers J, let $J(i)$ denote the set of integers in J that are congruent to $i \pmod d$. Thus for $i \in A \cup B \cup C$, the orbit of i under π is contained in $A(i) \cup B(i) \cup C(i)$.

We claim the orbit of i under π equals $A(i) \cup B(i) \cup C(i)$. To establish this, suppose the orbit of i intersects A, B and C in x, y, and z elements, respectively. From the definition of π, this means that

$$x(b+c) + y(c-a) + z(-a-b) = 0,$$

or, equivalently,

$$(x+y)(b+c)=(y+z)(a+b).$$

Clearly, $0\le x\le |A(i)|$, $0\le y\le |B(i)|$, and $0\le z\le |C(i)|$. The claim will follow if we can show that the only such integral solutions to the equation are $(x,y,z)=(0,0,0)$ and $(|A(i)|,|B(i)|,|C(i)|)$. (Since a nonempty cycle containing i exists, the latter possibility must actually be the solution.) Because $(a+b)/d$ and $(b+c)/d$ are relatively prime, it follows that $(a+b)/d$ divides $x+y$. Furthermore,

$$0\le x+y\le |A(i)|+|B(i)|=\frac{a+b}{d}.$$

This forces (x,y) to be $(0,0)$ or $(|A(i)|,|B(i)|)$. Similarly, $(y,z)=(0,0)$ or $(|B(i)|,|C(i)|)$. Combining results yields $(x,y,z)=(0,0,0)$ or $(|A(i)|,|B(i)|,|C(i)|)$, proving our claim.

Thus π has exactly d cycles, each comprising a congruence class modulo d. Writing $n=qd+r$ with $0\le r<d$, it is now easy to count that π has r cycles of length $q+1$ and $d-r$ cycles of length q. Therefore, the characteristic polynomial of T_π is $(x^{q+1}-1)^r(x^q-1)^{d-r}$. When $a=n-k$, $b=1$, and $c=k-1$, we have $d=\gcd(n+1,k)$, $r=d-1$, and characteristic polynomial $(x^{(n+1)/d}-1)^{d-1}(x^{(n+1)/d-1}-1)$.

Also solved by J. C. Binz (Switzerland) and the proposer. There were one incorrect and two incomplete solutions.

Sums of Reciprocals of Logarithms of Binomial Coefficients **June 1996**

1503. *Proposed by Nick Lord, Tonbridge School, Kent, England.*

Does the sequence $\left(\Sigma_{k=1}^{n-1}1/\ln\binom{n}{k}\right)_{n=2}^{\infty}$ converge?

Solution by Edward Schmeichel, San Jose State University, San Jose, California.

The sequence diverges to infinity. Since $\ln x$ is an increasing function, we have

$$\ln k!=\ln 2+\ln 3+\cdots+\ln k>\int_1^k \ln x\,dx>k(\ln k-1),$$

implying $k!>(k/e)^k$. Thus,

$$\binom{n}{k}<\frac{n^k}{k!}<\left(\frac{en}{k}\right)^k,$$

and so

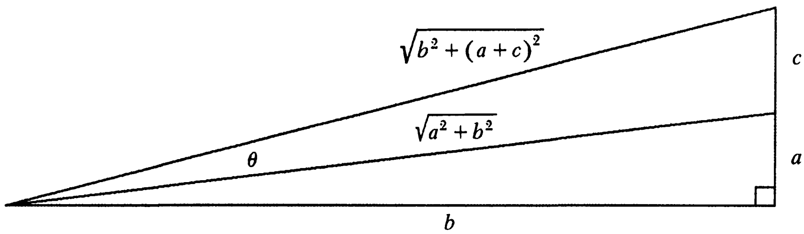
$$\frac{1}{\ln\binom{n}{k}}>\frac{1}{\ln\left(\frac{en}{k}\right)^k}=\frac{1}{k(1+\ln n-\ln k)}.$$

Differentiating, we see that $x(1+\ln n-\ln x)$ increases on $1\le x\le n$. Thus, we obtain

$$\sum_{k=1}^{n-1}\frac{1}{\ln\binom{n}{k}}>\sum_{k=1}^{n-1}\frac{1}{k(1+\ln n-\ln k)}>\int_1^n\frac{dx}{x(1+\ln n-\ln x)}=\ln(1+\ln n).$$

This latter expression goes to infinity as n does, proving the assertion.

Also solved by Con Amore Problem Group (Denmark), David Callan, and the proposer. There was one incorrect solution.



Answers

Solutions to the Quickies on page 224.

A865. For $j = 1, 2, \ldots, n$, define the random variable X_j to be 1 if the ball numbered j is correctly guessed and 0 if not. The expected number of correct guesses is the sum of the expected values of the X_j. Now observe that $X_j = 1$ if and only if, within the sequence of n draws from the urn, ball 1 precedes ball 2, which precedes ball $3, \ldots,$ which precedes ball j. Thus, the probability that $X_j = 1$ is $1/j!$. The expected number of correct guesses is $\frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}$, which approaches $e - 1$ as $n \to \infty$.

A866. The greatest lower bound and least upper bound are 0 and 1, respectively. Set

$$Q = \frac{a^2 + b^2 + ac}{\sqrt{(a^2+b^2)(a^2+b^2+c^2+2ac)}}.$$

Applying the law of cosines to the figure below, we have $Q = \cos\theta$. Since $0° < \theta < 90°$, it follows that $0 < Q < 1$. Setting $b = 1$, the values 0 and 1 are approached as $a \to 0$, $c \to \infty$, and $a = 1$, $c \to 0$, respectively.

A867. Let $\mathscr{A}$ denote the collection of $(n-1)$-element subsets of $\{1, 2, \ldots, kn\}$ and, for $j = 1$ to k, let

$$I_j := \{(j-1)n + 1, (j-1)n + 2, \ldots, jn\}.$$

Define ϕ: $\mathscr{A} \to \mathscr{A}$ as "increment each element by 1" with addition taken modulo n in such a way that each I_j is preserved. Clearly, ϕ is a permutation of $\mathscr{A}$ consisting entirely of (disjoint) n-cycles. Working modulo n, ϕ adds $n-1$ (or subtracts 1) from the sum of the elements of each $\{x_1, \ldots, x_{n-1}\} \in \mathscr{A}$. Hence this sum hits each residue class modulo n exactly once in each cycle (orbit) of ϕ. So

$$\left|\left\{\{x_1, \ldots, x_{n-1}\} \in \mathscr{A}: \sum_{j=1}^{n-1} x_j \equiv i \pmod{n}\right\}\right|$$

is the same for all i, $1 \le i \le n$, and the result follows. $\left(\text{The integrality of } \binom{kn}{n-1}/n\right.$ also follows from the identity $\left.\binom{kn}{n-1}/n = \binom{kn}{n} - (k-1)\binom{kn}{n-1}.\right)$

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.

Velleman, Daniel J., Fermat's Last Theorem and Hilbert's program, *Mathematical Intelligencer* 19 (1) (Winter 1997) 64–67.

Maybe, despite Wiles's proof being correct, Fermat's Last Theorem isn't true after all—maybe a counterexample will be found. "Should we say that what Wiles has established is not that Fermat's Last Theorem is true, but rather that *if* ZF [Zermelo-Fraenkel set theory] is consistent *then* Fermat's Last Theorem is true? ... Must we believe in the existence of the universe of all sets, and the truth of the ZF axioms for this universe, to be convinced that Wiles's proof can be trusted?" Author Velleman raises this question to remind us of the unresolved state of the foundations of mathematics.

Shurman, Jerry, *Geometry of the Quintic*, Wiley, 1997; xi + 200 pp, $39.95 (P). ISBN 0–471–13017–6.

This book reviews Felix Klein's *Lectures on the Icosahedron and Equations of the Fifth Degree* from a modern mathematical perspective. Marvelous mathematics enters, from classifying the groups of automorphisms of the Riemann sphere and finding and inverting generators for them, to the relationship between the icosahedral group and the quintic, to reduction of the general quintic to Brioschi form by radicals, plus Doyle and McMullen's 1989 iterative solution to the quintic. This is a book that helps the reader—advanced undergraduate, graduate, or faculty seminar participant—realize and appreciate the connections between branches of mathematics, in the context of solving an engaging problem.

Körner, T.W., *The Pleasures of Counting*, Cambridge University Press, 1996; x + 534 pp, $59.95, $34.95 (P). ISBN 0–521–56087–X, 0–521–56823–4.

"*Question* How do you tell whether a mathematician is an extrovert or an introvert? *Answer* Extrovert mathematicians look at *your* feet when they talk to you." (p. 226). Prof. Körner's book is extroverted in a more usual sense; it draws the reader into mathematics by starting from problems and experience in the real world and surrounds the problems with a rich context of history and perspective. A major source is warfare (strategies for submarine warfare, radar, Richardson's theory of arms races and his statistics of deadly quarrels, and the Enigma cipher machine) but other topics cast a wider net (cholera and statistics, biological scaling, the Lorentz transformation, classic algorithms, Braess's paradox, design of anchors, and the extinction of surnames, among others). He projects an aura of mathematics as an intellectual activity. The book is intended for "able" high-school students and first-year undergraduates, but it requires comfort with mathematical notation, tolerance for algebraic calculation, and (at a few well-advertised points) calculus. (The index is inadequate: None of "introvert," "extrovert," nor "anchor" appears; and the von Koch snowflake curve of pp. 215–219 cannot be found under "Koch," "snowflake," or "curve," but only under "length of curve" and the dubiously alphabetized "von Koch.")

Ivars Peterson's "MathLand," and Keith Devlin's "Devlin's Angle," both linked from MAA Online at `http://www.maa.org/`.

The MAA has devoted considerable effort to its Web site, which features two regular columns. Peterson's "Mathland" offers a weekly one- or two-page article on a mathematical topic, in the Martin Gardner style (i.e., no equations) and with references. Recent topics have included a new Mersenne prime, Tom Stoppard's play *Arcadia*, deceptive features of graphing calculator displays, and map coloring. "Devlin's Angle" is a monthly, featuring epimathematical topics, such as the myth that the value of pi was once legislated to be 3, a letter to a new college student, and how the year 2001 will not bring a computer like HAL in the film *2001: A Space Odyssey*. In short, MAA Online gives readers more of what writers Peterson (*Science News*) and Devlin (the MAA newsletter *Focus*) do well.

Devlin, Keith, Soft mathematics: The mathematics of people, `http://forum.swarthmore.edu/social/articles/softmath.short.html`. Are mathematicians turning soft? `http://www.maa.org/devlin/devlinangle_april.html`. *Goodbye, Descartes: The End of Logic and the Search for a New Cosmology of the Mind*, Wiley, 1997; x + 301 pp, \$27.95. ISBN 0–471–14216–6.

Editor Devlin is promoting the idea and use of *soft mathematics*, "a genuine attempt to blend mathematics with other approaches in trying to analyze or describe some phenomenon." He goes on to say that "Just as counterfeit mathematics is not mathematics, so too soft mathematics is not mathematics ... [but] can lead to real (hard) mathematics." His examples involve the use of mathematical notations or techniques because of their abstractness and conciseness. His motivation is that the social sciences are not mathematical sciences—put another way, mathematics has failed "in our attempts to understand the human world of people and minds"—hence the opening for mathematics of another texture. In *Goodbye, Descartes* (let me not repeat its apocalyptic subtitle), he "describes the entire history of Mankind's quest for a mathematical science of human reasoning and communication," concluding that "the existing techniques of logic and mathematics ... are inadequate for understanding the human mind."

Math Forum: A Virtual Center for Math Education on the Internet, at `http://forum.swarthmore.edu/`.

An outgrowth of its predecessor Geometry Forum, sponsor of the Usenet newsgroups `sci.math.geometry.*`, the Math Forum at Swarthmore College is a Web site funded by the National Science Foundation. The site intends to develop a "distributed, universal index ... for up-to-date, comprehensive access to all of the sites and individual pages available for mathematics education" (no doubt these phrases are taken directly from the grant application!). The home page offers access to geometry topics, humanistic mathematics, public understanding of mathematics, and more.

Arney, David C. (ed.), *Interdisciplinary Lively Application Projects (ILAPs)*, MAA, 1997; xi + 222 pp, \$27.50 (\$22 for MAA members). ISBN 0–88385–706–5.

This volume contains eight small-group project handouts in applications of mathematics, plus related materials (background, sample solutions, and a short student technical writing guide) and short articles about the ongoing effort to produce ILAPs. Each project takes from six to ten hours of student effort, and the backgrounds required in mathematics and in concepts from the application area are carefully specified. The topics of the eight projects are aerobic fitness, building a deck, landing from a parachute jump, flexibility of aircraft wings, backpacking to Pike's Peak, smog and inversions in the Los Angeles basin, designing bridge supports, and the diffusion of groundwater contaminants.

Bruss, F. Thomas, The fallacy of the two envelopes problem, *Mathematical Scientist* 21 (1996) 112–119.

A conundrum that has been making the rounds goes as follows: "Two envelopes contain respectively an amount of money and twice that much. You are offered one of the envelopes, drawn at random. Before you open it, you are offered the opportunity to switch and claim the other envelope instead; should you switch?" Let the first envelope contain \$S. The reasoning supporting the switch is that the other envelope contains either $\$2S$ or $\$S/2$, each with probability 1/2; so the expected value of its contents is $0.5(2S + S/2) = 1.25S > S$. Author Bruss points out the fallacy involved by carefully distinguishing the relevant sample space, and he describes an analogous deterministic fallacy. The two-envelopes fallacy is philosophical in nature, proceeding from using the same representation (notation) equivocally for two different concepts; but convincing mathematicians is one thing, and explaining it to your Aunt Mathilda is another.

Gallian, Joseph A., Error detection methods, *ACM Computing Surveys* 28 (3) (September 1996) 504–517.

Gallian presents in detail and concretely the use of check digits for error detection and correction, showing applications to the Universal Product Code (UPC), credit cards, banking, and blood banks. Many methods in use, such as modulus 9 and modulus 7 schemes (used on USPS money orders, VISA travelers checks, airline tickets, and FedEx and UPS packages), are distinctly inferior—they do not detect even all single-digit errors. Only German bank notes use a noncommutative system, based on the dihedral group D_{10}; this scheme detects all single-digit errors and all transposition errors involving adjacent digits. Why not a check digit for phone numbers? That might prevent the half dozen "wrong number" calls that you get every week at home (and the other half dozen at the office).

Pickover, Clifford A., *The Loom of God: Mathematical Tapestries at the Edge of Time*, Plenum, 1997; 292 pp, \$29.95. ISBN 0-306-45411-4.

"Mathematics is the loom upon which God weaves the fabric of the universe." How are mathematics and religion related? Author Pickover has tackled a tough topic and written a very engaging book. He begins with Pythagoreanism, explores numerology and kabala, finds fractals in mandalas and in quipus, investigates the geometry of Stonehenge, and considers "mathematical" arguments for the existence of God (including one by Gödel). Happily, there is no mention of the modern mysticism of the Great Pyramid; in fact, Egypt and Babylon are scarcely mentioned. Each chapter begins with an episode of science fiction (people from the future visit the past) and ends with a section on "The Science behind the Science Fiction." Curiously, God gets deleted from the famous quotation by Kronecker (*"Die ganze Zahl schuf der liebe Gott, alles Uebrige ist Menschenwerk"*) in Pickover's weak (and ungrammatical) rendering of it as "The primary source of all mathematics are the integers" (p. 81).

Artifact. *Time* 149 (18) (5 May 1997) 26.

Roger Penrose is suing Kimberly-Clark, manufacturers of toilet paper with embossings similar to a Penrose pattern. The pattern is practical—indentings along it quilt the paper and make it seem more bulky, despite there being less paper. "When ... the population of Great Britain [is] invited ... to wipe their bottoms on what appears to be the work of a Knight of the Realm without his permission, then a last stand must be taken." Or at least a long sit. Meanwhile, descendants of Leibniz are planning to sue every mathematician, scientist, and engineer for unauthorized use of the integral sign ....

Hubbard, Barbara Burke, *The World According to Wavelets: The Story of a Mathematical Technique in the Making*, A K Peters, 1996; xix + 264 pp, $34. ISBN 1–56881–047–4.

Suppose that you are a science writer who "carefully avoided all math courses" in college but is asked to write a popular account of a new development in mathematics—say, for example, wavelets. At what mathematical level should you pitch the writing? Author Hubbard answers that question by providing a "plain English" (no formulas) account about wavelets, focusing on the intuition, applications, and people involved in developing the field; these chapters take up about 40% of the book. In the account are embedded short boxes that tempt the reader to turn to a mathematical explanation of details, to be found in short sections in a part of the book entitled "Beyond Plain English." This "hypertext" approach works well. An appendix notes mathematical symbols used, offers a few proofs, gives references to other books, and tells where to get wavelets software.

Devlin, Keith, *Mathematics: The Science of Patterns*, Scientific American Library, 1997; viii + 216 pp, $19.95 (P). ISBN 0–7167–6022–3. Stein, Sherman, *Strength in Numbers: Discovering the Joy and Power of Mathematics in Everyday Life*, Wiley, 1996; xiii + 272 pp, $24.95. ISBN 0–471–15252–8. Wells, David, *You Are a Mathematician: A Wise and Witty Introduction to the Joy of Numbers*, Wiley, 1995; viii + 424 pp, $24.95. ISBN 0–471–18077–7. Adams, William J., *Get a Grip on Your Math* and *Get a Firmer Grip on Your Math*, Kendall/Hunt, 1996; xiii + 256 pp, $18.95 (P), vii + 290 pp, $18.95 (P).

Whenever I visit a bookstore, I check the section on mathematics, which is usually less than half a shelf at the bottom of the one bookcase devoted to the sciences. Next time you visit a bookstore, you may find some of the books listed above, which are all by mathematicians and all intended for the general reader. I am curious how the general reader (GR) would choose among them; which should I recommend? Would GR opt for Devlin: the slimmest, no exercises and few equations, with utterly beautiful color illustrations (the other books have none)? This book invites you to dip in at almost any page that you open it to. Would GR instead be impressed with the practical appeal of the Adams books, which investigate statistics, probability, optimization, and applications of mathematics in an informal way (only *Firmer Grip* has exercises)? Perhaps GR's interest would be caught by the chapter titles of Stein's book (e.g., "All There Is to Know about Fractions," "How to Read Mathematics," "A Fresh Look at Kindergarten"), which treats epimathematical topics (reform of mathematics education, what computers can/can't do, myths about mathematics), reviews elementary mathematics in interesting fashion, and undertakes a gentle introduction to calculus. The GR who opens Wells's book will realize right away that it will make for an active read, since boxed problems appear every few pages (with solutions at the end of each chapter). The book is the longest and the print is notably the smallest, but this is the book for those who want to learn by doing.

Bukiet, Bruce, Elliotte Rusty Harold, and José Luis Palacios, A Markov chain approach to baseball, *Operations Research* 45 (1) (January–February 1997) 14–23.

Columns by Ian Stewart in *Scientific American* in 1996 have revived interest in analysis of the game of Monopoly as a Markov chain (see "Take a walk on the Boardwalk," by Stephen D. Abbott and Matt Richey, in *College Mathematics Journal* 28 (3) (May 1997) 162–171). Now, on to baseball! Authors Bukiet *et al.* use a Markov chain to evaluate the performance of a baseball team, to determine the effect of each player on team performance (and hence the anticipated effect of player trades), and to find the optimal batting order. Are you wondering what their predictions are for the league-leaders in the 1997 season? American League: Yankees (East), Indians (Central), Mariners (West); National League: Marlins or Braves (tie; Marlins win playoff) (East), Cardinals (Central), Padres (West).

NEWS AND LETTERS

Twenty-Fifth Annual USA Mathematical Olympiad – Problems and Solutions

1. Prove that the average of the numbers $n \sin n^\circ$ $(n = 2, 4, 6, \ldots, 180)$ is $\cot 1^\circ$.

Solution. All arguments of trigonometric functions will be in degrees. We need to prove

$$2\sin 2 + 4\sin 4 + \cdots + 178\sin 178 = 90\cot 1, \tag{1}$$

which is equivalent to

$$2\sin 2 \cdot \sin 1 + 2(2\sin 4 \cdot \sin 1) + \cdots + 89(2\sin 178 \cdot \sin 1) = 90\cos 1. \tag{2}$$

Using the identity $2\sin a \cdot \sin b = \cos(a-b) - \cos(a+b)$, we find

$$\begin{aligned}
&2\sin 2 \cdot \sin 1 + 2(2\sin 4 \cdot \sin 1) + \cdots + 89(2\sin 178 \cdot \sin 1)\\
&\quad = (\cos 1 - \cos 3) + 2(\cos 3 - \cos 5) + \cdots + 89(\cos 177 - \cos 179)\\
&\quad = \cos 1 + \cos 3 + \cos 5 + \cdots + \cos 175 + \cos 177 - 89\cos 179\\
&\quad = \cos 1 + (\cos 3 + \cos 177) + \cdots + (\cos 89 + \cos 91) - 89\cos 179\\
&\quad = \cos 1 + 89\cos 1 = 90\cos 1,
\end{aligned}$$

so (1) is true.

Second Solution. This solution uses complex numbers. Let $\operatorname{Re} z$ and $\operatorname{Im} z$ denote the real part and the imaginary part, respectively, of the complex number z. We use the relation $\sin n = \operatorname{Im} e^{i\pi n/180}$ together with the summation formula

$$x + 2x^2 + \cdots + nx^n = \frac{x + x^2 + \cdots + x^n - nx^{n+1}}{1-x} = \frac{x - x^{n+1}}{(1-x)^2} - \frac{nx^{n+1}}{1-x} \qquad (x \neq 1).$$

Let $\omega = e^{i\pi/90} = \cos 2 + i\sin 2$. Then $\omega^{90} = -1$ and

$$\operatorname{Im}\frac{\omega}{(1-\omega)^2} = \operatorname{Im}\left(\frac{1}{\omega^{1/2} - \omega^{-1/2}}\right)^2 = 0.$$

Thus

$$\begin{aligned}
2\sin 2 + 4\sin 4 + \cdots + 178\sin 178 &= 2\operatorname{Im}(\omega + 2\omega^2 + 3\omega^3 + \cdots + 89\omega^{89})\\
&= 2\operatorname{Im}\left[\frac{\omega+1}{(1-\omega)^2} + \frac{89}{1-\omega}\right] = 2\operatorname{Im}\left[\frac{2\omega + (1-\omega)}{(1-\omega)^2} + \frac{89}{1-\omega}\right]\\
&= 2\operatorname{Im}\frac{90}{1-\omega} = 180\operatorname{Im}\frac{\omega^{-1/2}}{\omega^{-1/2} - \omega^{1/2}} = 90\cot 1.
\end{aligned}$$

2. For any nonempty set S of real numbers, let $\sigma(S)$ denote the sum of the elements of S. Given a set A of n positive integers, consider the collection of all distinct sums $\sigma(S)$ as S ranges over the nonempty subsets of A. Prove that this collection of sums can be partitioned into n classes so that in each class, the ratio of the largest sum to the smallest sum does not exceed 2.

Solution. Let $A = \{a_1, a_2, \ldots, a_n\}$ where $a_1 < a_2 < \cdots < a_n$. For $i = 1, 2, \ldots, n$ let $s_i = a_1 + a_2 + \cdots + a_i$ and take $s_0 = 0$. All the sums in question are less than or equal to s_n, and if σ is one of them, we have

$$s_{i-1} < \sigma \leq s_i \tag{1}$$

for an appropriate i. Divide the sums into n classes by letting C_i denote the class of sums satisfying (1). We claim that these classes have the desired property. To establish this, it suffices to show that (1) implies

$$\frac{1}{2} s_i < \sigma \leq s_i. \tag{2}$$

Suppose (1) holds. The inequality $a_1 + a_2 + \cdots + a_{i-1} = s_{i-1} < \sigma$ shows that the sum σ contains at least one addend a_k with $k \geq i$. Then since then $a_k \geq a_i$, we have

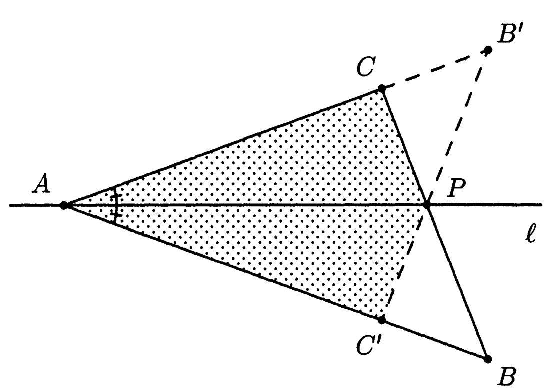
$$s_i - \sigma < s_i - s_{i-1} = a_i \leq a_k \leq \sigma,$$

which together with $\sigma \leq s_i$ implies (2).

Note. The result does not hold if 2 is replaced by any smaller constant c. To see this, choose n such that $c < 2 - 2^{-(n-1)}$ and consider the set $\{1, \ldots, 2^{n-1}\}$. If this set is divided into n subsets, two of $1, \ldots, 2^{n-1}, 1 + \cdots + 2^{n-1}$ must lie in the same subset, and their ratio is at least $(1 + \cdots + 2^{n-1})/(2^{n-1}) = 2 - 2^{-(n-1)} > c$.

3. Let ABC be a triangle. Prove that there is a line ℓ (in the plane of triangle ABC) such that the intersection of the interior of triangle ABC and the interior of its reflection $A'B'C'$ in ℓ has area more than 2/3 the area of triangle ABC.

Solution. Let a, b, c be the lengths of the sides BC, CA, AB, respectively; without loss of generality $a \leq b \leq c$. Choose ℓ to be the angle bisector of $\angle A$. Let P be the intersection of ℓ with BC. Since $AC \leq AB$, the intersection of triangles ABC and $A'B'C'$ is the disjoint union of two congruent triangles, APC and APC'.



Considering BC as a base, triangles APC and ABC have equal altitudes, so their areas are in the same ratio as their bases:

$$\frac{\text{Area}(APC)}{\text{Area}(ABC)} = \frac{PC}{BC}.$$

Since AP is the angle bisector of $\angle A$, we have $BP/PC = c/b$, so

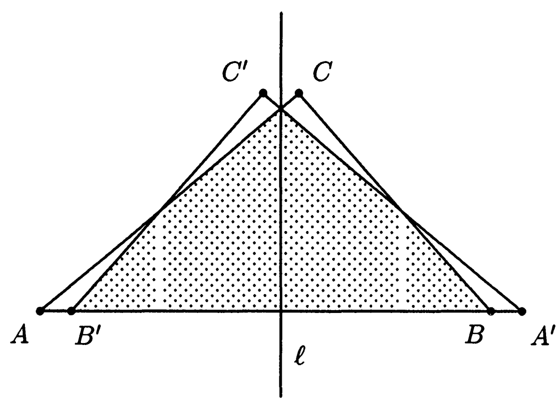
$$\frac{PC}{BC} = \frac{PC}{BP+PC} = \frac{1}{c/b+1}.$$

Thus it suffices to prove

$$\frac{2}{c/b+1} > \frac{2}{3}. \tag{1}$$

But $2b \geq a+b > c$ by the triangle inequality, so $c/b < 2$ and thus (1) holds.

Note. Let $\mathcal{F}$ denote the figure given by the intersection of the interior of triangle ABC and the interior of its reflection in ℓ. Another approach to the problem involves finding the maximum attained for Area($\mathcal{F}$)/Area(ABC) by taking ℓ from the family of lines perpendicular to AB.



By choosing the best alternative between the angle bisector at A and the optimal line perpendicular to BC,

$$\frac{\text{Area}(\mathcal{F})}{\text{Area}(ABC)} > \frac{2}{1+\sqrt{2}} = 2(\sqrt{2}-1) = 0.828427\ldots$$

can be attained. This constant is in fact the best possible.

4. An n-term sequence $(x_1, x_2, \ldots, x_n)$ in which each term is either 0 or 1 is called a ***binary sequence of length n***. Let a_n be the number of binary sequences of length n containing no three consecutive terms equal to 0, 1, 0 in that order. Let b_n be the number of binary sequences of length n that contain no four consecutive terms equal to 0, 0, 1, 1 or 1, 1, 0, 0 in that order. Prove that $b_{n+1} = 2a_n$ for all positive integers n.

Solution. We refer to the binary sequences counted by (a_n) and (b_n) as "type A" and "type B", respectively. For each binary sequence $(x_1, x_2, \ldots, x_n)$ there is a corresponding binary sequence $(y_0, y_1, \ldots, y_n)$ obtained by setting $y_0 = 0$ and

$$y_i = x_1 + x_2 + \cdots + x_i \mod 2, \quad i = 1, 2, \ldots, n. \tag{1}$$

(Addition mod 2 is defined as follows: $0+0 = 1+1 = 0$ and $0+1 = 1+0 = 1$.) Then

$$x_i = y_i + y_{i-1} \mod 2, \quad i = 1, 2, \ldots, n,$$

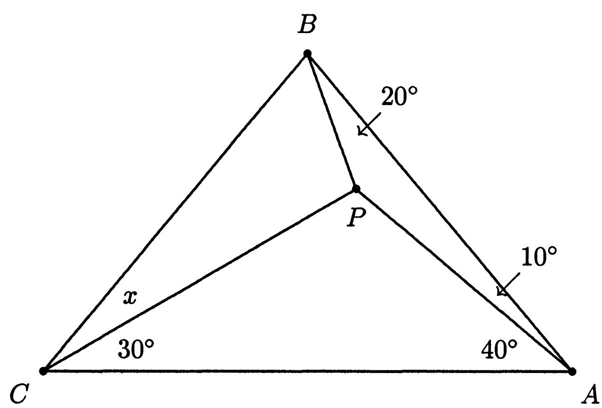
and it is easily seen that (1) provides a one-to-one correspondence between the set of all binary sequences of length n and the set of binary sequences of length $n+1$ in which the first term is 0. Moreover, the binary sequence $(x_1, x_2, \ldots, x_n)$ has three consecutive terms equal to 0, 1, 0 in that order if and only if the corresponding sequence $(y_0, y_1, \ldots, y_n)$ has four consecutive terms equal to 0, 0, 1, 1 or 1, 1, 0, 0 in that order, so the first is of type A if and only if the second is of type B. The set of type B sequences of length $n+1$ in which the first term is 0 is exactly half the total number of such sequences, as can be seen by means of the mapping in which 0's and 1's are interchanged.

5. Triangle ABC has the following property: there is an interior point P such that $\angle PAB = 10°$, $\angle PBA = 20°$, $\angle PCA = 30°$, and $\angle PAC = 40°$. Prove that triangle ABC is isosceles.

Solution. All angles will be in degrees. Let $x = \angle PCB$. Then $\angle PBC = 80 - x$. By the Law of Sines,

$$1 = \frac{PA}{PB}\frac{PB}{PC}\frac{PC}{PA} = \left(\frac{\sin\angle PBA}{\sin\angle PAB}\right)\left(\frac{\sin\angle PCB}{\sin\angle PBC}\right)\left(\frac{\sin\angle PAC}{\sin\angle PCA}\right)$$

$$= \frac{\sin 20 \sin x \sin 40}{\sin 10 \sin(80-x)\sin 30} = \frac{4\sin x \sin 40 \cos 10}{\sin(80-x)}.$$



The identity $2\sin a \cdot \cos b = \sin(a-b) + \sin(a+b)$ now yields

$$1 = \frac{2\sin x(\sin 30 + \sin 50)}{\sin(80-x)} = \frac{\sin x(1 + 2\cos 40)}{\sin(80-x)},$$

so

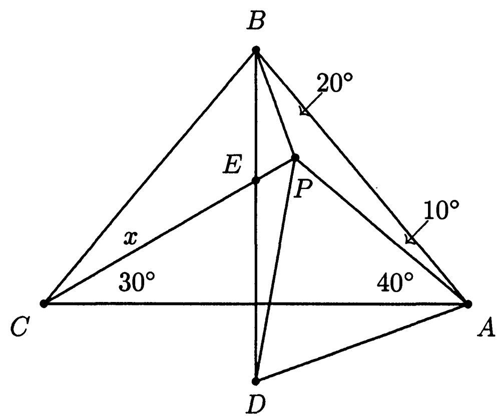
$$2\sin x \cos 40 = \sin(80-x) - \sin x = 2\sin(40-x)\cos 40.$$

This gives $x = 40 - x$ and thus $x = 20$. It follows that $\angle ACB = 50 = \angle BAC$, so triangle ABC is isosceles.

Second Solution. Let D be the reflection of A across the line BP. Then triangle APD is isosceles with vertex angle

$$\angle APD = 2(180 - \angle BPA) = 2(\angle PAB + \angle ABP) = 2(10+20) = 60,$$

so it is equilateral. Also, $\angle DBA = 2\angle PBA = 40$. Since $\angle BAC = 50$, we have $DB \perp AC$.



Let E be the intersection of DB with CP. Then

$$\angle PED = 180 - \angle CED = 180 - (90 - \angle ACE) = 90 + 30 = 120,$$

so $\angle PED + \angle DAP = 180$. We deduce that the quadrilateral $APED$ is cyclic, and therefore $\angle DEA = \angle DPA = 60$. Finally, we note that $\angle DEA = 60 = \angle DEC$. Since $AC \perp DE$, we find that A and C are symmetric across the line DE, which implies that $BA = BC$, as desired.

6. Determine (with proof) whether there is a subset X of the integers with the following property: for any integer n there is exactly one solution of $a + 2b = n$ with $a, b \in X$.

Solution. Yes, there is such a subset. Note that $(r, s) = (0,0), (1,0), (0,1), (1,1)$ gives $r + 2s = 0, 1, 2, 3$, respectively. Thus if the problem is restricted to the non-negative integers, it is clear that the set of numbers whose representation in base 4 contains only the digits 0 and 1 has the desired property. To accommodate the negative integers as well, we alter this construction slightly by switching to "base -4". That is, we express a given integer as $\sum_{i=0}^{k} c_i(-4)^i$, with $c_i \in \{0, 1, 2, 3\}$. Let X be the set of numbers whose representation uses only the digits 0 and 1. In view of the observation made at the beginning, the proof that X has the desired property is complete once we show that every integer has a unique representation in base -4.

Existence and Uniqueness. Given an integer n, choose k such that

$$0 \leq n + 3(4 + 4^3 + 4^5 + \cdots + 4^{2k-1}) \leq 4^{2k+1} - 1.$$

Then by means of the usual base 4 expansion, write

$$n + 3(4 + 4^3 + 4^5 + \cdots + 4^{2k-1}) = \sum_{i=0}^{2k} c_i \, 4^i.$$

Setting $d_{2i} = c_{2i}$ and $d_{2i-1} = 3 - c_{2i-1}$ gives

$$\sum_{i=0}^{2k} d_i(-4)^i = \sum_{i=0}^{k} c_{2i} \, 4^{2i} - \sum_{i=1}^{k} (3 - c_{2i-1}) 4^{2i-1} = n.$$

Thus we have existence. To show base -4 representations are unique, let (c_i) and (d_i) be two distinct finite sequences of elements from $\{0, 1, 2, 3\}$, and let j the smallest integer such that $c_j \neq d_j$. Then

$$\sum_{i=0}^{k} c_i(-4)^i \not\equiv \sum_{i=0}^{k} d_i(-4)^i \pmod{4^j},$$

so the two numbers represented by (c_i) and (d_i) are distinct.

Thirty-Seventh Annual International Mathematical Olympiad – Problems

1. We are given a positive integer r and a rectangular board $ABCD$ with dimensions $|AB| = 20, |BC| = 12$. The rectangle is divided into a grid of 20×12 unit squares. The following moves are permitted on the board: one can move from one square to another only if the distance between the centers of the two squares is $\sqrt{r}$. The task is to find a sequence of moves leading from the square with A as a vertex to the square with B as a vertex.

(a) Show that the task cannot be done if r is divisible by 2 or 3.

(b) Prove that the task is possible when $r = 73$.

(c) Is there a solution when $r = 97$?

2. Let P be a point inside triangle ABC such that

$$\angle APB - \angle ACB = \angle APC - \angle ABC.$$

Let D, E be the incenters of triangles APB, APC, respectively. Show that AP, BD, CE meet at a point.

3. Let S denote the set of nonnegative integers. Find all functions f on S taking values in S such that

$$f(m + f(n)) = f(f(m)) + f(n) \qquad \forall m, n \in S.$$

4. The positive integers a and b are such that the numbers $15a+16b$ and $16a-15b$ are both squares of positive integers. What is the least possible value that can be taken on by the smaller of these two squares?

5. Let $ABCDEF$ be a convex hexagon such that AB is parallel to DE, BC is parallel to EF, and CD is parallel to FA. Let R_A, R_C, R_E denote the circumradii of triangles FAB, BCD, DEF, respectively, and let P denote the perimeter of the hexagon. Prove that

$$R_A + R_C + R_E \geq \frac{P}{2}.$$

6. Let p, q, n be positive integers with $p + q < n$. Let $(x_0, x_1, \ldots, x_n)$ be an $(n+1)$-tuple of integers satisfying the following conditions:

(a) $x_0 = x_n = 0$.

(b) For each i with $1 \leq i \leq n$, either $x_i - x_{i-1} = p$ or $x_i - x_{i-1} = -q$.

Show that there exist indices $i < j$ with $(i, j) \neq (0, n)$, such that $x_i = x_j$.

Notes

The 1996 USA Mathematical Olympiad was prepared by Titu Andreescu, Elgin Johnston, Jim Propp, Cecil Rousseau (chair), Richard Stong, and Paul Zeitz.

The top eight students on the 1996 USAMO were (in alphabetical order):

Carl J. Bosley	Topeka, KS
Christopher C. Chang	Palo Alto, CA
Nathan G. Curtis	Alexandria, VA
Michael R. Korn	Arden Hills, MN
Carl A. Miller	Bethesda, MD
Josh P. Nichols-Barrer	Newton Center, MA
Alexander H. Saltman	Austin, TX
Daniel P. Stronger	New York, NY

Christopher Chang was the winner of the Greitzer-Klamkin award, given to the top scorer on the USAMO. Members of the USA team at the 1996 IMO (Mumbai, India) were Carl Bosley, Christopher Chang, Nathan Curtis, Michael Korn, Carl Miller, and Alexander Saltman.

The training program to prepare the USA team for the IMO (the Mathematical Olympiad Summer Program) was held at the University of Nebraska, Lincoln, NE. Titu Andreescu, Elgin Johnston, Kiran Kedlaya, and Paul Zeitz served as instructors, assisted by Jeremy Bem and Jonathan Weinstein.

The booklet *Mathematical Olympiads 1996* presents additional solutions to problems on the 25th USAMO and solutions to the 37th IMO. This booklet is available from:

Dr. Walter Mientka
Department of Mathematics
University of Nebraska
Lincoln, NE 68588-0658.

Such a booklet has been published every year since 1976. Copies are $5.00 for each year 1976-1996.

The USA Mathematical Olympiad, participation of the US team in the International Mathematical Olympiad, and the sequence of examinations leading to qualification for these olympiads are under the administration of the M.A.A. Committee on American Mathematical Competitions, and these activities are sponsored by eight organizations of professional mathematicians. For further information about this sequence of examinations, contact the Executive Director of the Committee, Professor Mientka, at the above address.

This report was prepared by Cecil Rousseau, The University of Memphis.

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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