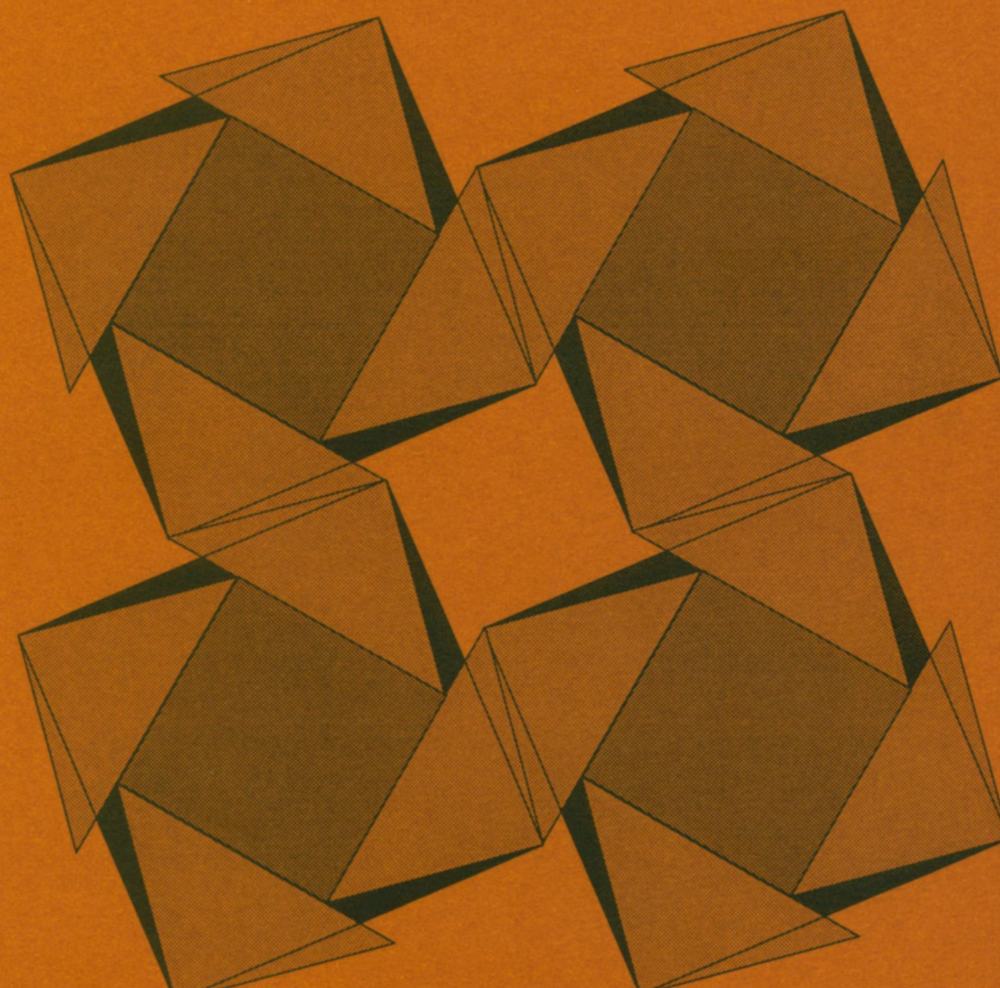
Vol. 72, No. 5, December 1999

MATHEMATICS MAGAZINE



A Pythagorean pattern (see p. 407)

- Olivier and Abel on Series Convergence
- Apollonian Cubics
- Tangent Records, π, and the Meaning of Life

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 71, pp. 76–78, 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 Frank Farris, Editor, Department of Mathematics and Computer Science, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053-0290. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

AUTHORS

Michael Goar received a B.S. degree in physics from UCLA in 1980 and taught mathematics, physics, and computer science at Mayfield High School in Las Cruces, New Mexico, before returning to school to earn a master's degree in mathematics at New Mexico State University in 1997. He currently teaches mathematics at Ft. Collins High School in Ft. Collins, Colorado, and uses mathematics projects to help motivate classroom material. This paper came about through an assignment in a course on the use of original source material in the classroom, taught by David Pengelley and Reinhard Laubenbacher at NMSU.

Paris Pamfilos received his Ph.D. from Cologne University, Germany in 1979. He held visiting positions at the University of Bonn, the University of Essen, and the University of Cologne. Since 1979 he has been at the University of Crete, in Greece. His research interests include differential geometry, Lie groups, mechanics, Euclidean geometry, scientific computing, and programming.

Apostolos Thoma received his Ph.D. from Purdue University, West Lafayette, Indiana, in 1989. He has held visiting positions at the University of Crete, Iraklion, and the Max-Planck Institut für Mathematik, Bonn. Since 1994 he has been at the University of Ioannina in Greece. His research interests include algebraic geometry and commutative and computational algebra. Euclidean geometry has been his favorite topic since his high school years.

Ira Rosenholtz graduated from Brandeis University and received his Ph.D. in topology from the University of Wisconsin, under the supervision of Joseph M. Martin. In addition to loving his terrific family and the teaching and learning of mathematics, he enjoys reading mysteries and playing slow pitch softball and chess. He is also the author of a forthcoming book on Infinity, which is about even bigger things than are mentioned in his article. He is still searching for more World Records, but he knows absolutely nothing about the Meaning of Life.

Vol. 72, No. 5, December 1999

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.

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

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

Printed in the United States of America

ARTICLES

Olivier and Abel on Series Convergence: An Episode from Early 19th Century Analysis

MICHAEL GOAR
Ft. Collins High School
Ft. Collins, CO 80525

1. Introduction

Calculus and introductory analysis courses are based upon a theoretical framework whose historical development is too often ignored. In reordering and repackaging these theoretical results, authors and instructors often present material in a succession that is natural for an experienced mathematician, but less so for the student. In many cases, the development of a mathematical concept may be better illuminated by studying original source material to *observe* the process of intuitive notions being replaced by more sophisticated ones as the weaknesses of the former are exposed by counterexamples.

Students who encounter this material have the opportunity to witness the difficulties once faced by mathematicians in coming to terms with the meaning of such things as limits and convergence—at the same time as the students' ideas about these concepts are maturing. This historic struggle often parallels students' own difficulties, and enhances their appreciation of their own thinking while deepening their understanding of the underlying mathematics.

We will describe a misstep made while attempting to simplify the process of determining whether an infinite series converges. This episode dates to the 1820s, when the field of analysis was emerging as a formalization of ideas and methods used in calculus, which began its development much earlier. Mathematics historians now regard much of the 19th century as a period of "rigorization" in analysis, where standards of proof, along with a notation we would recognize today, gradually became more widely employed. The misstep came in the form of a simple, universal series convergence test proposed by a mathematician named Louis Olivier, which was subsequently invalidated by the young Norwegian prodigy Niels Abel. The misconception at the heart of this ill-fated conjecture, and the skillfully crafted counterargument forwarded by Abel, make this exchange an excellent study example for students in courses that treat series convergence.

2. Setting the stage

We begin with the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$, a standard example that conflicts with students' initial concepts of convergence. Proofs of its divergence vary, but many use a comparison argument put forward by Jakob Bernoulli around 1690, which compares the harmonic series to another positive-termed series which is clearly less, term by term [**1**]:

$$\sum a_n = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \cdots + \frac{1}{16} + \frac{1}{17} + \cdots$$

$$\sum b_n = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{16} + \frac{1}{32} + \cdots$$

Since Σb_n contains infinitely many groupings of terms that sum to $\frac{1}{2}$, the series diverges. Thus the larger series Σa_n diverges too, by comparison.

A similar proof was given by Cauchy in *Cours d'analyse* (1821), which compiled many of the series convergence tests we recognize in textbooks today, including the root test, the ratio test, the logarithm test, and what is often called Cauchy's condensation test (for a nonnegative, monotone decreasing sequence a_k, $\Sigma_{k=1}^{\infty} a_k$ converges if and only if $\Sigma_{k=0}^{\infty} 2^k a_{2^k}$ converges). Cauchy writes:

> ...before effecting the summation of any series, I had to examine in which cases the series can be summed, or, in other words, what are the conditions of their convergence; and on this topic I have established general rules which seem to me to merit some attention [**2**].

With so many convergence tests available, each carrying its own set of conditions, one sees a clear motivation for mathematicians then to look for a smaller set of more universal tests, and, in particular, any test that established both necessary and sufficient conditions for convergence. There is a modern parallel for students: After having been convinced in one way or another that the harmonic series truly diverges, and after having seen that the p-series $\Sigma_{n=1}^{\infty} \frac{1}{n^p}$ converges for $p > 1$, students might reasonably conclude that the harmonic series forms a sort of "boundary" case with which other potentially convergent series of positive terms could be compared. Would a series whose partial sums increase more slowly than those of the harmonic series then necessarily converge?

Just such a claim was made by Louis Olivier in his paper *Remarques sur les series infinies et leur convergence* (Remarks on infinite series and their convergence), published in 1827 in the *Journal für die reine und angewandte Mathematik*, also known as *Crelle's Journal*. Olivier asserts that a series whose terms are positive (or whose terms can be grouped so that each group has a positive sum) and approach zero more rapidly than those of a harmonic series will converge, and vice versa. He states this by proposing a criterion based upon the limit of the product of n and the general term a_n.

Following is an excerpt from Olivier's original paper [**3**], and then a translation. (See FIGURE 1.)

> Therefore if one finds in an infinite series, the product of the n^{th} term, or of the n^{th} group of terms which keep the same sign, by n, is zero, for $n = \infty$, one can regard precisely this situation as an indicator that the series is convergent, and conversely, the series is not convergent if the product $n \cdot a_n$ is nonzero for $n = \infty$.

Notice that it was customary at that time to write "$n = \infty$" where we would write "$n \to \infty$" today. It is interesting to consider how infinite quantities were viewed by those working in this period, a time in which the ε and δ characterizations of limits that are now standard were just emerging in mathematical literature. Olivier continues

Donc si l'on trouve, que dans une série infinie, le produit du n^{me} terme, ou du n^{me} des groupes de termes qui conservent le même signe, par n, est zéro, pour $n = \infty$, on peut regarder cette seule circonstance comme une marque, que la série est convergente; et réciproquement, la série ne peut pas être convergente, si le produit $n.a_n$ n'est pas nul pour $n = \infty$.

Nous allons appliquer ce criterium de la convergence des séries infinies à quelques exemples.

6.

Exemples.

I. Dans la série

$$1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} \ldots\ldots + \frac{1}{n} + R.$$

le n^{me} terme a_n est $\frac{1}{n}$. Donc $n.a_n = 1$ n'est pas 0. Donc la série n'est pas convergente.

II. Dans la série

$$1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \ldots\ldots + \frac{1}{n} - \frac{1}{n+1} + R.$$

ou n est un nombre impair quelconque, le n^{me} des groupes de termes, qui conservent le même signe, est $a_n = \frac{1}{2n-1} - \frac{1}{2n} = \frac{1}{2n(2n-1)}$. Donc $n.a_n = \frac{1}{2(2n-1)}$. Ce produit est zéro pour $n = \infty$. Donc la série est convergente.

III. Dans la série

$$1 + \frac{1}{2^p} + \frac{1}{3^p} \ldots\ldots + \frac{1}{n^p} + R,$$

le produit $n.a_n$ est égal à $\frac{1}{n^{p-1}}$. Cette quantité est nulle pour $n = \infty$, si $p > 1$. Donc la série est convergente, si $p > 1$.

FIGURE 1

Excerpt from Olivier's 1827 article.

by applying his convergence test to a few examples. Note the use of "R" to indicate the sum of the remaining terms, or "tail," of each series.

We will apply this criterion for the convergence of infinite series to several examples.

I. For the series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + R,$$

the n^{th} term a_n is $\frac{1}{n}$. Thus $n \cdot a_n = 1$ is not 0. Thus the series is not convergent.

In the following example, pairs of terms are grouped, with each group positive, so that the series satisfies Olivier's criterion.

II. For the series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots + \frac{1}{n} - \frac{1}{n+1} + R,$$

where n is an arbitrary odd number, the n^{th} group of terms, which keep the same sign, is $a_n = \frac{1}{2n-1} - \frac{1}{2n} = \frac{1}{2n(2n-1)}$. Thus $n \cdot a_n = \frac{1}{2(2n-1)}$. This product is zero for $n = \infty$. Thus the series is convergent.

III. For the series

$$1+\frac{1}{2^p}+\frac{1}{3^p}\cdots+\frac{1}{n^p}+R,$$

the product $n \cdot a_n$ is equal to $\frac{1}{n^{p-1}}$. This quantity is null for $n=\infty$, if $p>1$. Thus the series is convergent, if $p>1$.

3. A reply from Abel

Before the mid-1800s it was common to see mathematical arguments advanced solely on the weight of examples or by "proofs" that would be better described as appeals to intuition. This state of affairs, in 1826, inspired the young Norwegian mathematician Abel to write:

> I shall devote all my efforts to bring light into the immense obscurity that reigns today in Analysis. It so lacks any plan or system, that one is really astonished that there are so many people who devote themselves to it—and, still worse, it is devoid of any rigor [**1**].

Niels Henrik Abel grew up in the small village of Findo, Norway, during a time of great local economic difficulty. His interest in mathematics is credited largely to a remarkable teacher, Bernt Michael Holmboe, who inspired the young Abel to study the works of great mathematicians, including Newton, Euler, and Lagrange. Holmboe patiently guided Abel to assimilate the contents of Gauss's *Disquisitiones Arithmeticae* and to resolve to learn mathematics from original sources produced by such masters. Abel's prodigious success with his studies, complemented by Holmboe's encouragement and financial support, led him to complete his degree at the age of 19 and eventually earn a small stipend to support travel and study in France and Germany, then the center of progress in analysis. In Berlin, Abel met A. L. Crelle, who was publishing the first research journal devoted solely to pure and applied mathematics. This encounter benefited each, as Crelle received excellent papers from Abel, while the young mathematician gained a medium of publication [**4**].

Volume 3 of *Crelle's Journal* carried five submissions from Abel, the first entitled "*Note sur le memoire de M. L. Olivier No. 4 du second tome de ce journal, ayant pour titre 'remarques sur les series infinies et leur convergence.'*" (Note on the memoir of Mr. L. Olivier, No. 4 in the second volume of this journal, titled 'Remarks on infinite series and their convergence.') Abel offers a counterexample to Olivier's claim and then proves that *no* function of n can be used in the type of limit test for series convergence proposed by Olivier. Abel's construction is both illuminating and instructive; the following translation is from the original French [**5**]:

> One finds on page 34 of this memoir the following theorem for recognizing if a series is convergent or divergent:
>
> "If one finds in an infinite series, the product of the n^{th} term, or n^{th} group of terms which keep the same sign, by n, is zero, for $n=\infty$, one can regard precisely this situation as an indicator that the series is convergent, and conversely, the series is not convergent if the product $n \cdot a_n$ is nonzero for $n=\infty$."
>
> The latter part of this theorem is very true, but the first does not appear to be.

For example the series

$$\frac{1}{2\log 2}+\frac{1}{3\log 3}+\frac{1}{4\log 4}+\cdots+\frac{1}{n\log n}$$

is divergent while $n\cdot a_n=\frac{1}{\log n}$ is zero for $n=\infty$.

In fact, the natural logarithms[1] in question are always less than the numbers themselves minus 1, that is, one always has $\log(1+x)<x$. If $x>1$,

$$\log(1+x)=x-x^2\left(\frac{1}{2}-\frac{1}{3}x\right)-x^4\left(\frac{1}{4}-\frac{1}{5}x\right)\cdots$$

thus also in this latter case $\log(1+x)<x$ because $\frac{1}{2}-\frac{1}{3}x$, $\frac{1}{4}-\frac{1}{5}x,\ldots$ are all positive.

Abel then continues his proof of the divergence of $\Sigma\frac{1}{n\log n}$ by cleverly finding a lower bound for each of its terms.

By letting $x=\frac{1}{n}$, this gives

$$\log\left(1+\frac{1}{n}\right)<\frac{1}{n}\quad\text{or, equivalently,}\quad\log\left(\frac{1+n}{n}\right)<\frac{1}{n},$$

or[2]

$$\log(1+n)<\frac{1}{n}+\log n=\left(1+\frac{1}{n\log n}\right)\log n:$$

thus

$$\log\log(1+n)<\log\log n+\log\left(1+\frac{1}{n\log n}\right).$$

But since $\log(1+x)<x$, one has $\log\left(1+\frac{1}{n\log n}\right)<\frac{1}{n\log n}$; thus, by virtue of the preceding expression,

$$\log\log(1+n)<\log\log n+\frac{1}{n\log n}.$$

This inequality in hand, Abel sums over all $n>1$, which, after canceling, leaves the partial sum of his series bounded below by an expression that is unbounded as $n\to\infty$.

---

[1]The French term is "logarithmes hyperbolique," or *hyperbolic logarithms*. The reader may wish to consider why this name was used.

[2]In the original paper, the following line was erroneously printed as

$$\log(1+n)<\frac{1}{n}-\log n=\left(1+\frac{1}{n\log n}\right)\log n.$$

We assume the sign error is due to the publisher, not to Abel.

By letting successively $n = 2, 3, 4, \ldots$ one finds

$$\log\log 3 < \log\log 2 + \frac{1}{2\log 2},$$

$$\log\log 4 < \log\log 3 + \frac{1}{3\log 3},$$

$$\log\log 5 < \log\log 4 + \frac{1}{4\log 4},$$

$$\ldots$$

$$\log\log(1+n) < \log\log n + \frac{1}{n\log n}.$$

Therefore, upon taking the sum,

$$\log\log(1+n) < \log\log 2 + \frac{1}{2\log 2} + \frac{1}{3\log 3} + \frac{1}{4\log 4} \cdots + \frac{1}{n\log n}.$$

But $\log\log(1+n) = \infty$ for $n = \infty$; thus the sum of the proposed series

$$\frac{1}{2\log 2} + \frac{1}{3\log 3} + \frac{1}{4\log 4} + \cdots + \frac{1}{n\log n}$$

is infinitely large and consequently the series diverges. The theorem announced in the citation is thus at fault in this case.

After validating his counterexample, Abel extends his argument further, illustrating that a scheme like Olivier's fails not only when taking the limit of $n \cdot a_n$, but in general. He considers the factor n in $n \cdot a_n$ as a specific case of the more general expression $\varphi(n) \cdot a_n$, where the function $\varphi(n)$ (which Abel writes as φn) is used to perform this type of test.

In general, one can demonstrate that it is impossible to find a function φn such that any series $a_0 + a_1 + a_2 + a_3 + \cdots + a_n$, where we suppose all terms are positive, should be convergent if $\varphi n \cdot a_n$ is zero for $n = \infty$, and divergent in the contrary case. This is what one can make clear with the help of the following theorem.

If the series $a_0 + a_1 + a_2 + a_3 + \cdots + a_n + \cdots$ is divergent, then

$$\frac{a_1}{a_0} + \frac{a_2}{a_0 + a_1} + \frac{a_3}{a_0 + a_1 + a_2} + \cdots + \frac{a_n}{a_0 + a_1 + \cdots + a_{n-1}} + \cdots$$

is so too. In fact, upon remarking that the quantities $a_0, a_1, a_2, \ldots$ are positive, one can, by virtue of the theorem $\log(1+x) < x$, demonstrate the following,

$$\log(a_0 + a_1 + a_2 + \cdots + a_n) - \log(a_0 + a_1 + a_2 + \cdots + a_{n-1}),$$

namely $\log\left(1 + \frac{a_n}{a_0 + a_1 + a_2 + \cdots + a_{n-1}}\right)$, is less than $\frac{a_n}{a_0 + a_1 + a_2 + \cdots + a_{n-1}}$. Thus when taking successively $n = 1, 2, 3, \ldots,$

$$\log(a_0 + a_1) - \log a_0 < \frac{a_1}{a_0},$$

$$\log(a_0 + a_1 + a_2) - \log(a_0 + a_1) < \frac{a_2}{a_0 + a_1},$$

$$\log(a_0 + a_1 + a_2 + a_3) - \log(a_0 + a_1 + a_2) < \frac{a_3}{a_0 + a_1 + a_2},$$

$$\ldots$$

$$\log(a_0 + a_1 + \cdots + a_n) - \log(a_0 + a_1 + \cdots + a_{n-1}) < \frac{a_n}{a_0 + a_1 + \cdots + a_{n-1}},$$

and, upon taking the sum,

$$\log(a_0 + a_1 + \cdots + a_n) - \log a_0 < \frac{a_1}{a_0} + \frac{a_2}{a_0 + a_1} + \cdots + \frac{a_n}{a_0 + a_1 + \cdots + a_{n-1}}.$$

But if the series $a_0 + a_1 + a_2 + \cdots + a_n$ is divergent, its sum is infinite, and the logarithm of this sum is so likewise, so the sum of the series

$$\frac{a_1}{a_0} + \frac{a_2}{a_0 + a_1} + \cdots + \frac{a_n}{a_0 + a_1 + \cdots + a_{n-1}}$$

is also infinitely large, and this series is consequently divergent if the series $a_0 + a_1 + a_2 + \cdots + a_{n-1}$ is.

Note the similarity between this argument and the one Abel uses to show the divergence of $\Sigma \frac{1}{n \log n}$. Finally, Abel uses the theorem he has just demonstrated (if a series $a_0 + a_1 + a_2 + a_3 + \cdots + a_n + \cdots$ diverges, then

$$\frac{a_1}{a_0} + \frac{a_2}{a_0 + a_1} + \frac{a_3}{a_0 + a_1 + a_2} \cdots + \frac{a_n}{a_0 + a_1 + \cdots + a_{n-1}} + \cdots$$

also diverges) to show that any series convergence test like Olivier's will lead to a contradiction.

This now stated, we suppose that φn be a function of n such that the series $a_0 + a_1 + a_2 + \cdots + a_n + \cdots$ is convergent or divergent according as $\varphi n \cdot a_n$ is zero or not for $n = \infty$. Then the series

$$\frac{1}{\varphi 1} + \frac{1}{\varphi 2} + \frac{1}{\varphi 3} + \frac{1}{\varphi 4} + \cdots + \frac{1}{\varphi n} + \cdots$$

will be divergent, and the series

$$\frac{1}{\varphi 2 \cdot \frac{1}{\varphi 1}} + \frac{1}{\varphi 3\left(\frac{1}{\varphi 1} + \frac{1}{\varphi 2}\right)} + \frac{1}{\varphi 4\left(\frac{1}{\varphi 1} + \frac{1}{\varphi 2} + \frac{1}{\varphi 3}\right)} + \cdots$$
$$+ \frac{1}{\varphi n\left(\frac{1}{\varphi 1} + \frac{1}{\varphi 2} + \frac{1}{\varphi 3} + \cdots + \frac{1}{\varphi(n-1)}\right)}$$

convergent; because in the first, one has $a_n \cdot \varphi n = 1$, and in the second, $a_n \cdot \varphi n = 0$ for $n = \infty$.

(The reader might consider why, in the second series, $a_n \cdot \varphi(n) \to 0$ as $n \to \infty$, as Abel claims.) At the end we see Abel apply his argument to Olivier's criterion (in which case $\varphi(n) = n$), encountering the same contradiction as he does in the general case.

Now following the theorem established before, the second series is necessarily divergent, whenever the first is, thus a function φn such as assumed here, doesn't exist. By taking $\varphi n = n$, the two series in question become

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots$$

and

$$\frac{1}{2 \cdot 1} + \frac{1}{3(1 + \frac{1}{2})} + \frac{1}{4(1 + \frac{1}{2} + \frac{1}{3})} + \cdots + \frac{1}{n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}\right)} + \cdots$$

which consequently are both divergent.

4. Conclusion

Counterexamples and proof by contradiction are standard tools for distinguishing necessary from sufficient conditions in deductive reasoning. Here we see a beautiful piece of mathematics employing both to clarify a misconception. It is interesting from a historical perspective to note that Abel proves a version of what we now call the Abel–Dini theorem, established in its general form in 1867 by U. Dini **[6]**.

THEOREM. *If* $\sum_{n=1}^{\infty} d_n$ *is an arbitrary divergent series of positive terms, and*

$$D_n = d_1 + d_2 + \cdots + d_n$$

are its partial sums, the series

$$\sum_{n=1}^{\infty} a_n \equiv \sum_{n=1}^{\infty} \frac{d_n}{D_n^{\alpha}}$$

converges when $\alpha > 1$ *and diverges when* $\alpha \leq 1$.

In his reply to Olivier, Abel proves a result similar to the case where $\alpha = 1$.

Although $\lim_{n\to\infty} n \cdot a_n = 0$ does not imply convergence of Σa_n, as claimed by Olivier in 1827, a partial converse holds for a series of positive, monotone decreasing terms:

$$\sum a_n \text{ converges} \Rightarrow \lim_{n\to\infty} n a_n = 0.$$

This is what Abel probably meant by remarking that "the latter part of this theorem is very true...".

Is there a universal test for convergence? Does a series exist that converges or diverges more slowly than any other? These and other fundamental questions about the convergence of positive-termed series were ultimately resolved by the close of the 19th century, combining work of many mathematicians.

Acknowledgment. Source material for this paper was drawn from a set of notes and readings prepared by Dr. Otto Bekken of Agder Högskolan, in Kristiansand, Norway. I would like to thank Dr. Bekken for sharing his love for and knowledge of mathematics history, and for his help and inspiration toward this work. I would also like to thank Dr. David Pengelley at New Mexico State University for his efforts to motivate current and future teachers to prepare teaching materials based upon original sources, for his help with translating, and for his patience in reviewing this paper.

EXERCISES.

1. Use the relationship $\log(1+x) < x$ to show that if a_n is positive, then

$$\log(a_0 + a_1 + a_2 + \cdots + a_n) - \log(a_0 + a_1 + a_2 + \cdots + a_{n-1}) < \frac{a_n}{a_0 + a_1 + a_2 + \cdots + a_{n-1}}.$$

2. Show that if the series

$$\frac{1}{\varphi 1} + \frac{1}{\varphi 2} + \frac{1}{\varphi 3} + \frac{1}{\varphi 4} + \cdots + \frac{1}{\varphi n} + \cdots$$

is divergent, and if Olivier's criterion holds for $\varphi(n)$, then the series

$$\frac{1}{\varphi 2\cdot\frac{1}{\varphi 1}}+\frac{1}{\varphi 3\left(\frac{1}{\varphi 1}+\frac{1}{\varphi 2}\right)}+\frac{1}{\varphi 4\left(\frac{1}{\varphi 1}+\frac{1}{\varphi 2}+\frac{1}{\varphi 3}\right)}+\cdots$$
$$+\frac{1}{\varphi n\left(\frac{1}{\varphi 1}+\frac{1}{\varphi 2}+\frac{1}{\varphi 3}+\cdots\frac{1}{\varphi(n-1)}\right)}$$

converges.

3. Show that the series

$$\frac{1}{2\cdot 1}+\frac{1}{3\left(1+\frac{1}{2}\right)}+\frac{1}{4\left(1+\frac{1}{2}+\frac{1}{3}\right)}+\cdots+\frac{1}{n\left(1+\frac{1}{2}+\frac{1}{3}+\cdots\frac{1}{n-1}\right)}+\cdots$$

diverges without using the theorem proven by Abel.

4. Prove that if $\{a_n\}$ is both positive and decreasing, and Σa_n converges, then $\lim_{n\to\infty} na_n = 0$.

REFERENCES

1. E. Hairer and G. Wanner, *Analysis by Its History*, Springer-Verlag, Berlin, Germany, 1996.
2. I. Grattan-Guiness, *Development of the Foundations of Mathematical Analysis from Euler to Riemann*, M.I.T. Press, Cambridge, MA, 1970.
3. L. Olivier, Remarques sur les series infinies et leur convergence, *Journal für die reine und angewandte Mathematik*, 2 (1827), 31–44.
4. J. Maxfield and M. Maxfield, *Abstract Algebra and Solution by Radicals*, W.B. Saunders Company, Philadelphia, PA, 1971.
5. N. Abel, Note sur le memoire de Mr. L. Olivier No. 4 du second tome de ce journal, ayant pour tire 'Remarques sur les series infinies et leur convergence,' *Journal für die reine und angewandte Mathematik*, 3 (1828), 79–81.
6. K. Knopp, *Theory and Application of Infinite Series*, Blackie and Son Ltd., London, UK, 1944.
7. O. Bekken, Abel and Convergence Tests, presented at the quadrennial meeting of the International Study Group on the Relations between History and Pedagogy of Mathematics (1992), Toronto, Canada, August 1992.

Apollonian Cubics: An Application of Group Theory to a Problem in Euclidean Geometry

PARIS PAMFILOS
University of Crete
Iraklion 71409
Greece

APOSTOLOS THOMA
University of Ioannina
Ioannina 45110
Greece

Introduction

Elliptic curves (nonsingular plane cubic curves) have a natural group structure. The so-called *chord and tangent addition rule* is based on the fact that, in general, a straight line intersects a cubic curve in three points. This group structure is one of the main tools in the study of *arithmetic* properties of elliptic curves ([**6**], [**9**]). Elliptic curves play a central role in the proof of Fermat's last theorem by A. Wiles [**14**] and they enter even in areas like cryptography, with Lenstra's elliptic curve algorithm [**7**]. The purpose of this article is to apply this simple group structure to study *geometric* properties of a particular class of cubic curves, arising in the following problem of Euclidean metric geometry (see FIGURE 1):

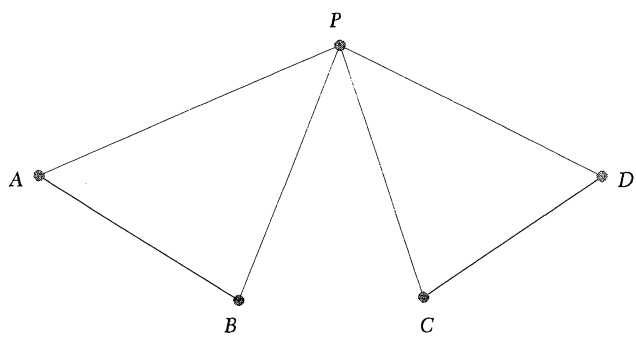


FIGURE 1

PROBLEM 1. Given two line segments AB and CD in the plane, find the locus of the points P such that the angles viewing AB and CD from P are equal or supplementary.

The locus turns out to be the union of two cubic curves (see FIGURE 4) with a number of very interesting algebraic and geometric properties. We call cubic curves like these *Apollonian cubics*, after the great geometer Apollonius of Perga (ca. 260–190 B.C.), who first considered a special case of this problem: the Apollonian circle of two segments BA and AC on the same line is part of the locus of the points P viewing the segments BA and AC under equal angles.

In 1852 Jacob Steiner stated Problem 1 and proved that the locus is described by a pair of cubics, both of which pass through the four points A, B, C, and D **[10]**. He was apparently unaware of the work of Van Rees who, in 1829, studied the same family of cubic curves in another context **[12]**. In 1915 Gomes Teixeira **[11]** remarked that these cubics are precisely the curves studied by Van Rees in their *normal form*

$$x(x^2+y^2)=a(x^2+y^2)-bx-cy.$$

The problem also attracted the interest of the geometers Brocard, Chasles, Dandelin, Darboux, Quetelet, and Salmon ([**3**], [**10**], [**11**]).

Cubic curves and the chord-tangent group structure

Cubics are plane curves defined by an equation $f(x, y)=0$, where $f(x, y)$ is a polynomial of third degree in the variables x and y. Many excellent books treat cubic plane curves, either from the arithmetic point of view ([**6**], [**9**]), or as a special class of algebraic plane curves ([**2**], [**5**], [**8**], [**13**]).

A cubic curve is called *irreducible* if the polynomial $f(x, y)$ is irreducible (i.e., it is not the product of other polynomials of lower degrees). A point (x_0, y_0) of an irreducible cubic is called *singular* if both partial derivatives at the point are zero, i.e., $f_x(x_0, y_0)=f_y(x_0, y_0)=0$. A nonsingular point is called *regular*; it has a uniquely defined tangent line. An irreducible cubic may have at most one singular point; in this case the cubic is called *singular*. There are two kinds of singular points: a *node* has two different tangents and a *cusp* has one double tangent. FIGURE 9 shows a singular cubic with a node.

In general, a line through two regular points of an irreducible cubic intersects the cubic in a regular point, making three points in total. The tangent line to a regular point of an irreducible cubic intersects the cubic again in a regular point and we have again three intersection points if we count the tangent point twice. It may happen that the tangent to a point intersects the cubic again in the same point. This point counts three times in the intersection of the cubic and the tangent line. Such points are called *flex points*. They are characterized, as in calculus, as the points where the curvature changes sign. (Note that irreducible cubics have at most *nine* flex points in the complex projective plane, of which at most three have real number coordinates.) The line joining two flex points intersects the cubic in a third flex point, as illustrated in FIGURE 2.

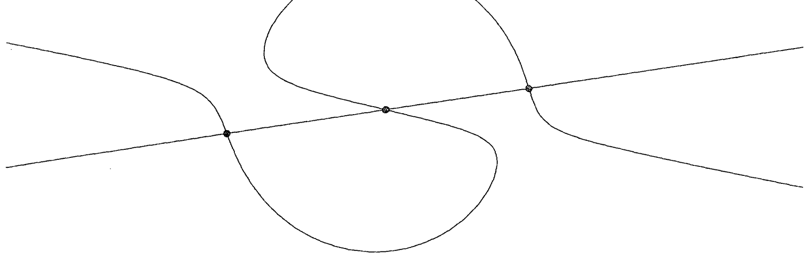


FIGURE 2
The line joining two flex points intersects the cubic in a third flex point.

Now let c be an irreducible plane cubic curve and O a flex point of c. The following chord-tangent law is used to define the group structure on the set of regular points of c, with O as the identity element: In order to add two different regular points A and B consider the chord through A, B that meets the curve in a third point P. The line joining O and P meets the curve at a third point which is defined to be the sum $A+B$, as shown in FIGURE 3.

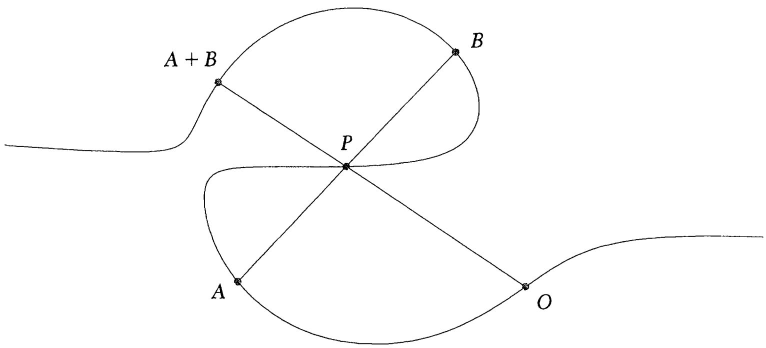


FIGURE 3
The addition law in a cubic.

To add a regular point C to itself, consider the tangent to the curve at C; it meets the curve in a third point Q. The line joining O and Q meets the curve in a third point, which is defined to be $C+C$ (or $2C$).

This addition turns the set of regular points of the curve into an abelian group, with O as the identity element (**[2]**, **[8]**, **[13]**). The geometry and this group structure on a cubic curve blend nicely, as the following theorem shows:

THEOREM 1. *Let $P_1, \dots, P_{3n}$ be any $3n$ regular points on a cubic. The points $P_1, \dots, P_{3n}$ are the intersection points with a curve of degree n if and only if $P_1 + \cdots + P_{3n} = O$.*

In particular, three regular points X, Y, and Z of the cubic are collinear if and only if $X+Y+Z=O$. In Theorem 1 we count *all* points of intersection, including points with complex coordinates and points at infinity, and some points may be counted more than once. For example, for the tangent line at a point C and its other intersection point, Q, we have $2C+Q=O$; for the tangent line at a flex point F, we have $3F=O$.

We include proofs in what follows since in most cases they are as elegant as the theorems, and they show how well geometry and algebra combine. Also, we will consider only the cases that the Apollonian cubics are irreducible. Note that for certain positions of the four points A, B, C, D, the associated Apollonian cubics are reducible. The reducible Apollonian cubics are either a circle and a line through its center or an equilateral hyperbola and the line at infinity.

Equations of Apollonian cubics

Translating Problem 1 into Cartesian coordinates and factoring shows that the equation of the locus is the product of two cubic polynomials. Indeed, let $A=(x_1, y_1)$, $B=(x_2, y_2)$, $C=(x_3, y_3)$, $D=(x_4, y_4)$, and

$$Q_i = (y-y_i)\prod_{i\neq j}(x-x_j) - (x-x_i)\prod_{i\neq j}(y-y_j),$$

where $1 \leq i,j \leq 4$. The equation of the locus is quite symmetrical:

$$[Q_1 - Q_2 - Q_3 + Q_4][Q_1 - Q_2 + Q_3 - Q_4] = 0. \tag{1}$$

The left side of equation (1) appears to be a product of two quartic polynomials, but in each bracket the fourth degree terms cancel, so each factor has degree three.

Thus the locus is the union of two cubic curves (see FIGURE 4), which we call Apollonian cubics. Apollonian cubics may have one or two connected components. Each cubic in FIGURE 4 has two components, while the cubic in FIGURE 5 has only one component. If an Apollonian cubic has two components, then one is bounded and the other is unbounded. Note also that in each cubic curve, some points view AB and CD under equal angles and some points of the same curve view AB and CD under supplementary angles. For example, in the cubic of FIGURE 4, for which the unbounded component passes through the points B and D, the points of the arc (BD) and the shorter arc (AC) view the segments AB and CD under supplementary angles. On the other hand, the points of the longer arc (AC) and the two unbounded arcs with endpoints B or D, view AB and CD under equal angles.

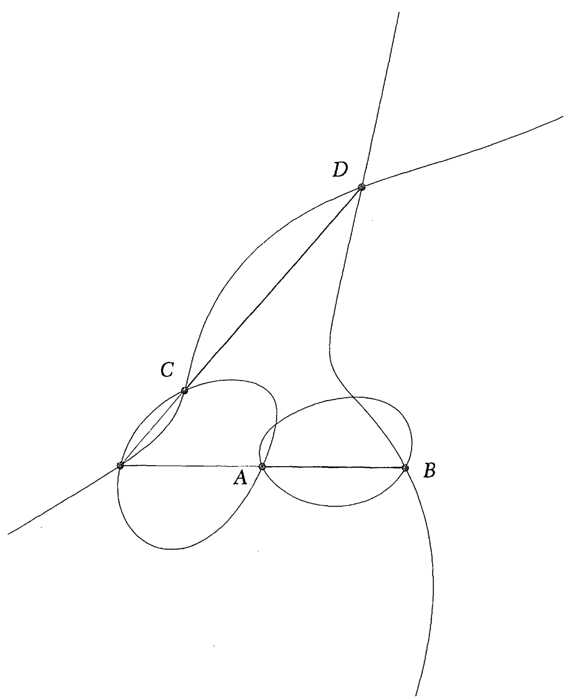


FIGURE 4
The two Apollonian cubics $AP[AD, BC]$ and $AP[AC, BD]$.

But why does the locus break into two cubics? *Does a geometric property distinguish the cubics in the two factors of equation* (1)?

The answer comes from observing the form of the two polynomials. Notice that interchanging B and C leaves the equation $Q_1 - Q_2 - Q_3 + Q_4 = 0$ invariant. This means that the points of this cubic view, not only the segments AB and CD, but also the segments AC and BD under equal or supplementary angles. We specify this cubic using the symbol $AP[AD, BC]$.

DEFINITION. The Apollonian cubic, denoted by $AP[AD, BC]$, is the cubic c such that for every point P of c,

(i) P views the line segments AB and CD under equal or supplementary angles;
(ii) P views the line segments AC and BD under equal or supplementary angles.

The other cubic, with equation $[Q_1 - Q_2 + Q_3 - Q_4] = 0$, is the Apollonian cubic $AP[AC, BD]$, i.e., the locus of points P that view the pairs of segments AB, CD and AD, BC under equal or supplementary angles.

The following theorem shows that the four points A, B, C, and D lie on the Apollonian cubic $AP[AD, BC]$, as also the two cyclic points at infinity, $I = [1, i, 0]$ and $I' = [i, 1, 0]$ (the two common points of all circles in the complex projective plane), and are related by means of the group structure.

THEOREM 2. *The four points A, B, C, D defining the Apollonian cubic $c = AP[AD, BC]$, the intersection point E of the lines AB and CD, the intersection point F of the lines AC and BD, and the two cyclic points at infinity, I and I', all lie on the curve. They satisfy the equations*

$$C = B + O_1, \quad D = A + O_1, \quad \text{and} \quad F = E + O_1,$$

where O_1 is one of the (at most) three points of order two on the cubic.

Proof. Let $L_{ij} = (y - y_i)(x - x_j) - (x - x_i)(y - y_j)$ and $K_{ij} = (x - x_i)(x - x_j) + (y - y_i)(y - y_j)$. Then $L_{ij} = 0$ is the equation of the line through the points (x_i, y_i) and (x_j, y_j), and $K_{ij} = 0$ is the equation of the circle whose diameter is the line segment $(x_i, y_i)(x_j, y_j)$. Then the equation $[Q_1 - Q_2 - Q_3 + Q_4] = 0$ can be written also in the form:

$$L_{12} K_{34} - L_{34} K_{12} = 0 = L_{13} K_{24} - L_{24} K_{13}.$$

The point $A = (x_1, y_1)$ is on the line L_{12} and on the circle K_{12}, so $L_{12}(x_1, y_1) = 0$ and $K_{12}(x_1, y_1) = 0$. This implies that $(L_{12} K_{34} - L_{34} K_{12})(x_1, y_1) = 0$. This means that A belongs to $AP[AD, BC]$. Similarly we see that the points B, C, and D, the intersection point E of the lines L_{12} and L_{34}, the intersection point F of the lines L_{13} and L_{24}, and the two cyclic points at infinity, I and I', belong to $AP[AD, BC]$.

Applying Theorem 1 to the lines L_{12}, L_{34}, L_{13}, and L_{24} gives

$$A + B + E = O; \quad C + D + E = O; \quad A + C + F = O; \quad B + D + F = O.$$

Eliminating E and F, we get $2(C - B) = O$ and $2(D - A) = O$. In group theory, these mean that the points $C - B$ and $D - A$ have order at most two. Geometrically, these mean that the tangents of c at the points $C - B$ and $D - A$ pass through the zero point O, since $2(C - B) + O = O$ and $2(D - A) + O = O$. But O is one of the flexes, $B \neq C$ and $A \neq D$, so there are at most three such different points, say O_1, O_2, O_3 (see **[13]** and FIGURE 8). Thus, $C = B + O_i$ and $D = A + O_j$. But $A + C = B + D$, so $O_i = O_j$. From now on we denote this point by O_1. Concerning the ambiguity in the choice of O_1, see the remark after Theorem 5 and also FIGURE 8. Substituting $B + O_1$ for C in $A + C + F = O$, and using $A + B + E = O$, we get $F = E + O_1$. □

The relations among the points A, B, C, and D in Theorem 2 can also be expressed as $B = C + O_1$ and $A = D + O_1$, since the point O_1 has order two ($2O_1 = O$ implies $O_1 = -O_1$).

The following theorem is quite similar to the chord-tangent theorem for circles. Its proof is typical of the way one uses the group structure of the cubic to get results in geometry.

THEOREM 3 (CHORD-TANGENT THEOREM). *Let $c = AP[AD, BC]$ be an irreducible Apollonian cubic. The angle between the chord AB and the tangent AX to c at A is equal or supplementary to the angle with which the point A views the segment CD.*

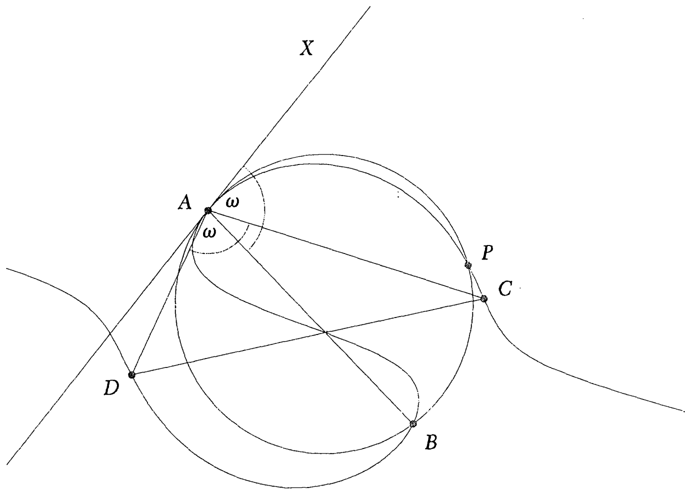


FIGURE 5
The chord-tangent theorem.

Proof. Let X be a point on the tangent line to c at the point A. Consider the circle that is tangent to the line AX at A and passes through B; see FIGURE 5. Let P be the other intersection point of the circle with c. Then $\angle XAB = \angle APB$ or $\angle XAB + \angle APB = \pi$. The point P lies on c, therefore $\angle APB = \angle CPD$ or $\angle APB + \angle CPD = \pi$. Since P, A, and B are intersection points of the circle and the cubic, Theorem 1 implies that

$$2A + B + P + I + I' = O. \qquad (2)$$

Here A is counted twice because the circle is tangent there to the cubic. The points I and I' are the *cyclic points*, contained in every circle of the complex projective plane. They have the property that every conic that passes through them and any three non-collinear real points is a circle. By adding $2O_1 = O$ and equation (2) we have also $A + (A + O_1) + (B + O_1) + P + I + I' = O$. Therefore the points A, $A + O_1 = D$, $B + O_1 = C$, and P lie also on a circle, so $\angle CPD = \angle CAD$ or $\angle CPD + \angle CAD = \pi$. We conclude that $\angle XAB = \angle CAD$ or $\angle XAB + \angle CAD = \pi$. □

The following proposition gives an extra point H on the cubic $AP[AD, BC]$, which will play a crucial role in our proof of Van Rees's theorem. Consider four points in the plane A, B, C, D and let E, F be respectively, the intersection points of the lines AB, CD and AC, BD.

> The four circles passing respectively through the points $\{A, B, F\}$, $\{C, D, F\}$, $\{E, A, C\}$, and $\{E, B, D\}$ have a common point (the *Miquel point*).

This is Miquel's theorem, discovered A. Miquel in 1838 [4]. The proof of the following proposition is also a simple algebraic proof of Miquel's theorem.

PROPOSITION. *Let $c = AP[AD, BC]$ be an irreducible Apollonian cubic, and let E, F be respectively, the intersection points of the lines AB, CD, and AC, BD. The four circles passing respectively through the points $\{A, B, F\}$, $\{C, D, F\}$, $\{E, A, C\}$ and $\{E, B, D\}$ intersect at a point H that belongs to the Apollonian cubic $c = AP[AD, BC]$.*

Proof. Every Apollonian cubic $AP[AD, BC]$ has at infinity the two cyclic points I, I' and a third point, denoted by L, and defined by the direction of its asymptote.

Applying Theorem 1 to the line at infinity, the line AC, and the circle $\{A, B, F\}$ we have respectively the three equations:

$$I + I' + L = O, \quad A + C + F = O, \quad \text{and} \quad A + B + F + H + I + I' = O,$$

where H is the other intersection point of the circle with the cubic $AP[AD, BC]$. Substituting $B = C + O_1$ in the last and canceling terms using the first two, we get $H = L + O_1$. The proof follows, since H is independent of the circle $\{A, B, F\}$. □

Van Rees's theorem

The following theorem (Van Rees, 1829 [**12**]) expresses the main property of the Apollonian cubics. The Apollonian cubic $AP[AD, BC]$ has been defined as the locus of the points viewing the pairs of segments AB, CD and AC, BD with equal or supplementary angles. Van Rees's theorem demonstrates that the points of the same cubic $AP[AD, BC]$ *also view infinitely many other pairs of segments with equal or supplementary angles*. In fact, if A', B' are *any* two points of $AP[AD, BC]$, then the points of this cubic view the pairs of segments $A'B', C'D'$ and $A'C', B'D'$ with equal or supplementary angles, where $C' = B' + O_1$ and $D' = A' + O_1$.

THEOREM 4 (VAN REES). *Let $c = AP[AD, BC]$ be an irreducible Apollonian cubic, let A' and B' be two arbitrary regular points on the cubic, and let $C' = B' + O_1$, $D' = A' + O_1$. Then the two Apollonian cubics $AP[AD, BC]$ and $AP[A'D', B'C']$ are identical.*

Proof. We apply an old (1764) and powerful theorem due to Bezout ([**1**], [**2**], [**13**]): *If two plane curves have more intersection points than the product of the degrees of their defining polynomials, then they have a common component.* Here we also count points with complex coordinates and points at infinity. Therefore, to prove that two cubics, one of them irreducible, are identical it is enough to prove that they have at least ten ($= 3 \cdot 3 + 1$) points in common. Whenever addition is used in the following proof, it is addition in the group of c.

Let P be the third intersection point of the line B, B' with the cubic c. Then $P + B + B' = O$ implies (adding $2O_1 = O$) $P + C + (B' + O_1) = O$. Thus $C' = B' + O_1$ is the third intersection point of the line PC with c (see FIGURE 6).

Consider the Apollonian cubic $AP[AD, B'C']$ (points viewing AB' and DC' under equal or supplementary angles, and also AC' and DB'). We claim that c and $AP[AD, B'C']$ have the following ten points in common:

$$P, \quad A, \quad B', \quad C', \quad D, \quad -A - B', \quad (-A - B') + O_1, \quad H, \quad I, \quad I'.$$

The above ten points are all points of c. The point P belongs to $c = AP[AD, BC]$, so it views AB and CD under equal or supplementary angles, as it does also for AC and BD. But B' is on the line PB and C' on the line PC. Therefore, P views AB' and

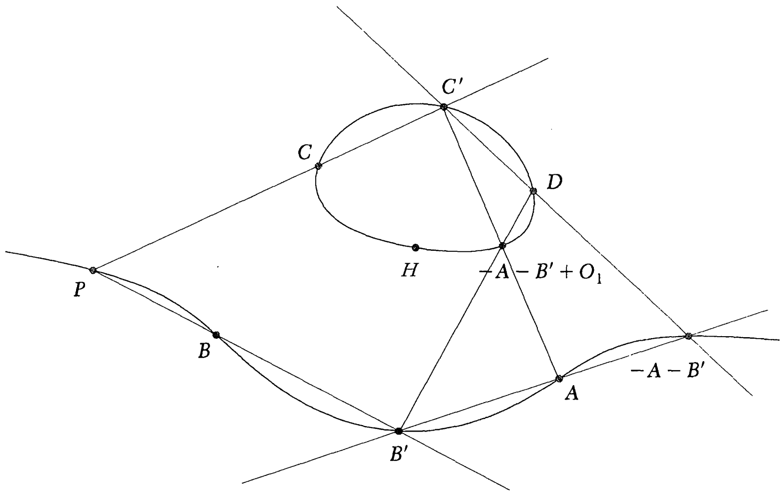


FIGURE 6
The proof of Van Rees's theorem.

$C'D$ under equal or supplementary angles and also AC' and $B'D$, which means that it belongs to the Apollonian cubic $AP[AD, B'C']$.

From Theorem 2, the points A, B', C', D, I, and I' are points of the Apollonian cubic $AP[AD, B'C']$, as also the intersection point of the lines AB' and $C'D$. The same holds for the intersection point of the lines AC' and $B'D$. But the last two points are the points $-A-B'$ and $R = -A - B' + O_1$ of c.

From the Proposition, the circles $\{AB'R\}$ and $\{DC'R\}$ intersect at a point belonging to the Apollonian cubic $AP[AD, B'C']$, the corresponding Miquel point. Let H be the Miquel point of c. Using the group structure on c we have from $A + (B' + O_1) + R = O$, $L + I + I' = O$ and $H = L + O_1$ that $A + B' + R + H + I + I' = O$, which means that H belongs to the circle $\{AB'R\}$, and similarly on $\{DC'R\}$. Therefore the two Miquel points are the same.

We conclude that the two cubics c and $AP[AD, B'C']$ have at least 10 points in common; by Bezout's theorem, they are identical. The same argument shows that $AP[AD, B'C']$ is identical with $AP[A'D', B'C']$.

Note that the proof is valid under the assumption that the ten points P, A, B', C', D, $-A-B'$, $(-A-B')+O_1$, H, I, and I' are all different. This is always true except for finitely many points B'. In these cases one has first to go through the above argument for an intermediate point B'' (i.e., proving $AP[AD, BC] = AP[AD, B''C'']$ $= AP[B''C'', B'C'] = AP[A'D', B'C']$), or one can complete the proof by a continuity argument. □

As an application of Van Rees's theorem we give the following theorem (see FIGURE 7), which also explains the right choice of the point O_1 in Theorem 2 (see FIGURE 8).

THEOREM 5 (FOUR TANGENTS). *Let c be an irreducible Apollonian cubic with two real components, P be any point on the unbounded component of c, and A be one of the four points of contact of the four tangents from P to c. Then*

(i) *the other three points of contact are $A + O_1$, $A + O_2$ and $A + O_3$;*
(ii) *the angles viewing the two components from P are equal or supplementary;*
(iii) *the line $A, A + O_1$ is orthogonal to the line $A + O_2, A + O_3$.*

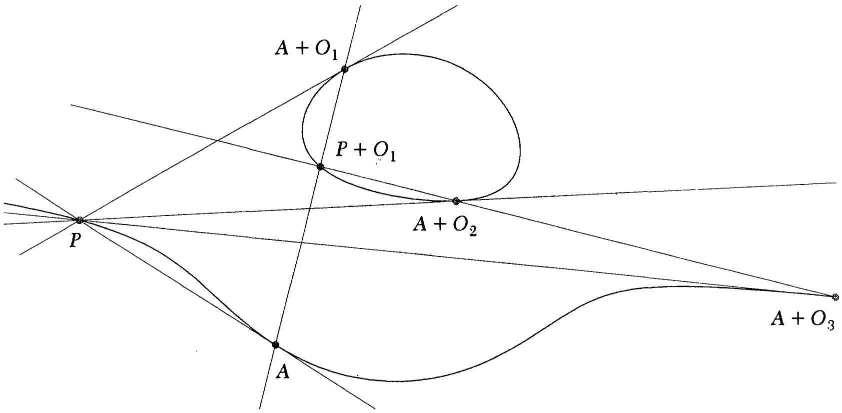


FIGURE 7
The four tangent theorem.

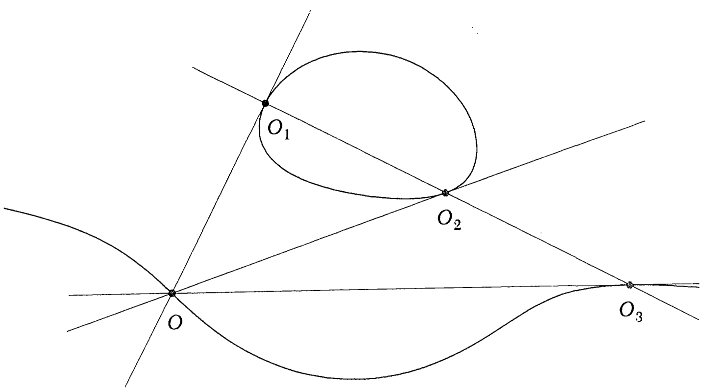


FIGURE 8
The right choice of O_1.

Proof. (i) Let Y be one of the other three points of contact. Then we have $2Y = 2A = -P$ or, equivalently, $2(Y - A) = O$. This means that $Y - A$ is a point of order two, so $Y = A + O_i$, for some $i \in \{1,2,3\}$.

(ii) For the points $A' = A$ and $B' = A + O_2$ we have $C' = B' + O_1 = A + O_2 + O_1 = A + O_3$ (note that $O_1 + O_2 + O_3 = O$ and each O_i has order two) and $D' = A + O_1$. The result follows by applying the Theorem 4.

(iii) Adding to the equation $2A + P = O$, the equations $2O_1 = O$ and $O_1 + O_2 + O_3 = O$, we get $A + (A + O_1) + (P + O_1) = O$ and $(A + O_2) + (A + O_3) + (P + O_1) = O$. Therefore $A(A + O_1)$ and $(A + O_2)(A + O_3)$ intersect at $P + O_1$. Take an arbitrary point X on c. Then $\angle X(P + O_1)A = \angle(X + O_1)(P + O_1)(A + O_1)$ or $\angle X(P + O_1)A + \angle(X + O_1)(P + O_1)(A + O_1) = \pi$. But A, $P + O_1$, and $A + O_1$ are collinear, so $A(A + O_1)$ is the interior or exterior bisector of the angle $\angle X(P + O_1)(X + O_1)$. Similarly $(A + O_2)(A + O_3)$ is the interior or exterior bisector of the angle $\angle X(P + O_1)(X + O_1)$. Since the two lines are different, one is the interior and the other the exterior bisector of the angle $\angle X(P + O_1)(X + O_1)$; the orthogonality result follows. □

REMARK. When $P = O$ (which is a flex point of the cubic), the four contact points are O, O_1, O_2, and O_3 (see FIGURE 8). According to Theorem 5, O_2O_3 is orthogonal to OO_1, and this determines O_1 uniquely among the points of order two. Note that the three points O_1, O_2, and O_3 are collinear (see **[13]**). In the cases where c has just one component or a node, there is just one real point of order two.

Singular Apollonian cubics

For certain positions of the points A, B, C, and D we have singular Apollonian cubics (see FIGURE 9). For example, it is easy to see from the equation of the curve $AP[AD, BC]$, that when $B = C$ the curve has a singularity at the point B which is a node and the two tangents at B are orthogonal. Actually the two tangents are the interior and exterior bisectors of the angle $\angle ABD$. All singular irreducible Apollonian cubics are of this kind. In fact, if the Apollonian cubic $c = AP[AD, BC]$ is singular and S is the singular point of c, then the two Apollonian cubics $AP[AD, BC]$ and $AP[AD, SS]$ are the same.

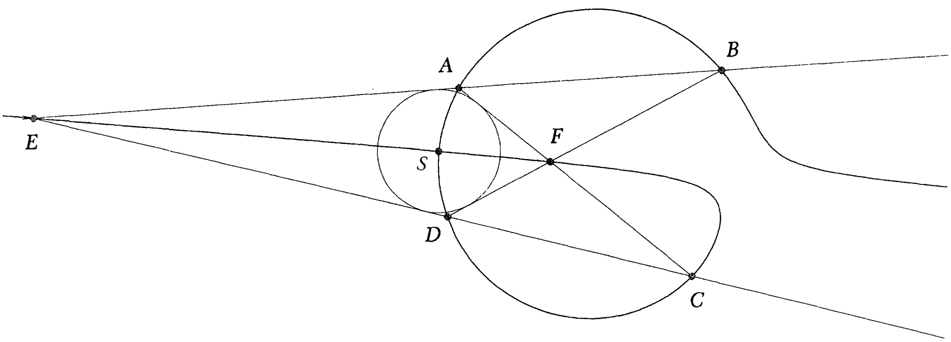


FIGURE 9
A singular Apollonian cubic.

It is interesting that there is a geometric property of the points A, B, C, D that makes the Apollonian cubic $AP[AD, BC]$ singular: *The irreducible cubic $c = AP[AD, BC]$ is singular if and only if there is a circle inscribed in the complete quadrilateral of the four lines AB, AC, BD, and CD*, as is illustrated in FIGURE 9. The center S of the circle coincides with the singular point of the cubic.

Further study

We have developed a computer program for those who wish to pursue further study of the Apollonian cubics, or want to experiment with drawing Apollonian cubics for various positions of the points A, B, C, and D. The program is available at `http://www.maa.org/pubs/mm_supplements/index.html`.

It includes a design-engine for the Apollonian cubics, exposition of the relevant theory, additional figures and exercises.

Acknowledgment. The authors thank the referees for valuable comments and suggestions.

REFERENCES

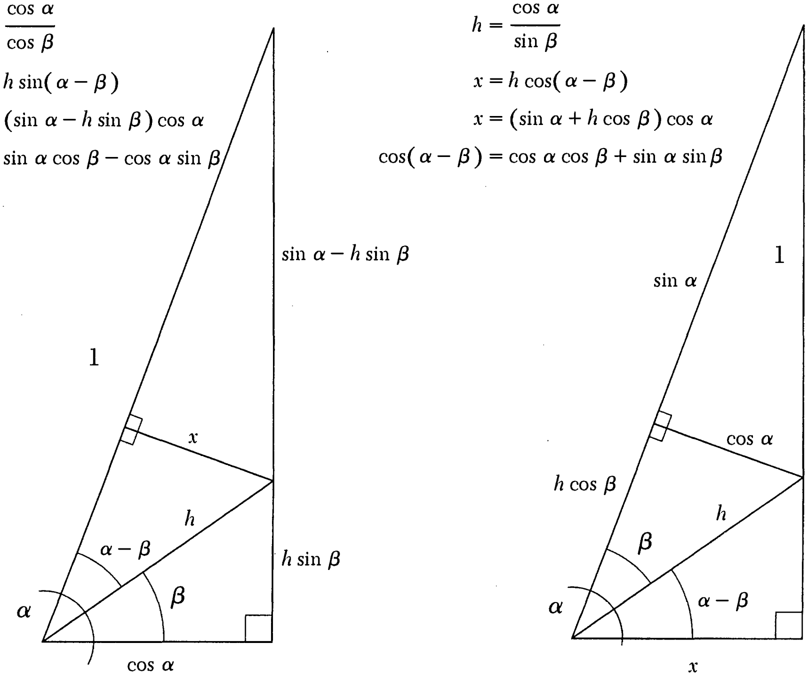
1. E. Bezout, Sur le degré des équations résultantes de l'évanouissement des inconnus, *Memoires présentes par divers savants à l'Academie des sciences de l'Institut de France*, (1764).
2. E. Brieskorn and H. Knörrer, *Plane Algebraic Curves*, Birkhäuser Verlag, Basel, Switzerland, 1986.
3. M. H. Brocard, Solutions de Questions Proposées, *Nouvelles Annales de Mathématiques*, Serie III, 15 (1915), p. 138.
4. H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited*, Math. Assoc. of America, Washington, DC, 1967.
5. W. Fulton, *Algebraic Curves: An Introduction to Algebraic Geometry*, W. A. Benjamin, Inc., New York, NY, 1969.
6. D. Husemöller, *Elliptic Curves*, Graduate Texts in Mathematics 111, Springer-Verlag, New York, NY, 1987.
7. H. W. Lenstra, Factoring integers with elliptic curves, *Annals of Math.* 126 (1987), pp. 649–673.
8. M. Reid, *Undergraduate Algebraic Geometry*, London Mathematical Society Student Texts 12, Cambridge University Press, Cambridge, UK, 1988.
9. J. Silverman, J. Tate, *Rational Points on Elliptic Curves*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, NY, 1992.
10. J. Steiner, Aufgaben und Lehrsätze, *Gesammelte Abhandlungen*, Bd II, p. 485, Chelsea, New York, NY, 1971.
11. G. Texeira, A propos de la question 1491, *Nouvelles Annales de Mathématiques*, Serie III, 15 (1915), p. 362.
12. Van Rees, Memoire sur les focales, *Correspondance mathématique et physique*, Quetelet, V (1829), pp. 361–378.
13. R. Walker, *Algebraic Curves*, Springer-Verlag, New York, NY, 1978.
14. A. Wiles, Modular elliptic curves and Fermat's Last Theorem, *Annals of Math.* 141 (1995), pp. 443–551.

Proof Without Words: Geometry of Subtraction Formulas

See `http://www.maa.org/pubs/mm_supplements/index.html` for Geometry of Addition Formulas (using the same diagram).

$$h = \frac{\cos\alpha}{\cos\beta}$$
$$x = h\sin(\alpha-\beta)$$
$$x = (\sin\alpha - h\sin\beta)\cos\alpha$$
$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$$

$$h = \frac{\cos\alpha}{\sin\beta}$$
$$x = h\cos(\alpha-\beta)$$
$$x = (\sin\alpha + h\cos\beta)\cos\alpha$$
$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$$



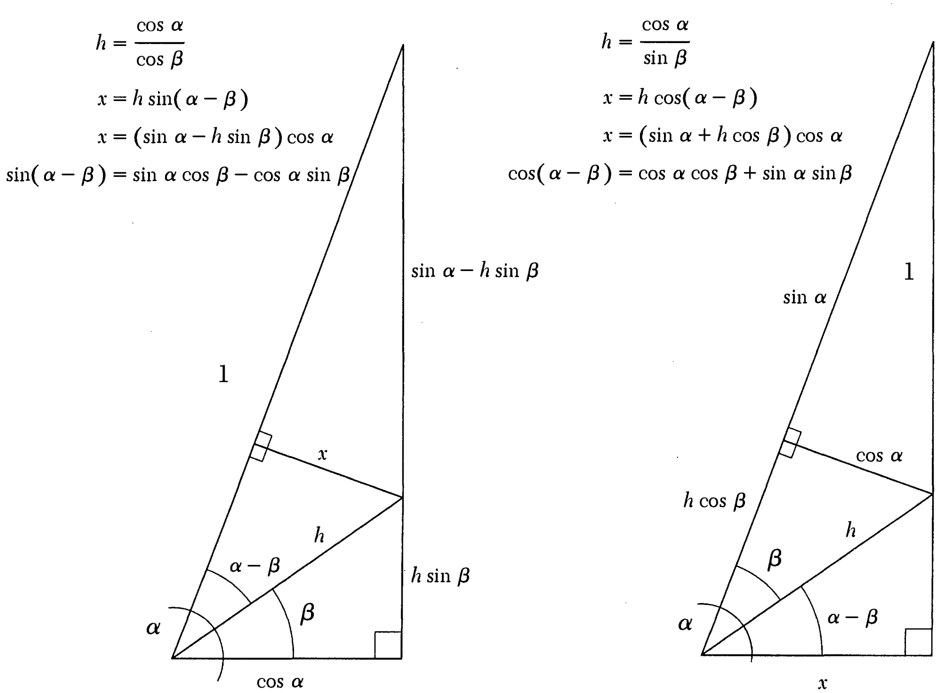
—Leonard M. Smiley
University of Alaska
Anchorage, AK 99508-8186

REFERENCES

1. E. Bezout, Sur le degré des équations résultantes de l'évanouissement des inconnus, *Memoires présentes par divers savants à l'Academie des sciences de l'Institut de France*, (1764).
2. E. Brieskorn and H. Knörrer, *Plane Algebraic Curves*, Birkhäuser Verlag, Basel, Switzerland, 1986.
3. M. H. Brocard, Solutions de Questions Proposées, *Nouvelles Annales de Mathématiques*, Serie III, 15 (1915), p. 138.
4. H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited*, Math. Assoc. of America, Washington, DC, 1967.
5. W. Fulton, *Algebraic Curves: An Introduction to Algebraic Geometry*, W. A. Benjamin, Inc., New York, NY, 1969.
6. D. Husemöller, *Elliptic Curves*, Graduate Texts in Mathematics 111, Springer-Verlag, New York, NY, 1987.
7. H. W. Lenstra, Factoring integers with elliptic curves, *Annals of Math.* 126 (1987), pp. 649–673.
8. M. Reid, *Undergraduate Algebraic Geometry*, London Mathematical Society Student Texts 12, Cambridge University Press, Cambridge, UK, 1988.
9. J. Silverman, J. Tate, *Rational Points on Elliptic Curves*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, NY, 1992.
10. J. Steiner, Aufgaben und Lehrsätze, *Gesammelte Abhandlungen*, Bd II, p. 485, Chelsea, New York, NY, 1971.
11. G. Texeira, A propos de la question 1491, *Nouvelles Annales de Mathématiques*, Serie III, 15 (1915), p. 362.
12. Van Rees, Memoire sur les focales, *Correspondance mathématique et physique*, Quetelet, V (1829), pp. 361–378.
13. R. Walker, *Algebraic Curves*, Springer-Verlag, New York, NY, 1978.
14. A. Wiles, Modular elliptic curves and Fermat's Last Theorem, *Annals of Math.* 141 (1995), pp. 443–551.

Proof Without Words: Geometry of Subtraction Formulas

See `http://www.maa.org/pubs/mm_supplements/index.html` for Geometry of Addition Formulas (using the same diagram).



—Leonard M. Smiley
University of Alaska
Anchorage, AK 99508-8186

Tangent Sequences, World Records, π, and the Meaning of Life: Some Applications of Number Theory to Calculus

IRA ROSENHOLTZ
Eastern Illinois University
Charleston, IL 61920

Introduction

It is easy to see that the series

$$\sum_{n=1}^{\infty} \frac{\sin(n)}{n^2}$$

converges. Less obvious is the fact that

$$\sum_{n=1}^{\infty} \frac{\sin(n)}{n}$$

also converges (though not absolutely). Relevant are the boundedness of $\sin(n)$ in the first case and the boundedness of $\sum_{n=1}^{N} \sin(n)$ in the second. It is natural to ask, "Do the series

$$\sum_{n=1}^{\infty} \frac{\tan(n)}{n}, \quad \sum_{n=1}^{\infty} \frac{\tan(n)}{n^2}, \quad \sum_{n=1}^{\infty} \frac{\tan(n)}{n^3}, \dots$$

converge?" A simpler related question is, "Do the terms of these series go to zero?" Even here, the answer is not obvious. While it is relatively easy to see that $\tan(n)$ eventually gets large in absolute value for certain integers n, it is not at all clear how $\tan(n)/n^k$ behaves for a fixed k. These tangent sequences are the subject of this article. We obtain the following three main results:

1. $\lim_{n \to \infty} \frac{\tan(n)}{n}$ does not exist;
2. $\lim_{n \to \infty} \frac{\tan(n)}{n^8} = 0$;
3. $|\tan(n)|/n^2$ is "small" (in a sense that will soon be made clear) for all n having fewer than *8,000,000 digits*!

It all involves some pretty applications of number theory. And the author knows of no other instance in which knowing several million digits of π is actually *useful*.

Preliminaries

In order for $|\tan(n)|$ to be "large," n must be "close to" an odd multiple (say m) of $\pi/2$. But then n/m must be "very close" to $\pi/2$. The best rational approximations to an irrational number are the "convergents" of its continued fraction expansion. So before proceeding with our results, let's take a moment and review some of the basics and notation (which is not standard in the literature) concerning continued fractions.

We'll write a continued fraction as

$$[a_1; a_2, a_3, \dots] = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}$$

where all of the a_k's are integers, and all are positive except possibly a_1. The *convergents* are the rational numbers

$$\frac{p_1}{q_1} = a_1, \quad \frac{p_2}{q_2} = a_1 + \frac{1}{a_2}, \quad \frac{p_3}{q_3} = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3}}, \quad \text{etc.}$$

Each irrational number α has a unique such continued fraction expansion $[a_1; a_2, a_3, \dots]$ such that $\alpha = \lim p_k/q_k$. Conversely, each such infinite sequence $[a_1; a_2, a_3, \dots]$ has convergents whose limit is an irrational number. We can thus identify α with its continued fraction expansion $[a_1; a_2, a_3, \dots]$.

Using this notation, $\pi/2 = [1; 1, 1, 3, 31, 1, 145, 1, \dots]$, and the first few convergents are shown in Table 1:

TABLE 1: Some convergents for $\pi/2$

k	a_k	p_k/q_k
1	1	1/1
2	1	2/1
3	1	3/2
4	3	11/7
5	31	344/219
6	1	355/226
7	145	51819/32989
8	1	52174/33215

Relationships among α, the a_k's, p_k's, and q_k's, and some well-known facts about these sequences which we will need throughout this article, are on exhibit in the following Gallery:

A Gallery of Facts About the Continued Fraction Expansion of an Irrational Number α

(a) $\alpha_1 = \alpha$; $a_n = [\alpha_n]$; $\alpha_{n+1} = 1/(\alpha_n - a_n)$.

(b) $p_{k+1} = a_{k+1} p_k + p_{k-1}$; $q_{k+1} = a_{k+1} q_k + q_{k-1}$.

(c) $p_1 < p_2 < p_3 < \dots$ and $q_1 \le q_2 < q_3 < \dots$.

(d) $p_k q_{k+1} - p_{k+1} q_k = (-1)^k$.

(e) $|p_k/q_k - \alpha| < 1/q_k^2$.

(f) p_k/q_k $(k > 1)$ is the "best rational approximation" to α in the sense that any fraction that is closer to α than p_k/q_k must have larger denominator.

(g) If $|p/q - \alpha| < 1/2q^2$ with p/q reduced and $q > 0$, then, for some k, $p/q = p_k/q_k$; that is, p/q must be a convergent for α.

(h) $1/q_k(q_k + q_{k+1}) < |p_k/q_k - \alpha| < 1/q_k q_{k+1}$.

Facts (a)–(e) may be found in almost any number theory book having a chapter on continued fractions. For facts (f)–(h), see, e.g., **[6]** and **[5]**.

First main result

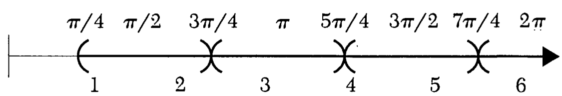
We begin with our first main result:

THEOREM 1. $\lim_{n\to\infty} \frac{\tan(n)}{n}$ *does not exist.*

We will show this in two steps: First, we'll show that if this limit exists, then it must be 0. Then we'll show the less obvious fact that this limit is not 0.

To prove the limit (if it exists) must be 0, we construct an increasing sequence of positive integers n_k so that $\tan(n_k)/n_k \to 0$. Let k be a positive integer, and let $n_k = [k\pi]$. Then $|n_k - k\pi| < 1 < \pi/3$, i.e., n_k differs from a multiple of π by less than $\pi/3$. Thus $|\tan(n_k)| < \sqrt{3}$ and $|\tan(n_k)|/n_k < \sqrt{3}\,/n_k$, and we are done.

Here is another reason why the limit of $\tan(n)/n$ would have to be 0. Consider the open intervals of radius $\pi/4$ centered at positive integer multiples of $\pi/2$:



While two consecutive integers may lie in the same one of these intervals (and often do), at least one of every *three* consecutive positive integers must lie in a different and adjacent one of these intervals. That is, at least one of every three consecutive positive integers, say N, lies within $\pi/4$ of some integer multiple of π. Thus $|\tan(N)|/N < 1/N$, and it is again easy to construct a sequence of positive integers N for which $\tan(N)/N \to 0$.

To prove that the limit of $\tan(n)/n$ *cannot* be 0, we'll construct an increasing sequence of integers p for which $\tan(p)/p$ is bounded away from 0. Let p_1/q_1, p_2/q_2, $p_3/q_3, \ldots$ be the convergents of the continued fraction expansion for $\pi/2$. Gallery Fact (d) implies that at least one of q_k or q_{k+1} must be odd. Let p/q be one of these fractions with q odd and $q > 1$. Then, from Gallery Fact (e), $|p/q - \pi/2| < 1/q^2$, and then $|p - q\pi/2| < 1/q < \pi/4$. Since q is odd, the cotangent function is differentiable on the interval $[q\pi/2 - \pi/4, q\pi/2 + \pi/4]$, so we may apply the mean value theorem:

$$|\cot(p)| = |\cot(p) - \cot(q\pi/2)| = |p - q\pi/2|\,|-\csc^2(z)|,$$

for some z between p and $q\pi/2$. (The first equality holds because q is odd.) But on this interval, $\csc^2(z) \le 2$, so $|\cot(p)| \le 2|p - q\pi/2|$. Therefore,

$$|\tan(p)| \ge \frac{1}{2|p - q\pi/2|} > \frac{q}{2} \ge \frac{p}{4},$$

because $p_k/q_k \le 2$ for all k. Thus, using Gallery Fact (c), there are infinitely many positive integers p so that $|\tan(p)|/p > 1/4$. This completes the proof.

Second main result

The next result is mostly a corollary of the hard work of others, so some historical background is in order. In 1953, K. Mahler **[7]** showed that $|\pi - p/q| > 1/q^{42}$ for all integers p, q with $q \ge 2$, and went on to indicate that 42 can be replaced by 30 if q exceeds some integer Q. The 30 was improved to 20 by M. Mignotte in 1974 **[8]**. In

1984, the Chudnovsky brothers **[2]** reduced the 20 to 14.65. Five years later, Borwein, Borwein, and Bailey **[1]** commented that the 14.65 irrationality estimate due to the Chudnovskys "is certainly not the best possible. ... It is likely that 14.65 should be replaced by $2+\epsilon$ for any $\epsilon > 0$. Almost all transcendental numbers satisfy such an inequality." In 1993, M. Hata **[4]** improved the 14.65 to 8.02. (See **[4]** for a more detailed history.)

THEOREM 2. $\lim_{n\to\infty} \dfrac{\tan(n)}{n^8} = 0.$

Proof. Hata's result states that there is a positive integer Q so that if $q \geq Q$ and r is any integer, then $|\pi - r/q| > 1/q^{8.02}$. In particular, letting $r = 2p$ gives $|\pi - 2p/q| > 1/q^{8.02}$. Multiplying by q and dividing by 2 yields the inequality

$$|q\pi/2 - p| > \frac{1}{2q^{7.02}}.$$

Now let p be a positive integer, and q be the unique positive integer so that $|p - q\pi/2| \leq \pi/4$. We may assume without loss of generality that q is "sufficiently large." If q is even, then $|\tan(p)| < 1$, so $|\tan(p)|/p^8 < 1/p^8$, which is very small. If q is odd, then we may apply the mean value theorem as in the proof above and obtain

$$|\cot(p)| = |p - q\pi/2||-\csc^2(z)| \geq 1|p - q\pi/2|.$$

It follows that

$$|\tan(p)| \leq \frac{1}{|p - q\pi/2|} < 2q^{7.02}.$$

Since $|p - q\pi/2| \leq \pi/4$ implies $q \leq p$, we see that $|\tan(p)|/p^8 < 2/p^{0.98}$, which is small. This completes the proof of Theorem 2.

The preceding argument, slightly modified, shows that if the aforementioned conjecture of Bailey and the Borweins is true, then $\tan(n)/n^{1+\epsilon} \to 0$ as $n \to \infty$ (in particular, $\tan(n)/n^2 \to 0$).

Third main result

How does someone go about showing that $|\tan(n)|/n^2$ is "small" for all n having at most several million digits? Certainly not by trying each and every one of these n's—no computer in the universe can do that. The key is to know which n's to look at. Here is the idea: If $\tan(n)/n^2$ does not approach zero as $n \to \infty$, then $\tan(n)/n$ must be unbounded. If this is the case, then there must be infinitely many $|\tan(n)|/n$ records. A positive integer N is called a $|\tan(n)|/n$ *record* if $|\tan(N)|/N > |\tan(n)|/n$ for all positive integers $n < N$. The following theorem tells us where to look for these records.

THEOREM 3. *If N is a $|\tan(n)|/n$ record, then N must be the numerator of a convergent of the continued fraction expansion for $\pi/2$.*

We will show this in several steps. Suppose n is a positive integer, and let m be the unique positive integer such that $|n - m\pi/2| \leq \pi/4$. As before, if m is even, then n is within $\pi/4$ of a multiple of π, so $|\tan(n)| < 1$, and so, for these n, $|\tan(n)|/n < 1 < |\tan(1)|/1$. So these n are not $|\tan(n)|/n$ records. Therefore we may assume m is odd, say $m = 2j+1$. Then $\cot(m\pi/2 - n) = \cot(\pi j + \pi/2 - n) = \cot(\pi/2 - n) = \tan(n)$. Rather than use the mean value theorem this time, we will get more information from the series expansion for the cotangent:

$$\cot(x) = \frac{1}{x} - \frac{x}{3} - \frac{x^3}{45} - \frac{2x^5}{945} - \cdots$$
$$= \frac{1}{x} + \sum_{k=1}^{\infty} \frac{2^{2k}(-1)^k B_{2k} x^{2k-1}}{(2k)!},$$

where B_{2k} is the $2k$-th Bernoulli number, and all the coefficients of the last sum are negative (see, e.g., **[9]**). The function

$$g(x) = \begin{cases} \dfrac{\cot(x) - 1/x}{x} & \text{if } x \neq 0 \\ -\dfrac{1}{3} & \text{if } x = 0 \end{cases}$$

is continuous on the interval $[-\pi/4, \pi/4]$, and (because all coefficients are negative) attains its maximum absolute value at the endpoints $\pm\pi/4$. Thus, for $x \neq 0$,

$$\frac{1}{|x|} - \alpha|x| \leq |\cot(x)| < \frac{1}{|x|},$$

where $\alpha = |g(\pm\pi/4)| < 0.348$, and hence

$$\frac{1}{|n - m\pi/2|} - .348|n - m\pi/2| < |\tan(n)| < \frac{1}{|n - m\pi/2|}.$$

We have proved the following result:

THEOREM 4. *If n is a positive integer, m is the unique positive integer such that $|n - m\pi/2| \leq \pi/4$, and m is odd, then*

$$\frac{1}{|n - m\pi/2|} - .348|n - m\pi/2| < |\tan(n)| < \frac{1}{|n - m\pi/2|}.$$

Thus, if m is odd and $|n - m\pi/2|$ is very small, $|\tan(n)|$ is very close to $1/|n - m\pi/2|$.

Theorem 4 lets us estimate $|\tan(n)|/n$ without having to calculate the tangent. Unfortunately, if m and n are huge integers, calculating the reciprocal of $n|n - m\pi/2|$ is still a relatively "expensive" computation, especially if it must be done for many n. We shall obtain a useful "cheap" estimation result as a corollary to Theorem 4, but first we return to the proof of Theorem 3.

After noting that both 1 and 11 are $|\tan(n)|/n$ records, that both are numerators of convergents for $\pi/2$, and that $|\tan(11)|/11 \approx 20.54 > 2$, it suffices to prove two Lemmas.

LEMMA 1. *Under the hypotheses of Theorem 4, if n/m (with m positive) is a reduced fraction and $|\tan(n)|/n > 2$, then n/m must be a convergent for $\pi/2$.*

Proof. By assumption and Theorem 4, we have

$$2 < \frac{|\tan(n)|}{n} < \frac{1}{n|n - m\pi/2|}.$$

This implies $|n/m - \pi/2| < 1/(2mn)$. Also, $|n - m\pi/2| < \pi/4$ implies $n \geq m$, so $|n/m - \pi/2| < 1/(2m^2)$. By Gallery Fact (g), n/m must be a convergent for $\pi/2$.

It remains to consider the case where n/m is not reduced. For example, $|\tan(33)|/33 \approx 2.28$, but 33 is not the numerator of a convergent for $\pi/2$. However, the value of m corresponding to $n = 33$ is 21, and $33/21 = 11/7$ which *is* a convergent, and $|\tan(33)|/33 < |\tan(11)|/11 \approx 20.54$. We show this situation is typical.

LEMMA 2. *Suppose m and n are positive integers, with m odd and $|n - m\pi/2| < \pi/4$. Assume $n = pd$ and $m = qd$ where p, q, and d are integers, with $d > 1$. Then $|\tan(p)|/p > d^2|\tan(n)|/n$. In particular, $|\tan(p)|/p > |\tan(n)|/n$.*

(So we can't set any records if n/m is not reduced.)

Proof. Our series representation for $\cot(x)$ implies that

$$|\cot(x)| = \frac{1}{|x|} - \frac{1}{3}|x| - \frac{1}{45}|x|^3 - \frac{2}{945}|x|^5 - \cdots$$

for $|x| \le \pi/4$. Therefore, after noticing that d and q must be odd and that $|p - q\pi/2| < \pi/4$, we have

$$\begin{aligned}\frac{|\tan(p)|}{p} &= \frac{|\cot(q\pi/2 - p)|}{p}\\ &= \frac{1}{p|q\pi/2 - p|} - \frac{|q\pi/2 - p|}{3p} - \frac{|q\pi/2 - p|^3}{45p} - \frac{2|q\pi/2 - p|^5}{945p}\cdots\end{aligned}$$

and

$$\begin{aligned}d^2\frac{|\tan(n)|}{n} &= d^2\frac{|\cot(m\pi - n)|}{n}\\ &= d^2\left(\frac{1}{n|m\pi/2 - n|} - \frac{|m\pi/2 - n|}{3n} - \frac{|m\pi/2 - n|^3}{45n} - \frac{2|m\pi/2 - n|^5}{945n}\cdots\right)\\ &= \frac{1}{p|q\pi/2 - p|} - \frac{d^2|q\pi/2 - p|}{3p} - \frac{d^4|q\pi/2 - p|}{45p} - \frac{2d^6|q\pi/2 - p|^5}{945p}\cdots.\end{aligned}$$

Thus,

$$\begin{aligned}&\frac{|\tan(p)|}{p} - d^2\frac{|\tan(n)|}{n}\\ &\quad = \frac{(d^2 - 1)|q\pi/2 - p|}{3p} + \frac{(d^4 - 1)|q\pi/2 - p|^3}{45p} + \frac{2(d^6 - 1)|q\pi/2 - p|^5}{945p} + \cdots\\ &\quad > 0.\end{aligned}$$

This completes the proof of Theorem 3.

We now return to the question of "cheaply" estimating $|\tan(p_k)|/p_k$, and derive the following estimation result as a consequence of Theorem 4.

COROLLARY 1. *If p_k/q_k is a convergent for $\pi/2 = [a_1; a_2, a_3 \ldots]$ with q_k odd, then $|\tan(p_k)|/p_k$ is within $2/3$ of $2(a_{k+1} + 1)/\pi$.*

Thus, the search for $|\tan(n)|/n$ records is more or less equivalent to searching for record a_{k+1}'s with q_k odd. The point of this last estimate is that if you are looking for $|\tan(n)|/n$ records, and you are willing to tolerate an error of 2/3, then this estimate allows you to limit the computations necessary after the a_k's have been computed: You need not calculate the tangent, nor the reciprocal of $p_k|p_k - q_k\pi/2|$, nor even the q_k's. You need only keep track of the parity of the q_k's (for which the q_k's themselves are unnecessary), and the p_k's if you wish—and I did. Unfortunately, I wasn't aware of this corollary until very late in the game.

Proof of Corollary 1. First, note that

$$q_{k+1} - \frac{.348}{q_{k+1}} < |\tan(p_k)| < q_{k+1} + q_k; \tag{A}$$

this follows immediately from Theorem 4 and Gallery Fact (h). Second, applying Gallery Facts (b) and (c) to the inequalities in (A) gives

$$\begin{aligned} a_{k+1}q_k &\leq (a_{k+1}q_k + q_{k-1}) - 1 = q_{k+1} - 1 < q_{k+1} - \frac{.348}{q_{k+1}} \\ &< |\tan(p_k)| < q_{k+1} + q_k = (a_{k+1}q_k + q_{k-1}) + q_k < (a_{k+1} + 2)q_k. \end{aligned}$$

Thus

$$\frac{a_{k+1}q_k}{p_k} < \frac{|\tan(p_k)|}{p_k} < \frac{(a_{k+1}+2)q_k}{p_k},$$

which implies

$$|\tan(p_k)|/p_k \text{ is within } q_k/p_k \text{ of } (a_{k+1}+1)q_k/p_k. \tag{B}$$

Next, note that Gallery Fact (h) implies

$$\left|\frac{q_k}{p_k} - \frac{2}{\pi}\right| = \frac{2\left|p_k - q_k\frac{\pi}{2}\right|}{\pi p_k} < \frac{2}{\pi p_k q_{k+1}}. \tag{C}$$

To complete the proof of Corollary 1, we calculate:

$$\begin{aligned} \left|\frac{|\tan(p_k)|}{p_k} - \frac{2(a_{k+1}+1)}{\pi}\right| &\leq \left|\frac{|\tan(p_k)|}{p_k} - \frac{q_k(a_{k+1}+1)}{p_k}\right| \\ &\quad + \left|\frac{q_k(a_{k+1}+1)}{p_k} - \frac{2(a_{k+1}+1)}{\pi}\right| \\ &< \frac{q_k}{p_k} + (a_{k+1}+1)\left|\frac{q_k}{p_k} - \frac{2}{\pi}\right| \qquad \text{(by (B))} \\ &< \frac{q_k}{p_k} + \frac{2(a_{k+1}+1)}{\pi p_k q_{k+1}} \qquad \text{(by (C))} \\ &< \frac{q_k}{p_k} + \frac{2(a_{k+1}+1)}{3p_k(a_{k+1}+1)} \\ &\left(\text{since } q_{k+1} = a_{k+1}q_k + q_{k-1} > a_{k+1} + 1 \text{ and } \frac{2}{\pi} < \frac{2}{3}\right) \\ &= \frac{q_k}{p_k} + \frac{2}{3p_k} \\ &< \frac{2}{\pi} + \frac{2}{\pi p_k q_{k+1}} + \frac{2}{3p_k} \qquad \text{(by (C))} \\ &< \frac{2}{\pi} + \frac{4}{3p_k}. \end{aligned}$$

The last quantity is less than $2/3$ if and only if $p_k > 2\pi/(\pi - 3) \approx 44.375\ldots$, i.e., for $k \geq 5$. But the cases where $k < 5$ and q_k odd (namely, $k = 1$, 2, and 4) are true by inspection. This completes the proof.

To summarize: If we desire to find $|\tan(n)|/n$ records, we need only look at numerators of those convergents of the continued fraction expansion of $\pi/2$ having odd denominators.

Searching for records

I began my search with the aid of the multidigit precision of *Maple* V. (One should be very skeptical of such calculations, but our estimation results provided comforting checks and balances.) Using over 120,000 digits of π, I was able to find the first eight $|\tan(n)|/n$ records in the tables below. This wasn't bad, since it implied that $|\tan(n)|/n^2$ is "small" for all $n \leq 10^{60,000}$. My computer's lack of memory prevented me from going further. But I was hungry for more.

I tried using *Mathematica* 3.0, but seemed to be having little luck. Finally, I wrote to Professor Stan Wagon of Macalester College, who has written on both *Mathematica* and π. He, in turn, contacted Dr. Mark Sofroniou of Wolfram Research; together, they obtained four new $|\tan(n)|/n$ records (and confirmed my old records) using a beta version of *Mathematica*. I find their results impressive, to say the least. With their help, I was able to use over 16 million digits of π, but found no new records. This still shows that $|\tan(n)|/n^2$ is "small" for all $n \leq 10^{8,000,000}$. So, for anyone interested in HUGE integers and tiny real numbers, here are the results. (The quotes on "Theorem" and "Corollary" and "Proof" are used because the results assume the accuracy of both hardware and software.)

"THEOREM" 5.

$|\tan(n)|/n$ records for $n \leq 10^{8,000,000}$			
k	p_k	a_{k+1}	$|\tan(p_k)|/p_k$
1	1	1	≈ 1.56
4	11	31	≈ 20.54
118	$\approx 1.32 \times 10^{60}$	84	≈ 54.52
136	$\approx 1.40 \times 10^{69}$	116	≈ 74.77
315	$\approx 2.37 \times 10^{154}$	873	≈ 556.31
3727	$\approx 1.86 \times 10^{1940}$	4319	≈ 2750.20
3763	$\approx 5.53 \times 10^{1961}$	16555	≈ 10539.85
15503	$\approx 6.94 \times 10^{7992}$	38112	≈ 24263.75
153396	$\approx 1.72 \times 10^{78922}$	67828	≈ 43181.13
156559	$\approx 1.79 \times 10^{80545}$	358274	≈ 228085.42
984404	$\approx 1.37 \times 10^{506889}$	372743	≈ 237296.33
1119377	$\approx 5.37 \times 10^{576450}$	16186423	≈ 10304597.53

The fifth line of this table means, for example, that p_{315} is a number near 2.37×10^{154}, $|\tan(p_{315})|/p_{315}$ is the new record maximum, a number near 556.31, and that $|\tan(n)|/n$ is less than or equal to this value for all $n < p_{3727} \approx 1.86 \times 10^{1940}$, at which point a new record is set.

I must admit that contemplating the gigantic p_k's makes me feel like a little kid again. To get a feel for the size of these numbers, the exact value of the 8th record (corresponding to $k = 15503$) takes about *3 full pages* to print! You can imagine what the later records must look like.

Applying Theorem 5 to $\tan(n)/n^2$, we obtain the following results:

"COROLLARY" 2.

Estimates for $\lvert\tan(n)\rvert/n^2$ for $n \le 10^{8,000,000}$		
For	$\lvert\tan(n)\rvert/n^2$ is at most	
$1 \le n < p_4$	$\lvert\tan(p_1)\rvert/(p_1)^2$	(≈ 1.56)
$p_4 \le n < p_{118}$	$\lvert\tan(p_4)\rvert/(p_4)^2$	(≈ 1.87)
$p_{118} \le n < p_{136}$	$\lvert\tan(p_{118})\rvert/(p_{118})^2$	$(\approx 4.14 \times 10^{-59})$
$p_{136} \le n < p_{315}$	$\lvert\tan(p_{136})\rvert/(p_{136})^2$	$(\approx 5.33 \times 10^{-68})$
$p_{315} \le n < p_{3727}$	$\lvert\tan(p_{315})\rvert/(p_{315})^2$	$(\approx 2.34 \times 10^{-152})$
$p_{3727} \le n < p_{3763}$	$\lvert\tan(p_{3727})\rvert/(p_{3727})^2$	$(\approx 1.48 \times 10^{-1937})$
$p_{3763} \le n < p_{15503}$	$\lvert\tan(p_{3763})\rvert/(p_{3763})^2$	$(\approx 1.90 \times 10^{-1958})$
$p_{15503} \le n < p_{153396}$	$\lvert\tan(p_{15503})\rvert/(p_{15503})^2$	$(\approx 3.50 \times 10^{-7989})$
$p_{153396} \le n < p_{156559}$	$\lvert\tan(p_{153396})\rvert/(p_{153396})^2$	$(\approx 2.51 \times 10^{-78918})$
$p_{156559} \le n < p_{984404}$	$\lvert\tan(p_{156559})\rvert/(p_{156559})^2$	$(\approx 1.27 \times 10^{-80540})$
$p_{984404} \le n < p_{1119377}$	$\lvert\tan(p_{984404})\rvert/(p_{984404})^2$	$(\approx 1.72 \times 10^{-506884})$
$p_{1119377} \le n \le 10^{8,000,000}$	$\lvert\tan(p_{1119377})\rvert/(p_{1119377})^2$	$(\approx 1.92 \times 10^{-576444})$

"Proof." Smaller numerators and larger denominators make for smaller fractions. In particular, if M and N are consecutive $|\tan(n)|/n$ records, and $M \le n < N$, then

$$\frac{|\tan(n)|}{n^2} = \frac{|\tan(n)|/n}{n} \le \frac{|\tan(M)|/M}{n} \le \frac{|\tan(M)|}{M^2}.$$

Thus, for $n \le 10^{8,000,000}$, $|\tan(n)|/n^2 < 11{,}000{,}000/n$.

Conclusions

If $\lim_{n\to\infty} \frac{|\tan(n)|}{n^2} \ne 0$, then it sure is taking its time giving any indication of this. Of course, the tables above don't *prove* anything (except perhaps that $|\tan(n)|/n^2$ is small for all $n \le 10^{8 \text{ million}}$!). As Richard K. Guy observes in **[3]**: "You can't tell by looking," and "There aren't enough small numbers to meet the many demands made of them." But perhaps a *conjecture* is in order. And if, in fact, $\lim_{n\to\infty} \frac{|\tan(n)|}{n^2} \ne 0$, wouldn't this make a wonderful addition to Guy's lists of eventually failing patterns?

No wonder calculus texts don't include this limit in the exercises!

Readers may be interested in similar limits involving other trigonometric functions. The study of $|\cot(n)|/n$ records is similar, and slightly simpler, than that of $|\tan(n)|/n$ records. Except for the trivial record $n = 1$, records occur only at integers that are numerators of convergents of π (rather than $\pi/2$), and the denominator need not be odd. The largest record for $n \le 10^{8,000,000}$ occurs at index $k = 11504932$, where $p_k \approx 4.83 \times 10^{5929967}$, $a_{k+1} = 878783625$, and $|\cot(p_k)|/p_k$ is within $1/3$ of $(a_{k+1}+1)/\pi \approx 279725515.97$. Of course, $|\sec(n)|/n$ records and $|\csc(n)|/n$ records follow from the $|\tan(n)|/n$ and $|\cot(n)|/n$ records, respectively, using trigonometric identities.

We leave the proof of our final corollary as an exercise for the interested reader.

"COROLLARY" 3. $$\sum_{n=1}^{10^{8,000,000}} \frac{|\tan(n)|}{n^3} < 3.$$

Acknowledgment. I'd like to thank my wife, Susan, for bearing with me as always, my son, Jeremy, for buying me a delightful book on π for my birthday, and my daughter, Ruth, for pointing out the following passage from a FORTRAN manual:

> The primary purpose of the DATA statement is to give names to constants; instead of referring to π as 3.141592653589793 at every appearance, the variable PI can be given that value with a DATA statement and used instead of the longer form of the constant. This also simplifies modifying the program should the value of pi change.

No comment.

Thanks to Stan Wagon, Mark Sofroniou, and Jamie Peterson for their generous help with *Mathematica*, and to Peter Borwein for pointing out reference [**4**]. Thanks to my colleagues at Eastern for lending me their ears (as well as some of their "memory").

Finally, thanks to the editor, Paul Zorn, and the referees for many helpful suggestions, and for suffering through multiple versions of this article. It truly was (and still is!) a work in progress.

REFERENCES

1. J. M. Borwein, P. B. Borwein, and D. H. Bailey, Ramanujan, modular equations, and approximations to π, or, How to compute one billion digits of π, *Amer. Math. Monthly* 96 (1989), 201–219.
2. D. V. Chudnovsky and G. V. Chudnovsky, Padé and Rational Approximations to Systems of Functions and their Arithmetic Applications, Lecture Notes in Mathematics #1052, Springer-Verlag, Berlin, Germany, 1984, 37–84.
3. R. K. Guy, The strong law of small numbers, *Amer. Math. Monthly* 95 (1988), 697–712.
4. M. Hata, Rational approximations to π and some other numbers, *Acta Arithmetica* 63 (1993), 335–349.
5. A. Ya. Khinchin, *Continued Fractions*, Dover, New York, NY, 1997, 1–50.
6. W. J. LeVeque, *Fundamentals of Number Theory*, Dover, New York, NY, 1996, 226–238.
7. K. Mahler, On the approximation of π, *Nederl. Akad. Wetensch. Proc. Ser. A*, 56 (1953), 30–42.
8. M. Mignotte, Approximations rationnelles de π et quelques autres nombres, *Bull. Soc. Math. France Mém.* 37 (1974), 121–132.
9. W. Scharlau and H. Opolka, *From Fermat to Minkowski; Lectures on the Theory of Numbers and Its Historical Development*, Springer-Verlag, New York, NY, 1985, 19 ff.

NOTES

Raising the Roots

AL CUOCO
Center for Mathematics Education
EDC
55 Chapel Street
Newton, MA 02458

Without knowing the roots of a given polynomial equation, how do you find the equation whose roots are some fixed power of the roots of the original?

This simple question can connect many topics from high school and undergraduate mathematics. This paper takes one path through the question, showing applications of ideas from algebra and linear algebra and previewing some more advanced topics.

Quadratic equations Let's start with degree 2, and suppose our quadratic equation is written as $x^2 - 2abx + b^2 = 0$. If its roots are α_1 and α_2, then

$$\alpha_1 + \alpha_2 = 2ab \quad \text{and} \quad \alpha_1\alpha_2 = b^2.$$

An equation whose roots are α_1^k and α_2^k is

$$x^2 - \left(\alpha_1^k + \alpha_2^k\right)x + \alpha_1^k\alpha_2^k = 0$$

so the problem comes down to finding $\alpha_1^k + \alpha_2^k$ and $\alpha_1^k\alpha_2^k$ in terms of the coefficients of the original equation.

Well, the product is easy:

$$\alpha_1^k\alpha_2^k = (\alpha_1\alpha_2)^k = b^{2k}.$$

So the object of the game is to express $\alpha_1^k + \alpha_2^k$ in terms of a and b. Let's look for a recursion.

$$\begin{aligned}\alpha_1^2 + \alpha_2^2 &= (\alpha_1 + \alpha_2)(\alpha_1 + \alpha_2) - 2\alpha_1\alpha_2\\ &= (2ab)^2 - 2b^2\\ &= 2b^2(2a^2 - 1)\end{aligned}$$

$$\begin{aligned}\alpha_1^3 + \alpha_2^3 &= (\alpha_1 + \alpha_2)\left(\alpha_1^2 + \alpha_2^2\right) - \alpha_1\alpha_2(\alpha_1 + \alpha_2)\\ &= 2ab\left(2b^2(2a^2 - 1)\right) - b^2(2ab)\\ &= 2b^3\left(2a(2a^2 - 1) - a\right)\\ &= 2b^3(4a^3 - 3a)\end{aligned}$$

$$\begin{aligned}\alpha_1^4+\alpha_2^4 &= (\alpha_1+\alpha_2)(\alpha_1^3+\alpha_2^3)-\alpha_1\alpha_2(\alpha_1^2+\alpha_2^2)\\ &= 2ab(2b^3(4a^3-3a))-b^2(2b^2(2a^2-1))\\ &= 2b^4(2a(4a^3-3a)-(2a^2-1))\\ &= 2b^4(8a^4-8a^2+1)\\ &\vdots\qquad\vdots\end{aligned}$$

Inductively, we see that $\alpha_1^k+\alpha_2^k=2b^kt_k(a)$, where $t_k(a)$ is a polynomial in a of degree k. In fact,

$$t_1(a)=a,\quad t_2(a)=2a^2-1,\quad t_3(a)=4a^3-3a,\quad \text{and}\quad t_4(a)=8a^4-8a^2+1.$$

Furthermore, for $k>2$, we have

$$\begin{aligned}2b^kt_k(a) &= \alpha_1^k+\alpha_2^k\\ &= (\alpha_1+\alpha_2)(\alpha_1^{k-1}+\alpha_2^{k-1})-\alpha_1\alpha_2(\alpha_1^{k-2}+\alpha_2^{k-2})\\ &= 2ab(2b^{k-1}t_{k-1}(a))-b^2(2b^{k-2}t_{k-2}(a))\\ &= 2b^k(2at_{k-1}(a)-t_{k-2}(a)),\end{aligned}$$

so that $t_k(a)=2at_{k-1}(a)-t_{k-2}(a)$, and we have a recursion for the t_k:

$$t_k(a)=\begin{cases}a & \text{if } k=1;\\ 2a^2-1 & \text{if } k=2;\\ 2at_{k-1}(a)-t_{k-2}(a) & \text{if } k>2.\end{cases}$$

These are the *Chebyshev polynomials* (one usually starts with $t_0(a)=1$, an equation that makes sense in the current context). Much is known about these polynomials; one of their many beautiful properties is that they yield trigonometric identities. This can be seen as follows:

Let $\theta\in\mathbb{R}$, $\alpha=\cos\theta+i\sin\theta$, and $\beta=\cos\theta-i\sin\theta$. Then α and β are roots of $x^2-2ax+1$, where $a=\cos\theta$. So,

$$\begin{aligned}2t_k(a) &= \alpha^k+\beta^k\\ &= (\cos\theta+i\sin\theta)^k+(\cos\theta-i\sin\theta)^k\\ &= 2\cos k\theta.\end{aligned}$$

Since $a=\cos\theta$, we have $t_k(\cos\theta)=\cos k\theta$, so that the t_k provide a machine for generating the double, triple, quadruple, ... angle formulas for cosine:

$$\begin{aligned}\cos 2\theta &= t_2(\cos\theta)=2\cos^2\theta-1\\ \cos 3\theta &= t_3(\cos\theta)=4\cos^3\theta-3\cos\theta\\ \cos 4\theta &= t_3(\cos\theta)=8\cos^4\theta-8\cos^2\theta+1\\ &\vdots\qquad\vdots\end{aligned}$$

Higher degrees Next, we let the degree increase and keep k constant. Suppose $a_0+a_1x+a_2x^2+\cdots+a_nx^n=f(x)$ has roots $\{\alpha_1,\dots,\alpha_n\}$. How do you find an equation whose roots are the α_i^k?

The situation is more complicated here, because there's more to worry about than the sum and the product of the roots. But the question should be answerable. Since

$$f(x)=a_n(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n), \tag{1}$$

the coefficients of f are the *elementary symmetric functions* of the roots:

$$\alpha_1 + \alpha_2 \cdots + \alpha_n = -\frac{a_{n-1}}{a_n}$$

$$\alpha_1\alpha_2 + \alpha_1\alpha_3 \cdots + \alpha_{n-1}\alpha_n = \frac{a_{n-2}}{a_n}$$

$$\sum_{h<i<j} \alpha_h\alpha_i\alpha_j = -\frac{a_{n-3}}{a_n}$$

$$\vdots \qquad \vdots$$

$$\alpha_1\alpha_2 \cdots \alpha_n = (-1)^n\frac{a_0}{a_n}$$

and to find an equation satisfied by the α_i^k, we'd have to express these symmetric functions of the α_i^k in terms of the a_i. A simple one is the constant term:

$$\alpha_1^k\alpha_2^k \dots \alpha_n^k = \left((-1)^n\frac{a_0}{a_n}\right)^k.$$

Newton gave a formula for the sum of the k^{th} powers of the roots in terms of the a_i. And the *theorem on symmetric functions* says that *any* symmetric function of the roots (including the ones we care about) can be expressed as a polynomial in the a_i (for details, see [**1**]). The expressions can get quite complex. It turns out that there is another approach that allows the calculation of the equation whose roots are powers of the roots of a given equation without recourse to symmetric functions. It works like this:

Suppose we want to find the equation satisfied by the squares of the roots of $f(x)=0$. Replace x by $-x$ in (1):

$$f(-x) = a_n(-x-\alpha_1)(-x-\alpha_2)\dots(-x-\alpha_n). \tag{2}$$

Multiply this together with (1):

$$f(x)f(-x) = \pm a_n^2(x^2-\alpha_1^2)(x^2-\alpha_2^2)\dots(x^2-\alpha_n^2). \tag{3}$$

But $f(x)f(-x)$ is a polynomial in x^2, say $g(x^2)$. Now replace x^2 by y on both sides of (3), and you get

$$g(y) = \pm a_n^2(y-\alpha_1^2)(y-\alpha_2^2)\dots(y-\alpha_n^2).$$

So $g(y)=0$ is an equation whose roots are the α_i^2.

For the cubes of the roots, we use a similar process:

Suppose, as before, that $f(x) = a_0 + a_1x + \cdots + a_nx^n$ has zeros $\alpha_1, \dots, \alpha_n$. To simplify the calculations, group the terms of f by powers of 3, and let

$$A = a_0 + a_3x^3 + a_6x^6 + \cdots; \quad B = a_1 + a_4x^3 + a_7x^6 + \cdots;$$
$$C = a_2 + a_5x^3 + a_8x^6 + \cdots.$$

Then $f(x) = A + Bx + Cx^2$. Now let $\omega = e^{2\pi i/3}$. Replace x by ωx and $\omega^2 x$ to get three equations:

$$f(x) = A + Bx + Cx^2 = (x-\alpha_1)(x-\alpha_2)\dots(x-\alpha_n); \tag{4}$$

$$f(\omega x) = A + B\omega x + C\omega^2x^2 = (\omega x-\alpha_1)(\omega x-\alpha_2)\dots(\omega x-\alpha_n); \tag{5}$$

$$f(\omega^2 x) = A + B\omega^2 x + C\omega x^2 = (\omega^2 x-\alpha_1)(\omega^2 x-\alpha_2)\dots(\omega^2 x-\alpha_n). \tag{6}$$

I've used here the fact that A, B, and C are invariant under the substitutions (because they are polynomials in x^3). Multiply (4), (5), and (6) together to get:

$$\begin{aligned} f(x)f(\omega x)f(\omega^2 x) &= A^3 + (B^3 - 3ABC)x^3 + C^3 x^6 \\ &= \pm(x^3 - \alpha_1^3)(x^3 - \alpha_2^3)\ldots(x^3 - \alpha_n^3). \end{aligned} \tag{7}$$

This is a polynomial in x^3, so you can substitute y for x^3 on both sides, and you have a polynomial whose roots are the α_i^3.

EXAMPLE. Suppose $f(x) = -156 + 124x - 75x^2 + 35x^3 - 9x^4 + x^5$. The roots of f are $\{3, \pm 2i, 3 \pm 2i\}$. To get the equation whose roots are the squares of these, we want

$$f(x)f(-x) = 24336 + 8024x^2 - 247x^4 - 123x^6 + 11x^8 - x^{10}.$$

Put $y = x^2$ to get a fifth-degree polynomial whose roots are the squares of the roots of f (as can be checked by direct substitution):

$$24336 + 8024y - 247y^2 - 123y^3 + 11y^4 - y^5.$$

For the cubes, write f as $f(x) = A + Bx + Cx^2$, where

$$A = -156 + 35x^3, \quad B = 124 - 9x^3, \quad \text{and} \quad C = -75 + x^3.$$

Then, our "norm equation" (7) says that

$$f(x)f(\omega x)f(\omega^2 x) = A^3 + (B^3 - 3ABC)x^3 + C^3 x^6,$$

which simplifies to

$$-3796416 + 109504x^3 - 59895x^6 + 1775x^9 - 9x^{12} + x^{15}.$$

Putting $y = x^3$, we get a polynomial whose roots are the cubes of the roots of f (a fact one can check by direct substitution):

$$-3796416 + 109504y - 59895y^2 + 1775y^3 - 9y^4 + y^5.$$

Doing it in general To form the equation whose roots are the kth powers of the roots of f, we'd have to form the product of all the "conjugates" of f:

$$f(x) \cdot f(\zeta x) \cdot f(\zeta^2 x) \cdot f(\zeta^3 x) \cdots f(\zeta^{k-1} x),$$

where $\zeta = \cos\frac{2\pi}{k} + i \sin\frac{2\pi}{k}$. This process of forming all the conjugates of a polynomial and multiplying these together leads to messy calculations with complex numbers. Fortunately, there's more classical mathematics that can help out, allowing us to work entirely with polynomials over the original coefficient ring.

A useful idea in algebra is to let an element of a system "act on" the system, looking at the element as both a member of the system and a function on the system and using the algebra in the system to define function application. Understanding this action often leads to an understanding of the actor. In our case, suppose

$$f(x) = A_0 + A_1 x + A_2 x^2 + \cdots + A_{k-1} x^{k-1},$$

where the $A_i = A_i(x^k)$ are polynomials in x^k, invariant under the substitutions we are about to make, and consider its "first" conjugate:

$$f(\zeta x) = A_0 + A_1 \zeta x + A_2 \zeta^2 x^2 + \cdots + A_{k-1} \zeta^{k-1} x^{k-1}.$$

Call this thing β_1. Let's look at the effect of β_1 on the powers of ζ:

$$\begin{aligned}
\beta_1 \cdot 1 &= A_0 + A_1 \zeta x + A_2 \zeta^2 x^2 + \cdots + A_{k-1} \zeta^{k-1} x^{k-1} \\
\beta_1 \cdot \zeta &= A_{k-1} x^{k-1} + A_0 \zeta + A_1 \zeta^2 x + A_2 \zeta^3 x^2 + \cdots + A_{k-2} \zeta^{k-1} x^{k-2} \\
\beta_1 \cdot \zeta^2 &= A_{k-2} x^{k-2} + A_{k-1} \zeta x^{k-1} + A_0 \zeta^2 + A_1 \zeta^3 x + \cdots + A_{k-3} \zeta^{k-1} x^{k-3} \\
\vdots &= \vdots \ \ddots \\
\beta_1 \cdot \zeta^{k-1} &= A_1 x + A_2 \zeta x^2 + A_3 \zeta^2 x^3 + \cdots + A_0 \zeta^{k-1}.
\end{aligned}$$

Write this as a matrix equation:

$$\beta_1 \begin{pmatrix} 1 \\ \zeta \\ \zeta^2 \\ \zeta^3 \\ \vdots \\ \zeta^{k-1} \end{pmatrix} = \begin{pmatrix} A_0 & A_1 x & A_2 x^2 & A_3 x^3 & \cdots & A_{k-1} x^{k-1} \\ A_{k-1} x^{k-1} & A_0 & A_1 x & A_2 x^2 & \cdots & A_{k-2} x^{k-2} \\ A_{k-2} x^{k-2} & A_{k-1} x^{k-1} & A_0 & A_1 x & \cdots & A_{k-3} x^{k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_1 x & A_2 x^2 & A_3 x^3 & A_4 x^4 & \cdots & A_0 \end{pmatrix} \begin{pmatrix} 1 \\ \zeta \\ \zeta^2 \\ \zeta^3 \\ \vdots \\ \zeta^{k-1} \end{pmatrix}.$$

Subtracting gives

$$\begin{pmatrix} A_0 - \beta_1 & A_1 x & A_2 x^2 & A_3 x^3 & \cdots & A_{k-1} x^{k-1} \\ A_{k-1} x^{k-1} & A_0 - \beta_1 & A_1 x & A_2 x^2 & \cdots & A_{k-2} x^{k-2} \\ A_{k-2} x^{k-2} & A_{k-1} x^{k-1} & A_0 - \beta_1 & A_1 x & \cdots & A_{k-3} x^{k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_1 x & A_2 x^2 & A_3 x^3 & A_4 x^4 & \cdots & A_0 - \beta_1 \end{pmatrix} \begin{pmatrix} 1 \\ \zeta \\ \zeta^2 \\ \zeta^3 \\ \vdots \\ \zeta^{k-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Because this equation takes place in an integral domain, the determinant of the left-hand matrix is zero.

Similarly, if $0 \le j \le k-1$, and

$$\beta_j = f(\zeta^j x) = A_0 + A_1 \zeta^j x + A_2 \zeta^{2j} x^2 + \cdots + A_{k-1} \zeta^{(k-1)j} x^{k-1},$$

then

$$\beta_j \begin{pmatrix} 1 \\ \zeta^j \\ \zeta^{2j} \\ \zeta^{3j} \\ \vdots \\ \zeta^{(k-1)j} \end{pmatrix} = \begin{pmatrix} A_0 & A_1 x & A_2 x^2 & A_3 x^3 & \cdots & A_{k-1} x^{k-1} \\ A_{k-1} x^{k-1} & A_0 & A_1 x & A_2 x^2 & \cdots & A_{k-2} x^{k-2} \\ A_{k-2} x^{k-2} & A_{k-1} x^{k-1} & A_0 & A_1 x & \cdots & A_{k-3} x^{k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_1 x & A_2 x^2 & A_3 x^3 & A_4 x^4 & \cdots & A_0 \end{pmatrix} \begin{pmatrix} 1 \\ \zeta^j \\ \zeta^{2j} \\ \zeta^{3j} \\ \vdots \\ \zeta^{(k-1)j} \end{pmatrix}.$$

So

$$\begin{vmatrix} A_0-\beta_j & A_1x & A_2x^2 & A_3x^3 & \cdots & A_{k-1}x^{k-1} \\ A_{k-1}x^{k-1} & A_0-\beta_j & A_1x & A_2x^2 & \cdots & A_{k-2}x^{k-2} \\ A_{k-2}x^{k-2} & A_{k-1}x^{k-1} & A_0-\beta_j & A_1x & \cdots & A_{k-3}x^{k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_1x & A_2x^2 & A_3x^3 & A_4x^4 & \cdots & A_0-\beta_j \end{vmatrix} = 0$$

and $\{\beta_0, \dots, \beta_{k-1}\}$ (that is, the conjugates of f) are all roots of the following polynomial equation in t:

$$\begin{vmatrix} A_0-t & A_1x & A_2x^2 & A_3x^3 & \cdots & A_{k-1}x^{k-1} \\ A_{k-1}x^{k-1} & A_0-t & A_1x & A_2x^2 & \cdots & A_{k-2}x^{k-2} \\ A_{k-2}x^{k-2} & A_{k-1}x^{k-1} & A_0-t & A_1x & \cdots & A_{k-3}x^{k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_1x & A_2x^2 & A_3x^3 & A_4x^4 & \cdots & A_0-t \end{vmatrix} = 0. \qquad (8)$$

This equation is a polynomial in t with coefficients that are polynomials in x. Its roots are the conjugates of f. But we want the *product* of these conjugates. So, we want the constant term of the left side of (8) (up to a sign). But you get the constant term by putting $t=0$. In other words, the product of f and all its conjugates is the following "circulant" (see **[2]** and **[3]** for more on circulants):

$$\begin{vmatrix} A_0 & A_1x & A_2x^2 & A_3x^3 & \cdots & A_{k-1}x^{k-1} \\ A_{k-1}x^{k-1} & A_0 & A_1x & A_2x^2 & \cdots & A_{k-2}x^{k-2} \\ A_{k-2}x^{k-2} & A_{k-1}x^{k-1} & A_0 & A_1x & \cdots & A_{k-3}x^{k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_1x & A_2x^2 & A_3x^3 & A_4x^4 & \cdots & A_0 \end{vmatrix}$$

This will be a polynomial in x^k (a nice exercise), and putting $y = x^k$ produces a polynomial whose roots are the kth powers of the roots of f.

EXAMPLE. If $f(x) = A + Bx + Cx^2$ (where A, B, and C are polynomials in x^3), then the product of the conjugates of f is

$$\begin{vmatrix} A & Bx & Cx^2 \\ Cx^2 & A & Bx \\ Bx & Cx^2 & A \end{vmatrix} = A^3 + (B^3 - 3ABC)x^3 + C^3x^6$$

as before.

For the fourth powers of the roots, write f as $f(x) = A + Bx + Cx^2 + Dx^3$, where the coefficients are polynomials in x^4. The desired polynomial can be obtained from the determinant:

$$\begin{vmatrix} A & Bx & Cx^2 & Dx^3 \\ Dx^3 & A & Bx & Cx^2 \\ Cx^2 & Dx^3 & A & Bx \\ Bx & Cx^2 & Dx^3 & A \end{vmatrix} = A^4 + (-B^4 + 4AB^2C - 2A^2C^2 - 4A^2BD)x^4$$

$$+ (C^4 - 4BC^2D + 2B^2D^2 + 4ACD^2)x^8 - D^4x^{12}$$

by replacing x^4 by y. ■

Acknowledgment. This work was supported in part by NSF grant DUE 9450731. I first heard of the problem through Bill Gosper (via Dick Askey). The idea in the previous section was inspired by something one of my high school students, Jan Nelson, did a long time ago. My high school *teacher*, Frank Kelley, just celebrated his 71st birthday, and he's still inspiring young people to study mathematics.

REFERENCES

1. E. Artin, *Galois Theory*, University of Notre Dame Press, London, UK, 1971.
2. A. C. Aitken, *Determinants and Matrices*, Oliver and Boyd, London, UK, 1942.
3. W. S. Burnside and A. W. Panton, *The Theory of Equations*, Volume II, Dover, New York, NY, 1960.

Invariants Under Group Actions to Amaze Your Friends

DOUGLAS E. ENSLEY
Shippensburg University
Shippensburg, PA 17257

Introduction This paper presents some simple magic tricks that work by themselves based upon mathematical principles. There are whole books devoted to this subject (see, e.g., [**2**], [**3**], or [**5**]), but this article specifically explores card effects that exploit invariants under the (group) action of mixing the cards. Brent Morris' recent book [**4**] is a wonderful exposition of the mathematics behind some of the effects that take advantage of the groups generated by perfect shuffles. This article orthogonally explores tricks where a *spectator* is allowed to mix the cards. The underlying theme is that a large permutation group leaves an audience with the feeling that the cards are being mixed while leaving an interesting set of invariants under the group action that can be used to perform a (hopefully) startling effect.

TV magic This trick was performed on a recent television program: A volunteer from the audience was handed the four aces from a deck of cards while the performer turned his back. (The reader may wish to take four playing cards, one from each suit, and follow along.) The magician then gave the following instructions:

1. Stack the four cards face-up with the heart at the bottom, then the club, then the diamond, and finally the spade.
2. Turn the spade (the uppermost card) face down.
3. Perform any of the following operations as many times and in any order that you wish:
   (a) Cut any number of cards from the top to the bottom.
   (b) Turn the top two cards over as one.
   (c) Either turn the entire stack over or do not—your choice.
4. Turn the topmost card over, then turn the top two cards over as one, and then turn the top three cards over as one.

At this point, the prestidigitator correctly divines that the club is the only card facing the opposite way from the others. As long as the audience member correctly followed the above directions, the magician is sure to be right.

Acknowledgment. This work was supported in part by NSF grant DUE 9450731. I first heard of the problem through Bill Gosper (via Dick Askey). The idea in the previous section was inspired by something one of my high school students, Jan Nelson, did a long time ago. My high school *teacher*, Frank Kelley, just celebrated his 71st birthday, and he's still inspiring young people to study mathematics.

REFERENCES

1. E. Artin, *Galois Theory*, University of Notre Dame Press, London, UK, 1971.
2. A. C. Aitken, *Determinants and Matrices*, Oliver and Boyd, London, UK, 1942.
3. W. S. Burnside and A. W. Panton, *The Theory of Equations*, Volume II, Dover, New York, NY, 1960.

Invariants Under Group Actions to Amaze Your Friends

DOUGLAS E. ENSLEY
Shippensburg University
Shippensburg, PA 17257

Introduction This paper presents some simple magic tricks that work by themselves based upon mathematical principles. There are whole books devoted to this subject (see, e.g., [**2**], [**3**], or [**5**]), but this article specifically explores card effects that exploit invariants under the (group) action of mixing the cards. Brent Morris' recent book [**4**] is a wonderful exposition of the mathematics behind some of the effects that take advantage of the groups generated by perfect shuffles. This article orthogonally explores tricks where a *spectator* is allowed to mix the cards. The underlying theme is that a large permutation group leaves an audience with the feeling that the cards are being mixed while leaving an interesting set of invariants under the group action that can be used to perform a (hopefully) startling effect.

TV magic This trick was performed on a recent television program: A volunteer from the audience was handed the four aces from a deck of cards while the performer turned his back. (The reader may wish to take four playing cards, one from each suit, and follow along.) The magician then gave the following instructions:

1. Stack the four cards face-up with the heart at the bottom, then the club, then the diamond, and finally the spade.
2. Turn the spade (the uppermost card) face down.
3. Perform any of the following operations as many times and in any order that you wish:
   (a) Cut any number of cards from the top to the bottom.
   (b) Turn the top two cards over as one.
   (c) Either turn the entire stack over or do not—your choice.
4. Turn the topmost card over, then turn the top two cards over as one, and then turn the top three cards over as one.

At this point, the prestidigitator correctly divines that the club is the only card facing the opposite way from the others. As long as the audience member correctly followed the above directions, the magician is sure to be right.

Fair play and an easier trick This trick is somewhat surprising because it seems to involve what we will refer to as "fair play." That is, what appears to one person to be mixing cards is actually preserving all of the properties that a second person cares about. Hence the spectator feels that the process is fair, which makes the outcome surprising. Another simple trick that illustrates this is one of the first card tricks that any child learns.

> The magician has a spectator choose a card, memorize it, and return it to the top of the deck. He then allows the spectator to cut the cards as many time as she would like. The magician spreads the cards face up and announces the chosen card.

This works because under the *action* of cutting the cards, the adjacency of pairs of cards is *invariant*. Thus, by remembering the name of the bottom card on the deck (which is secretly glimpsed while the spectator is looking at her own card), the magician can spot the chosen card as the one in front of the secret card in the face-up spread.

The group in this case is the subgroup H of S_{52} generated by the cyclic-shift permutation

$$\sigma = (1 \quad 2 \quad 3 \quad 4 \quad \ldots \quad 51 \quad 52)$$

and the invariant is the set A of adjacent pairs of cards in the deck (where the top and bottom cards are considered an adjacent pair). It is clear that A is not changed by any action in H, so since H appears to the spectator to be mixing the cards but the invariant set A can be used to find a chosen card, this is an effective magic trick.

Analysis of the original problem We next address how to represent the state of the cards for the original trick in this paper. We have a strange sort of permutation where each value has an orientation (face-up versus face-down) as well as a position in the deck. We will represent such a permutation as a "colored" permutation of $1,2,3,4$ with the convention that underlined type represents "face down" for the cards. A typical set of decisions in the trick might go as follows, where the deck starts off face up with a heart (4) at the bottom, then a club (3), then a diamond (2), and then a spade (1).

The original deck is represented as	$1,2,3,4$
(i) Turning the spade (the uppermost card) face down gives us	$\underline{1},2,3,4$
(ii) Cutting *two* cards from the top to the bottom gives us	$3,4,\underline{1},2$
(iii) Turning the top two cards over as one yields	$\underline{4},\underline{3},\underline{1},2$
(iv) Cutting *three* cards from the top to the bottom makes this	$2,\underline{4},\underline{3},\underline{1}$
(v) Turning the top two cards over as one again gives us	$4,\underline{2},\underline{3},\underline{1}$
(vi) Turning the entire stack over yields	$1,3,2,\underline{4}$
(vii) Turning the topmost card over,	$\underline{1},3,2,\underline{4}$
then the top two cards over as one,	$\underline{3},1,2,\underline{4}$
and then the top three cards over as one respectively yields finally	$\underline{2},\underline{1},3,\underline{4}$

Note that in the final arrangement Card 3 (the club) is turned up while the other three cards are turned down, so the trick does work with these choices... as if there were ever any doubt. This trick is interesting mathematically in that it really seems that every permutation of the cards could be achieved using the steps for mixing, but clearly this is not so. There are $2^4 \cdot 4! = 384$ ways we can arrange $1,2,3,4$ and underline each number or not, but the decisions at Step 3 in the trick will prevent many of these from happening. So the first question is to characterize the outcomes we can and cannot get.

Formally we could represent the permutations that act on the cards as the subgroup H of S_8 (thinking of the four cards' front-and-back pairs as eight objects) generated by the permutations that constitute the three operations in the trick that require decisions, namely cutting one card, turning two cards as one, and flipping over the entire packet. These will be denoted κ, τ, and ϕ respectively and are given below. We will continue to use our more visual notation of letters in two styles to express these mappings. Here the underline merely means that the particular card was reversed from its original orientation.

$$\kappa = ABCD \mapsto BCDA \quad \tau = ABCD \mapsto \underline{BA}CD \quad \phi = ABCD \mapsto \underline{DCBA}$$

The first proposition states in essence that at the end of each step in the trick there will always be 1 or 3 face down cards. Let C_0 denote the set of arrangements of the cards that have 1 or 3 face down cards. Notice that the first step of the trick causes the packet to be in C_0. With four cards in hand it is easy to check that each of the three mappings generating H leaves C_0 fixed. That is,

PROPOSITION 1. *C_0 is invariant under the action of H.*

The next observation is that, in the end, it does not matter which cards are faced up or down, just that the club (3) is faced differently than the others. It does not require a lot of experimentation to realize that sometimes the club is the only card faced down and sometimes it is the only card faced up. Hence, we change our representation to only underline the card (singular by the preceding proposition) that is faced differently in the deck, and the above process looks like this. The deck starts off face up with a heart (4) at the bottom, then a club (3), then a diamond (2), and then a spade (1).

The original deck is represented as		$1, 2, 3, 4$
(i)	Turning the uppermost card face down gives us	$\underline{1}, 2, 3, 4$
(ii)	Cutting *two* cards gives us	$3, 4, \underline{1}, 2$
(iii)	Turning the top two cards over as one yields	$4, 3, 1, \underline{2}$
(iv)	Cutting *three* of cards makes this	$\underline{2}, 4, 3, 1$
(v)	Turning the top two cards over as one again gives us	$\underline{4}, 2, 3, 1$
(vi)	Turning the entire stack over yields	$1, 3, 2, \underline{4}$
(vii)	Turning the topmost card over, then the top two cards over as one, and then the top three cards over as one yields finally	$2, 1, \underline{3}, 4$

Using this representation, we now realize that there are $4 \cdot 4! = 96$ outcomes, although we can still never generate this many of them with the decisions that the spectator is allowed to make. This representation also allows us to make clear the next proposition regarding invariance. Let C_1 denote the arrangements of the packet of cards so that the number 3 card is two cards away from the wrong-way (underlined) card.

PROPOSITION 2. *C_1 is invariant under the action of H.*

Proof. A packet of cards p in C_1 must originally look like one of the following, where here an underlined letter indicates a card reversed from the rest of the packet:

$$3, A, \underline{B}, C \quad C, 3, A, \underline{B} \quad \underline{B}, C, 3, A \quad A, \underline{B}, C, 3$$

The following table shows what happens to each of the mappings κ, τ, and ϕ act on a packet p.

p	$\kappa(p)$	$\tau(p)$	$\phi(p)$
$3, A, \underline{B}, C$	$A, \underline{B}, C, 3$	$A, 3, B, \underline{C}$	$C, \underline{B}, A, 3$
$C, 3, A, \underline{B}$	$3, A, \underline{B}, C$	$3, C, \underline{A}, B$	$\underline{B}, A, 3, C$
$\underline{B}, C, 3, A$	$C, 3, A, \underline{B}$	$\underline{C}, B, 3, A$	$A, 3, C, \underline{B}$
$A, \underline{B}, C, 3$	$\underline{B}, C, 3, A$	$B, \underline{A}, C, 3$	$3, C, \underline{B}, A$

In every case, the property that defines C_1 *is preserved.*

The last step in the trick is to turn over the topmost card, then the top two cards (as one), and then the top three cards (as one). To represent this action we will use the notation $\mapsto$ to express the execution of each of the three steps. We will use an asterisk to denote that the card has changed its orientation. This final operation on the entire packet is then represented

$$\begin{aligned} ABCD &\mapsto A^*BCD \\ &\mapsto B^*ACD \\ &\mapsto C^*A^*BD \end{aligned}$$

PROPOSITION 3. *If the packet starts with the club two places away from the wrong-way card, then the club will be the wrong-way card after the final operation.*

Proof. Since the original packet starts off with the club two away from the wrong-way card, then we need to consider just the cases where (1) A or C is the club, or (2) B or D is the club. In case (1), the operations above will reverse both the club and the wrong-way card resulting in the club being the wrong-way card. In case (2), the operations will reverse only the two cards that are neither the club nor the wrong-way card resulting in the club being the wrong-way card.

A solitaire game Here is a simple game of solitaire that can be played with a deck of cards in your hands—we used to play a version of this game on car trips since it does not require a table top. It is equivalent to the game "Even Up," which was recently analyzed in **[1]**. The deck is held face up and fanned through with the player removing pairs of cards of the same color whenever they occur adjacent in the deck. Of course, the removal of adjacent pairs may create other adjacent pairs which will also have to be removed. The game ends when there are no more adjacent same-colored pairs to remove. Winning the game means having no cards left at the end. This can be turned into a magic trick as follows:

> The magician calls upon two spectators to each take half the deck of cards and shuffle them independently. They then merge their stacks (alternatively dealing the cards into a single stack) and play the solitaire game. To the amazement of everyone, they find they have won.

The only real trick is that the magician really does split the deck *in half*, meaning that not only does each spectator get 26 cards but also each spectator gets 13 red cards and 13 black cards. If each spectator has 13 red cards and 13 black cards, then they can shuffle their respective halves until they are blue in the face and when they merge them, the solitaire game will be guaranteed to be a winner.

To see why this is true, imagine one spectator's cards came from a blue-backed deck while the other's cards came from a green-backed deck. The final deck will alternate colors of their backs even though the faces of the cards are fairly shuffled.

Let us decide that the green-backed cards assume the even positions while the blue-backed cards are in the odd positions. As the solitaire game is played, cards are removed from the deck in adjacent pairs which share the same face-color. Hence at any point in the game, (i) the number of blue-backed red cards is equal to the number of green-backed red cards, (ii) the number of blue-backed black cards is equal to the number of green-backed black cards, and (iii) the blue-backed and green-backed cards alternate. Given these three properties that remain invariant as the game is played, it is impossible that the game should ever end in a loss. This is true because a losing final position must consist of cards alternating in face colors, and property (iii) then dictates that the red cards and black cards should have different back colors contrary to properties (i) and (ii).

In terms of invariants under group actions, the permutation group H at work here is the (large!) subgroup of S_{52} generated by permutations that shuffle the odd positions among themselves and the even positions among themselves. Letting C denote the decks of cards that will lead to a solitaire win, we can state the conclusion of the previous discussion as follows.

PROPOSITION 4. *C is invariant under the action of H.*

The reason that this establishes that the trick will work is that the obvious winning arrangement p_0 which has all red cards in the top half of the deck and all black cards in the bottom half of the deck is in C. As a computational aside, this analysis also tells us that the probability of winning this solitaire game with a fairly shuffled deck of cards is simply the probability that the deck is in an order obtainable as in the magic trick. This probability is

$$\frac{\binom{26}{13}^2}{\binom{52}{26}} \approx 0.218\,.$$

Unfortunately, this probability is a bit high for this trick to be really amazing since people familiar with the game could decide that the magician was just lucky. It is the sort of trick that would be more effective if done with several pairs of spectators simultaneously.

Conclusions Many magic tricks, particularly those using cards or those involving mentalism, use a set of procedures to make the spectator feel he is making free choices, when in reality the results of these choices are all equivalent for the magician's purposes.

REFERENCES

1. A. T. Benjamin and J. J. Quinn, Unevening the odds of "Even Up," this MAGAZINE 72 (1999), 145–146.
2. M. Gardner, *Mathematics, Magic and Mystery*, Dover Publications, New York, NY, 1956.
3. R. V. Heath, *Mathemagic: Magic, Puzzles, and Games with Numbers*, Dover Publications, New York, NY, 1953.
4. S. B. Morris, *Magic Tricks, Card Shuffling, and Dynamic Computer Memory*, The Mathematical Association of America, Washington, DC, 1998.
5. W. Simon, *Mathematical Magic*, Dover Publications, New York, NY, 1993.

On Groups That Are Isomorphic to a Proper Subgroup

SHAUN FALLAT
CHI-KWONG LI
DAVID LUTZER
DAVID STANFORD
College of William and Mary
Williamsburg, VA 23187-8795

Introduction When is a group isomorphic to a proper subgroup of itself? Clearly, no finite group can have this property, but what about $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, the familiar additive groups of integers, rational, real, and complex numbers? What about $\mathbb{R}^n$, the additive group of n-dimensional vectors? What about the multiplicative groups of non-zero rational, real, or complex numbers? What about the multiplicative group $T = \{z \in \mathbb{C} : |z| = 1\}$ of complex numbers with modulus one? What about your own favorite infinite group from the first modern algebra class?

These easily stated questions are very special cases of an important problem in group theory (namely to determine whether or not two groups are isomorphic) and can be the basis for classroom discussion in an introductory modern algebra course as soon as the notions of group, subgroup, and isomorphism have been introduced. Further, such questions can be posed again as new algebraic constructions (e.g., product groups and quotient groups) are introduced, and they have analogues for the other familiar algebraic structures (rings, fields) often found in undergraduate modern algebra. In addition, the isomorphic subgroup question provides a valuable way to get students to think about the familiar groups $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$ in a non-trivial context. Finally, the question can be the basis for open-ended student projects in such a course.

Textbooks develop standard techniques for showing that two groups are *not* isomorphic. Perhaps one is cyclic and the other is not. Perhaps the groups have different cardinalities. Perhaps the two groups have a different number of elements of some order k. On the other hand, for students in an introductory course, showing that two groups *are* isomorphic usually means constructing a specific isomorphism. As a result, many of the examples of isomorphic groups given in introductory courses are transparently isomorphic.

The goal of this note is to illustrate how the proper subgroup question can be used in an introductory course, and to show how ideas from linear algebra can be used in an introductory modern algebra course to exhibit non-transparently isomorphic groups. We will answer the questions posed in the opening paragraph, and suggest further projects that might be of interest to students. We do not claim novelty for the results below, nor do we give the most general statements or the best possible proofs of the results. (An elegant classical reference for related material is **[2]**.)

Some easy examples Clearly, if a group G is isomorphic to a proper subgroup of itself, then $|G|$, the cardinality of G, must be infinite. However, being infinite is not enough, because easy examples show that some infinite groups are, and others are not, isomorphic to proper subgroups of themselves.

EXAMPLE 1. The additive group $\mathbb{Z}$ of all integers is isomorphic to the subgroup of even integers under the isomorphism $f(x) = 2 * x$.

EXAMPLE 2. The additive group $\mathbb{Q}$ of all rational numbers is *not* isomorphic to a proper subgroup of itself because every homomorphism $f: \mathbb{Q} \to \mathbb{Q}$ has the form $f(x) = f(1) * x$, so that that every isomorphism from $\mathbb{Q}$ into $\mathbb{Q}$ is actually an isomorphism onto $\mathbb{Q}$.

QUESTION 1. What about the direct sum group $\mathbb{Q} \oplus \mathbb{Q}$? Is it isomorphic to a proper subgroup of itself? What about $\mathbb{Q} \oplus \mathbb{Z}$? (The answers are "No" and "Yes," respectively.)

QUESTION 2. What about the quotient group $\mathbb{Q}/\mathbb{Z}$? It is not isomorphic to a proper subgroup of itself because if g is an isomorphism from $\mathbb{Q}/\mathbb{Z}$ into itself and if $A(k) = \{x \in \mathbb{Q}/\mathbb{Z}: x \text{ has order } k\}$, then $g[A(k)] \subseteq A(k)$. But then, $A(k)$ being finite and g being one-to-one, we have $g[A(k)] = A(k)$. Because $\mathbb{Q}/\mathbb{Z} = \bigcup\{A(k): k \geq 1\}$, g must be onto.

Next we show that some familiar multiplicative groups provide many different examples of groups that are isomorphic to proper subgroups of themselves. We consider the multiplicative groups $\mathbb{Q}^+, \mathbb{Q}^*, \mathbb{R}^+, \mathbb{R}^*$ consisting, respectively, of all positive rationals, non-zero rationals, positive reals, and non-zero reals.

PROPOSITION 1. *The multiplicative groups $\mathbb{Q}^+$, $\mathbb{Q}^*$, $\mathbb{R}^+$, and $\mathbb{R}^*$ are all isomorphic to proper subgroups of themselves.*

Proof. Define $f: \mathbb{Q}^* \to \mathbb{Q}^*$ by $f(x) = x^3$. This f is an isomorphism from $\mathbb{Q}^*$ onto a proper subgroup of itself, and when restricted to $\mathbb{Q}^+$, f provides an isomorphism from $\mathbb{Q}^+$ onto a proper subgroup of itself. Next, the function $g(x) = e^x$ is an isomorphism from the additive group $\mathbb{R}$ onto the multiplicative group $\mathbb{R}^+$. Hence, by Theorem 1 (to follow), we have that $\mathbb{R}^+$ is isomorphic to a proper subgroup of itself. Because the multiplicative group $\mathbb{R}^*$ is the internal direct sum of $\mathbb{R}^+$ and the two element group $T_0 = \{-1, 1\}$ with its usual multiplication, we see that $\mathbb{R}^*$ is isomorphic to a proper subgroup of itself. (See also Exercise 1, below.)

It is natural to wonder whether the groups in Proposition 1 are really different. They are. Cardinality arguments show that the rational and real groups in Proposition 1 are distinct. To distinguish between $\mathbb{Q}^+$ and $\mathbb{Q}^*$, ask how many solutions the equation $x^2 = 1$ has in each group. The same question distinguishes between $\mathbb{R}^+$ and $\mathbb{R}^*$. Compare this with Proposition 4, below.

More complicated examples and $\mathbb{Q}$-linearity Certain groups are able to carry more than just group structure. For example, the additive groups $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{R}^n$ are also vector spaces over the field $\mathbb{Q}$, and group homomorphisms of these groups are easily seen to be $\mathbb{Q}$-linear mappings as well. Those facts allow students to use some of the ideas that they encountered in their first linear algebra course, namely the notions of spanning sets, bases, and dimension. Of course, one must now deal with vector spaces that are infinite dimensional over $\mathbb{Q}$, and one must distinguish between finite, countable, and uncountable dimensional spaces. That added complication allows students to review the finite-dimensional proofs they encountered in linear algebra, to see whether basis theory still works in more general spaces. By mimicking the finite-dimensional proofs, and using the observation that the $\mathbb{Q}$-linear span of an infinite set S has cardinality $|\mathbb{Q}| * |S| = |S|$, strong students would be able to prove:

PROPOSITION 2. *Two vector spaces V and W over the field $\mathbb{Q}$ are $\mathbb{Q}$-linearly isomorphic if and only if V and W have bases over $\mathbb{Q}$ of the same cardinality.*

There is a second result that is useful in exploiting the $\mathbb{Q}$-linear structure of certain groups, but one that is rarely mentioned in introductory textbooks, probably because it depends on what Halmos called "transfinite trickery" [**1**, p. 13].

PROPOSITION 3. *Any linearly independent set in any vector space V is contained in a basis for V. In particular, there is a basis B for the $\mathbb{Q}$-vector space $\mathbb{R}$ with $1 \in B$.*

Undergraduates have no problem understanding the statement of Proposition 3. Our experience suggests that most are willing to accept the statement without insisting on a proof. For the others, Proposition 3 could be the basis for an outside reading project on Zorn's lemma. Students who know that $\mathbb{Q}$ is countable while $\mathbb{R}$ is not will be able to prove that the basis B for $\mathbb{R}$ over $\mathbb{Q}$ is infinite, a fact that we need below. Using Propositions 2 and 3, one can prove:

THEOREM 1. *The additive group $\mathbb{R}$ is isomorphic to a proper subgroup of itself.*

Proof. Using Proposition 3 choose any basis B for $\mathbb{R}$ over $\mathbb{Q}$. Then B is infinite. Choose distinct $b(1), b(2), \dots$ in B and define $s: B \to B$ by $s(b(n)) = b(n+1)$ and $s(b) = b$ if $b \in B - \{b(n): n \geq 1\}$. Then extend s over $\mathbb{R}$ in a $\mathbb{Q}$-linear way. The resulting $\mathbb{Q}$-vector space isomorphism is the required group isomorphism from $\mathbb{R}$ onto a proper subgroup of itself.

Students might be tempted to think that additional examples of groups that are isomorphic to proper subgroups could be obtained from other familiar groups such as $\mathbb{R}^n$ and $\mathbb{C}$. Such examples do exist, but they are not *new* examples because Proposition 2 yields:

PROPOSITION 4. *Each of the additive groups $\mathbb{R}^n, \mathbb{C}^n, \mathbb{R}[X] = \{p(X): p$ is a polynomial with coefficients in $\mathbb{R}\}$, and $\mathbb{C}[X] = \{p(X): p$ is a polynomial with coefficients in $\mathbb{C}\}$ is isomorphic to the additive group $\mathbb{R}$.*

Proof. Starting with a basis B for $\mathbb{R}$ over $\mathbb{Q}$, one can show that each of these groups is a $\mathbb{Q}$-vector space with a basis of cardinality $|B|$. Now apply Proposition 2 to conclude that the groups listed in this proposition are $\mathbb{Q}$-linearly isomorphic, and hence group isomorphic.

A slightly less familiar but very important group is the multiplicative group $T = \{z \in \mathbb{C}: |z| = 1\}$, where $|z|$ denotes the usual absolute value of the complex number z. In a moment we will need to know that the function $h(x) = e^{2\pi i x}$ induces a group isomorphism from $\mathbb{R}/\mathbb{Z}$ onto T.

Consider $\mathbb{C}^*$, the multiplicative group of all non-zero complex numbers. This group is not isomorphic to any of the groups considered above (namely, the multiplicative groups $\mathbb{Q}^+$, $\mathbb{Q}^*$, $\mathbb{R}^+$, $\mathbb{R}^*$, and the additive groups listed in Proposition 4) because it contains two elements of order three (i.e., nontrivial solutions of the equation $x^3 = 1$) while none of the other groups has this property. Is $\mathbb{C}^*$ isomorphic to a proper subgroup of itself? The answer is "yes." The most elementary proof that we know uses the linear algebra tools from the previous section as a start, and then makes careful use of the isomorphism theorems for groups. In this case, one can give a concrete example of a proper subgroup of $\mathbb{C}^*$ to which $\mathbb{C}^*$ is isomorphic, and the result is somewhat counter-intuitive.

THEOREM 2. *The multiplicative groups $\mathbb{C}^*$ and T are isomorphic.*

Proof. Using Proposition 3, choose a basis B for $\mathbb{R}$ as a $\mathbb{Q}$-vector space, with $1 \in B$. Because B is infinite, we can write $B = B_1 \cup B_2$ where $B_1 \cap B_2 = \emptyset$, $|B_i| = |B|$ and

$1 \in B_1$. For each $b \in B$ let Q_b be the $\mathbb{Q}$-vector space $\{q * b : q \in \mathbb{Q}\}$. Then Proposition 2 yields an isomorphism f from the $\mathbb{Q}$-vector space $\mathbb{R}$ onto $(\oplus\{Q_b : b \in B_1\}) \oplus (\oplus\{Q_b : b \in B_2\})$ that sends the number $1 \in \mathbb{R}$ to the vector $1 \in Q_1 \subset R_1$ where $R_i = \oplus\{Q_b : b \in B_i\}$. By Proposition 2, each R_i is $\mathbb{Q}$-linearly isomorphic to $\mathbb{R}$.

Now think of the $\mathbb{Q}$-vector spaces above as additive groups. Then f is a group isomorphism from $\mathbb{R}$ onto $R_1 \oplus R_2$, and $(R_2, +)$ is group isomorphic to $(\mathbb{R}^+, *)$ under the exponential function. Writing $Z_1 = f[\mathbb{Z}]$ and writing $\cong$ to denote group isomorphism, we have $\mathbb{R}/\mathbb{Z} \cong R_1/Z_1 \oplus R_2 \cong (T, *) \oplus (\mathbb{R}^+, *)$. But the usual polar representation of non-zero complex numbers establishes a group isomorphism between $T \oplus (\mathbb{R}^+, *)$ and $\mathbb{C}^*$. Thus $T \cong \mathbb{R}/\mathbb{Z} \cong T \oplus \mathbb{R}^+ \cong \mathbb{C}^*$, as claimed.

More exercises for undergraduates The ideas in this paper can be the basis for exploration projects in a first modern algebra course. In this section, we give a few examples of questions that an instructor might pose at various stages of the course.

EXERCISE 1. *Direct sums and proper subgroups*. Suppose G and H are groups, one of which is isomorphic to a proper subgroup of itself. Then so is $G \oplus H$. What about the converse? If $G \oplus H$ is isomorphic to a proper subgroup of itself, must the same be true of one of G and H?

EXERCISE 2. *Which groups are $\mathbb{Q}$-linear spaces*? The central idea in our paper is that many familiar groups are, in fact, $\mathbb{Q}$-linear vector spaces. Characterize all Abelian groups that are $\mathbb{Q}$-linear vector spaces.

EXERCISE 3. *Counting morphisms*. How many group homomorphisms exist from the additive group $\mathbb{Q}$ into itself? From the additive group $\mathbb{R}$ into itself? From the multiplicative groups considered in Section 3 into themselves? How many group isomorphisms exist in each case?

EXERCISE 4. *Fields and subfields*. Which fields are field isomorphic to proper subfields of themselves? One can show that the usual fields $\mathbb{Q}$ and $\mathbb{R}$ are *not* isomorphic to proper subfields of themselves, but that there are fields lying between $\mathbb{Q}$ and $\mathbb{R}$ that are isomorphic to proper subfields of themselves. One approach is to prove that the identity function is the only field isomorphism from either $\mathbb{Q}$ or $\mathbb{R}$ into itself. One can also show that the usual field $\mathbb{C}$ of complex numbers is isomorphic to a proper subfield of itself. In addition, unlike the situation for $\mathbb{Q}$ and $\mathbb{R}$, there are many field automorphisms of $\mathbb{C}$. See [3] for an elegant discussion.

EXERCISE 5. *Quotient groups*. When is a group G isomorphic to a non-trivial quotient group of itself? (By a "nontrivial quotient group of G" we mean a quotient group G/H where $|H| > 1$.)

REFERENCES

1. P. Halmos, *Finite Dimensional Vector Spaces*, D. Van Nostrand, New York, NY, 1958.
2. I. Kaplansky, *Infinite Abelian Groups*, University of Michigan Press, Ann Arbor, MI, 1954.
3. P. Yale, Automorphisms of the complex numbers, this MAGAZINE 39 (1966), 135–141.

Integration of Radial Functions

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

Introduction Most advanced calculus texts justify double integration in polar coordinates by a geometric argument which may be convincing for many but inadequate for others. Even to deduce it from a general "change of variables" theorem (beyond the scope of most such texts) requires some guile because the polar coordinates change-of-variable map is not globally C^1 invertible. Analogous remarks apply to spherical coordinates. However, the simplest and most memorable applications of these techniques involve the integration of *radial* functions over *disks* or *balls* centered at the origin.

The aim of this note is to discuss the following theorem, which, I hope, may be useful for some instructors of multivariable calculus. The proof is simple in the sense that it depends on little more than the Darboux criterion of Riemann integrability and the fact that the set $\{x \in \mathbb{R}^n : |x| \leq 1\}$ has n-dimensional volume (Jordan content or Lebesgue measure); we denote that volume by V_n. (We write $|x| = (x_1^2 + \cdots + x_n^2)^{1/2}$ for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.)

We will show how this theorem can be used to compute V_n and to prove the important fact that

$$\int_{\mathbb{R}^n} e^{-|x|^2/2}\, dx = (2\pi)^{n/2} \quad \text{for all} \quad n \in \mathbb{N}.$$

More sophisticated "Lebesgue" versions of most of what is presented here, together with related results, can be found in [**1**, pp. 75–77] and [**2**, pp. 393–396].

THEOREM. *Suppose that* $0 \leq a < b$, *and set* $A = \{x \in \mathbb{R}^n : a \leq |x| \leq b\}$. *Let* $g : [a, b] \to \mathbb{R}$ *be Riemann integrable on* $[a, b]$, *and let* $f(x) = g(|x|)$ *for* $x \in A$. *Then* f *is Riemann integrable on* A, *and*

$$\int_A f = nV_n \int_a^b g(r) r^{n-1} dr.$$

Proof. We assume, without embarrassment, that $n \geq 2$ and that $g(r) \geq 0$ for $a \leq r \leq b$. Given $\epsilon > 0$, choose $a = r_0 < r_1 < r_2 < \cdots < r_N = b$ and $0 \leq m_k \leq M_k$ for $1 \leq k \leq N$, such that

$$r_k - r_{k-1} < \epsilon \quad \text{for} \quad 1 \leq k \leq N; \tag{1}$$

$$m_k \leq g(r) \leq M_k \quad \text{for} \quad r_{k-1} \leq r \leq r_k, \quad 1 \leq k \leq N;$$

$$\sum_{k=1}^{N} (M_k - m_k)(r_k - r_{k-1}) < \epsilon. \tag{2}$$

For $k = 1, \dots, N$, let $A_k = \{x \in A : r_{k-1} < |x| \leq r_k\}$. Note that $m_k \leq f(x) \leq M_k$ for $x \in A_k$, and that the n-dimensional volume of A_k is $V_n(r_k^n - r_{k-1}^n)$. Hence the upper and lower Riemann integrals of f over A satisfy

$$\sum_{k=1}^{N} m_k V_n(r_k^n - r_{k-1}^n) \leq \underline{\int_A} f \leq \overline{\int_A} f \leq \sum_{k=1}^{N} M_k V_n(r_k^n - r_{k-1}^n).$$

But, by the mean value theorem, we have

$$nr_{k-1}^{n-1}(r_k - r_{k-1}) \le r_k^n - r_{k-1}^n \le nr_k^{n-1}(r_k - r_{k-1})$$

for $1 \le k \le N$. It follows that

$$L_\epsilon \le \underline{\int_A} f \le \overline{\int_A} f \le U_\epsilon,$$

where $L_\epsilon = nV_n \sum_{k=1}^n m_k r_{k-1}^{n-1}(r_k - r_{k-1})$ and $U_\epsilon = nV_n \sum_{k=1}^n M_k r_k^{n-1}(r_k - r_{k-1})$. On the other hand,

$$\sum_{k=1}^N m_k r_{k-1}^{n-1}(r_k - r_{k-1}) \le \int_a^b g(r) r^{n-1} dr \le \sum_{k=1}^N M_k r_k^{n-1}(r_k - r_{k-1}).$$

We conclude that $\underline{\int_A} f$, $\overline{\int_A} f$, and $nV_n \int_a^b g(r) r^{n-1}\, dr$ all belong to the interval $[L_\epsilon, U_\epsilon]$ —and this is so for every $\epsilon > 0$.

To complete the proof it suffices to demonstrate that $U_\epsilon - L_\epsilon \to 0$ as $\epsilon \to 0^+$. To this end, first note that if $0 \le m \le M$, $0 \le r \le R \le b$, $R - r < \epsilon$, and $n \ge 2$. Then, by another appeal to the mean value theorem, we find

$$\begin{aligned} 0 &\le MR^{n-1} - mr^{n-1} = (M-m)R^{n-1} + m(R^{n-1} - r^{n-1}) \\ &\le (M-m)R^{n-1} + m(n-1)R^{n-2}(R-r) \le (M-m)b^{n-1} + M(n-1)b^{n-2}\epsilon. \end{aligned}$$

Thus, for $0 < \epsilon < 1$ and with $M = \max\{M_1, \dots, M_N\}$, we have

$$\begin{aligned} 0 &\le (nV_n)^{-1}(U_\epsilon - L_\epsilon) = \sum_{k=1}^N \left[M_k r_k^{n-1} - m_k r_{k-1}^{n-1}\right](r_k - r_{k-1}) \\ &\le \sum_{k=1}^N \left[(M_k - m_k)b^{n-1} + M_k(n-1)b^{n-2}\epsilon\right](r_k - r_{k-1}) \\ &\le b^{n-1} \sum_{k=1}^N (M_k - m_k)(r_k - r_{k-1}) + M(n-1)b^{n-2}\epsilon \sum_{k=1}^N (r_k - r_{k-1}) \\ &\le b^{n-1}\epsilon + M(n-1)b^{n-2}\epsilon(b-a) \end{aligned}$$

according to (1) and (2). Hence $U_\epsilon - L_\epsilon \to 0$ as $\epsilon \to 0^+$. ■

Examples. Suppose that $n \in \mathbb{N}$ and $\gamma \in \mathbb{R}$. If $0 < a < b$, then

$$\begin{aligned} \int_{a \le |x| \le b} |x|^\gamma d(x_1, \dots, x_n) &= nV_n \int_a^b r^{\gamma+n-1}\, dr \\ &= \begin{cases} nV_n \log\left(\frac{b}{a}\right) & \text{if } \gamma + n = 0; \\ nV_n \dfrac{b^{\gamma+n} - a^{\gamma+n}}{\gamma+n} & \text{if } \gamma + n \ne 0. \end{cases} \end{aligned}$$

Thus $\int_{0 < |x| \le b} |x|^\gamma\, d(x_1, \dots, x_n)$ diverges if $\gamma \le -n$ and converges to $\frac{nV_n b^{\gamma+n}}{\gamma+n}$ if $\gamma > -n$. Similarly, $\int_{a \le |x|} |x|^\gamma\, d(x_1, \dots, x_n)$ diverges if $\gamma \ge -n$ and converges to $\frac{-nV_n a^{\gamma+n}}{\gamma+n}$ if $\gamma < -n$.

These simple observations imply the following result:

CONVERGENCE CRITERIA. *Suppose that* $n \in \mathbb{N}$, $f:\mathbb{R}^n \to \mathbb{R}$, *and (for simplicity of exposition) assume that f is continuous.*

(i) *If there exist* a, $A > 0$ *and* $\beta > n$ *such that* $|f(x)| \leq \frac{A}{|x|^\beta}$ *for* $|x| \geq a$, *then* $\int_{\mathbb{R}^n} f(x)\,dx$ *is absolutely convergent.*

(ii) *If* $\lim_{|x|\to\infty} |x|^\alpha f(x) = c$ *for some* $\alpha \in \mathbb{R}$ *and some nonzero* $c \in \mathbb{R}$, *then f is integrable on* $\mathbb{R}^n$ *(in fact, absolutely integrable) if and only if* $\alpha > n$.

For example, $\int_{\mathbb{R}^n}(1+|x|^2)^{-\alpha}\,dx$ is convergent if and only if $\alpha > n/2$. More importantly, $\int_{\mathbb{R}^n} e^{-\lambda|x|^2}\,dx$ converges for each $n \in \mathbb{N}$ and every $\lambda > 0$—please read on.

The trace as an integral Suppose that $2 \leq n \in \mathbb{N}$. It is not difficult to see that for $1 \leq i, j \leq n$,

$$\int_{|x|\leq 1} x_i x_j \, d(x_1,\dots,x_n) = 0 \quad \text{if } i \neq j, \text{ and}$$

$$\int_{|x|\leq 1} x_i^2 \, d(x_1,\dots,x_n) = \int_{|x|\leq 1} x_j^2 \, d(x_1,\dots,x_n).$$

Thus, for $1 \leq i \leq n$,

$$\int_{|x|\leq 1} x_i^2 \, d(x_1,\dots,x_n) = \frac{1}{n}\int_{|x|\leq 1} |x|^2 \, d(x_1,\dots,x_n) = \frac{V_n}{n+2}.$$

Now suppose that $A = (a_{ij})$ is a real $n \times n$ matrix, and let q denote the quadratic form defined by

$$q(x) = xAx^t = \sum_{i,j=1}^{n} a_{ij} x_i x_j,$$

for $x = (x_1,\dots, x_n) \in \mathbb{R}^n$. Then

$$\int_{|x|\leq 1} q(x)\,dx = \sum_{i,j=1}^{n} a_{ij}\int_{|x|\leq 1} x_i x_j \, d(x_1,\dots,x_n)$$
$$= \sum_{i=1}^{n} a_{ii}\int_{|x|\leq 1} x_i^2 \, d(x_1,\dots,x_n).$$

Thus, if $\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$ denotes the *trace* of A, we have

$$\mathrm{tr}(A) = (n+2)V_n^{-1}\int_{|x|\leq 1} xAx^t \, d(x_1,\dots,x_n).$$

The integral of the Gaussian For $n \in \mathbb{N}$ and $R > 0$, let

$$I_n(R) = \int_{|x|\leq R} e^{-|x|^2/2}\, d(x_1,\dots,x_n) \tag{3}$$

and note that

$$\int_{-R}^{R}\cdots\int_{-R}^{R} e^{-|x|^2/2}\,dx_1\dots dx_n = \left(\int_{-R}^{R} e^{-t^2/2}\,dt\right)^n = I_1(R)^n. \tag{4}$$

Since the integrand is everywhere positive and

$$\{x\in\mathbb{R}^n:|x|\le R\}\subseteq[-R,R]^n\subseteq\{x\in\mathbb{R}^n:|x|\le\sqrt{n}R\},$$

we have

$$I_n(R)\le I_1(R)^n\le I_n(\sqrt{n}R)\quad\text{for}\quad n\in\mathbb{N}\quad\text{and}\quad R>0. \tag{5}$$

But

$$I_2(R)=2V_2\int_0^R e^{-r^2/2}r\,dr=2\pi\left(-e^{-r^2/2}\right)\Big|_{r=0}^R=2\pi\left(1-e^{-R^2/2}\right)$$

for $R>0$. Hence

$$2\pi=\lim_{R\to\infty}I_2(R)=\int_{\mathbb{R}^2}e^{-|x|^2/2}\,dx.$$

It then follows from (5) that $2\pi=\lim_{R\to\infty}I_2(R)=\lim_{R\to\infty}I_1(R)^2$, so that

$$\int_{-\infty}^{\infty}e^{-t^2/2}\,dt=\lim_{R\to\infty}I_1(R)=\sqrt{2\pi}.$$

By appealing to (5) once again we conclude that $\lim_{R\to\infty}I_n(R)=\lim_{R\to\infty}I_1(R)^n$, i.e.,

$$\int_{\mathbb{R}^n}e^{-|x|^2/2}\,dx=(2\pi)^{n/2}$$

for every $n\in\mathbb{N}$.

Rational multiples of π According to (3) and the Theorem, for $n\in\mathbb{N}$ and $R>0$,

$$\begin{aligned}I_n(R)&=nV_n\int_0^R e^{-r^2/2}r^{n-1}\,dr=V_n\int_0^R e^{-r^2/2}\frac{d}{dr}(r^n)\,dr\\&=V_n\left(r^ne^{-r^2/2}\big|_0^R+\int_0^R e^{-r^2/2}r^{n+1}\,dr\right),\end{aligned}$$

so that

$$(2\pi)^{n/2}=\int_{\mathbb{R}^n}e^{-|x|^2/2}\,dx=nV_n\int_0^\infty e^{-r^2/2}r^{n-1}\,dr=V_n\int_0^\infty e^{-r^2/2}r^{n+1}\,dr.$$

But then, for all $n\in\mathbb{N}$, $(2\pi)^{n+2/2}=(n+2)V_{n+2}\int_0^\infty e^{-r^2/2}r^{n+1}\,dr$. Hence, for all $n\in\mathbb{N}$,

$$\frac{(2\pi)^{n/2}}{V_n}=\frac{(2\pi)^{n+2/2}}{(n+2)V_{n+2}},\quad\text{or}\quad V_{n+2}=\frac{2\pi}{n+2}V_n.$$

Since $V_1=2$ and $V_2=\pi$, it follows that $V_{2k}=\frac{\pi^k}{k!}$ and $V_{2k-1}=\frac{k!4^k\pi^{k-1}}{(2k)!}$ for all $k\in\mathbb{N}$.

It may be amusing to note that $\sum_{k=1}^{\infty}V_{2k}=e^{\pi}+e^{i\pi}$.

REFERENCES

1. Gerald B. Folland, *Real Analysis*, John Wiley & Sons, New York, NY, 1984.
2. Karl R. Stromberg, *An Introduction to Classical Real Analysis*, Wadsworth, Belmont, CA, 1981.

Is 0.999... = 1?

FRED RICHMAN
Florida Atlantic University
Boca Raton, FL 33431

Few mathematical structures have undergone as many revisions or have been presented in as many guises as the real numbers. Every generation reexamines the reals in the light of its values and mathematical objectives. [3]

Arguing whether 0.999... is equal to 1 is a popular sport on the newsgroup sci.math. It seems to me that people are often too quick to dismiss the idea that these two numbers might be different. The issues here are closely related to Zeno's paradox, and to the notion of potential infinity versus actual infinity. Also at stake is the orthodox view of the nature of real numbers.

One argument for the equality goes like this. Set $x = 0.999\ldots$, multiply both sides by 10 to get $10x = 9.999\ldots$, then subtract the first equation from the second. The result is $9x = 9$, so $x = 1$. Essentially you are observing that $9x + x = 9 + x$, which is true, and then concluding that $9x = 9$. That's a valid inference, *if* x is cancellable.

But one man's proof is another man's *reductio ad absurdum*. Although most everyone will agree that the above argument shows that if x is cancellable, then $x = 1$, the believer and the skeptic differ in their interpretation of what this means. The believer, quite reasonably, takes for granted that you can cancel x, and regards the argument as a proof of the equality. For the skeptic, who considers the equality to be false, the argument is a proof that you cannot cancel x (if you could, then the equality would hold). So the skeptic must adopt the position that subtraction of real numbers is not always possible.

The skeptic would say that $9x$ is equal to $8.999\ldots$, not 9, using the usual algorithm for multiplication: In each digit position we multiply 9 times 9, and add to that a carry of 8 from the position to the right, except at the position before the decimal point, where we simply get the carry of 8. The skeptic considers the number $8.999\ldots$ to be different from 9, just as $0.999\ldots$ is different from 1, even though $8.999\ldots + x = 9 + x$.

Here is an even simpler argument for the equality: Multiply the equation $1/3 = 0.333\ldots$ by 3 to get $1 = 0.999\ldots$. The multiplication step is pretty hard to fault, so a skeptic must challenge the first equation. This simple argument gets its force from the fact that most people have been trained to accept the first equation without thinking.

Yet a third kind of argument is that

$$0.999\ldots = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \cdots,$$

and the sum of the geometric series on the right is

$$\frac{9/10}{1 - 1/10} = 1.$$

A skeptic who accepts the series interpretation could say that $0.999\ldots$ *converges* to 1, or that it is equal to 1 *in the limit*, but is not *equal* to 1. Standard usage is ambiguous as to whether the expression on the right stands for the series or for its limit. The fact that we use that notation whether or not the series converges argues in favor of the series interpretation. Also, we talk about the rate of convergence of such expressions.

So some distinction between convergence and equality in the present case might well be appropriate.

Perhaps the situation is that some real numbers, like $\sqrt{2}$, can only be approximated, whereas others, like 1, can be written exactly, but can also be approximated. So 0.999... is a series that approximates the exact number 1. Of course this dichotomy depends on what we allow for approximations. For some purposes we might allow any rational number, but for our present discussion the terminating decimals—the decimal fractions—are the natural candidates. These can only approximate $1/3$, for example, so we don't have an exact expression for $1/3$.

This might be a good point to say a word about infinity, potential and actual. What do those three dots in 0.999... signify? Simply the unending possibility of writing down more nines. That's what *potential infinity* is. *Actual infinity* is the idea that all those nines have been written down. Potential infinity invites us to consider 0.999... as a process. Actual infinity invites us to think of 0.999... as a completed object. People who think of 0.999... as a process—a series rather than a limit—are not so tempted to equate it with 1.

The most famous exponent of potential infinity, as opposed to actual infinity, was Aristotle, who said, in Book III of his *Physics* **[6]**, "In no other sense, then, does the infinite exist; but it does exist in this sense, namely potentially," and, "my argument ... denies that the infinite can exist in such a way as to be actually infinite." Gauss **[4]** considered "the use of an infinite quantity as a completed one" as "never permissible in mathematics." On the other hand, Georg Cantor **[1]**, who set the tone for twentieth century mathematics, believed that "each potential infinite ... presupposes an actual infinite."

Zeno's paradox of dichotomy has to do with the idea that, when you approach an object, you repeatedly halve your distance from it, with the result that you never reach it. Here is a decimal version of the paradox appropriate to the issue at hand: If you travel from 0 to 1, you must go successively through the points 0.9, 0.99, 0.999, 0.9999, and so on. You can never reach 1 because you never finish visiting all those intermediate points. Of course this paradox has been refuted many times, but that is an essential part of being a paradox: there are compelling arguments both for and against it. As Benson Mates said **[7]**, "it is possible to have impeccable arguments for both sides of a contradiction." Zeno's paradox has survived for thousands of years. The controversy about $0.999\ldots = 1$ is one of its aspects.

Decimal numbers What kind of setting would support the skeptic's view? By a (nonnegative) *decimal number*, I mean an infinite string of digits with a decimal point, like 1247.4215347528.... As usual, we don't allow the string to start with a 0 unless the decimal point comes immediately after. There are a couple of standard notational conventions. An infinite string of 6's (without a decimal point) is denoted by $\bar{6}$. So the number at issue, 0.999..., is denoted by $0.\bar{9}$. The number $120.45\bar{0}$ is said to *terminate*, and is denoted simply by 120.45, while $120.\bar{0}$ is denoted by 120 (with no decimal point).

The decimal numbers are ordered in the standard way. Line up the decimal points and compare corresponding digits. At the first place where the digits differ, the number with the bigger digit is the bigger number. So 999.999... is less than 1247.421..., because the initial 1 of the latter number is the first place where the digits differ, while 1247.430... is bigger than 1247.421..., because the 3 in the former number is bigger than the corresponding digit 2 in the latter, and that is the first place they differ. In particular, $0 = 0.000\ldots$ is the smallest decimal number, and $0.999\ldots < 1$.

How do you add two decimal numbers? There is a problem because carries can propagate over arbitrarily long stretches, and we can't start adding at the far right! But the carry can never be bigger than 1, so if the sum of the two digits in a given place is less than 9, or if it is greater than 9, then we can compute the digits in the sum up to, but not necessarily including, that place. If the sum of the digits is *exactly* 9 from some place on, then there will be no carry at, or past, that place.

Digression The question as to whether there is a carry into a given place cannot be decided by a finite computation. That means that you can't necessarily *compute* the decimal expansion of a sum from the decimal expansions of its addends. For example, suppose we have a number x whose decimal expansion starts out $0.05555\ldots$. So x is close to $0.0\overline{5}$, but we can't be sure it is equal to $0.0\overline{5}$ because we only know as many digits in its expansion as we care to compute. It may be that $x \leq 0.0\overline{5}$, or it may be that $x > 0.0\overline{5}$. What is the first digit after the decimal point in the expansion of $0.0\overline{4} + x$? It is 0 if $x \leq 0.0\overline{5}$, and it is 1 if $x > 0.0\overline{5}$. We may not have enough information, even after computing the first million places, to determine which of these alternatives holds.

The fact that you can't compute the decimal expansion of a sum from the decimal expansions of its addends is a well known phenomenon that was noticed by Turing. In a fully constructive treatment of the real numbers, this is often stated by saying (informally) that not every positive real number has a decimal expansion. More precisely, there is no known constructive proof that every positive real number has a decimal expansion.

What is the nature of this algebraic structure—decimal numbers under addition—that we have defined? You can check that the addition is commutative and associative, and that there is an identity, 0. The cancellable elements are precisely the terminating decimals, because $0.\overline{9} + x = 1 + x$ for all nonterminating x.

Is $1/3 = 0.\overline{3}$? Clearly, if a sum is cancellable, then each addend is cancellable, so there is no decimal number x such that $x + x + x = 1$. That is, $1/3$ is not a decimal number. More generally, no nonterminating decimal number x can satisfy an equation of the form $mx = n$ with m and n positive integers.

What about multiplication of decimal numbers? This is more complicated than addition, but can also be described as a natural extension of the way we multiply decimal fractions. We saw a little of that in the description of how to multiply 9 times $0.999\ldots$ to get $8.999\ldots$. However, it is convenient to define multiplication in terms of cuts, and we will want to look at cuts in any case for the insight they give into the controversy.

Dedekind cuts Dedekind cuts are usually defined in the ring of rational numbers, but if we are interested in decimal numbers, we will want to work with a different ring. Let D be any dense subring of the rational numbers. That is, D is any subring of the rational numbers other than the ring of integers. We have in mind the ring of *decimal fractions*, those rational numbers that can be expressed with denominator a power of 10. A *Dedekind cut* in D may be defined as a nonempty proper subset S of D such that if $x < y$ and $y \in S$, then $x \in S$.

This is essentially Dedekind's definition in [**2**]. Dedekind then identified the cut $\{x \in D : x < r\}$ with the cut $\{x \in D : x \leq r\}$, for each r in D, saying they were "only unessentially different." A similar move, made for example in [**8**, Definition 1.4], is to restrict ourselves to Dedekind cuts that do not have a greatest element, so $\{x \in D : x \leq r\}$ is not considered to be a cut. Why do that? Precisely to rule out the existence of distinct numbers $0.\overline{9}$ and 1. Indeed, $0.\overline{9}$ corresponds to the cut

$\{x \in D : x < 1\}$ while 1 corresponds to the cut $\{x \in D : x \leq 1\}$. In general, we may identify an element d in D with the cut $\{x \in D : x \leq d\}$ (we call these *principal cuts*). So we see that in the traditional definition of the real numbers, the equation $0.\bar{9} = 1$ is built in from the beginning. That is why anyone who challenges that equation is, in fact, challenging the traditional formal view of the real numbers.

If D is the ring of decimal fractions, then each decimal number ξ gives rise to a Dedekind cut

$$\{x \in D : x \leq \xi\}$$

in D. Note that this cut contains 0. Conversely, any Dedekind cut S in the ring of decimal fractions, that contains 0, is associated with a unique decimal number ξ as follows. For fixed m, the largest element of S that is of the form $n/10^m$ gives the digits in ξ up to the m-th place to the right of the decimal point. For example, the nonterminating decimal number 3.1415926535... corresponds to the cut consisting of all those decimal fractions r such that $r < 3$, or $r < 3.1$, or $r < 3.14$, or $r < 3.141$, or ... .

Let cut D denote the set of all Dedekind cuts in D. Define the sum of two cuts in the usual way

$$u + v = \{x + y : x \in u \text{ and } y \in v\}.$$

It is easily shown that the commutative and associate laws hold, and that the principal cut $\{x \in D : x \leq 0\}$ is an additive identity. So cut D is a *monoid*. The elements of D, in the guise of principal cuts, form a subgroup of this monoid. In fact, D consists precisely of the (additively) cancellable elements of cut D. This is because

$$u + d = \{x + d : x \in u\}$$

while if $1^- = \{x \in D : x < 1\}$, then $u + 1^- = u + 1$ for any cut u that is not principal. However, the nonprincipal cuts are cancellable among themselves, and are closed under addition, so they also form a subgroup of cut D. This group may be identified with the traditional real numbers, as Rudin **[8]** does with cuts in the rational numbers. Recall that any traditional positive real number has a unique *nonterminating* decimal expansion. Note that $0^- = \{x \in D : x < 0\}$ is the identity element of the group of nonprincipal cuts.

The order on cut D is given by inclusion of cuts. The *weakly positive* cuts are those that contain the rational number 0. These correspond exactly to the decimal numbers if D is the ring of decimal fractions. The *product* of two weakly positive cuts u and v is defined to be $\{st : s \in u \text{ and } t \in v\}$. This multiplication on weakly positive cuts shows how to multiply any two decimal numbers. It's straightforward to show that the associative, commutative, and distributive laws hold. So the decimal numbers form a *positive, totally ordered, commutative semiring* in the sense of **[5]**.

The picture here is the traditional real numbers, in the form of nonprincipal cuts, living uneasily together with the ring D, in the form of principal cuts. For each element d of D, there is a traditional real number d^- just below it, and $u + d^- = u + d$ for each traditional real number u. That, for traditionalists, is a complete description of the additive structure of cut D. Note that $d^- = d + 0^-$.

Clearly $0.\bar{9} = 1 + 0^-$, so 0^- is sort of a negative infinitesimal. On the other hand, you can't solve the equation $0.\bar{9} + X = 1$ because, in cut D, the sum of a traditional real with any real is a traditional real.

Another point of view We looked at Dedekind cuts in order to describe multiplication of decimal numbers, and to see another way of describing the decimal numbers themselves. Because we looked at cuts in the ring of decimal fractions, both positive and negative, we got some numbers in addition to the decimal numbers. While the number 1^- corresponds to the decimal number $0.\bar{9}$, there is no decimal number corresponding to $(-1)^-$, which is the cut $\{x \in D : x < -1\}$. Nor is there a decimal number corresponding to -1 itself.

However, cut D can be characterized in the following way (for D the nonnegative decimal fractions). It contains the decimal numbers, and all the decimal fractions, both positive and negative. Each element of cut D can be written as a difference $u - d$ of a decimal number u and a (nonnegative) decimal fraction d. We may think of cut D as being obtained from the decimal numbers by adjoining the negative decimal fractions, and taking sums. This construction is legitimate because the decimal fractions are cancellable in D.

Instead of extending the decimal numbers to include additive inverses of those decimal numbers that are cancellable under addition, we could extend them to include multiplicative inverses of those decimal numbers that are cancellable under multiplication. These are exactly the positive decimal fractions, because $(0.\bar{9})x = x$ whenever x is a nonterminating decimal number. This construction is an instance of forming a semiring of fractions; see **[5]**. It is not hard to verify that the result is (isomorphic to) the weakly positive elements of cut $\mathbb{Q}$, where $\mathbb{Q}$ is the ring of rational numbers.

Open problems Although we can introduce negative decimal fractions, negative numbers in general present a serious problem because we don't have cancellation in cut D. We can't simply write them as additive inverses of positive numbers. Moreover, we have no interpretation for the number $-3.14159265\ldots$ because this represents a process of approximation from *above*, -3.14159 being greater than $-3.14159265\ldots$, whereas in cut D all real numbers are approximated from *below*. Of course we could just introduce symbols like $-3.14159265\ldots$, but it's not clear how to get a satisfactory coherent system that incorporates them.

Because of this, multiplication of *arbitrary* real numbers is also a serious problem, if for no other reason than that we don't know how to multiply -1 by $3.14159265\ldots$. Even in the traditional approach, multiplication is awkward. The elegant treatment of addition is replaced by an ugly division into cases: one defines how to multiply positive numbers, and extends to negative numbers according to the usual rules **[8**, pp. 7–8**]**.

REFERENCES

1. Cantor, Georg, *Über die verschiedenen Ansichten in Bezug auf die actualunendlichien Zahlen*, 1886, quoted in Hallett, Michael, *Cantorian Set Theory and Limitation of Size*, Oxford University Press, Oxford, UK, 1988.
2. Dedekind, Richard, Continuity and irrational numbers, (1872), in *Essays on the Theory of Numbers*, Dover, New York, NY, 1963.
3. Faltin, F., N. Metropolis, B. Ross, and G.-C. Rota, The real numbers as a wreath product, *Advances in Math.*, 16 (1975), 278–304.
4. Gauss, Carl Friedrich, *Briefwechsel zwischen C. F. Gauss und H. C. Schumacher*, 1860, quoted in Dauben, Joseph Warren, *Georg Cantor, His Mathematics and Philosophy of the Infinite*, Princeton University Press, Princeton, NJ, 1990.
5. Golan, Jonathan S., *The Theory of Semirings with Applications in Mathematics and Theoretical Computer Science*, Wiley, New York, NY, 1991.
6. Heath, Thomas, *Mathematics in Aristotle*, Oxford University Press, Oxford, UK, 1970.
7. Mates, Benson, *Skeptical Essays*, University of Chicago Press, Chicago, IL, 1981.
8. Rudin, Walter, *Principles of Mathematical Analysis*, McGraw-Hill, New York, NY, 1964.

Math Bite: On the Definition of Collineation

Let V and W be vector spaces over some field $\mathbb{F}$, and suppose that $\dim V \geq 2$. A *collineation* is a one-to-one function $f: V \to W$ such that if $x, y, z \in V$ are collinear, then $f(x), f(y), f(z)$ are collinear. The definition says, in essence, that a collineation sends lines to lines. But this is not the precise content of the definition. For one thing, the definition requires that a collineation be one-to-one, which is apparently stronger than simply preserving collinearity. For another, at least on casual inspection, the definition does not require the image of a line to be a line. It would suffice that the image of a line be a *subset* of a line—although this turns out not to occur. (Indeed, one version of the Fundamental Theorem of Projective Geometry states that, for $\mathbb{F} = \mathbb{R}$, every collineation is an affine map.) Given these observations, it is natural to wonder just how far the idea of sending lines to lines goes towards characterizing collineations. The answer is: *it almost does.*

THEOREM. *Suppose that $f: V \to W$ is a function for which the image of every line in V is a line in W. Then either f is a collineation or $f(V)$ has dimension 1.*

Proof. First observe that f sends (affine) subspaces to (affine) subspaces. Indeed, if A is a subspace of V and if $x, y \in f(A)$ are distinct points, then $x = f(a)$ and $y = f(b)$ for some distinct $a, b \in A$. Then if L is the unique line through a and b, we have $L \subseteq A$, so the line $f(L)$ through x and y is contained in $f(A)$.

Suppose that $\dim f(V) > 1$ and that $x, y \in V$ with $f(x) = f(y)$. Let L_{xy} be the line through x and y. By hypothesis, there exists z with $f(z) \notin f(L_{xy})$. Let A be a plane through x, y, and z. Let L be the line through $f(x)$ and $f(z)$, and let L_{xz} (resp. L_{yz}) be the line through x and z (resp. y and z). So $f(L_{xz}) = f(L_{yz}) = L$. Note that dim $f(A) > 1$, since $f(A)$ contains the plane through $f(L_{xy})$ and $f(z)$. Choose any line L' in $f(A)$ that is disjoint from L. Choose a line $\hat{L}$ in A with $f(\hat{L}) = L'$. Since L and L' are disjoint, L_{xz} and $\hat{L}$ must be disjoint. Hence, since they both belong to A, L_{xz} and $\hat{L}$ are parallel. Similarly, L_{yz} and $\hat{L}$ are parallel. Hence L_{xz} and L_{yz} are parallel. Since z belongs to both lines, $L_{xz} = L_{yz}$. So we have two distinct lines, L_{xz} and L_{xy}, both of which contain x and y. Hence $x = y$, and the proof is complete.

In fact, there do exist functions that send lines to lines but are not collineations. Suppose that V is a real normed vector space with norm $\|\cdot\|$. Consider $f: V \to \mathbb{R}$, $f(x) = \|x\| \cdot \sin(\|x\|)$. Since the codomain has dimension 1, the only thing that needs verifying here is that the image of every line is the entire codomain, and not just part of it.

—G. CAIRNS, G. ELTON, P. J. STACEY
LA TROBE UNIVERSITY
MELBOURNE, AUSTRALIA 3083

Density of the Images of Integers Under Continuous Functions with Irrational Periods

SINING ZHENG
JIANGCHEN CHENG
Dalian University of Technology
Dalian 116024
China

1. Introduction We begin with the sequence $\{\sin n\}$, $n = 1, 2, \ldots$. Since this divergent sequence is bounded, it has a convergent subsequence. To establish the existence of a subsequence that converges to zero, it suffices to show that, for any integer $k > 0$, there are integers n_k and m_k such that $|n_k - m_k\pi| < 1/k$. If $[x]$ denotes the integer part of a real number x, then the fractional part is $x - [x]$, and $x - [x] \in (0, 1)$ for any irrational number x. Clearly, there are at least two integers p_k and q_k among any $(k+1)$ distinct integers such that

$$|p_k\pi - [p_k\pi] - (q_k\pi - [q_k\pi])| = |(p_k - q_k)\pi - ([p_k\pi] - [q_k\pi])| < 1/k,$$

since the fractional part of any integer multiple of π is in $(0, 1)$. Thus, we get $|n_k - m_k\pi| < 1/k$ with $m_k = p_k - q_k$, $n_k = [p_k\pi] - [q_k\pi]$.

It is interesting that the range of $\sin x$ is full of accumulation points of the set $\{\sin n\}$. Indeed, for every $\alpha \in [-1, 1]$ there is a subsequence $\{n_k\}$ of $\{n\}$ such that $\lim_{k\to\infty} \sin n_k = \alpha$. In fact, if θ is an irrational number and $n = 1, 2, \ldots$, then $n\theta - [n\theta]$ is dense in the interval $(0, 1)$ by Kronecker's theorem **[1, 3]** and uniformly distributed in $(0, 1)$ by Weyl's principle **[2]**. So $2n\pi - [2n\pi] + k$ is dense and uniformly distributed in the interval $(k, k+1)$ for each integer k, which implies the density of the sequence $\{\sin n\}$ in the closed unit interval. The analogous conclusions hold for general continuous functions with irrational periods. In this paper we propose a *constructive* procedure, motivated by the continued fraction algorithm, for obtaining such subsequences.

2. A recursive procedure on irrational numbers The following two lemmas describe a useful recursive procedure for any given irrational number L. We get two sequences $\{n_k\}$ and $\{m_k\}$ of integers such that $m_k L + n_k$ tends to zero monotonically as $k \to \infty$.

LEMMA 1. *Let L be any irrational number greater than* 1, *and suppose that* $0 < x_1 < x_0$ *and* $x_0/x_1 = L$. *Then*

$$x_{n+2} = x_n - [x_n/x_{n+1}]x_{n+1} \tag{1}$$

is well defined for all $n = 0, 1, 2, \ldots$ (*that is,* x_n *is never zero*), *and*

$$0 < x_{n+2} < x_n/2 \tag{2}$$

for all n.

Proof. By assumption, $0 < x_1 < x_0$ and x_0/x_1 is irrational.

Assume that $0 < x_{k+1} < x_k$, with x_k/x_{k+1} irrational. By (1), $x_{k+2} = x_k - [x_k/x_{k+1}]x_{k+1}$, and $x_{k+2}/x_{k+1} - [x_k/x_{k+1}]$ is irrational and $0 < x_{k+2} < x_{k+1}$. By induction, x_{n+1}/x_{n+2} is irrational, with $0 < x_{n+2} < x_{n+1}$ for all $n = 0, 1, 2, \ldots$. Moreover,

$$x_{n+2} = x_n - [x_n/x_{n+1}]x_{n+1} \le x_n - x_{n+1} < x_n - x_{n+2}.$$

Thus $x_{n+2} < x_n/2$ for $n = 0, 1, 2, \ldots$. □

REMARK. Inequality (2) and the inequality $0 < x_{n+2} < x_{n+1}$ imply that the sequence $\{x_n\}$ (defined by (1)) tends monotonically to zero as $n \to \infty$.

LEMMA 2. *For each* x_k *defined by* (1) *we can find integers* m_k *and* n_k *such that* $x_k = m_k L + n_k$, *with* $m_k = (-1)^k |m_k|$ *and* $n_k = (-1)^{k-1}|n_k|$ *for* $k = 2, 3, \ldots$.

Proof. We find the m_k and n_k inductively. For simplicity, put $x_0 = L > 1$ and $x_1 = 1$. Then

$$x_2 = x_0 - [x_0/x_1]x_1 = L - [L] = m_2 L + n_2$$

with $m_2 = 1$ and $n_2 = -[L] = -|n_2|$. We have

$$\begin{aligned} x_3 = x_1 - [x_1/x_2]x_2 &= 1 - [x_1/x_2](L - [L]) \\ &= -[x_1/x_2]L + 1 + [x_1/x_2][L] = m_3 L + n_3, \end{aligned}$$

with $m_3 = -[x_1/x_2] = -|m_3|$ and $n_3 = 1 + [x_1/x_2][L] = |n_3|$.

Now assume that m_k, n_k, m_{k+1}, and n_{k+1} have been found, i.e., that

$$x_k = m_k L + n_k, \quad m_k = (-1)^k |m_k|, \quad n_k = (-1)^{k-1}|n_k|,$$

and

$$x_{k+1} = m_{k+1} L + n_{k+1}, \quad m_{k+1} = (-1)^{k+1}|m_{k+1}|, \quad n_{k+1} = (-1)^k |n_{k+1}|.$$

Then

$$\begin{aligned} x_{k+2} &= x_k - [x_k/x_{k+1}]x_{k+1} = m_k L + n_k - [x_k/x_{k+1}](m_{k+1}L + n_{k+1}) \\ &= (m_k - [x_k/x_{k+1}]m_{k+1})L + n_k - [x_k/x_{k+1}]n_{k+1} = m_{k+2}L + n_{k+2}, \end{aligned}$$

where

$$\begin{aligned} m_{k+2} &= m_k - [x_k/x_{k+1}]m_{k+1} = (-1)^k |m_k| - (-1)^{k+1}|m_{k+1}|[x_k/x_{k+1}] \\ &= (-1)^{k+2}(|m_k| + |m_{k+1}|[x_k/x_{k+1}]) = (-1)^{k+2}|m_{k+2}| \end{aligned}$$

and

$$\begin{aligned} n_{k+2} &= n_k - [x_k/x_{k+1}]n_{k+1} = (-1)^{k-1}|n_k| - (-1)^k |n_{k+1}|[x_k/x_{k+1}] \\ &= (-1)^{k+1}(|n_k| + |n_{k+1}|[x_k/x_{k+1}]) = (-1)^{k+1}|n_{k+2}|. \end{aligned}$$

This completes the proof by induction. □

3. Constructive results for density of the images of integers Now we state our constructive results. First we consider $f(x) = \sin x$, one of the simplest functions with irrational periods.

THEOREM 1. *There exists a subsequence* $\{p_k\}$ *of the sequence* $\{p\}$ *of natural numbers such that* $\lim_{k\to\infty} \sin p_k = 0$.

Proof. Take $L = \pi$ in Lemmas 1 and 2. Let $p_k = |n_k|$ (n_k is determined in Lemma 2). Then

$$|\sin p_k| = |n_k| = |\sin(x_k - m_k \pi)| = |\sin x_k|.$$

By Lemma 1, x_k tends to zero monotonically as $k \to \infty$, so $\lim_{k\to\infty} \sin p_k = 0$, and the proof is complete. □

In fact, *every* point of the interval $[-1, 1]$ is an accumulation point of $\{\sin n\}$.

THEOREM 2. *For every* $\alpha \in [-1,1]$ *there exists a subsequence* $\{p_k\}$ *of* $\{p\}$ *such that* $\lim_{k\to\infty} \sin p_k = \alpha$.

Proof. The case $\alpha = 0$ is Theorem 1. Now put $L = 2\pi$.

Case (i): $0 < \alpha \le 1$. We have $A := \arcsin \alpha \in (0, \pi/2]$. Take $r_k = A - [A/x_k]x_k$, with x_k defined by (1). We can assume $A > x_k$ since x_k tends to zero. Obviously, $0 \le r_k \le x_k$. By Lemma 2, $x_{2k-1} = 2m_{2k-1}\pi + n_{2k-1}$ with $n_{2k-1} = (-1)^{2k}|n_{2k-1}| = |n_{2k-1}| > 0$. Put $p_k = [A/x_{2k-1}]n_{2k-1}$, and $q_k = [A/x_{2k-1}]m_{2k-1}$. Then $p_k \in \mathbb{N}$ and $A - r_{2k-1} = [A/x_{2k-1}]x_{2k-1} = p_k + 2q_k\pi$. So

$$\sin p_k = \sin(A - r_{2k-1} - 2q_k\pi) = \sin(A - r_{2k-1}).$$

Thus $\lim_{k\to\infty} \sin p_k = \sin A = \alpha$, since r_{2k-1} tends to zero as $k \to \infty$.

Case (ii): $-1 \le \alpha < 0$. We have $A = \arcsin \alpha \in [-\pi/2, 0)$. Take $r_k = |A| - [|A|/x_k]x_k$, with x_k defined by (1), and assume $0 < x_k < |A|$. It follows from Lemma 2 that

$$x_{2k} = 2m_{2k}\pi + n_{2k} = 2m_{2k} + (-1)^{2k-1}|n_{2k}| = 2m_{2k}\pi - |n_{2k}|.$$

If we write

$$p'_k = [|A|/x_{2k}]n_{2k} \quad \text{and} \quad q_k = [|A|/x_{2k}]m_{2k},$$

then $|A| - r_{2k} = p'_k + 2q_k\pi$ and $p_k = -p'_k \in \mathbb{N}$. Thus

$$\sin p_k = \sin(-p'_k) = -\sin(|A| - r_{2k} - 2q_k\pi) = -\sin(|A| - r_{2k})$$

and $\lim_{k\to\infty} \sin p_k = -\sin|A| = \sin A = \alpha$. The proof is complete. □

REMARK. If we take $A = \pi - \arcsin \alpha$ in Theorem 2, then $\pi/2 \le A \le 3\pi/2$, and therefore the three cases of $\alpha = 0$, $0 < \alpha \le 1$, and $-1 \le \alpha < 0$ can be unified and treated by the procedure used for the case when $0 < \alpha \le 1$.

The results above are easily extended to any continuous function with irrational period.

THEOREM 3. *Let* $f(x)$ *be a continuous function with irrational period* T. *Then, for any point* α *in the range of* $f(x)$, *there exists a subsequence* $\{p_k\}$ *of* $\{p\}$ *such that* $\lim_{k\to\infty} f(x_{p_k}) = \alpha$.

Proof. We assume without loss of generality that $(0, T]$ is a periodic interval. Then, for any point α in the range of $f(x)$, there is $A \in (0, T]$ such that $f(A) = \alpha$.

If $T > 1$, take $L = T$ and let x_k be given by (1). If we define $r_k = A - [A/x_k]x_k$, we can prove the theorem as in case (ii) of the proof of Theorem 2.

If $0 < T \le 1$, then there is $n_T \in \mathbb{N}$ such that $(n_T - 1)T < 1 < n_T T$. If we take $L = n_T T$, the preceding argument applies. The proof is complete. □

Acknowledgment. Supported by NEC and NNSF of China.

REFERENCES

1. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, UK, 1981.
2. L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley-Interscience, New York, NY, 1974.
3. I. Niven, *Irrational Numbers*, Carus Math. Monographs, No. 11, Mathematical Association of America, Washington, DC, 1956.

Bold Play Is Best: A Simple Proof

RICHARD ISAAC
Lehman College, CUNY
Bronx, NY 10468-1589

Introduction In the classical ruin problem, a gambler with initial capital i dollars plays against an adversary with initial capital $a-i$ dollars. Here i and a are positive integers, $i<a$. The gambler wins a dollar with probability p and loses a dollar with probability $q=1-p$; the game is repeated until one of the players goes broke. (Part of the analysis of the problem shows the game cannot go on forever.) Another interpretation of this model is that of a gambler playing a game in a casino (the adversary) where the gambler plays until she goes broke or until she wins a fixed predetermined amount, at which time she quits. Given a total fixed capital a, we are interested in the probability q_i that the gambler starting with initial capital i, $1 \le i \le a-1$, is ruined.

This problem is often discussed in a first course in probability and introduces the student to the ideas of random walks and Markov chains. Our main reference is the classic book of Feller [**3**, chapter 14]; see also [**1**], [**4**], and [**5**]. We will regard the sequence of the gambler's fortune after each play as a random walk in the interval $[0, a]$, with absorbing barriers at 0 and a. The probability of ruin is the probability of hitting 0 before hitting a.

Using a difference equation approach and some algebra, the following solution is derived in the case $p \neq 1/2$ (see, e.g., [**3**]):

$$q_i = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^a - 1}. \tag{1}$$

This is valid for $1 \le i \le a-1$, and even for the boundary values 0 and a if the intuitively reasonable definitions $q_0 = 1$, $q_a = 0$ are made. Thus a gambler with 100 dollars in her pocket, playing repeated games of craps ($p \approx .493$) and determined to win 10 dollars or go broke in the attempt, will be ruined with approximate probability .253.

Now what happens if the gambler changes the stakes: instead of betting a dollar at each play, she bets a fixed s dollars? Here s must be possible in the sense that both i and a must be divisible by s. For example, the gambler with 100 dollars who wants to win 10 dollars can bet at 1, 2, 5, or 10 dollar stakes. What does changing the stakes do to the gambler's probability of ruin? What is the gambler's best strategy to minimize the probability of ruin?

From the random walk point of view, changing the stakes means the unit of fortune has changed, so the gambler with 100 dollars determined to win 10 dollars betting at 10 dollar stakes should have the same probability of ruin as the gambler with 10 dollars planning to quit after winning a dollar, and playing at one dollar stakes. Define $q_i(s)$ as the probability of the gambler's ruin, given that she starts with i dollars and bets at fixed s dollar stakes. For ease of notation, set $r = \frac{q}{p}$ in (1) and then note that this relation, together with the above remarks on changing stakes, shows

$$q_i(s) = \frac{r^{\frac{a}{s}} - r^{\frac{i}{s}}}{r^{\frac{a}{s}} - 1}. \tag{2}$$

Let us assume from now on that *the gambler plays an unfavorable game, that is,* $p < 1/2$. This, after all, is the typical situation for a gambler at a casino. It turns out, surprisingly to many, that the gambler minimizes her ruin probability by playing *boldly*, that is, by playing at the highest stakes possible. Feller **[3]** states this fact about bold play, but only proves it in the case of doubling or halving bets. Chung **[1]** and Ross **[5]** discuss the gambler's ruin problem, but omit any mention of what happens when the stakes are changed. In my own book **[4]** the result is stated but not proved. At a more advanced perspective, Dubins and Savage **[2]** prove the optimality of bold play in the unfavorable game case in a much more general context than the one considered here, but their proofs require a lot of high-powered mathematical machinery. Yet a simple proof of this interesting fact in the classical case outlined above, although not completely trivial, is short and depends only on some elementary calculus. Here's how it goes.

The result Consider the right-hand side of equation (2) as a continuous function of s for fixed r, i, and a, and $s > 0$. If this function can be proved to be decreasing in s, the optimality of bold play follows. The obvious approach is to differentiate the right-hand side of (2) with respect to s and see that the result is negative. But the obvious approach gives a rather messy expression, not immediately seen to be negative. Instead, we take the derivative in two steps, and use the chain rule to arrive at the conclusion.

THEOREM. *If* $p < 1/2$, *the function*

$$q_i(s) = \frac{r^{\frac{a}{s}} - r^{\frac{i}{s}}}{r^{\frac{a}{s}} - 1}$$

is a decreasing function of s *on* $s > 0$.

Proof. Let $w = r^{\frac{i}{s}}$. Since $p < 1/2$ it follows that $w > 1$. Let $ki = a$. Since $i < a$ we must have $k > 1$. Now rewrite $q_i(s)$ as

$$q_i(s) = \frac{w^k - w}{w^k - 1} = 1 + \frac{1 - w}{w^k - 1}. \tag{3}$$

Consider the derivative $dq_i(s)/dw$; we will show this derivative is positive on $w > 1$. This derivative can be found by taking the derivative of the second term of the right-hand side of (3). Using this term and the quotient rule for taking derivatives, the sign of $dq_i(s)/dw$ is seen to be determined by the quantity

$$-(w^k - 1) - (1 - w) \cdot kw^{k-1}. \tag{4}$$

We claim that this expression is positive for all $w > 1$. To see this, observe that the positivity of (4) is equivalent to the inequality

$$h(w) := (k - 1)w^k - kw^{k-1} > -1. \tag{5}$$

Consider the function $h(w)$ on the interval $w > 1$. We get

$$h'(w) = k(k-1)(w^{k-1} - w^{k-2}),$$

which is positive because $w > 1$ and $k > 1$. Thus $h(w)$ is an increasing function on $w > 1$. But note that $h(1) = -1$; this observation completes the proof of the inequality (5). Consequently, the derivative $dq_i(s)/dw$ is positive on $w > 1$. Moreover, from the definition of w it is clear that w is a decreasing function of s, that is $dw/ds < 0$, so that

$$\frac{dq_i(s)}{ds} = \frac{dq_i(s)}{dw} \cdot \frac{dw}{ds} < 0,$$

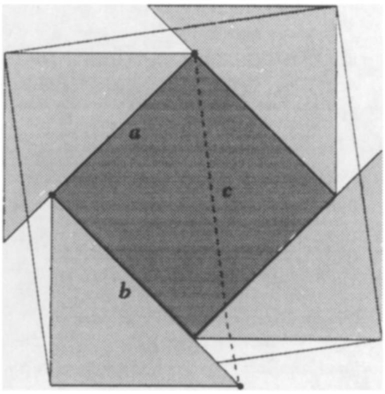
and the proof of the theorem is complete.

Acknowledgment. We thank the referee for several useful suggestions.

REFERENCES

1. K. L. Chung, *Elementary Probability Theory with Stochastic Processes*, Springer–Verlag, New York, NY, 1979.
2. L. E. Dubins and L. J. Savage, *How to Gamble if You Must: Inequalities for Stochastic Processes*, McGraw-Hill, New York, NY, 1965.
3. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, third edition, Wiley, New York, NY, 1968.
4. R. Isaac, *The Pleasures of Probability*, Springer-Verlag, New York, NY, 1995.
5. S. Ross, *A First Course in Probability*, Macmillan, New York, NY, 1988.

Proof Without Words: $a^2 + b^2 = c^2$



—Poo-sung Park
Seoul National University
Shillim-Dong San 56–1, Kwanak-Gu
Seoul, Korea

Consider the function $h(w)$ on the interval $w > 1$. We get

$$h'(w) = k(k-1)(w^{k-1} - w^{k-2}),$$

which is positive because $w > 1$ and $k > 1$. Thus $h(w)$ is an increasing function on $w > 1$. But note that $h(1) = -1$; this observation completes the proof of the inequality (5). Consequently, the derivative $dq_i(s)/dw$ is positive on $w > 1$. Moreover, from the definition of w it is clear that w is a decreasing function of s, that is $dw/ds < 0$, so that

$$\frac{dq_i(s)}{ds} = \frac{dq_i(s)}{dw} \cdot \frac{dw}{ds} < 0,$$

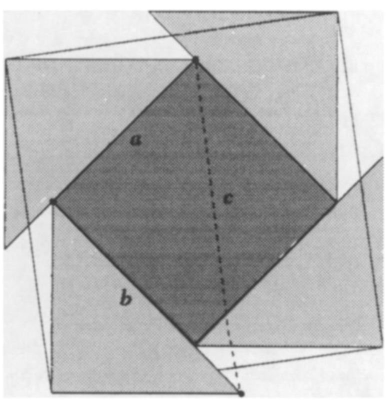
and the proof of the theorem is complete.

Acknowledgment. We thank the referee for several useful suggestions.

REFERENCES

1. K. L. Chung, *Elementary Probability Theory with Stochastic Processes*, Springer–Verlag, New York, NY, 1979.
2. L. E. Dubins and L. J. Savage, *How to Gamble if You Must: Inequalities for Stochastic Processes*, McGraw-Hill, New York, NY, 1965.
3. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, third edition, Wiley, New York, NY, 1968.
4. R. Isaac, *The Pleasures of Probability*, Springer-Verlag, New York, NY, 1995.
5. S. Ross, *A First Course in Probability*, Macmillan, New York, NY, 1988.

Proof Without Words: $a^2 + b^2 = c^2$



—Poo-sung Park
Seoul National University
Shillim-Dong San 56–1, Kwanak-Gu
Seoul, Korea

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 May 1, 2000.

1584. *Proposed by Ira Rosenholtz, Eastern Illinois University, Charleston, Illinois.*

Let n be a positive integer, and let $\triangle_n$ be the set of ordered triples of positive integers which are the side lengths of a nondegenerate triangle of perimeter n. Show that the cardinality of $\triangle_n$ is a triangular number.

1585. *Proposed by Shahin Amrahov, Ankara, Turkey.*

Prove that the number

$$\sum_{n=1}^{1998}\left(\sum_{k=1}^{n} k^{1998}\right)1997^{n-1}$$

is not a perfect square.

1586. *Proposed by Gerald A. Edgar, Ohio State University, Columbus, Ohio.*

Let w be a nonnegative, continuous, and nonincreasing function on $[0,\infty)$. Let g be a nonnegative, continuous function on $[0,\infty)$. For a given $\alpha \in (0,1)$, assume that

$$\alpha x g(x) \le \int_0^x \min\{w(t), g(x)\}\, dt \qquad \text{for all } x > 0.$$

(a) Show that there is a positive constant c_α, independent of w and g, such that

$$\int_0^\infty g(x)\,dx \le c_\alpha \int_0^\infty w(x)\,dx.$$

(b)* Find the smallest possible value of c_α.

(* Neither the proposer nor the editors have provided a solution to (b). Solvers of only (a) will be acknowledged.)

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.*

1587. *Proposed by Kevin Ferland, Bloomsburg University, Bloomsburg, Pennsylvania, and Florian Luca, Czech Academy of Science, Prague, Czech Republic.*

Consider constructions using straightedge and compass. Prove or disprove the following:
(a) Given any ellipse, the foci can be constructed.
(b) Given any hyperbola, the foci and asymptotes can be constructed.
(c) Given any parabola, the focus and directrix can be constructed.

1588. *Proposed by Emeric Deutsch, Polytechnic University, Brooklyn, New York.*

Let $a = (a_0, a_1, a_2, \dots)$ be any sequence of complex numbers. Define the sequence transformation T by $T(a) = (b_0, b_1, b_2, \dots)$, where

$$b_n = \begin{vmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} & a_n \\ -1 & a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \\ 0 & -1 & a_0 & \cdots & a_{n-3} & a_{n-2} \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & a_0 & a_1 \\ 0 & 0 & 0 & \cdots & -1 & a_0 \end{vmatrix}.$$

Find a determinant expression for the nth term of the sequence $T^{(q)}(a)$, where q is a positive integer. (Here $T^{(q)}$ denotes the q-fold composition of T.)

Quickies

Answers to the Quickies are on page 414

Q895. *Proposed by Herbert E. Salzer, Brooklyn, New York.*

If m and n are relatively prime integers, prove that

$$m^3n^2 + m^2n^3 + m^3n + 2m^2n^2 + mn^3 \quad \text{and} \quad m^2n + mn^2 + mn + m + n$$

are relatively prime.

Q896. *Proposed by Norman Schaumberger, Professor Emeritus, Bronx Community College, New York, New York.*

If a, b, and c are positive, prove that

$$\left(a^{1/b}b^{1/c}c^{1/a}\right)^{a+b+c} \geq \left(a^b b^c c^a\right)^{1/a+1/b+1/c}.$$

Solutions

Periodic Sequences of Sums **December 1998**

1559. *Proposed by Joaquín Gómez Rey, I. B. "Luis Buñuel," Alcorcón, Madrid, Spain.*

For what complex numbers z is the sequence $(a_n(z))_{n \geq 0}$ defined by

$$a_n(z) = \sum_{k=0}^{n} \binom{n+k}{2k} z^k$$

periodic?

Solution by Paul K. Stockmeyer, Department of Computer Science, College of William and Mary, Williamsburg, Virginia.

The values of z that make the sequence $(a_n(z))_{n\geq 0}$ periodic are exactly the numbers $2\cos(2\pi r)-2$ where r is a rational number in the interval $0\leq r<1/2$. The period is the denominator of r when reduced to lowest terms.

It is a straightforward task to confirm that the given sequence satisfies the recurrence relation $a_n(z)-(z+2)a_{n-1}(z)+a_{n-2}(z)=0$, with characteristic polynomial $p(x)=x^2-(z+2)x+1$. This polynomial has a repeated root when $z=0$ and when $z=-4$. In the first case, we have $a_n(0)=1$ for all $n\geq 0$, which is clearly periodic. In the second case, we have $a_n(-4)=(-1)^n(2n+1)$, which is not.

Otherwise, the characteristic polynomial has two distinct roots x_1 and x_2, which implies that

$$a_n(z)=c_1x_1^n+c_2x_2^n,$$

where the constants c_1 and c_2 are determined by the values of $a_0(z)$ and $a_1(z)$. Noting that $x_1x_2=1$ and that $a_0(z)=1\neq 0$, we see that this sequence will be periodic if and only if the numbers x_1 and x_2 are roots of unity, of the form $e^{2\pi ri}$ for some rational r. Setting $p(x)=0$ for such a value of x and solving for z yields $z=e^{2\pi ri}+e^{-2\pi ri}-2=2\cos(2\pi r)-2$. From the symmetry of the cosine function, we can restrict our attention to the range $0\leq r\leq 1/2$. The endpoint $r=0$ corresponds to $z=0$, where we have seen that the sequence is periodic; the endpoint $r=1/2$ corresponds to $z=-4$, where we have seen that it is not.

Also solved by J. C. Binz (Switzerland), Con Amore Problem Group (Denmark), Daniele Donini (Italy), Harry Kiesel, Heinz-Jürgen Seiffert (Germany), Michael Vowe (Switzerland), and the proposer. There was one incorrect solution.

Areas of Triangles Inscribed in a Circle **December 1998**

1560. *Proposed by Wu Wei Chao, Guang Zhou Normal University, Guang Zhou City, Guang Dong Province, China.*

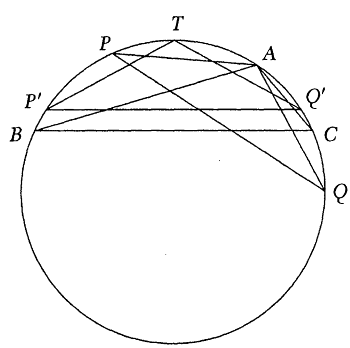
Points A, B, C, P, Q, and T lie on a circle and satisfy $AB>AC$, T is on the same side of BC as A with $TB=TC$, and $AP=AQ=\sqrt{AB\cdot AC}$. Let $[ABC]$ denote the area of $\triangle ABC$, and so forth.

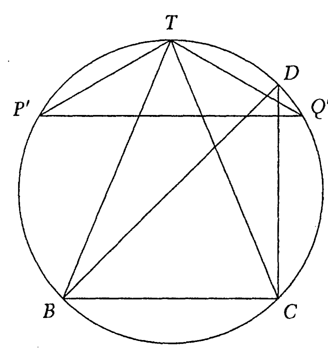
(a) If $\angle BAC\geq 90°$, prove that $[ABC]>[APQ]$.

(b) If $\sqrt{AB\cdot AC}\leq BC$, prove that $[TBC]>[APQ]$.

Solution by Peter Y. Woo, Biola University, La Mirada, California.

Let O be the center of the circle, which we may assume has radius 1. Let P' and Q' be the two points on the circle for which $TP'=TQ'=AP=AQ$. Then $[TP'Q']=[APQ]$, so that we may consider $\triangle TP'Q'$ in place of $\triangle APQ$.





By the arithmetic mean-geometric mean inequality, $TP' < (AB + AC)/2$. On the other hand,

$$\begin{aligned}\frac{AB+AC}{2} &= \sin\frac{\angle AOB}{2} + \sin\frac{\angle AOC}{2}\\ &= 2\sin\frac{\angle AOB + \angle AOC}{4}\cos\frac{\angle AOB - \angle AOC}{4}\\ &= TB\cos\frac{\angle AOB - \angle AOC}{4} < TB.\end{aligned}$$

Therefore, $TP' < TB$ and $\angle P'TQ' > \angle BTC = \angle BAC$.

(a) When $\angle BAC \geq 90°$, we have $\sin\angle BAC > \sin\angle P'TQ'$, hence

$$[ABC] = \frac{1}{2}AB \cdot AC \sin\angle BAC > \frac{1}{2}TP' \cdot TQ' \sin\angle P'TQ' = [TP'Q'].$$

(b) We have

$$[TBC] = \frac{1}{2}TB \cdot TC \sin\angle BTC = 4\sin\frac{\angle BTC}{2}\cos^3\frac{\angle BTC}{2}.$$

Simple calculus shows that this area increases as $\angle BTC$ increases from 0 to $\pi/3$ and then decreases as $\angle BTC$ increases $\pi/3$ to π. Thus, if $\angle BTC \geq \pi/3$, it is clear that $[TBC] > [TP'Q']$. If $\angle BTC < \pi/3$, let $D \neq B$ on the circle satisfy $CD = BC$. Then the given $TP' \leq BC$ implies $\angle P'TQ' \geq \angle BCD > \pi/3$, so that $[P'TQ'] \leq [BCD]$. However, $[TBC] > [BCD]$ because the former has the longer altitude to their common side BC. We again conclude $[TBC] > [TP'Q']$.

Also solved by Victor Y. Kutsenok, Hieu D. Nguyen, and the proposer.

Nonnegative Linear Combinations of Products **December 1998**

1561. *Proposed by Emre Alkan, student, University of Wisconsin, Madison, Wisconsin.*

Let $a_1, \ldots, a_k$ be pairwise relatively prime, positive integers. Determine the largest integer not expressible in the form

$$x_1 a_2 a_3 \cdots a_k + x_2 a_1 a_3 \cdots a_k + \cdots + x_k a_1 a_2 \cdots a_{k-1},$$

for some nonnegative integers $x_1, \ldots, x_k$.

Solution by Daniele Donini, Bertinoro, Italy.

The answer is $M = (k-1)a_1 a_2 \cdots a_k - (a_2 a_3 \cdots a_k + a_1 a_3 \cdots a_k + \cdots + a_1 a_2 \cdots a_{k-1})$.

Suppose first that $M = x_1 a_2 a_3 \cdots a_k + x_2 a_1 a_3 \cdots a_k + \cdots + x_k a_1 a_2 \cdots a_{k-1}$ for some nonnegative integers $x_1, \ldots, x_k$. Then

$$(k-1)a_1 a_2 \cdots a_k = (x_1 + 1)a_2 a_3 \cdots a_k + (x_2 + 1)a_1 a_3 \cdots a_k + \cdots + (x_k + 1)a_1 a_2 \cdots a_{k-1}.$$

It follows that a_i divides $x_i + 1$ and so $x_i + 1 \geq a_i$. Thus,

$$(k-1)a_1 a_2 \cdots a_k \geq a_1 a_2 a_3 \cdots a_k + a_2 a_1 a_3 \cdots a_k + \cdots + a_k a_1 a_2 \cdots a_{k-1} = k a_1 a_2 \cdots a_k,$$

a contradiction.

We now have to show that every integer $N > M$ can be expressed in the given form. We prove this by induction on k. The initial step is the case $k = 2$. Because a_1 and a_2

are relatively prime, the set $\{0, a_2, 2a_2, \ldots, (a_1-1)a_2\}$ is a complete set of residues modulo a_1. It follows that there exist integers x_1 and x_2 with $0 \le x_1 < a_1$ such that $N = x_1 a_2 + x_2 a_1$. Furthermore,

$$x_2 = \frac{N - x_1 a_2}{a_1} > \frac{M - (a_1 - 1)a_2}{a_1} = -1$$

implies $x_2 \ge 0$.

For the inductive step, suppose the result is true for $k-1$ with $k-1 \ge 2$. By considering two cases—at least two of $a_2, \ldots, a_k$ greater than 1 and at most one of $a_2, \ldots, a_k$ greater than 1—it is easy to check that

$$(k-1) - \sum_{i=1}^{k} 1/a_i \ge 1 - 1/a_1 - 1/(a_2 \cdots a_k),$$

from which it follows immediately upon multiplication by $a_1 a_2 \cdots a_k$ that

$$M \ge a_1 a_2 \cdots a_k - (a_2 a_3 \cdots a_k + a_1).$$

By the proof in the case $k = 2$, we can find integers x_1 and y_1 with $0 \le x_1 < a_1$ such that

$$N = x_1 a_2 a_3 \cdots a_k + y_1 a_1.$$

Then

$$\begin{aligned} y_1 &= \frac{N - x_1 a_2 a_3 \cdots a_k}{a_1} > \frac{M - x_1 a_2 a_3 \cdots a_k}{a_1} \\ &= (k-2) a_2 a_3 \cdots a_k - (a_3 a_4 \cdots a_k + \cdots + a_2 a_3 \cdots a_{k-1}) + \frac{(a_1 - x_1 - 1) a_2 a_3 \cdots a_k}{a_1} \\ &\ge (k-2) a_2 a_3 \cdots a_k - (a_3 a_4 \cdots a_k + \cdots + a_2 a_3 \cdots a_{k-1}). \end{aligned}$$

Thus, we may apply the case $k-1$ to write

$$y_1 = x_2 a_3 a_4 \cdots a_k + \cdots + x_k a_2 a_3 \cdots a_{k-1}$$

for nonnegative integers $x_2, \ldots, x_k$. Substitution yields

$$N = x_1 a_2 a_3 \cdots a_k + x_2 a_1 a_3 \cdots a_k + \cdots + x_k a_1 a_2 \cdots a_{k-1}.$$

Also solved by J. C. Binz (Switzerland), Jean Bogaert (Belgium), Con Amore Problem Group (Denmark), Kathleen E. Lewis, Brian W. McEnnis, Paul J. Zwier, and the proposer. There were two incorrect solutions.

A Tangent and Cosine Identity **December 1998**

1562. *Proposed by John Wickner, student, University of St. Thomas, St. Paul, Minnesota, and Scott Beslin and Valerio De Angelis, Nicholls State University, Thibodaux, Louisiana.*

Prove that

$$\tan\left(\frac{1}{4}\tan^{-1}4\right) = 2\left(\cos\frac{6\pi}{17} + \cos\frac{10\pi}{17}\right).$$

(The equality symbol was inadvertently omitted from the original proposal.)

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

Two applications of the half-angle formula

$$\tan\left(\frac{1}{2}\theta\right) = \frac{\sqrt{1+\tan^2\theta}-1}{\tan\theta}, \quad 0<\theta<\pi/2,$$

and some algebra show that

$$\tan\left(\frac{1}{4}\tan^{-1}4\right) = \frac{\sqrt{34+2\sqrt{17}}-\sqrt{17}-1}{4}.$$

Now let $\omega = e^{2\pi i/17}$. Then

$$\omega^{17}=1 \quad \text{and} \quad \omega^{16}+\omega^{15}+\cdots+\omega+1=0.$$

Let $a_1 = \omega+\omega^2+\omega^4+\omega^8+\omega^9+\omega^{13}+\omega^{15}+\omega^{16}$ and $a_2 = \omega^3+\omega^5+\omega^6+\omega^7+\omega^{10}+\omega^{11}+\omega^{12}+\omega^{14}$. Then a_1 and a_2 are real numbers whose sum is -1 and whose product is -4. Note that $1, \omega, \omega^2, \ldots, \omega^{16}$ are vertices of a regular heptadecagon, from which it is easily seen that $a_2<0$ (and consequently $a_1>0$). It follows that $a_2 = (-1-\sqrt{17})/2$. Now let $b_1 = \omega^3+\omega^5+\omega^{12}+\omega^{14}$ and $b_2 = \omega^6+\omega^7+\omega^{10}+\omega^{11}$. Then b_1 and b_2 are real numbers whose sum is a_2 and whose product is -1. Because $b_2<0$, we find that

$$b_1 = \frac{\sqrt{34+2\sqrt{17}}-\sqrt{17}-1}{4}.$$

The identity follows from

$$b_1 = 2\,\mathrm{Re}(\omega^3+\omega^5) = 2\left(\cos\frac{6\pi}{17}+\cos\frac{10\pi}{17}\right).$$

Also solved by Tewodros Amdeberhan, Michel Bataille (France), Rich Bauer, Brian D. Beasley, Nirdosh Bhatnagar, J. C. Binz (Switzerland), Jean Bogaert (Belgium), John Christopher, Charles R. Diminnie, Matt Foss, Hans Kappus (Switzerland), William A. Newcomb, Richard Parris, Michael J. Semenoff, Albert Stadler (Switzerland), Michael Woltermann, Paul J. Zwier, and the proposers.

Closed Partitions of Fields **December 1998**

1563. *Proposed by Wu Wei Chao, Guang Zhou Normal University, Guang Zhou City, Guang Dong Province, China.*

For a given field F, classify all possible partitions of F into finitely many equivalence classes such that each class is closed under addition and multiplication by *distinct* elements in the class.

Solution by the editors.

The possible partitions are (i) a single class; (ii) F finite, one class a subfield of F or $\{0, 1, -1\}$ or $\{0, x\}$ for some $x \in F$, and the rest singletons; (iii) F of characteristic 2 with classes $F - \{0\}$ and $\{0\}$; and (iv) F finite of characteristic 2 with one class the non-zero elements of a subfield of F, perhaps one of the form $\{0, x\}$ for some $x \in F$, and the rest singletons.

All of the above are easily shown to satisfy the conditions of the problem. To show no other possibilities exist, let $\bar{x}$ denote the class of x.

Suppose first that F has characteristic 0. Fix $n \in \{2, 3, \ldots\}$. Then an infinite number of $x, x/n, x/n^2, \ldots$ must be in the same class. If x/n^j and x/n^k, $j<k$, are

in the same class, then by adding x/n^k a total of $(m-1)n^{k-j}$ times to x/n^j, we see that $mx/n^j \in \overline{x/n^j}$. This class must contain both x and arbitrary mx/n, so that $\mathbb{Q}_{>0}x \subset \bar{x}$. From $(-x)(-2x)=x(2x)$, we see that $\mathbb{Q}^{\times}x \subset \bar{x}$. Additive closure implies $0 = x+(-x) \in \bar{x}$, so that the partition has a single class.

Now suppose that F has odd characteristic p. Suppose $\bar{x}$ is not $\{0,1,-1\}$, $\{0,x\}$, or a singleton. Now, excluding these cases, $\bar{x} \subset \{0,x,-x\}$ implies $-x^2 = 0$, x, or $-x$, a contradiction. Thus, there exists $y \in \bar{x} - \{0,x,-x\}$. If we keep adding y beginning with x, we see that either $x-y = x+(p-1)y \in \bar{x}$ or $x+ky=y$ for some $k \in \{3,4,\ldots,p-2\}$. (We rule out $k=0,1,2$ by our assumption on y.) Because $x+y \in \bar{x}$, in the former case $2x=(x+y)+(x-y) \in \bar{x}$. In the latter case, $x+(k-2)y = -y$, so that $-y \in \bar{x}$. But again $x-y \in \bar{x}$ and $2x \in \bar{x}$. We conclude that $\bar{x}$ is closed under addition. Additive closure implies $0 \in \bar{x}$ so that such a class $\bar{x}$ is unique. For this class, $-x^2 = x(-x) \in \bar{x}$, hence $x^2 \in \bar{x}$, and $\bar{x}$ is closed under multiplication. Finally, because $F - \bar{x}$ is finite, so are $\bar{x}$ and F. Thus, we have case (i) or case (ii).

Finally, suppose that F has characteristic 2. Observe that $\bar{x} \cup \{0\}$ is closed under addition. We show that $x^2 \in \bar{x}$, implying that $\bar{x}$ is closed under multiplication, unless $\bar{x}$ is $\{x\}$ or $\{0,x\}$. We may assume $x \neq 0,1$. Choose $y \in \bar{x} - \{0,x\}$. Then $x+y$, xy, and $x(x+y) = x^2+xy$ are all in $\bar{x}$. Therefore, $x^2 = (x^2+xy)+xy \in \bar{x}$. We next show that $1 \in \bar{x}$ for such a special class, implying that such a class is unique. This is clear if x has finite multiplicative order. If x has infinite order, then some pair of distinct elements $1+x^j$ and $1+x^k$ are in the same class, as is their sum. Because $(1+x^j)+(1+x^k) = x^j+x^k \in \bar{x}$, this class is $\bar{x}$. Therefore, $1 = (1+x^j)+x^j \in \bar{x}$. As in the odd characteristic case, F must be finite. We conclude that for this special class $\bar{x} \cup \{0\}$ must be a subfield of F. Because the complement of this subfield is finite, we are left with cases (i), (ii), (iii), or (iv).

Answers

Solutions to the Quickies on page 409

A895. If m and n are relatively prime then so are their sum and product, $m+n$ and mn. Iterating this two more times yields the desired conclusion.

A896. Using the weighted arithmetic mean-geometric mean inequality we have

$$\frac{a+b+c}{ab+bc+ca} = \frac{1/b}{1/a+1/b+1/c}\cdot\frac{1}{a} + \frac{1/c}{1/a+1/b+1/c}\cdot\frac{1}{b} + \frac{1/a}{1/a+1/b+1/c}\cdot\frac{1}{c}$$
$$\geq \left(\frac{1}{a}\right)^{\frac{1/b}{1/a+1/b+1/c}}\left(\frac{1}{b}\right)^{\frac{1/c}{1/a+1/b+1/c}}\left(\frac{1}{c}\right)^{\frac{1/a}{1/a+1/b+1/c}}.$$

Reciprocation yields

$$\left(a^{1/b}b^{1/c}c^{1/a}\right)^{\frac{1}{1/a+1/b+1/c}} \geq \frac{ab+bc+ca}{a+b+c}.$$

A second application of the weighted arithmetic mean-geometric mean inequality leads to

$$\frac{ab+bc+ca}{a+b+c} = \frac{b}{a+b+c}a + \frac{c}{a+b+c}b + \frac{a}{a+b+c}c$$
$$\geq a^{b/(a+b+c)}b^{c/(a+b+c)}c^{a/(a+b+c)} = \left(a^b b^c c^a\right)^{1/(a+b+c)}.$$

The claim follows from combining these last two inequalities.

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.

McKay, Brendan, Dror Bar-Natan, Maya Bar-Hillel, and Gil Kalai, Solving the Bible Code puzzle, *Statistical Science* 14 (2) (1999) 149–173.

Utterly demolishes claims by Witztum et al. (Equidistant letter sequences in the Book of Genesis, *Statistical Science* 9 (1994) 429–438) and by Michael Drosnin (*The Bible Code*, Simon and Schuster, 1997) that there are equidistant letter codes in the Hebrew Book of Genesis. The search specifications were inadequately specific, so that choices by Witztum et al. "'tuned' their method to their data, thus invalidating their statistical test."

Finch, Steven, Unsolved Mathematics Problems, `http:www.mathsoft.com/asolve` .
Eppstein, David, The Geometry Junkyard: Open Problems, `http://www.ics.uci.edu/~eppstein/junkyard/open.html` .

These Web sites are rich collections of essays on scores of unsolved mathematical problems, together with links to many sites about other problems. Some examples are tiling problems of various kinds; finding the number of self-avoiding rook walks on a chessboard; finding your way out of a forest; optimizing moat-crossing; and the existence of a triangle with sides, medians, altitudes, and area all rational—plus links to pages on more named mathematical constants than you can imagine. The site managers welcome your contributions.

Joyce, David, Euclid's *Elements* Online,
`http://aleph0.claru.edu/~djoyce/java/elements/elements.html` .

This site contains a text of Euclid's *Elements* together with historical and mathematical comments (many from Heath's famous edition). The diagrams included are Java applets in which the reader can vary parameters of the situation and observe how elements of the diagram respond.

Strang, Gilbert, The discrete cosine transform, *SIAM Review* 41 (1) (1999) 135–147. Berry, Michael W., Zlatko Drmač, and Elizabeth R. Jessup, Matrices, vector spaces, and information retrieval, *SIAM Review* 41 (2) (1999) 335–362.

SIAM Review has a new look and a new organizational structure. The new Education section provides modules about tools in applied mathematics, written "more in the style of a textbook section rather than a research article." Strang's module on the DCT begins at a fast clip but is comfortably informal, is packed with perspective, and has the reader wishing at many points to hear even more. Unexpectedly, the motivation comes not at the beginning but at the end, where Strang explains how the JPEG compression algorithm for images uses DCT-2. Berry et al. offer a more leisurely tour of automated information-retrieval strategies based on orthogonal factorizations of matrices (QR decomposition, singular value decomposition), though these methods have not yet been tested in search engines.

Gerdes, Paulus, *Geometry from Africa: Mathematical and Educational Explorations*, MAA, 1999; xv + 210 pp, (P). ISBN 0–88385–715–4. Zaslavsky, Claudia, *Africa Counts: Number and Pattern in African Cultures*, 3rd ed., Lawrence Hill, 1999; xv + 352 pp, $16.95 (P). ISBN 1–55652–350–5. Eglash, Ron, *African Fractals: Modern Computing and Indigenous Design*, Rutgers University Press, 1999; xi + 258 pp, $60, $25 (P). ISBN 0–8135–2614–0, 0–8135–2613–2.

All three of these books exhibit geometrical ideas in African cultures, as manifested in carvings, paintings, weavings, baskets, mats, and decorations of pots. Gerdes also considers how geometrical patterns lead to the Pythagorean theorem, how to use crafts (particularly weaving) to explore mathematical ideas, and the Chokwe tradition of drawing unicursal figures in the sand. Zaslavsky includes number words and gestures plus games and magic squares. She give updates on pp. 290ff and notes text revisions on pp. 295–297; in particular, the chapter bibliographies have been updated. The book by Eglash was reviewed in this MAGAZINE in June 1999.

Cromwell, Peter R., *Polyhedra*, Cambridge University Press, 1997; xii + 451 pp. ISBN 0–521–55432–2. Simon, Lewis, Bennett Arnstein, and Rona Gurkewitz, *Modular Origami Polyhedra*, rev. and enl. ed., Dover, 1999; iv + 59 pp, $5.95 (P). ISBN 0–486–40476–5.

As Cromwell notes, there are three kinds of books that treat polyhedra: recreational mathematics books with a chapter on basic properties and perhaps Euler's formula, books on polytopes in arbitrary dimensions, and guides to model-making. In the third category, the book by Simon et al. shows how to make polyhedra from several pieces of paper folded the same way. Cromwell's own book is a beautifully-illustrated chronological survey about polyhedra up to about 1900, with a few topics omitted (Steinitz's theorem, Alexandrov's theorem, duality). Included are regular and semiregular polyhedra, star polyhedra, Euler's formula, rigidity, symmetry types, coloring, transitivity, and much more.

Stigler, Stephen M., *Statistics on the Table: The History of Statistical Concepts and Methods*, Harvard University Press, 1999; ix + 488 pp. ISBN 0–674–83601–4.

This collection of essays differs from the author's *The History of Statistics: The Measurement of Uncertainty before 1900* (Harvard University Press, 1986) in being topic-oriented rather than chronological. All the essays apparently appeared over the past 25 years in similar form. Stigler considers the development of social and behavioral statistics, Galtonian ideas, seventeenth-century explorers, who discovered what, and the relation between statistics and standards, including origins of the name "normal" for the familiar continuous distribution. (But the index will not help you find easily where the name "Gaussian" is discussed in Ch. 14.) If you teach statistics, you will greatly enjoy—and learn from—this book.

Hallinan, Peter L., Gaile G. Gordon, A.L. Yuille, Peter Giblin, and David Mumford, *Two- and Three-Dimensional Patterns of the Face*, A K Peters, 1999; viii + 262 pp. ISBN 1–56881–087–3.

This book, filled with many figures and photographs, surveys approaches to face recognition. The mathematics involved includes Bayesian probability, linear filters, singular value decompositions, vector calculus, and differential geometry.

Turner, Paul E., and Lin Chao, Prisoner's dilemma in an RNA virus, *Nature* 398 (1 April 1999) 441–443.

The authors find Prisoner's Dilemma interactions in the evolution of fitness of viruses coinfecting the same cell.

NEWS AND LETTERS

Acknowledgments

Along with our associate editors, the following referees have assisted the MAGAZINE *during the past year. We thank them for their time and care.*

Aczél, János, *Waterloo, Canada*

Andrews, Peter, *Eastern Illinois University, Charleston, IL*

Apostol, Tom, *California Institute of Technology, Pasadena, CA*

Askey, Richard, *University of Wisconsin, Madison, WI*

Axler, Sheldon, *San Francisco State University, San Francisco, CA*

Bailey, Herb, *Rose-Hulman Institute of Technology, Terre Haute, IN*

Barbeau, Edward, *University of Toronto, Toronto, Canada*

Barksdale, Jr., James, *Western Kentucky University, Bowling Green, KY*

Barr, Thomas, *Rhodes College, Memphis, TN*

Baxley, John, *Wake Forest University, Winston-Salem, NC*

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

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

Beineke, Lowell, *Indiana University-Purdue University, Fort Wayne, IN*

Berndt, Bruce, *University of Illinois, Urbana, IL*

Bivens, Irl, *Davidson College, Davidson, NC*

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

Bridger, Mark, *Newton Upper Falls, MA*

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

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

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

Colley, Susan, *Oberlin College, Oberlin, OH*

Cowen, Carl, *Purdue University, West Lafayette, IN*

Crannell, Annalisa, *Franklin and Marshall College, Lancaster, PA*

Culp-Ressler, Wendell, *Franklin and Marshall College, Lancaster, PA*

Deeba, Elias, *University of Houston, Downtown, Houston, TX*

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

Dobbs, David, *University of Tennessee, Knoxville, TN*

Duncan, John, *University of Arkansas, Fayetteville, AR*

Emert, John, *Ball State University, Muncie, IN*

Feil, Todd, *Denison University, Granville, OH*

Feroe, John, *Vassar College, Poughkeepsie, NY*

Firey, William, *Corvallis, OR*

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

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

Flanders, Harley, *Jacksonville Beach, FL*

Foster, Lorraine, *California State University, Northridge, CA*

Fraser, Craig, *University of Toronto, Victoria College, Toronto, Canada*

Fredricks, Gregory, *Lewis and Clark College, Portland, OR*

Gallian, Joseph, *University of Minnesota, Duluth, MN*

Giblin, Peter, *University of Liverpool, Liverpool, England*

Gillman, Leonard, *Austin, TX*

Gloor, Philip, *St. Olaf College, Northfield, MN*

Goldfeather, Jerome, *University of Memphis, Memphis, TN*

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

Grunbaum, Branko, *University of Washington, Seattle, WA*

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

Gulick, Denny, *University of Maryland, College Park, MD*

Guy, Richard, *University of Calgary, Calgary, Canada*

Hahn, Liang–Shin, *Irvine, CA*

Haunsperger, Deanna, *Carleton College, Northfield, MN*

Heuer, Gerald, *Concordia College, Moorhead, MN*

Hirschhorn, Mike, *University of New South Wales, Sydney, New South Wales, Australia*

Hodges, Laurent, *Iowa State University, Ames, IA*

Honsberger, Ross, *University of Waterloo, Waterloo, Canada*

Hungerford, Thomas, *Cleveland State University, Cleveland, OH*

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

Johnsonbaugh, Richard, *DePaul University, Chicago, IL*

Johnston, Elgin, *Iowa State University, Ames, IA*

Jordan, Jim, *Washington State University, Pullman, WA*

Kendig, Keith, *Cleveland State University, Cleveland, OH*

Kennedy, Steven, *Carleton College, Northfield, MN*

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

King, James, *University of Washington, Seattle, WA*

Klamkin, Murray, *University of Alberta, Edmonton, Canada*

Krause, Eugene, *University of Michigan, Ann Arbor, MI*

Kung, Sidney, *Jacksonville University, Jacksonville, FL*

Kuzmanovich, James, *Wake Forest University, Winston–Salem, NC*

Laatsch, Richard, *Miami University, Oxford, OH*

Larson, Dean, *Gonzaga University, Spokane, WA*

Larson, Loren, *St. Olaf College, Northfield, MN*

Liu, Andrew, *University of Alberta, Edmonton, Canada*

LoBello, Anthony, *Allegheny College, Meadville, PA*

Mallinson, Philip, *Phillips Exeter Academy, Exeter, NH*

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

Martin, Greg, *University of Toronto, Toronto, Canada*

McCartney, Philip, *Northern Kentucky University, Highland Heights, KY*

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

McDermot, Richard, *Meadville, PA*

Megginson, Robert, *University of Michigan, Ann Arbor, MI*

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

Morgan, Frank, *Williams College, Williamstown, MA*

Needham, Tristan, *University of San Francisco, San Francisco, CA*

Nelsen, Roger, *Lewis and Clark College, Portland, OR*

Nievergelt, Yves, *Eastern Washington University, Cheney, WA*

Nunemacher, Jeffrey, *Ohio Wesleyan University, Delaware, OH*

Osofsky, Barbara, *Rutgers University, New Brunswick, NJ*

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

Pinkham, Roger, *Stevens Institute of Technology, Hoboken, NJ*

Pinsky, Mark, *Northwestern University, Evanston, IL*

Pomerance, Carl, *University of Georgia, Athens, GA*

Post, Karel, *Eindhoven, The Netherlands*

Riddle, Larry, *Agnes Scott College, Decatur, GA*

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

Roth, Richard, *University of Colorado, Boulder, CO*

Sandefur, James, *Georgetown University, Washington, DC*

Saunders, Sam, *Kirkland, WA*

Saxe, Karen, *Macalester College, St. Paul, MN*

Schaumberger, Norman, *Hofstra University, Hempstead, NY*

Schwenk, Allen, *Western Michigan University, Kalamazoo, MI*

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

Stone, Alexander, *University of New Mexico, Albuquerque, NM*

Stone, Arthur, *Northeastern University, Boston, MA*

Straffin, Jr., Philip, *Beloit College, Beloit, WI*

Strogatz, Steven, *Cornell University, Ithaca, NY*

Stroyan, Keith, *University of Iowa, Iowa City, IA*

Wagon, Stan, *Macalester College, St. Paul, MN*

Ward, Lesley Ann, *Harvey Mudd College, Claremont, CA*

Watkins, John, *Colorado College, Colorado Springs, CO*

Wilf, Herbert, *University of Pennsylvania, Philadelphia, PA*

Williams, Kenneth, *Carleton University, Ottawa, Canada*

Index to Volume 72

AUTHORS

New Editor of *Mathematics Magazine*

Starting January 1, 2000, please submit new manuscripts to:

Frank Farris
Department of Mathematics and Computer Science
Santa Clara University
500 El Camino Real
Santa Clara, CA 95053-0290

Please read the editorial guidelines posted at `www.maa.org/pubs/mathmag.html`. In addition, we offer the following ideas for potential authors:

- Initial submission continues to be in a physical rather than electronic form. Should your article be accepted, we will ask you to provide a LaTeX file using one of the templates provided at our website. If this is impossible for you, a text file or common word-processor document is acceptable.
- Remember that a good expository article begins with an introduction that grabs the reader's attention and encourages him or her to keep reading.
- If you wish to provide any electronic complement to your article, including such things as color illustrations, Java applets, or animations, supply the URL of your draft site. If your article is accepted, complements will be hosted at `www.maa.org`.
- In the interest of respecting the time of our referees, we recommend a referee's appendix, not for publication, but to guide the referee. Please expand on statements such as, "A simple calculation shows ... ." It is often appropriate to suppress such things in exposition, but a referee might find the additional information a time-saver.
- We strongly recommend that you search the electronic database of *Mathematics Magazine* and the *College Mathematics Journal* for articles on subjects related to yours. Follow the link to this site from the address above. This should help to fill out your bibliography and avoid any duplication.

UNITED STATES POSTAL SERVICE™

Statement of Ownership, Management, and Circulation
(Required by 39 USC 3685)

1. Publication Title	2. Publication Number	3. Filing Date
Mathematics Magazine	0 0 2 5 _ 5 7 0 x	9/14/99
4. Issue Frequency	**5. Number of Issues Published Annually**	**6. Annual Subscription Price**
Bimothly, except July and August	5	$20

7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*
MAA, 1529 18th St. Washington, D.C. 20036
Contact Person: H. Waldman
Telephone: 202-387-5200

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

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*
Publisher *(Name and complete mailing address)*
MAA, 1529 18th St. Washington, D. C. 20036-1385

Editor *(Name and complete mailing address)*
Paul Zorn, St. Olaf College, Northfield, MN 55057

Managing Editor *(Name and complete mailing address)*
Harry Waldman

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

Full Name	Complete Mailing Address
The Mathematical Association of America	1529 18th St. Washington, D.C. 20036

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

Full Name	Complete Mailing Address

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

PS Form 3526, September 1995 *(See Instructions on Reverse)*

13. Publication Title	14. Issue Date for Circulation Data Below	
Mathematics Magazine	June 1999	
15. Extent and Nature of Circulation	**Average No. Copies Each Issue During Preceding 12 Months**	**Actual No. Copies of Single Issue Published Nearest to Filing Date**
a. Total Number of Copies *(Net press run)*	15,000	15,000
b. Paid and/or Requested Circulation (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales *(Not mailed)*	0	0
b. Paid and/or Requested Circulation (2) Paid or Requested Mail Subscriptions *(Include advertiser's proof copies and exchange copies)*	13,500	13,500
c. Total Paid and/or Requested Circulation *(Sum of 15b(1) and 15b(2))*	13,500	13,500
d. Free Distribution by Mail *(Samples, complimentary, and other free)*	1,50 0	1,500
e. Free Distribution Outside the Mail *(Carriers or other means)*	0	0
f. Total Free Distribution *(Sum of 15d and 15e)*	1,500	1,500
g. Total Distribution *(Sum of 15c and 15f)*	15,000	15,000
h. Copies not Distributed (1) Office Use, Leftovers, Spoiled	0	0
h. Copies not Distributed (2) Returns from News Agents	0	0
i. Total *(Sum of 15g, 15h(1), and 15h(2))*	15,000	15,000
Percent Paid and/or Requested Circulation *(15c / 15g x 100)*	90%	90%

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

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Donald J. Albers, Director of Publications
Date: 9/14/99

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

Instructions to Publishers

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

Failure to file or publish a statement of ownership may lead to suspension of second-class authorization.

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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