Vol. 81, No. 1, February 2008

MATHEMATICS MAGAZINE

$$\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2+\cdots}}}}} = \sqrt{3+\sqrt{3-\sqrt{3+\sqrt{3-\sqrt{3+\cdots}}}}}$$

$$= \sqrt{7-\sqrt{7+\sqrt{7-\sqrt{7+\sqrt{7-\cdots}}}}}$$

$$= \sqrt{12-2\sqrt{12+2\sqrt{12-2\sqrt{12+\cdots}}}}$$

$$= \sqrt{19-3\sqrt{19+3\sqrt{19-3\sqrt{19+\cdots}}}}$$

$$= ???$$

Infinitely Nested Radicals

- Synthetic Partial Fraction Decompositions
- A Brief History of Impossibility
- Polynomial Root Squeezing
- Paint It Black—A Combinatorial Yawp

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

Submit new manuscripts to Allen Schwenk, Editor, Mathematics Magazine, Department of Mathematics, Western Michigan University, Kalamazoo, MI, 49008. Manuscripts should be laser printed, with wide line spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should mail three copies and keep one copy. In addition, authors should supply the full five-symbol 2000 Mathematics Subject Classification number, as described in *Mathematical Reviews*.

What numbers can result from infinitely nested radicals like the ones displayed on our cover? Can all these varied examples really give the same value? The lead article by Zimmerman and Ho answers these questions, and lots more.

AUTHORS

Seth Zimmerman received his mathematical training at Dartmouth and Princeton. After two decades abroad he returned to teach in the Bay Area, where by good fortune his office at Evergreen Valley College was adjacent to Chungwu Ho's, leading to years of fruitful collaboration. He currently pursues his research in probability and entropy theory at his home by the shores of Puget Sound. His concurrent work includes a popular version of The Inferno of Dante Alighieri, and translations of the poems of Osip Mandelstam. He and his wife are dedicated practitioners of Qi Gong, and he enjoys playing cello and singing in the local chorale in Bellingham, Washington.

Chungwu Ho was born in China, moved to Taiwan in 1949, and came to the US in 1960. He received his training at University of Washington and MIT, where he worked under the guidance of James R. Munkres. He taught for 30 years at Southern Illinois University at Edwardsville, including several years as chair of the math and statistics department. Since retiring there in 2000, he has been teaching at Evergreen Valley College. Besides teaching, he writes poetry and plays the Chinese musical instrument, Chinghu, to accompany his wife in Chinese opera singing.

William A. Adkins received his PhD from the University of Oregon. Since 1980 he has been at Louisiana State University, where he is currently a professor of mathematics. He has recently completed a term as Notes Editor of the *American Mathematical Monthly*.

Mark Davidson is currently Professor of Mathematics at Louisiana State University. He received his PhD from the University of California, Irvine in 1984. His research interest is harmonic analysis. The subject of partial fraction decompositions came up while preparing a book with William Adkins on ordinary differential equations wherein the Laplace transform is a principle tool.

Jeff Suzuki became a mathematician because his results in laboratory science frequently contradicted the known laws of physics and chemistry. In graduate school he combined his interests in mathematics, history, and physics in a dissertation on the history of the stability problem in celestial mechanics, and has subsequently focused on the mathematics of the 17th and 18th centuries because, in his words, it was "the last time it was possible to know everything." Current research interests include the history of college mathematics, geometric solutions to cubic and quartic equations, and learning to play the guitar.

Matt Boelkins earned his BS from Geneva College, MS from Western Washington University, and his doctorate from Syracuse University, all in mathematics. He has served on the faculty of Grand Valley State University since 1998 and holds the rank of associate professor. A regular mentor of undergraduate research projects on open problems involving polynomial functions, this is his fourth paper coauthored with students. Matt's favorite way to spend a winter day off is an early morning hockey game, followed by Muskegon River steelhead fishing, topped off by a cozy evening at home with his family reading books by the fire.

Justin From attended Central College in Pella, IA, where he double majored in mathematics and philosophy and played on the varsity football team. In the summer of 2005, he participated in the Grand Valley State University REU program, during which the results in this paper were proved. Upon graduating from Central in 2006, he joined Teach for America in Houston, TX. Justin has plans to attend law school.

Samuel Kolins is a second year graduate student at Cornell University. He received his bachelor's degree from Bowdoin College as a double major in mathematics and physics. This article was part of the work that he did at the Grand Valley State University Research Experience for Undergraduates program during the summer of 2005. In his spare time Samuel enjoys biking and hiking in the Ithaca area.

Vol. 81, No. 1, February 2008

MATHEMATICS MAGAZINE

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for *MATHEMATICS MAGAZINE* to an individual member of the Association is $131. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 20% dues discount for the first two years of membership.)

Subscription correspondence and notice of change of address should be sent to the Membership/ Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to

MAA Advertising
1529 Eighteenth St. NW
Washington DC 20036

Phone: (866) 821-1221
Fax: (202) 387-1208
E-mail: advertising@maa.org

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

On Infinitely Nested Radicals

SETH ZIMMERMAN
Evergreen Valley College
zim1s1@msn.com

CHUNGWU HO
Evergreen Valley College and
Southern Illinois University at Edwardsville*
HoC@alum.mit.edu

Are these really real numbers?

Many students are intrigued by the unusual equation:

$$\sqrt{2+\sqrt{2+\sqrt{2+\cdots}}}=2.$$

Intermediate algebra students may have been shown that the equation

$$\sqrt{2+\sqrt{2+\sqrt{2+\cdots}}}=x,$$

when squared on both sides, leads to the "solution" $x = 2$. While they are likely to see this as no more than a clever trick, the suggestion that there are strange numbers awaiting exploration has been seeded. Further on, trigonometry students may know that there are specific equivalents for some finite "nests" of radicals:

$$\sqrt{2+\sqrt{2}}=2\cos\left(\frac{\pi}{8}\right),$$

$$\sqrt{2+\sqrt{2+\sqrt{2}}}=2\cos\left(\frac{\pi}{16}\right),$$

$$\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}}=2\cos\left(\frac{\pi}{32}\right),$$

$$\vdots$$

and

$$\sqrt{2-\sqrt{2}}=2\sin\left(\frac{\pi}{8}\right),$$

* *Current address:* Evergreen Valley College, San Jose, CA 95135

$$\sqrt{2-\sqrt{2+\sqrt{2}}}=2\sin\left(\frac{\pi}{16}\right),$$

$$\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2}}}}=2\sin\left(\frac{\pi}{32}\right),$$

$$\vdots$$

Thus, they may accept that the limits of both sequences exist, and are 2 and 0, respectively. What happens if the number 2 is replaced by some other positive number a. Will the limit always exist? It is well-known that the Golden Ratio, ϕ, can be written as such a limit

$$\phi=\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}}.$$

But is it possible to write any arbitrary integer, rational number, or indeed π or e as the limit of some sequence of nested radicals? And if an integer k is such a limit, how many different sequences of radicals will converge to k? Although there seems to be some revived interest in this topic [**4**], [**5**], previous research has not considered these questions [**1**]–[**8**]. In this paper we will make a systematic study of nested radicals, answering many such questions and suggesting further lines of research for the interested reader.

The radicals $\sqrt{a+\sqrt{a+\sqrt{a+\cdots}}}$

Let us start with the familiar expression $\sqrt{2+\sqrt{2+\sqrt{2+\cdots}}}$. In order to specify what is meant by an infinite sequence of radicals, we introduce a more precise definition: let $x_1=\sqrt{2}$, and $x_{n+1}=\sqrt{2+x_n}$ for each $n\geq 1$. This recursive definition clearly gives rise to the sequence $\sqrt{2}$, $\sqrt{2+\sqrt{2}}$, $\sqrt{2+\sqrt{2+\sqrt{2}}}$, ... We may then consider the process of taking infinitely many radicals as the limit of this sequence, provided that it exists. Using the Half Angle Formula and Mathematical Induction, it is not difficult to see that for each $n\geq 1$, $x_n=2\cos(\frac{\pi}{2^{n+1}})$. Thus, the sequence $\{x_n\}$ is a bounded, monotonically increasing sequence, and hence, $\lim_{n\to\infty}x_n$ exists. In fact, $\lim_{n\to\infty}x_n=\lim_{n\to\infty}2\cos(\frac{\pi}{2^{n+1}})=2$. We now consider the more general nested radical sequence

$$\sqrt{a},\quad \sqrt{a+\sqrt{a}},\quad \sqrt{a+\sqrt{a+\sqrt{a}}},\quad \ldots.$$

Since $a\leq 0$ will give rise either to the trivial sequence of zeroes or to sequences involving imaginary numbers, we will restrict our attention to the case where $a>0$. For any real number $a>0$, define $r_1(a)=\sqrt{a}$, and for each $n\geq 1$, define $r_{n+1}(a)=\sqrt{a+r_n(a)}$. When the value of a is clear from the context, we will simply write r_n for $r_n(a)$.

LEMMA 1. *The sequence $\{r_n\}$ is always a bounded, monotonically increasing sequence. Thus, the limit $\lim_{n\to\infty}r_n$ exists.*

Proof. Since $a>0$, $r_1=\sqrt{a}<\sqrt{a+\sqrt{a}}=r_2$. Now, if $r_n<r_{n+1}$, then $r_{n+1}=\sqrt{a+r_n}<\sqrt{a+r_{n+1}}=r_{n+2}$. Thus, $\{r_n\}$ is an increasing sequence by induction.

To show that the sequence $\{r_n\}$ is bounded, first consider the case where $a \geq 2$. In this case, $0 < r_1 = \sqrt{a} \leq a$. Now, if $0 < r_n \leq a$ for any positive integer n, it follows that $0 < r_{n+1} = \sqrt{a + r_n} \leq \sqrt{2a} \leq \sqrt{a^2} = a$. Thus, by induction, $0 < r_n \leq a$ for each n.

Next consider the case where $0 < a < 2$. We will show that $0 < r_n \leq 2$ for each n. First note that $r_1 = \sqrt{a} < \sqrt{2} < 2$. If $0 < r_n \leq 2$ for any positive integer n, it follows that $0 < r_{n+1} = \sqrt{a + r_n} \leq \sqrt{a+2} \leq \sqrt{4} = 2$. Thus, by induction, $0 < r_n \leq 2$ for each n. ■

We will let $r(a)$, or simply r when the value a is clear from the context, be the limit $\lim_{n\to\infty} r_n$. Applying the limit on both sides of the defining equation $r_{n+1} = \sqrt{a + r_n}$ and clearing the radical will lead to the quadratic equation $r^2 - r - a = 0$. Thus $r = \frac{1 \pm \sqrt{1+4a}}{2}$. Since each $r_n > 0$, $r \geq 0$. Also, since $a > 0$, we may conclude that

$$r = \frac{1 + \sqrt{1+4a}}{2}. \tag{1}$$

Since $r(a) = \frac{1+\sqrt{1+4a}}{2}$, the relation $a \mapsto r(a)$ is one-to-one, i.e., a specific limit $r(a) > 0$ can come only from a specific value of $a > 0$. Given that the sequence $\sqrt{a}$, $\sqrt{a + \sqrt{a}}$, $\sqrt{a + \sqrt{a + \sqrt{a}}}, \ldots$ looks quite ugly when a is irrational, we will restrict a to being rational.

What are the possible limits if a is rational? Since $a > 0$,

$$r(a) = \frac{1 + \sqrt{1+4a}}{2} > \frac{1+1}{2} = 1.$$

Thus, only a real number greater than 1 can be a limit. From the equation (1) above, we see that any limit $r(a)$ will have to be a root of the quadratic equation $x^2 - x - a = 0$. And thus a transcendental number such as π or e can *never* be the limit of a sequence of nested radicals $\sqrt{a}$, $\sqrt{a + \sqrt{a}}$, $\sqrt{a + \sqrt{a + \sqrt{a}}}, \ldots$ if a is rational.

THEOREM 1. *For each rational number $h > 1$, $h(h-1)$ is the unique rational number a such that $r(a) = h$.*

Proof. Let a rational number $h > 1$ be given. We simply let a be the rational number $h(h-1)$. Then $r(a) = \frac{1+\sqrt{1+4a}}{2} = \frac{1+\sqrt{1+4h(h-1)}}{2} = \frac{1+\sqrt{4h^2-4h+1}}{2} = \frac{1+2h-1}{2} = h$. The uniqueness of a follows from the observation that the relation $a \mapsto r(a)$ is one-to-one. ■

Note that if h is an integer > 1, $a = h(h-1)$ will also be a positive integer. Thus, Theorem 1 in particular says that every positive integer $k > 1$ is the limit of a sequence of nested radicals $\sqrt{a}$, $\sqrt{a + \sqrt{a}}$, $\sqrt{a + \sqrt{a + \sqrt{a}}}, \ldots$ for a unique positive integer $a = k(k-1)$. For example,

$$3 = \sqrt{6 + \sqrt{6 + \sqrt{6 + \cdots}}} \quad \text{(where } a = 3(3-1) = 6\text{)}$$

$$4 = \sqrt{12 + \sqrt{12 + \sqrt{12 + \cdots}}} \quad \text{(where } a = 4(4-1) = 12\text{)}$$

$$5 = \sqrt{20 + \sqrt{20 + \sqrt{20 + \cdots}}} \quad \text{(when } a = 5(5-1) = 20\text{).}$$

The radicals $\sqrt{a+b\sqrt{a+b\sqrt{a\cdots}}}$

Let's extend our investigation slightly. Consider the sequence

$$\sqrt{a}, \quad \sqrt{a+b\sqrt{a}}, \quad \sqrt{a+b\sqrt{a+b\sqrt{a}}}, \quad \dots$$

for some positive rational numbers a and b. This sequence can be defined recursively by letting $s_1(a,b) = \sqrt{a}$, and for each $n \geq 1$, letting $s_{n+1}(a,b) = \sqrt{a + b \cdot s_n(a,b)}$. Again, when the values of a and b are clear from the context, we will simply use s_n for $s_n(a,b)$.

Note that for a given pair of a and b, if we let $r_n = \frac{1}{b}s_n$, then $r_1 = \frac{1}{b}s_1 = \sqrt{\frac{a}{b^2}}$, and for each $n \geq 1$,

$$r_{n+1} = \frac{1}{b}s_{n+1} = \frac{1}{b}\sqrt{a + bs_n} = \sqrt{\frac{a}{b^2} + \frac{1}{b}s_n} = \sqrt{\frac{a}{b^2} + r_n}.$$

Thus, $\{r_n\}$ is the sequence of nested radicals

$$\sqrt{\frac{a}{b^2}}, \quad \sqrt{\frac{a}{b^2} + \sqrt{\frac{a}{b^2}}}, \quad \sqrt{\frac{a}{b^2} + \sqrt{\frac{a}{b^2} + \sqrt{\frac{a}{b^2}}}}, \quad \dots$$

considered above, or in our notation, $s_n(a,b) = b \cdot r_n(\frac{a}{b^2})$. We may then conclude from the above that for any positive rational numbers a and b, the nested sequence of radicals $s_n(a,b)$ always converges. We will let $s(a,b)$, or simply s, be the limit, $\lim_{n\to\infty} s_n(a,b)$. Since $s_n(a,b) = b \cdot r_n(\frac{a}{b^2})$ for each n, we have

$$s(a,b) = b \cdot r\left(\frac{a}{b^2}\right) = b \cdot \frac{1 + \sqrt{1 + 4\frac{a}{b^2}}}{2} = \frac{b + \sqrt{b^2 + 4a}}{2}.$$

We will again ask what numbers can be the limit of the sequence $\{s_n(a,b)\}$. Since $a > 0$, the limit $s(a,b)$ must be greater than the number b. Applying the limit on both sides of the defining relation $s_{n+1} = \sqrt{a + b \cdot s_n}$, we see that the limit s will have to be a root of the quadratic equation $x^2 - bx - a = 0$. Thus, once again a transcendental number can never be the limit of such a sequence of nested radicals.

Since $s_n(a,b) = b \cdot r_n(\frac{a}{b^2})$ for each n, a number h is the limit for the sequence $\{s_n(a,b)\}$ for some positive rational numbers a and b if and only if $\frac{h}{b}$ is the limit for the sequence $\{r_n(\frac{a}{b^2})\}$. By Theorem 1 above, this is true if and only if $\frac{a}{b^2} = \frac{h}{b}(\frac{h}{b} - 1)$, or $a = h(h-b)$.

The following theorems can all be readily proven from this, and we state them without proof.

THEOREM 2. *For each positive rational number h, there are infinitely many pairs of rational numbers a and b such that $h = \lim_{n\to\infty} s_n(a,b)$.*

The situation is particularly interesting when the numbers involved are all integers.

THEOREM 3. *For each positive integer $k > 1$, there are exactly $k-1$ pairs of integers a and b, where $0 < b < k$, and $a = k(k-b)$, such that $k = \lim_{n\to\infty} s_n(a,b)$.*

For example, if we take $k = 4$, we have

$$\sqrt{12+\sqrt{12+\sqrt{12+\cdots}}} = 4 \quad \text{(when } b = 1\text{)}$$

$$\sqrt{8+2\sqrt{8+2\sqrt{8+\cdots}}} = 4 \quad \text{(when } b = 2\text{)}$$

$$\sqrt{4+3\sqrt{4+3\sqrt{4+\cdots}}} = 4 \quad \text{(when } b = 3\text{).}$$

If b is not restricted to integer values, we have

THEOREM 4. *Let $k > 1$ be an integer. For each divisor q of k and each integer p, $0 < p < kq$, let $b = \frac{p}{q}$ and $a = k(k - b)$. Then $k = \lim_{n\to\infty} s_n(a, b)$. In fact, these are the only possible choices for a and b, when a is a positive integer and b a positive rational number.*

For example, for $k = 4$,

$$\sqrt{15+\frac{1}{4}\sqrt{15+\frac{1}{4}\sqrt{15+\cdots}}} = 4 \quad \text{(when } q = 4,\ p = 1\text{)}$$

$$\sqrt{14+\frac{1}{2}\sqrt{14+\frac{1}{2}\sqrt{14+\cdots}}} = 4 \quad \text{(when } q = 4,\ p = 2\text{)}$$

$$\sqrt{13+\frac{3}{4}\sqrt{13+\frac{3}{4}\sqrt{13+\cdots}}} = 4 \quad \text{(when } q = 4,\ p = 3\text{)}$$

$$\vdots$$

$$\sqrt{1+\frac{15}{4}\sqrt{1+\frac{15}{4}\sqrt{1+\cdots}}} = 4 \quad \text{(when } q = 4,\ p = 15\text{).}$$

We've now seen that by introducing the factor b we've expanded the uniqueness of the earlier nests enormously. Simply comparing the statements of Theorems 1 and 2 makes this clear.

The radicals $\sqrt{a-b\sqrt{a-b\sqrt{a-\cdots}}}$

Let's now take a further step by considering the nested radicals

$$\sqrt{a},\quad \sqrt{a-\sqrt{a}},\quad \sqrt{a-\sqrt{a-\sqrt{a}}},\quad \ldots$$

and then, more generally, the nested radicals

$$\sqrt{a},\quad \sqrt{a-b\sqrt{a}},\quad \sqrt{a-b\sqrt{a-b\sqrt{a}}},\quad \ldots.$$

Let's begin with the first sequence. We will let $u_1(a) = \sqrt{a}$, and for each $n \geq 1$, let $u_{n+1}(a) = \sqrt{a - u_n(a)}$. To avoid imaginary numbers, we need to require $a - \sqrt{a} \geq 0$, or $a \geq 1$. Since $a = 1$ leads to an alternating sequence of zeroes and ones, we will assume $a > 1$. If this condition is satisfied, then u_1 is a positive real number $\leq \sqrt{a}$. Using the recursive relation $u_{n+1}(a) = \sqrt{a - u_n(a)}$, it is easy to see that if u_n is a positive real number $\leq \sqrt{a}$ so is u_{n+1}. By induction we may show that the sequence of even terms $u_2, u_4, u_6, \ldots$ is increasing, and the sequence of odd terms $u_1, u_3, u_5, \ldots$ is decreasing. Since both sequences are bounded between 0 and $\sqrt{a}$, both of them converge. To see that these two sequences converge to the same limit, we first note that $u_{n+1} + u_n \geq \sqrt{a}$. This can be seen as follows. Since $u_{n+1} + u_n = \sqrt{a - u_n} + u_n$ and each u_n lies in the interval $[0, \sqrt{a}\,]$, we may consider the function $f(x) = \sqrt{a - x} + x$ on this interval. Using $f'(x)$, we may show that f is increasing on the interval $[0, a - \frac{1}{4}]$, reaching its maximum at $a - \frac{1}{4}$, and then decreases from this point on. Now, restrict our attention to $f(x)$ for x in the interval $[0, \sqrt{a}\,]$. Regardless whether the maximum point $a - \frac{1}{4}$ falls in the interval $[0, \sqrt{a}\,]$ or not, the minimum of f on this interval can only be at one of the end points of this interval. Comparing the values of f at these two end points, we conclude that the minimum value of f on this interval is at the left end point $x = 0$ with $\sqrt{a}$ its minimum value. Hence, $f(x) \geq \sqrt{a}$ for all x in this interval. In particular, $u_{n+1} + u_n \geq \sqrt{a}$. From this, we have

$$\begin{aligned} |u_{n+1} - u_n| &= \frac{|u_{n+1}^2 - u_n^2|}{u_{n+1} + u_n} \\ &= \frac{|a - u_n - a + u_{n-1}|}{u_{n+1} + u_n} \\ &\leq \frac{|u_n - u_{n-1}|}{\sqrt{a}}. \end{aligned}$$

With this and the assumption that $a > 1$, we may conclude that the two subsequences $u_2, u_4, u_6, \ldots$ and $u_1, u_3, u_5, \ldots$ converge to the same limit, and consequently, the sequence $u_n(a)$ is convergent for any real number $a > 1$. Let $u(a)$, or simply u, be the limit. As before, the limit $u(a)$, for a given a, must satisfy the quadratic equation $x^2 + x - a = 0$, and hence, $u(a) = \frac{-1+\sqrt{1+4a}}{2}$. We see that the limit is always a positive number. On the other hand, we will see below in Theorem 5 (with the special case $b = 1$) that any positive number $h > \frac{1}{\phi}$ can be such a limit by letting $a = h(h + 1)$, where ϕ is the Golden Ratio. Now, consider the nested radicals $\sqrt{a}$, $\sqrt{a - b\sqrt{a}}$, $\sqrt{a - b\sqrt{a - b\sqrt{a}}}, \ldots$. Since we have already considered the case $\sqrt{a + b\sqrt{a + b\sqrt{a \cdots}}}$, we will restrict our attention to $b > 0$. Let $v_1(a, b) = \sqrt{a}$, and for each $n \geq 1$, let $v_{n+1}(a, b) = \sqrt{a - b \cdot v_n(a, b)}$. Note that for each n, $v_n(a, b) = b \cdot u_n(\frac{a}{b^2})$. In particular, for the $v_n(a, b)$ to be real numbers, we need to have $\frac{a}{b^2} \geq 1$, or $a \geq b^2$. In the following theorem, we will show that any positive number h, under minor restrictions, can be the limit of a sequence $v_n(a, b)$ for some positive numbers a and b. Surprisingly, we will find that the Golden Ratio $\phi = \frac{1+\sqrt{5}}{2}$ is involved.

THEOREM 5. *Let h, a, and b be positive numbers. Then $h = \lim_{n\to\infty} v_n(a, b)$ if and only if*

(1) $0 < b < \phi \cdot h$, *and*

(2) $a = h(h + b)$.

Proof. Since for each n, $v_n(a, b) = b \cdot u_n(\frac{a}{b^2})$, the limit $v = \lim_{n\to\infty} v_n(a, b)$ exists if and only if $u = \lim_{n\to\infty} u_n(\frac{a}{b^2})$ exists and $v = b \cdot u$. But the value of $u(\frac{a}{b^2})$

is given by $\frac{-1+\sqrt{1+4(\frac{a}{b^2})}}{2}$. We may then conclude that $v(a,b)$, if it exists, must satisfy $v(a,b) = b\cdot\frac{-1+\sqrt{1+4(\frac{a}{b^2})}}{2} = \frac{-b+\sqrt{b^2+4a}}{2}$. Simple algebraic manipulation will show that $h = \frac{-b+\sqrt{b^2+4a}}{2}$ if and only if $a = h(h+b)$. This establishes condition 2.

As for condition 1, we have shown that the sequence $u_n(\frac{a}{b^2})$ is well defined, i.e., free from imaginary numbers, if and only if $a \geq b^2$. But now, $a = h(h+b)$ and the equality $a = b^2$ leads to the sequence which is alternately b and 0. The condition can be restated as $h(h+b) > b^2$, or $b^2 - hb - h^2 < 0$. For a given h, this means that the value of b must lie between the two roots of the quadratic equation $b^2 - hb - h^2 = 0$ in b; that is, $(\frac{1-\sqrt{5}}{2})h < b < (\frac{1+\sqrt{5}}{2})h$. Since $b > 0$, we have $0 < b < (\frac{1+\sqrt{5}}{2})h$, or $0 < b < \phi\cdot h$. This establishes condition 1. ■

COROLLARY 1. *For each positive integer k, there are integers a and b such that* $k = \sqrt{a - b\sqrt{a - b\sqrt{a - \cdots}}}$. *In fact, since a and b must satisfy the two conditions of Theorem 5, there are only finitely many such integers a and b.*

If k and a are integers and b is allowed to be a fraction, we may conclude that for each divisor q of k, and for each integer p such that $0 < \frac{p}{q} < \phi\cdot k$, if we let $b = \frac{p}{q}$ and $a = k(k+b)$, we will again have $k = \sqrt{a - b\sqrt{a - b\sqrt{a - \cdots}}}$.

For instance, for $k = 4$ and $q = 1$, since $\phi\cdot k \approx 6.472$, there are six possible values for $p = 1, 2, 3, \ldots, 6$ and the corresponding values for $b = \frac{p}{q} = p$ and a are $b = 1, 2, 3, \ldots, 6$ and $a = 20, 24, 28, \ldots, 40$, respectively. For $k = 4$ and $q = 2$, there are 12 possible values for $p = 1, 2, 3, \ldots, 12$ and the corresponding values for b and a are $b = 1/2, 1, 3/2, 2, \ldots, 6$ and $a = 18, 20, 22, 24, \ldots, 40$, respectively.

Alternating sequences

Having considered the nests of positive signs and nests of negative signs, we now consider nests with alternating signs. Specifically, we will consider the sequences

$$\sqrt{a - b\sqrt{a + b\sqrt{a - \cdots}}} \quad \text{and} \quad \sqrt{a + b\sqrt{a - b\sqrt{a + \cdots}}}, \qquad a, b > 0.$$

As we will see, the limits of these sequences depend on whether we start with a positive or negative sign. For our investigation, we need two recursively defined sequences. Let $x_1 = \sqrt{a + b\sqrt{a}}$ and $y_1 = \sqrt{a - b\sqrt{a}}$. For each $n \geq 1$, let $x_{n+1} = \sqrt{a + b\,y_n}$ and $y_{n+1} = \sqrt{a - bx_n}$. For what values of a and b will these sequences be defined and when will they be convergent? Note that there is no problem for the x_n's as long as the y_n's are positive. On the other hand, for $y_2 = \sqrt{a - b\sqrt{a + b\sqrt{a}}}$ to be real, the value of a will have to be greater than or equal to $b\sqrt{a + b\sqrt{a}}$. This also turns out to be sufficient for the sequences to be defined, as we now show.

LEMMA 2. *Let a and b be two positive real numbers. The sequences* $\{x_n\}$ *and* $\{y_n\}$ *are well-defined if and only if* $a > b\sqrt{a + b\sqrt{a}}$.

Proof. As we pointed out in the definition of y_2, the condition $a > b\sqrt{a + b\sqrt{a}}$ is necessary. To show this condition is also sufficient, consider two positive numbers a and b satisfying $a > b\sqrt{a + b\sqrt{a}}$. We will show sufficiency by proving the following stronger assertion:

Assertion. Let a and b be two positive real numbers such that $a > b\sqrt{a + b\sqrt{a}}$. Then, for each positive integer n, $0 < y_n \leq \sqrt{a} \leq x_n \leq \sqrt{a + b\sqrt{a}}$.

We first show that these inequalities are true for $n = 1$. We have

$$a > b\sqrt{a + b\sqrt{a}} > b\sqrt{a + 0} = b\sqrt{a}.$$

Since $y_1 = \sqrt{a - b\sqrt{a}}$, this means that $y_1 > 0$. Also,

$$y_1 = \sqrt{a - b\sqrt{a}} < \sqrt{a - 0} = \sqrt{a}.$$

Thus,

$$0 < y_1 \leq \sqrt{a}.$$

On the other hand, by definition, $x_1 \leq \sqrt{a + b\sqrt{a}}$. Thus, our assertion holds for $n = 1$.

Now, suppose that our assertion holds for a positive integer n. Since

$$x_n \leq \sqrt{a + b\sqrt{a}},$$

we have $a - bx_n \geq a - b\sqrt{a + b\sqrt{a}}$, which is a positive real number by our condition on the numbers a and b. Thus, the number $y_{n+1} = \sqrt{a - bx_n}$ is a well-defined positive real number. Furthermore, since both b and x_n are positive,

$$y_{n+1} = \sqrt{a - bx_n} < \sqrt{a - 0} = \sqrt{a}.$$

In addition, since $by_n > 0$ and $y_n \leq \sqrt{a}$, we have

$$\sqrt{a} = \sqrt{a - 0} < \sqrt{a + by_n} \leq \sqrt{a + b\sqrt{a}}.$$

This says that $\sqrt{a} \leq x_{n+1} \leq \sqrt{a + b\sqrt{a}}$. Thus, the assertion is also true for $n + 1$. By induction, this proves our assertion and, in consequence, that our sequences are well-defined. ■

What numbers a and b do in fact satisfy the inequality $a > b\sqrt{a + b\sqrt{a}}$ of Lemma 2? Dividing both sides by b^2 and squaring the two quantities, we can rewrite this inequality as $\left(\frac{a}{b^2}\right)^2 - \left(\frac{a}{b^2}\right) > \sqrt{\frac{a}{b^2}}$. Comparing the graphs of the functions $f(t) = t^2 - t$ and $g(t) = \sqrt{t}$ for $t \geq 0$, we see that there is a unique constant $c > 0$ such that $t^2 - t > \sqrt{t}$ if and only if $t > c$. Thus, a and b satisfy the condition $a > b\sqrt{a + b\sqrt{a}}$ if and only if $\frac{a}{b^2} > c$ or $a > b^2c$. Solving the equation $t^2 - t - \sqrt{t} = 0$, one may show that

$$c = \frac{1}{3}\left\{2 + \sqrt[3]{\frac{25}{2} - \frac{3}{2}\sqrt{69}} + \sqrt[3]{\frac{25}{2} + \frac{3}{2}\sqrt{69}}\right\} \approx 1.75488.$$

In the following theorem, we will show that the sequences x_n and y_n are both convergent. Assuming this for the moment, let $x = \lim_{n\to\infty} x_n$ and $y = \lim_{n\to\infty} y_n$. Taking the limit on both sides of the equations $x_{n+1} = \sqrt{a + by_n}$ and $y_{n+1} = \sqrt{a - bx_n}$, we have $x = \sqrt{a + by}$ and $y = \sqrt{a - bx}$. From these, one may easily show that the difference of these two limits, $x - y$ is always the number b. Thus, the "hybrid" sequence

$$\sqrt{a},\quad \sqrt{a-b\sqrt{a}},\quad \sqrt{a+b\sqrt{a-b\sqrt{a}}},\quad \sqrt{a-b\sqrt{a+b\sqrt{a-b\sqrt{a}}}},\quad \ldots$$

consisting of both x_n's and y_n's cannot be convergent. Furthermore, if $x = \lim_{n\to\infty} x_n$ and $y = \lim_{n\to\infty} y_n$, one may easily show that the two limits, x and y, have to satisfy the equations

$$x^4 - 2ax^2 + b^3x + a^2 - ab^2 = 0, \quad \text{and} \quad y^4 - 2ay^2 - b^3y + a^2 - ab^2 = 0,$$

respectively. Thus, for rational numbers a and b, the limits of the nested sequences

$$\sqrt{a-b\sqrt{a+b\sqrt{a-\cdots}}} \quad \text{and} \quad \sqrt{a+b\sqrt{a-b\sqrt{a+\cdots}}}$$

can never be transcendental numbers. We may now specify what numbers those limits can be.

THEOREM 6. *Let c be the constant specified above. For any positive real number k, let a and b be any real numbers such that*

(1) $0 < b < \frac{2k}{(\sqrt{4c-3})-1}$, *and*

(2) $a = k^2 + bk + b^2$,

then $\lim_{n\to\infty} y_n = k$ *and* $\lim_{n\to\infty} x_n = k + b$.

Proof.

(1) Let any positive constant k be given. We first show that if a and b are two numbers satisfying the conditions above, the condition $a > b\sqrt{a+b\sqrt{a}}$ will be satisfied. This can be seen as follows. It was shown above that a and b satisfy the condition $a > b\sqrt{a+b\sqrt{a}}$ if and only if $\frac{a}{b^2} > c$. Now, $a = k^2 + bk + b^2$. The inequality $\frac{a}{b^2} > c$ is equivalent to $(\frac{k}{b})^2 + (\frac{k}{b}) + 1 > c$, or $(\frac{k}{b})^2 + (\frac{k}{b}) + 1 - c > 0$. This expression is true if and only if the positive number $(\frac{k}{b})$ is greater than the positive root of the quadratic equation $x^2 + x + 1 - c = 0$, or

$$\frac{k}{b} > \frac{-1+\sqrt{4c-3}}{2}$$

which is equivalent to condition 1 above. Thus, if the two conditions in the theorem are satisfied, the sequences $\{x_n\}$ and $\{y_n\}$ are well-defined by Lemma 2 above.

(2) We now assume that a and b are two numbers satisfying the two conditions given in the theorem. We have in particular $a > b\sqrt{a+b\sqrt{a}}$. By the assertion in the proof of Lemma 2, each x_n and y_n is a well defined positive number and $\sqrt{a} \le x_n$ for each n. For any given positive number k and for each integer $n \ge 2$, since $y_{n+1} = \sqrt{a - bx_n}$ and $x_{n+1} = \sqrt{a + by_n}$,

$$\begin{aligned}
|y_{n+1} - k| &= \frac{|y_{n+1}^2 - k^2|}{|y_{n+1} + k|} \\
&\le \frac{|(a - bx_n) - k^2|}{k} \quad (\text{since } y_{n+1} = \sqrt{a - bx_n} > 0) \\
&= \frac{|(bk + b^2) - bx_n|}{k} \quad (\text{since } a = k^2 + bk + b^2) \\
&= \frac{b|(k+b)^2 - x_n^2|}{k((k+b) + x_n)}
\end{aligned}$$

$$= \frac{b|(k+b)^2 - (a+by_{n-1})|}{k|(k+b)+x_n|} \quad \text{(using again } a = k^2+bk+b^2\text{)}$$

$$= \frac{b|kb - by_{n-1}|}{k|(k+b)+x_n|}$$

$$\leq \frac{b^2|y_{n-1}-k|}{k|(k+b)+\sqrt{a}|} \quad \text{(now since } \sqrt{a} \leq x_n\text{)}$$

$$= \frac{1}{\frac{k}{b}(\frac{k}{b}+1+\sqrt{\frac{a}{b^2}})}|y_{n-1}-k|$$

for all x_n and y_n.

Since $\frac{k}{b} > \frac{-1+\sqrt{4c-3}}{2}$, where $c = 1.75488\ldots > 1.75$, $\frac{k}{b} > \frac{-1+\sqrt{4\times 1.75-3}}{2} = \frac{1}{2}$. Also, $\sqrt{\frac{a}{b^2}} = \sqrt{\frac{k^2+bk+b^2}{b^2}} > 1$. Thus, $\frac{1}{\frac{k}{b}(\frac{k}{b}+1+\sqrt{\frac{a}{b^2}})} < \frac{1}{\frac{1}{2}(\frac{1}{2}+1+1)} < 1$. From this, it follows that $\lim_{n\to\infty} y_n = k$. Now, since $x_n = \sqrt{a+by_{n-1}}$, taking the limit, we have $\lim_{n\to\infty} x_n = \sqrt{a+kb} = \sqrt{k^2+bk+b^2+bk} = k+b$. This completes the proof. ∎

Given that $c \approx 1.75488\ldots < 1.76$, $\frac{2k}{(\sqrt{4c-3})-1} > \frac{2k}{(\sqrt{4\times 1.76-3})-1} \approx 1.9802k$. For $b \leq 1.9k$ the possible values for the limits of x_n and y_n can now be summarized as follows:

(1) Any positive number k can be the limit of the sequence y_n. We need only choose two positive numbers a and b such that $b \leq 1.9k$, and $a = k^2+bk+b^2$. If k is rational, we might choose a and b to be rational as well. And if k is an integer, then a and b can also be integers. Thus, for instance, with $k=2$, $b=1$, and $a=7$,

$$2 = \sqrt{7-\sqrt{7+\sqrt{7-\sqrt{7+\cdots}}}};$$

with $k=2$, $b=2$, and $a=12$,

$$2 = \sqrt{12-2\sqrt{12+2\sqrt{12-2\sqrt{12+\cdots}}}};$$

with $k=2$, $b=3$, and $a=19$,

$$2 = \sqrt{19-3\sqrt{19+3\sqrt{19-3\sqrt{19+\cdots}}}}.$$

(2) Any positive number r can be the limit of the sequence x_n. This is the case because we may first choose a positive $k < r$ such that $r-k \leq k$, and then let $b = r-k$ and $a = k^2+bk+b^2$. Note that since $b = r-k < 1.9k$, we have $\lim_{n\to\infty} x_n = k+b = r$. By this construction, we may also claim that any positive rational number r can be the limit for a sequence x_n for some appropriate rational numbers a and b.

For instance, with $r=1$, $k=\frac{1}{2}$, $b=\frac{1}{2}$, and $a=\frac{3}{4}$,

$$1 = \sqrt{\frac{3}{4}+\frac{1}{2}\sqrt{\frac{3}{4}-\frac{1}{2}\sqrt{\frac{3}{4}+\frac{1}{2}\sqrt{\frac{3}{4}-\cdots}}}};$$

and with $r = 1, k = \frac{2}{3}, b = \frac{1}{3}$, and $a = \frac{7}{9}$,

$$1 = \sqrt{\frac{7}{9} + \frac{1}{3}\sqrt{\frac{7}{9} - \frac{1}{3}\sqrt{\frac{7}{9} + \frac{1}{3}\sqrt{\frac{7}{9} - \cdots}}}}.$$

However, if we require a, b, and $\lim_{n\to\infty} x_n$ all to be positive integers, then since $\lim_{n\to\infty} x_n = k + b > b \geq 1$, the limit $\lim_{n\to\infty} x_n$ will have to be greater than 1. On the other hand, for any integer $r \geq 2$, we may choose k, b, and a as above, except that k will now be a positive integer. Consequently, $b = r - k$ and $a = k^2 + bk + b^2$ will also be positive integers, and $\lim_{n\to\infty} x_n = k + b = r$ is an integer.

For example, with $r = 2, k = 1, b = 1$, and $a = 3$,

$$2 = \sqrt{3 + \sqrt{3 - \sqrt{3 + \sqrt{3 - \cdots}}}};$$

with $r = 3, k = 2, b = 1$, and $a = 7$,

$$3 = \sqrt{7 + \sqrt{7 - \sqrt{7 + \sqrt{7 - \cdots}}}};$$

with $r = 4, k = 3, b = 1$, and $a = 13$,

$$4 = \sqrt{13 + \sqrt{13 - \sqrt{13 + \sqrt{13 - \cdots}}}};$$

with $r = 4, k = 2, b = 2$, and $a = 12$,

$$4 = \sqrt{12 + 2\sqrt{12 - 2\sqrt{12 + 2\sqrt{12 - \cdots}}}};$$

with $r = 5, k = 4, b = 1$, and $a = 21$,

$$5 = \sqrt{21 + \sqrt{21 - \sqrt{21 + \sqrt{21 - \cdots}}}};$$

with $r = 5, k = 3, b = 2$, and $a = 19$,

$$5 = \sqrt{19 + 2\sqrt{19 - 2\sqrt{19 + 2\sqrt{19 - \cdots}}}};$$

with $r = 5, k = 3, b = 2$, and $a = 19$,

$$5 = \sqrt{19 + 3\sqrt{19 - 3\sqrt{19 + 3\sqrt{19 - \cdots}}}}.$$

Sets of nests: a broader view

Let's step back now from our detailed calculations and proofs concerning individual nests and look at all of this from another perspective. Indeed, let's literally distance ourselves from any specific nest and view the formation which it and its related nests create on the real line. Since much of the research here is still open, we'll discuss it somewhat informally.

Consider the set S_2, all nests of the form

$$\sqrt{2 \pm \sqrt{2 \pm \sqrt{2 \pm \sqrt{2 \pm \cdots}}}},$$

where the succession of $\pm$'s signifies all possible combinations of $+$ and $-$. Assuming each of these infinite nests has a limit, (which they almost certainly do, but which we have not yet proven,) we will indicate the nests and their limits by the same designation. The reader can see that the minimum of these nests $\sqrt{2 - \sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots}}}}$ with $+$'s continuing, equals zero, and the maximum $\sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots}}}}$ equals 2. All the others lie between 0 and 2. Three sets of questions present themselves:

1. Can two different nests be equal, i.e., can two different nests have the same limit? If not, how do we determine which is greater?
2. How big is the set S_2? Is it countable or uncountable?
3. What does the set look like? If it is uncountable, does it have a nonempty interior?

As for the first set of questions, two different nests can never be equal to each other. We can in fact determine which is greater as follows: Reading left to right, find the first position at which they differ in sign. One of them, s_+, has a '$+$' in this position, and the other, s_- has a '$-$'. To the left of this position they are exactly the same. If there are an even number of $-$'s to the left of this position, then $s_- < s_+$, but if there are an odd number of $-$'s, then $s_+ < s_-$. To answer the second set of questions, note that since two different nests in S_2 are never equal, there is a one-to-one correspondence between S_2 and the set of all the decimals between 0 and 1 as written in binary. We might let $+$ correspond to 0 and $-$ correspond to 1. Thus, S_2 is uncountable. The third set of questions is the most intricate, and calls for a paper of its own. At this stage of extensive but incomplete investigation, S_2 does not seem to have any interior. Indeed it does not even seem to be dense anywhere. Its points are separated by infinite sequences of "gaps" – intervals containing no points of S_2 – whose lengths decrease nearly geometrically. There are countably many such sequences, without any shared gaps, giving S_2 a complex, fractal-like structure. (The reader might like to verify that the largest such "gap" is the interval between

$$\sqrt{2 - \sqrt{2 - \sqrt{2 + \sqrt{2 + \cdots}}}}$$

(where the $+$'s continue indefinitely) and

$$\sqrt{2 + \sqrt{2 - \sqrt{2 + \sqrt{2 + \cdots}}}}$$

(again, the $+$'s continue indefinitely). This, and other similar discoveries, point to an unusual structure of the set S_2 yet to be specified. The same consideration can be given to sets S_a of the form

$$\sqrt{a \pm \sqrt{a \pm \sqrt{a \pm \sqrt{a \pm \cdots}}}}$$

with a being integral, rational, or simply real, as long as the nests themselves are real. Conjunctions of these sets, many of whose elements coincide, as we've seen above, create even more intricate sets, as yet unexplored.

Further research

There are two further interesting avenues of research which we've probed, (and no doubt many more that we haven't.) We might extend our earlier examples by considering those cases in which the signs inside the nested radicals alternate in "blocks." For instance, for a sequence in which the signs alternate in a block of 4 in the pattern $+$, $+$, $+$, $-$, one may define

$$x_0 = \sqrt{a}, \quad y_n = \sqrt{a + bx_n},\ z_n = \sqrt{a + by_n}, \quad w_n = \sqrt{a + bz_n},$$
$$\text{and} \quad x_{n+1} = \sqrt{a - bw_n}$$

for each natural number n. We invite the readers to find conditions on the numbers a and b so that each of the sequences x_n, y_n, z_n, and w_n will converge, and find the numbers which are their limits. While the precise solution is certainly formidable, we are assured that the limits of any such sequence must, once again, be an algebraic number and never transcendental.

We might also consider the curious correspondence between sequences of nested radicals and continued fractions: for any positive numbers a and b,

$$\sqrt{a + b\sqrt{a + b\sqrt{a + \cdots}}} = b + \cfrac{a}{b + \frac{a}{b + \cdots}}$$

since both limits satisfy the quadratic equation $x^2 - bx - a = 0$.

Likewise, for any positive numbers a and b such that $a \geq b^2$,

$$\sqrt{a - b\sqrt{a - b\sqrt{a - \cdots}}} = (-b) + \cfrac{a}{(-b) + \frac{a}{(-b) + \cdots}},$$

since both limits satisfy the quadratic equation $x^2 + bx - a = 0$. How far can this equivalence between nested radicals and continued fractions be extended?

These are some of the questions which remain to be explored.

Acknowledgment. We wish to thank the reviewers for their helpful suggestions.

REFERENCES

1. Jonathan M. Borwein and G. de Barra, Nested radicals, *Amer. Math. Monthly* **98** (1991) 735–739.
2. A. Herschfeld, On infinite radicals, *Amer. Math. Monthly* **42** (1935) 419–429.
3. D. J. Jones, Continued powers and a sufficient condition for their convergence, this MAGAZINE, **68** (1995) 387–392.
4. M. A. Nyblom, More nested square roots of 2, *Amer. Math. Monthly* **112** (2005) 822–825.
5. L. D. Servi, Nested square roots of 2, *Amer. Math. Monthly* **110** (2003) 326–330.
6. W. S. Sizer, Continued roots, this MAGAZINE **59** (1986) 23–27.
7. Eric W. Weisstein, *CRC Concise Encyclopedia of Mathematics*, Second Ed., Chapman & Hall/CRC, 2003, pp. 1999–2000.
8. B. Wong and M. McGuffin, The museum of infinite nested radicals, `www.dgp.toronto.edu/~mjmcguff/math/nestedRadicals.html`.

Synthetic Partial Fraction Decompositions

WILLIAM ADKINS
Louisiana State University
Baton Rouge, Louisiana 70803-4918
adkins@math.lsu.edu

MARK DAVIDSON
Louisiana State University
Baton Rouge, Louisiana 70803-4918
davidson@math.lsu.edu

The partial fraction decomposition of a general rational function over the real numbers has been routinely treated in calculus texts, where the procedure is normally taught, via the technique of undetermined coefficients. That is, students are told that a given rational function $p_0(s)/v(s)$ is a sum of terms (called partial fractions) of the form

$$\frac{A_j}{(s-\gamma)^j} \quad \text{and} \quad \frac{B_k s + C_k}{(s^2+as+b)^k},$$

where A_j, B_k, and C_k are real constants. The partial fractions are determined by the linear factors $s-\gamma$ and irreducible quadratic factors s^2+as+b of the denominator $v(s)$, where the powers j and k occur up to the multiplicity of the factors. After finding a common denominator and equating the numerators students are left with a system of linear equations to solve for the undetermined coefficients A_j, B_k, C_k. Even when the multiplicities are relatively small such a scheme quickly becomes unwieldy. Moreover, at this level the students are probably not conversant enough with linear algebra to be able to follow an explanation of why this procedure works.

As has been observed numerous times, there is an alternative algorithmic method for partial fraction decompositions that primarily involves repeated division by polynomials. This method is described in the classic texts by Chrystal [**2**] (see pages 151–159) and van der Waerden [**6**] (Pages 88–89). In versions that are appropriate for elementary calculus and differential equations classes, this method for partial fractions has been presented in 1943 by Boldyreff [**1**], in 1972 by Hamilton [**3**], and again in 1988 by Scott and Peeples [**5**]. None of these papers reference the others, so it is presumably a result that is frequently rediscovered, and surprisingly, not as well-known as it should be. The algorithmic method has the advantage that it is constructive (assuming the factorization of the denominator), recursive (meaning that only one coefficient at a time is determined), and self checking. The goal of this paper is to present the partial fraction decomposition algorithm in a format that is amenable to recursive hand calculations in calculus or differential equations classes. Additionally, the calculation scheme will be structured so as to employ only real number arithmetic and evaluation of polynomials with real coefficients. These are calculations that are done on the coefficients of the polynomials involved, and hence we refer to the method as synthetic partial fraction decomposition, due to the use of synthetic division by linear and quadratic polynomials.

The algorithmic decomposition into partial fractions is based on a simple lemma, which is most convenient to state in the form expressed by Scott and Peeples [**5**].

LEMMA 1. *Suppose $t(s) \in \mathbb{R}[s]$ is an irreducible real polynomial (so that it is either linear or quadratic) which divides the denominator $v(s)$ of a proper rational function $p_0(s)/v(s) \in \mathbb{R}(s)$. Hence $v(s) = t^n(s)q(s)$ where $t(s)$ and $q(s)$ are relatively prime. Then there are unique polynomials $C_1(s)$ and $p_1(s)$ with* $\deg C_1(s) <$

$\deg t(s)$ *such that*

$$\frac{p_0(s)}{v(s)} = \frac{p_0(s)}{t^n(s)q(s)} = \frac{C_1(s)}{t^n(s)} + \frac{p_1(s)}{t^{n-1}(s)q(s)}. \tag{1}$$

The polynomials $C_1(s)$ *and* $p_1(s)$ *are given by*

$$C_1(s) = [p_0(s)]/[q(s)] \qquad \text{and} \qquad p_1(s) = (p_0(s) - C_1(s)q(s))/t(s).$$

The quotient $C_1(s) = [p_0(s)]/[q(s)]$ *takes place in the congruence class field* $\mathbb{R}[s]/\langle t(s)\rangle$ *(which is isomorphic to* $\mathbb{R}$ *or* $\mathbb{C}$*), where the symbol* $[f(s)]$ *means the congruence class of the polynomial* $f(s)$ *in* $\mathbb{R}[s]/\langle t(s)\rangle$*; the formula for* $p_1(s)$ *uses division in* $\mathbb{R}[s]$.

The proof of this lemma is straightforward. Equation (1) is equivalent to the polynomial equation

$$p_0(s) = C_1(s)q(s) + t(s)p_1(s).$$

This equation can be solved (uniquely) for $C_1(s)$ since the equation $p_0(s) = Xq(s)$ is uniquely solvable in the congruence class field $\mathbb{R}[s]/\langle t(s)\rangle$ because $[q(s)] \neq [0] \in \mathbb{R}[s]/\langle t(s)\rangle$. Since the elements of $\mathbb{R}[s]/\langle t(s)\rangle$ are represented uniquely by polynomials of degree less than the degree of $t(s)$, we have shown the existence and uniqueness of $C_1(s)$, and $p_1(s)$ exists (and is unique) because of the congruence identity $[p_0(s) - C_1(s)q(s)] = [0]$.

In case $t(s) = s - \gamma$ is linear, then $C_1(s)$ is the constant $p_0(\gamma)/q(\gamma)$. In case $t(s) = s^2 + as + b$ is an irreducible quadratic then the polynomial $C_1(s)$ has degree at most 1 so it has the form $C_1(s) = As + B$, and in place of the congruence arithmetic described above, the coefficients can also be computed from the complex equation

$$A\gamma + B = \frac{p_0(\gamma)}{q(\gamma)}, \tag{2}$$

where $t(\gamma) = 0$, by taking into account that A and B are real. Thus, if desired, the congruence arithmetic can be avoided.

An application of Lemma 1 produces two items:

- the partial fraction of the form

$$\frac{C_1(s)}{t^n(s)},$$

  where $C_1(s)$ is a constant when $t(s)$ is linear and $C_1(s)$ is linear when $t(s)$ is quadratic, and
- a remainder term of the form

$$\frac{p_1(s)}{t^{n-1}(s)q(s)},$$

such that the rational function $p_0(s)/v(s)$ is the sum of these two pieces. We can now repeat the process on the new rational function $p_1(s)/(t^{n-1}(s)q(s))$, where the multiplicity of $t(s)$ in the denominator has been reduced by 1, and continue in this manner until we have removed completely $t(s)$ as a factor of the denominator. In this manner we have recursively produced a sequence, $\frac{C_1(s)}{t^n(s)}, \dots, \frac{C_n(s)}{t(s)}$, which we will refer to as the ***t*(*s*)-chain** for the rational function $p_0(s)/v(s)$. The following table summarizes the data obtained.

The $t(s)$-chain	
$\dfrac{p_0(s)}{t^n(s)q(s)}$	$\dfrac{C_1(s)}{t^n(s)}$
$\dfrac{p_1(s)}{t^{n-1}(s)q(s)}$	$\dfrac{C_2(s)}{t^{n-1}(s)}$
$\vdots$	$\vdots$
$\dfrac{p_{n-1}(s)}{q(s)}$	$\dfrac{C_n(s)}{t(s)}$
$\dfrac{p_n(s)}{q(s)}$	

From the table we get

$$\frac{p_0(s)}{t^n(s)q(s)} = \frac{C_1(s)}{t^n(s)} + \cdots + \frac{C_n(s)}{t(s)} + \frac{p_n(s)}{q(s)}.$$

By factoring another linear or quadratic term out of $q(s)$ the process can be repeated until one obtains the complete partial fraction decomposition.

The production of the $t(s)$-chain for the rational function $p_0(s)/v(s)$ involves two calculations. The calculation of $C_1(s) = [p_0(s)]/[q(s)]$ and the subsequent calculation of $p_1(s) = (p_0(s) - C_1(s)q(s))/t(s)$. We will describe the calculations separately for the two cases where $t(s)$ is linear and where $t(s)$ is an irreducible quadratic.

The linear case

Synthetic division. We begin with a reminder of synthetic division of a polynomial $f(s)$ by a linear term of the form $s - \gamma$. By the division algorithm we can write

$$f(s) = f^*(s)(s - \gamma) + d_0, \tag{3}$$

where $d_0 = f(\gamma)$. Suppose $f(s) = c_n s^n + \cdots + c_1 s + c_0$ and $f^*(s) = d_n s^{n-1} + \cdots + d_2 s + d_1$. Set $d_{n+1} = 0$. Then Equation (3) implies $c_k = d_k - \gamma d_{k+1}$ so that $d_k = c_k + \gamma d_{k+1}$, for $k = n, \ldots, 0$. If $i_k = \gamma d_{k+1}$, then $d_k = c_k + i_k$ gives a recursive scheme for computing $f^*(s)$ and $f(\gamma)$. Writing this recursion in tabular form gives the familiar version of synthetic division:

$$\begin{array}{c|ccccccc} \gamma & c_n & c_{n-1} & c_{n-2} & \cdots & c_2 & c_1 & c_0 \\ & i_n & i_{n-1} & i_{n-2} & \cdots & i_2 & i_1 & i_0 \\ \hline d_{n+1} & d_n & d_{n-1} & d_{n-2} & \cdots & d_2 & d_1 & \boxed{d_0} \end{array}$$

Note that the last line contains both the evaluation $f(\gamma) = d_0$, which we will indicate by enclosing it in a box, and the coefficients of the quotient polynomial $f^*(s) = (f(s) - f(\gamma))/(s - \gamma)$.

Synthetic PFD by a linear term. We now turn our attention to computing the $t(s)$-chain when $t(s) = s - \gamma$ is a linear term dividing the denominator of a rational function $p_0(s)/v(s)$ with multiplicity n. According to Lemma 1, we can write

$$\frac{p_0(s)}{v(s)} = \frac{p_0(s)}{(s-\gamma)^n q(s)} = \frac{A_1}{(s-\gamma)^n} + \frac{p_1(s)}{(s-\gamma)^{n-1}q(s)},$$

where $A_1 = p_0(\gamma)/q(\gamma)$ and $p_1(s) = (p_0(s) - A_1 q(s))/(s-\gamma)$. Notice that we can write

$$\begin{aligned} p_1(s) &= \frac{p_0(s) - A_1 q(s)}{s-\gamma} \\ &= \frac{p_0^*(s)(s-\gamma) + p_0(\gamma) - A_1(q^*(s)(s-\gamma)+q(\gamma))}{s-\gamma} \\ &= p_0^*(s) - A_1 q^*(s). \end{aligned}$$

Synthetic division of both $q(s)$ and $p_0(s)$ by $s-\gamma$ produces $q^*(s)$, $p_0^*(s)$, $q(\gamma)$ and $p_0(\gamma)$. Hence, these divisions produce both the coefficient $A_1 = p_0(\gamma)/q(\gamma)$ of the partial fraction $A_1/(s-\gamma)^n$ and the numerator $p_1(s) = p_0^*(s) - A_1 q^*(s)$ for the remainder term

$$\frac{p_1(s)}{(s-\gamma)^{n-1}q(s)}.$$

Now repeat the process until $s-\gamma$ is no longer a divisor of the denominator. This is the end of the $s-\gamma$ chain. Notice that at each stage the term $q(s)$ in the denominator remains the same, so it is only necessary to do the division for this term once. The leftover term is $p_n(s)/q(s)$. Also note that if $q(s)=1$ then $q^*(s)=0$ and $q(\gamma)=1$; thus, in this case, the algorithm reduces to a sequence of divisions by $s-\gamma$. This special case is considered in the note by Kung [**4**].

An outline of one iteration of the synthetic PFD method has the following form. We assume the initial computation of $q^*(s)$ and $u = q(\gamma)$ have been made. We will use the convention that a polynomial $p(s) = c_n s^n + \cdots + c_1 s + c_0$ will be denoted by listing its coefficients $p = c_n \ \ldots \ c_1 \ c_0$ in order of decreasing powers of s. The computation of A_1 and $p_1(s)$ is then summarized in the following scheme:

$$\begin{array}{r|l} \gamma & p_0 \\ & p_0^* \quad \boxed{r_1} \\ - & A_1 q^* \\ \hline & p_1 \end{array}$$

On line 1) put the coefficients of $p_0(s)$. On line 2) put the result of synthetic division by γ. The boxed term, r_1, is the remainder ($= p_0(\gamma)$) and determines $A_1 = r_1/u$. On line 3) put $-A_1$ times q^*. Then p_1, the sum of lines 2) and 3), goes on line 4).

We illustrate the synthetic PFD method by the following example.

EXAMPLE 2. Find the partial fraction decomposition for the rational function

$$\frac{2s^4 - 8s^3 - 10s^2 + 8s}{(s+1)^3(s-1)^3}.$$

We begin with $\gamma = -1$ and $q(s) = (s-1)^3 = s^3 - 3s^2 + 3s - 1$. Synthetic division by $s+1$ produces

$$\begin{array}{r|rrrr} -1 & 1 & -3 & 3 & -1 \\ & 0 & -1 & 4 & -7 \\ \hline & 1 & -4 & 7 & \boxed{-8} \end{array}$$

Therefore $q^* = 1 \quad -4 \quad 7$ and $u = -8$. Hence, $A_i = -\frac{r_i}{8}$. In the following table we provide the details to find the $(s+1)$-chain. Note that in the heading we put $-q^* = -1 \quad 4 \quad -7$ to facilitate the calculation of $-A_i q^*$.

The $(s + 1)$-chain

	A_i	$-q^*$						
	$\frac{r_i}{u} = \frac{r_i}{-8}$			-1	4	-7		
$\frac{2s^4 - 8s^3 - 10s^2 + 8s}{(s+1)^3(s-1)^3}$		$-1\rfloor$	2	-8	-10	8	0	p_0
			0	-2	10	0	-8	
	$\frac{1}{(s+1)^3}$		2	-10	0	8	$\boxed{-8}$	p_0^* $\boxed{r_1}$
				-1	4	-7		$-A_1 q^*$
$\frac{2s^3 - 11s^2 + 4s + 1}{(s+1)^2(s-1)^3}$		$-1\rfloor$	2	-11	4	1		p_1
			0	-2	13	-17		
	$\frac{2}{(s+1)^2}$		2	-13	17	$\boxed{-16}$		p_1^* $\boxed{r_2}$
			-2	8	-14			$-A_2 q^*$
$\frac{-5s + 3}{(s+1)(s-1)^3}$		$-1\rfloor$	0	-5	3			p_2
			0	0	5			
	$\frac{-1}{(s+1)}$		0	-5	$\boxed{8}$			p_2^* $\boxed{r_3}$
		1	-4	7				$-A_3 q^*$
$\frac{s^2 - 4s + 2}{(s-1)^3}$		1	-4	2				p_3

The denominator of the last entry in the first column is of the form $(s - \gamma)^n q(s)$ with $\gamma = 1$ and $q(s) = 1$. Thus the table may be extended to give the $(s - 1)$-chain on the remainder term $\frac{s^2 - 4s + 2}{(s-1)^3}$. Here are the details.

The $(s - 1)$-chain

	A_i	$-q^*$				
	$\frac{r_i}{u} = r_i$			0		
$\frac{s^2 - 4s + 2}{(s-1)^3}$		$1\rfloor$	1	-4	2	p_3
			0	1	-3	
	$\frac{-1}{(s-1)^3}$		1	-3	$\boxed{-1}$	p_3^* $\boxed{r_4}$
			0	0		$-A_4 q^*$
$\frac{s-3}{(s-1)^2}$		$1\rfloor$	1	-3		p_4
			0	1		
	$\frac{-2}{(s-1)^2}$		1	$\boxed{-2}$		p_4^* $\boxed{r_5}$
			0			$-A_5 q^*$
$\frac{1}{s-1}$			1			p_5

We put the chains together to get the complete Partial Fraction Decomposition:

$$\frac{2s^4 - 8s^3 - 10s + 8s}{(s+1)^3(s-1)^3}$$
$$= \frac{1}{(s+1)^3} + \frac{2}{(s+1)^3} - \frac{1}{s+1} - \frac{1}{(s-1)^3} - \frac{2}{(s-1)^2} + \frac{1}{s-1}.$$

The quadratic case

We now address the quadratic case. This case is admittedly more complicated but some familiar algebraic constructs will simplify the end calculations so that the quadratic synthetic PFD method will follow a pattern very similar to the linear case.

Quadratic synthetic division. We begin by a brief review of quadratic synthetic division. Let $f(s)$ be any polynomial and $t(s) = s^2 + as + b$ a fixed quadratic. Then we can write

$$f(s) = f^*(s)(s^2 + as + b) + r_1 s + r_0. \tag{4}$$

Now suppose

$$\begin{aligned} f(s) &= c_n s^n + c_{n-1}s^{n-1} + \cdots c_1 s + c_0 \\ \text{and} \quad f^*(s) &= d_n s^{n-2} + d_{n-1}s^{n-3} + \cdots d_3 s + d_2. \end{aligned}$$

If we set $d_{n+1} = d_{n+2} = 0$ then Equation 4 implies the following relations

$$\begin{aligned} c_k &= d_k + ad_{k+1} + bd_{k+2}, \quad k = 2, \ldots, n, \\ c_1 &= r_1 + ad_2 + bd_3 \\ c_0 &= r_0 + bd_2. \end{aligned}$$

Solving for $d_2, \ldots d_n$, r_1, and r_0 gives

$$\begin{aligned} d_k &= c_k + (-b \quad -a) \cdot (d_{k+2} \quad d_{k+1}) \quad k = 2, \ldots, n, \\ r_1 &= c_1 + (-b \quad -a) \cdot (d_3 \quad d_2) \\ r_0 &= c_0 + (-b \quad -a) \cdot (d_2 \quad d_1), \end{aligned}$$

where we set $d_1 = 0$ and, for use below, we set $d_0 = 0$. In these formulas the dot product replaces the usual product found in synthetic division by a linear term. If $i_k = (-b \quad -a) \cdot (d_{k+2} \quad d_{k+1})$ then quadratic synthetic division takes on the following schematic form:

$$\begin{array}{cc|cccccccc} -b & -a & c_n & c_{n-1} & c_{n-2} & \cdots & c_2 & c_1 & c_0 \\ \hline & & i_n & i_{n-1} & i_{n-2} & \cdots & i_2 & i_1 & i_0 \\ \cline{3-9} d_{n+2} & d_{n+1} & d_n & d_{n-1} & d_{n-2} & \cdots & d_2 & d_1 & d_0 \\ & & & & & & & \boxed{r_1} & \boxed{r_0} \end{array}$$

Working from left to right we assume $d_{n+2}, \ldots, d_{k+1}$ have been computed. Then $d_k = c_k + i_k$, for $k = n, \ldots, 2$. To continue the established pattern, we insert $d_1 = d_0 = 0$ and compute the remainder terms as $r_1 = c_1 + i_1$ and $r_0 = c_0 + i_0$. The remainder is put on a separate line. In examples, we will not write zero in for d_{n+2} and d_{n+1}. The following example should convey the ease of use.

EXAMPLE 3. Find the quotient and remainder in

$$\frac{s^5+3s^4+4s^3+4s^2-s+3}{s^2+2s+3}.$$

Quadratic synthetic division gives

−3	−2	1	3	4	4	−1	3
		0	−2	−5	−1	−3	−9
		1	1	−1	3	0	0
						−4	−6

from which it follows that

$$\frac{s^5+3s^4+4s^3+4s^2-s+3}{s^2+2s+3}=s^3+s^2-s+3+\frac{-4s-6}{s^2+2s+3}.$$

Multiplication by a linear term. In the algorithm that follows it will be useful to express the product of a polynomial by a linear term within the synthetic scheme. Suppose $f(s)=c_ns^n+\cdots+c_1s+c_0$ is a polynomial and $L(s)=as+b$ is a linear term. The product $L(s)f(s)$ can be computed by a sort of sliding dot product. Express

$$L(s)f(s)=d_{n+1}s^{n+1}+\cdots+d_1s+d_0.$$

Let $c_{n+1}=0$ and $c_{-1}=0$. Then a straightforward calculation gives

$$d_k=(b\ \ a)\cdot(c_k\ \ c_{k-1}),$$

for $k=n+1,\ldots,0$. For example, if $f(s)=2s^3-3s^2+s+4$ and $L(s)=2s-1$ then $(-1+2s)(2s^3-3s^2+s+4)$ would be written within the synthetic method as

$$\begin{pmatrix}-1 & 2\end{pmatrix}\cdot\begin{pmatrix}2 & -3 & 1 & 4\end{pmatrix}=\begin{pmatrix}4 & -8 & 5 & 7 & -4\end{pmatrix}$$

to give $(2s-1)(2s^3-3s^2+s+4)=4s^4-8s^3+5s^2+7s-4$.

Synthetic PFD by a quadratic term. We now describe the synthetic partial fraction decomposition by an irreducible quadratic. Assume $p_0(s)$ and $v(s)$ are real polynomials and that s^2+as+b is an irreducible quadratic that is a factor of $v(s)$ of multiplicity n. Then according to Lemma 1 and Equation (2) we have

$$\frac{p_0(s)}{v(s)}=\frac{p_0(s)}{(s^2+as+b)^nq(s)}=\frac{A_1+B_1s}{(s^2+as+b)^n}+\frac{p_1(s)}{(s^2+as+b)^{n-1}q(s)}.$$

The coefficients of the linear term can be computed via the complex equation

$$A_1+B_1\gamma=\frac{p_0(\gamma)}{q(\gamma)},$$

where γ is a complex root of s^2+as+b, and

$$p_1(s)=\frac{p_0(s)-(A_1+B_1s)q(s)}{s^2+as+b}.$$

We now describe how to determine both the linear term $A_1 + B_1 s$ and the new numerator $p_1(s)$ by means of the division algorithm in a manner similar to the linear case.

To find A_1, B_1, and $p_1(s)$ we apply quadratic synthetic division to $p_0(s)$ and $q(s)$ to get

$$\frac{p_0(s)}{s^2+as+b} = p_0^*(s) + \frac{r_1 s + r_0}{s^2+as+b}$$

and

$$\frac{q(s)}{s^2+as+b} = q^*(s) + \frac{u_1 s + u_0}{s^2+as+b}.$$

We can now write

$$p_1(s) = p_0^*(s) - (A_1 + B_1 s)q^*(s) + \frac{r_1 s + r_0 - (A_1 + B_1 s)(u_1 s + u_0)}{s^2+as+b}. \quad (5)$$

Let F_1 be the last term in Equation (5). Since $p_1(s)$ is a polynomial the numerator of F_1 must have $s^2 + as + b$ as a factor. Thus

$$F_1 = -B_1 u_1,$$

the coefficient of s^2 in the numerator. It also follows that γ is a root of the numerator. Thus

$$A_1 + B_1\gamma = \frac{r_1\gamma + r_0}{u_1\gamma + u_0}. \quad (6)$$

Noting (from the quadratic formula) that the complex conjugate of the root γ of $s^2 + as + b$ is $-\gamma - a$, and that $\gamma(\gamma + a) = -b$, we can determine A_1 and B_1 from Equation (6) by rationalizing the denominator:

$$\begin{aligned} A_1 + B_1\gamma &= \frac{r_1\gamma + r_0}{u_1\gamma + u_0} \cdot \frac{-u_1(\gamma + a) + u_0}{-u_1(\gamma + a) + u_0} \\ &= \frac{r_1 u_1 b - r_0 u_1 a + r_0 u_0 + (r_1 u_0 - r_0 u_1)\gamma}{b u_1^2 - a u_1 u_0 + u_0^2}. \end{aligned}$$

Let

$$M = \begin{bmatrix} b & 0 \\ -a & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} r_1 & r_0 \end{bmatrix}, \quad \text{and} \quad U = \begin{bmatrix} u_1 & u_0 \end{bmatrix}.$$

Then after a short calculation we find

$$A_1 = \frac{RMU^t}{UMU^t} \quad \text{and} \quad B_1 = \frac{RJU^t}{UMU^t}.$$

Let $L_1(s) = A_1 + B_1 s$. With L_1 and F_1 determined by A_1 and B_1 we can rewrite Equation (5) as

$$p_1(s) = p_0^*(s) - L_1(s)q^*(s) + F_1.$$

An outline of one iteration of the synthetic PFD method has the following schematic form. We assume the initial computation of $q^*(s)$ and the remainder $u_1 s + u_0$ have been made.

$$\begin{array}{rl|l} -b & -a & p_0 \\ & & p_0^* \\ & & \qquad \boxed{r_1\ r_0} \\ & - & L_1 q^* \\ & & F_1 \\ \hline & & p_1 \end{array}$$

On the first line we place p_0. We divide p_0 synthetically by $s^2 + as + b$ to get the quotient, p_0^*, and the remainder $\boxed{r_1\ r_0}$. Both L_1 and F_1 are determined according to the formulas above. Place $-L_1 q^*$ on the fourth line and F_1 on the fifth line. The sum of the second, fourth, and fifth lines is p_1.

We illustrate the synthetic PFD method in the following example.

EXAMPLE 4. Find the partial fraction decomposition for the rational function

$$\frac{2s^3 + 8}{(s^2 + s + 2)^3 (s^2 + 2s + 2)^2}.$$

Let $q(s) = (s^2 + 2s + 2)^2 = s^4 + 4s^3 + 8s^2 + 8s + 4$. Synthetic division by $s^2 + s + 2$ gives

$$\begin{array}{rr|rrrrr} -2 & -1 & 1 & 4 & 8 & 8 & 4 \\ & & 0 & -1 & -5 & -9 & -6 \\ \hline & & 1 & 3 & 3 & 0 & 0 \\ & & & & & \boxed{-1} & \boxed{-2} \end{array}$$

Thus $q^* = 1 \quad 3 \quad 3$ and $U = \begin{bmatrix} u_1 & u_0 \end{bmatrix} = \begin{bmatrix} -1 & -2 \end{bmatrix}$. With the notation as above we have

$$M = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}.$$

From this it follows that

$$A_i = \frac{-2r_1 - r_0}{4}$$

and

$$B_i = \frac{-2r_1 + r_0}{4}$$

and hence we get

$$F_i = \frac{-2r_1 + r_0}{4}$$

and

$$L_i(s) = \frac{-2r_1 - r_0}{4} + \frac{-2r_1 + r_0}{4} s.$$

For the purposes of exposition we will first construct the synthetic PFD table and use it to construct the $s^2 + s + 2$-chain. We also will write $-q^* = -1 \ -3 \ -3$ in the heading to facilitate the calculation $-L_i q^*$.

Synthetic calculations for the (s^2+s+2)-chain

F_i	L_i	$-q^*$										
$\frac{-2r_1+r_0}{4}$	$\frac{-2r_1-r_0}{4}+\frac{-2r_1+r_0}{4}s$					-1	-3	-3				
						-2	-1	2	0	0	8	p_0
								0	-2	-2	4	
								2	-2	0	0	p_0^*
4	$-2+4s$									$\boxed{-2}$	$\boxed{12}$	$\boxed{R}$
						-4	-10	-6	6			$-L_1q^*$
									4			F_1
				-2	-1	-4	-10	-4	8			p_1
						0	4	14	12			
						-4	-6					p_1^*
0	$-10+0s$							$\boxed{10}$	$\boxed{20}$			$\boxed{R}$
					10	30	30					$-L_2q^*$
							0					F_2
			-2	-1	10	26	24					p_2
					0	-10	-20					
					10	0	0					p_2^*
-7	$-9-7s$					$\boxed{16}$	$\boxed{-4}$					$\boxed{R}$
		7	30	48	27							$-L_3q^*$
					-7							F_3
		7	30	48	30							p_3

We can now read off the (s^2+s+2)-chain.

The (s^2+s+2)-chain

$\dfrac{2s^3+8}{(s^2+s+2)^3(s^2+2s+2)^2}$	$\dfrac{-2+4s}{(s^2+s+2)^3}$
$\dfrac{-4s^3-10s^2-4s+8}{(s^2+s+2)^2(s^2+2s+2)^2}$	$\dfrac{-10}{(s^2+s+2)^2}$
$\dfrac{10s^2+26s+24}{(s^2+s+2)(s^2+2s+2)^2}$	$\dfrac{-9-7s}{(s^2+s+2)}$
$\dfrac{7s^3+30s^2+48s+30}{(s^2+2s+2)^2}$	

The remainder term

$$\frac{7s^3+30s^2+48s+30}{(s^2+2s+2)^2}$$

has denominator of the form

$$(s^2+2s+2)^2q(s)$$

with $q(s)=1$. With this data we have $q^*=0$, $U=\begin{bmatrix}0 & 1\end{bmatrix}$, and

$$M=\begin{bmatrix}2 & 0\\ -2 & 1\end{bmatrix}.$$

It follows that $A_i = r_0$ and $B_i = r_1$ from which we get

$$F_i = 0 \quad \text{and} \quad L_i = r_0 + r_1 s.$$

The synthetic PFD table is

Synthetic calculations for the (s^2+2s+2)-chain								
F_i	L_i	$-q^*$						
0	$r_0 + r_1 s$	0						
		-2	-2	7	30	48	30	p_0
				0	-14	-46	-32	
				7	16	0	0	p_0^*
0	$-2+2s$					$\boxed{2}$	$\boxed{-2}$	$\boxed{R}$
				0	0			$-L_1 q*$
					0			F_1
				7	16			

From this table we get the $(s^2 + 2s + 2)$-chain:

The (s^2+2s+2)-chain	
$\dfrac{7s^3+30s^2+48s+30}{(s^2+2s+2)^2}$	$\dfrac{-2+2s}{(s^2+2s+2)^2}$
$\dfrac{16+7s}{s^2+2s+2}$	

The complete partial fraction can now be read off from the chains:

$$\frac{2s^3+8}{(s^2+s+2)^3(s^2+2s+2)^2} = \frac{-2+4s}{(s^2+s+2)^3} + \frac{-10}{(s^2+s+2)^2} + \frac{-9-7s}{(s^2+s+2)}$$
$$+ \frac{-2+2s}{(s^2+2s+2)^2} + \frac{16+7s}{s^2+2s+2}.$$

REFERENCES

1. Alexander W. Boldyreff, Decomposition of Rational Fractions into Partial Fractions, *National Mathematics Magazine* **17** (1943) 261–167.
2. G. Chrystal, *Algebra, Part I*, Chelsea Publishing Co., New York, 1964.
3. Hugh J. Hamilton, The Partial Fraction Decomposition of a Rational Function, this MAGAZINE **45** (1972) 117–119.
4. Sidney H. Kung, Partial Fraction Decomposition by Division, *College Math. J.* **37** (2006) 132–134.
5. Dan Scott and Donald R. Peeples, A Constructive Proof of the Partial Fraction Decomposition, *American Mathematical Monthly* **95** (1988) 651–653.
6. B. L. van der Waerden, *Modern Algebra, Volume I*, Frederick Ungar Publishing Co., New York, 1953.

A Brief History of Impossibility

JEFF SUZUKI
Brooklyn College
Brooklyn, NY
jeff_suzuki@yahoo.com

Sooner or later every student of geometry learns of three "impossible" problems:

1. Trisecting the angle: Given an arbitrary angle, construct an angle exactly one-third as great.
2. Duplicating the cube: Given a cube of arbitrary volume, find a cube with exactly twice the volume.
3. Squaring the circle: Given an arbitrary circle, find a square with the same area.

These problems originated around 430 BC at a time when Greek geometry was advancing rapidly. We might add a fourth problem: inscribing a regular heptagon in a circle. Within two centuries, all these problems had been solved (see [**3**, Vol. I, p. 218–270] and [**1**] for some of these solutions).

So if these problems were all solved, why are they said to be impossible? The "impossibility" stems from a restriction, allegedly imposed by Plato (427–347 BC), that geometers use no instruments besides the compass and straightedge. This restriction requires further explanation. For that, we turn to Euclid (fl. 300 BC), who collected and systematized much of the plane geometry of the Greeks in his *Elements*.

Euclid's goal was to develop geometry in a deductive manner from as few basic assumptions as possible. The first three postulates in the *Elements* are (in modernized form):

1. Between any two points, there exists a unique straight line.
2. A straight line may be extended indefinitely.
3. Given any point and any length, a circle may be constructed centered at the point with radius equal to the given length.

These three postulates correspond to the allowable uses of compass and straightedge: to draw a line that passes through two given points; to extend a given line segment indefinitely; and to draw a circle about any given point with any given radius. To solve a problem using compass and straightedge means to use only these operations, repeated a finite number of times. The construction's validity can then be proven using only the postulates of Euclidean geometry.

For example, consider the problem of duplicating the cube. In order to duplicate a cube with a side length of a, it is necessary to construct a line segment of length $\sqrt[3]{2}a$. One of the simpler solutions, presented by Menaechmus around 350 BC, is equivalent to locating the intersection point of the parabola $ay = x^2$ and the hyperbola $xy = 2a^2$; these two curves intersect at the point $\left(\sqrt[3]{2}a, \sqrt[3]{4}a\right)$. Since this solution requires the use of the hyperbola and parabola, it is not a compass and straightedge solution.

A more subtle problem occurs with the trisection problem. Suppose we wish to trisect $\angle BOE$, which we may assume to be the central angle of arc BE in a circle (see Figure 1). There are several *neusis* ("verging") solutions, one of which is the following. Draw BC parallel to OE and then draw CA with the property that $DA = OB$ (the radius of the circle). It is relatively easy to prove that $\angle DOA = \frac{1}{3}\angle BOE$ (a proof we will leave to the reader). We can accomplish this construction with compass and straightedge as follows: Open the compass to fixed length equal to the radius OB. Using C as a pivot, swing the straightedge around, using the compass to measure out a length OB from the point where the straightedge crosses the circle, until you find the point D where the length $DA = OB$.

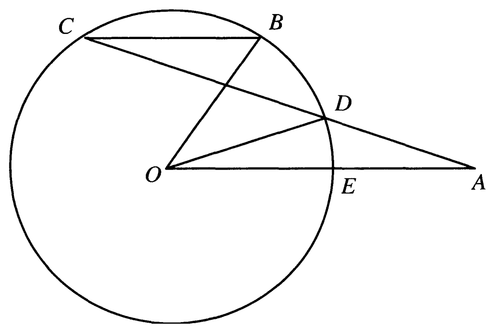


Figure 1 Neusis trisection of an angle.

There are at least two objections that can be raised to this "compass and straightedge" solution. First, the postulates only guarantee the existence of a line between two points, or the extension of an existing line; hence there is no guarantee that the line like CA, specified by a point C and a length DA, even exists. Second, the postulates only allow us to measure out a length by means of a circle of known center. This means we cannot measure the length DA equal to OB until we locate D. Thus, even though this solution uses compass and straightedge, it is not a compass and straightedge solution.

Even if we restrict ourselves to the canonical uses of the compass and straightedge, how can we distinguish between constructions that have never been done and those that are actually impossible? Before 1796, no compass and straightedge construction for a regular heptadecagon was known, but in that year Gauss discovered how to inscribe one in a circle. Might there be some as-yet-undiscovered means of trisecting an angle or duplicating the cube using compass and straightedge? In 1837 an obscure French mathematician named Pierre Wantzel (1814–1848) proved this could not be: cube duplication and angle trisection are in fact impossible, as is constructing a regular heptagon or squaring the circle. In the following we'll trace the steps leading up to Gauss's construction of the heptadecagon and Wantzel's proof of impossibility.

Descartes

The first important step towards proving certain constructions impossible was taken by René Descartes (1596–1650) in his *The Geometry* (1637). Descartes's key insight was that by identifying the lengths of line segments with real numbers, one could restate a geometric problem as an algebraic one, express the solution symbolically, then convert the algebraic expression into a geometric construction procedure.

In order to take this last step, we must develop an arithmetic of lines. Let AB and CD be two line segments (where we will assume CD is shorter than AB). Compass and straightedge techniques from the *Elements* allow us to find line segments that correspond to the sum $AB + CD$, difference $AB - CD$, and $q \cdot AB$ (for any positive rational q). The problem arises when trying to interpret the product $AB \cdot CD$. Euclid and others identified this product as the rectangle whose adjacent sides were equal in length to AB and CD. This would mean the arithmetic of line segments was not closed under multiplication; moreover, it would make the division of two line segments impossible to define.

Descartes realized that the theory of proportions could be used to identify the product of two line segments with another line segment, provided we had a line segment of unit length. Imagine two lines intersecting at B at any angle whatsoever, and say we wish to multiply BD by BC. Mark off BA equal to the unit, and join AC (see Figure 2). Draw DE parallel to AC. Then triangles BAC, BDE are similar, and we have the ratio $BE : BD = BC : BA$. This corresponds to the equality of the two products $BE \cdot BA = BC \cdot BD$. Since BA is equal to the unit, we can thus identify the line segment BE with the product $BC \cdot BD$. Thus the product of two line segments is another line segment. Division can be handled in virtually the same way.

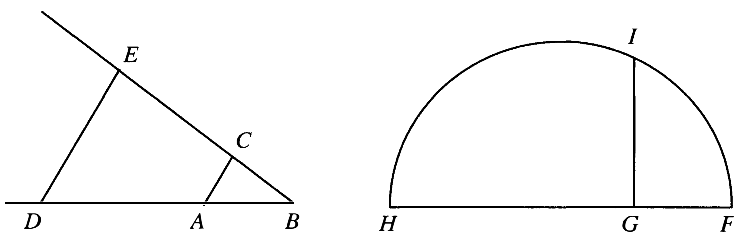


Figure 2 Multiplication and roots.

Proposition 14 of Book II of the *Elements* gives the construction technique for finding square roots (literally the side of a square equal in area to a given rectangle), which Descartes modified to extract square roots [**7**, p. 5]. Suppose we wish to find the square root of GH. Extend GH by GF equal to the unit, then draw the circle with FH as its diameter. The perpendicular GI will equal the square root of GH (see Figure 2).

Suppose we begin with a line segment AB (which we can take to be our unit). If we can construct a line of length $k \cdot AB$ using the above techniques, we say that k is a constructible number (and $k \cdot AB$ is a constructible line segment). In general, k is a constructible number if it is rational, or the root of a quadratic equation with constructible coefficients. A figure is constructible if all the line segments required for its construction are constructible. Moreover, given a constructible figure, any line segment we can obtain from it (e.g., the diagonal of a square) is constructible. For example, if we could square the circle, then $\sqrt{\pi}$ would be constructible; equivalently, if $\sqrt{\pi}$ is inconstructible, squaring the circle is impossible.

This identification of a geometric problem with an algebraic problem allows us to phrase the problem of constructibility in terms of the roots of a specific equation: if the root is a constructible number, the corresponding geometric problem can be solved using compass and straightedge alone. Duplicating the cube would allow us to find a line of length $\sqrt[3]{2}$, which is a root of the equation $x^3 - 2 = 0$. Constructing a regular n-gon would allow us to find a line of length $\sin \frac{2\pi}{n}$, which is the imaginary part of one of the roots of $x^n - 1 = 0$ (because of this, the problem of finding the roots of $x^n - 1 = 0$ is also known as the cyclotomy problem).

Trisection of an angle corresponds to a cubic equation as follows. Given a circle with center O and unit radius, with central angle AOC equal to 3θ. We wish to find point B on the circle where angle BOC is equal to θ. If we drop AD and BE perpendicular to OC, we have $AD = \sin 3\theta$, $BE = \sin\theta$. These quantities are related through the identity:

$$\sin 3\theta = 3\sin\theta - 4\sin^3\theta$$

Since angle AOC is the given angle, then $\sin 3\theta$ is a known quantity which we can designate as l. Thus if the real roots of $l = 3x - 4x^3$ are not constructible, trisection of the corresponding angle is impossible.

Vandermonde and Lagrange

The next step towards answering the constructibility problem came from the work of Alexandre-Théophile Vandermonde (1735–1796) and Joseph Louis Lagrange (1736–1813). Vandermonde's [**8**], presented to the Paris Academy in 1770, and Lagrange's [**6**], presented to the Berlin Academy in 1771, examined why general solutions to equations of degree 3 and 4 existed. Both came to the same conclusion independently: Our ability to solve these equations is due to the fact that we can find the value of certain expressions of the roots without knowing the roots themselves.

To understand their methods, consider the quadratic equation $x^2 - px + q = 0$, with roots $x = a$ and $x = b$. Thus $p = a + b$ and $q = ab$. Next, take any function of the roots of this equation. Some functions, such as $f(r_1, r_2) = r_1 + r_2$, have the same value regardless of which root we regard as r_1 and which root we regard as r_2; these are called symmetric functions. It was widely believed (though not proven until the middle of the nineteenth century) that every symmetric function of the roots of a polynomial could be expressed as a rational function of the coefficients. In this case, $f(a, b) = f(b, a) = p$.

On the other hand, consider a function like $g(r_1, r_2) = r_1 - r_2$. Depending on which root we call r_1 and which root we call r_2, g might take on one of two possible values, $a - b$ or $b - a$. In order to find the values of this non-symmetric function of the roots, Lagrange let the k distinct values be the roots of a kth degree equation. In our example, the two values of g would be the roots of:

$$(y - (a - b))(y - (b - a)) = y^2 - (a - b)^2$$

A little algebra shows us that $(a - b)^2 = (a + b)^2 - 4ab$. Since we know the values of $a + b$ and ab, we can determine, even without knowing the roots, that

$$(a - b)^2 = p^2 - 4q.$$

Thus the two different values of $a - b$ will be the two roots of $y^2 - (p^2 - 4q) = 0$. Hence $a - b = \sqrt{p^2 - 4q}$ or $-\sqrt{p^2 - 4q}$. It will make no difference which we choose; for example, we might let $a - b = \sqrt{p^2 - 4q}$. To solve for a and b separately, we need a second equation, which we can obtain from the coefficients: $a + b = p$. These two equations give us the system:

$$a + b = p \qquad a - b = \sqrt{p^2 - 4q}$$

Hence $a = \frac{p+\sqrt{p^2-4q}}{2}$, and $b = \frac{p-\sqrt{p^2-4q}}{2}$.

Both Vandermonde and Lagrange considered the problem of finding the nth roots of unity, which would be the roots of $x^n - 1 = 0$. Lagrange noted the correspondence between the roots of $x^n - 1 = 0$ and the cyclotomy problem; further, he observed that if n is prime, all of the roots can be generated by the successive powers of any root except $x = 1$. This allowed him to write equations relating the roots; solving the equations would give the roots of unity. Lagrange used his method to find the roots of unity for $n = 3$ through $n = 6$ (all of which can be found using only square roots), while Vandermonde found the roots of unity up to $n = 11$ using similar methods.

Gauss

According to legend, Carl Friedrich Gauss (1777–1855) discovered the constructibility of the regular heptadecagon in 1796; this inspired him to choose mathematics as his future field of study, despite the indifferent reception of his discovery by A. G. Kästner at Göttingen. Gauss's main contribution to the problem of cyclotomy was inventing a method of splitting the roots of unity into sets where the sum of the roots in each set was the root of an equation with determinable coefficients. He described his method in [**2**], where the solution to the cyclotomy problem appeared as one application of the theory of quadratic residues. While Gauss's discovery was unprecedented, it was a straightforward, albeit clever, application of the ideas of Lagrange and Vandermonde.

The nth roots of unity are solutions to the equation $x^n - 1 = 0$. Obviously any root r must satisfy $r^n = 1$. If n is the least power of r that is equal to 1, then r is said to be a primitive nth root of unity. For example, the roots of $x^4 - 1 = 0$ are ± 1, $\pm i$. Since $1^1 = 1$ and $(-1)^2 = 1$, then neither 1 nor -1 is primitive. On the other hand, the least power of i or $-i$ that gives 1 is the fourth power; thus i and $-i$ are primitive roots, and their powers will generate all the roots; for example:

$$i, \quad i^2 = -1, \quad i^3 = -i, \quad i^4 = 1$$

In general, as Lagrange noted, if n is prime, then there are $n - 1$ primitive roots of unity.

As we note above, constructibility of the regular n-gon corresponds to constructibility of the roots of $x^n - 1 = 0$. We'll illustrate Gauss's general method by finding the 5th roots of unity. These would be solutions to the equation $x^5 - 1 = 0$. There is one (non-primitive) root $x = 1$. Removing a factor of $x - 1$ we obtain the equation

$$x^4 + x^3 + x^2 + x + 1 = 0$$

which is called the cyclotomic equation. All primitive fifth roots r must satisfy this equation.

Gauss considered a sequence whose first term is a primitive root, and where each term is some (constant) power of the previous term. For example, if we take r and cube it repeatedly, we obtain:

$$r, r^3, r^9, r^{27}, r^{81} \ldots$$

Since r is a root of $x^5 - 1 = 0$, then $r^5 = 1$. Hence the above sequence simplifies to r, r^3, r^4, r^2, $r, \ldots$, and all the roots appear in this sequence. On the other hand, suppose we take r and repeatedly raise it to the fourth power, obtaining the sequence:

$$r, r^4, r^{16}, r^{64}, \ldots$$

In this case, the only distinct members of the sequence are r and r^4.

Note that the remaining roots, r^2 and r^3, are the squares of the two distinct terms of this last sequence: $(r)^2 = r^2$, and $(r^4)^2 = r^8 = r^3$. More generally, suppose n is prime and r is a primitive nth root of unity. Gauss showed that our sequence of powers will have k distinct elements, where k is a divisor of $n - 1$. Moreover the remaining roots (if k is not equal to $n - 1$) can be separated into sets of k distinct elements, each of which is a power of a root of the original set.

For example, consider the $n = 7$ case, and a primitive root p. The sequence

$$p, p^6, p^{36}, p^{216}, \ldots$$

contains only two distinct roots, p and p^6. The squares of these are p^2, $p^{12} = p^5$, and the cubes are p^3, p^4. Thus the six roots have been partitioned into three sets, $\{p, p^6\}$, $\{p^2, p^5\}$, and $\{p^3, p^4\}$.

Note that the decomposition is not unique; for example, the sequence

$$p, p^2, p^4, p^8, \ldots$$

contains three distinct roots, p, p^2, and p^4; the remaining roots are the cubes of these roots and the six roots will be partitioned into two sets, $\{p, p^2, p^4\}$, and $\{p^3, p^5, p^6\}$.

Returning to the $n = 5$ case, we have split the roots into two sets: $\{r, r^4\}$ and $\{r^2, r^3\}$. Gauss then considered the sum of the roots in each set (designating these sums as "periods"), and let the sums be the roots of an equation:

$$\begin{aligned}(y - (r + r^4))(y - (r^2 + r^3)) &= y^2 - (r^4 + r^3 + r^2 + r)y + (r + r^4)(r^2 + r^3) \\ &= y^2 - (r^4 + r^3 + r^2 + r)y + (r^3 + r^4 + r^6 + r^7) \\ &= y^2 - (r^4 + r^3 + r^2 + r)y + (r^4 + r^3 + r^2 + r) \\ &= y^2 + y - 1\end{aligned}$$

where we made use of the fact that r satisfies the equation $x^4 + x^3 + x^2 + x + 1 = 0$. Hence the two periods $r + r^4$ and $r^2 + r^3$ correspond to the two roots of the quadratic equation $y^2 + y - 1 = 0$. We find the roots are $y = \frac{-1 \pm \sqrt{5}}{2}$.

One of these roots corresponds to $r + r^4$, and the other corresponds to $r^2 + r^3$. In principle it makes no difference which we assign to $r + r^4$, though in practice it is convenient if r is the principal fifth root of unity $\cos \frac{2\pi}{5} + i \sin \frac{2\pi}{5}$. Gauss noted that we could find this root numerically and see which of the two roots of $y^2 + y - 1 = 0$ was equal to $r + r^4$. Alternatively, we might note that $r + r^4$ will have a positive real component; hence $r + r^4 = \frac{-1 + \sqrt{5}}{2}$.

To find r, we can construct a quadratic equation with r and r^4 as roots:

$$(z - r)(z - r^4) = z^2 - (r + r^4)z + r^5 = z^2 - \left(\frac{-1 + \sqrt{5}}{2}\right)z + 1$$

Note that the coefficients of this equation are constructible numbers; hence its roots will also be constructible. These roots are:

$$z = \frac{\left(\frac{-1+\sqrt{5}}{2}\right) \pm \sqrt{\left(\frac{-1+\sqrt{5}}{2}\right)^2 - 4}}{2}$$

One of these will be the principal fifth root of unity, and the other will be its fourth power. Since the principal fifth root of unity is equal to $\cos \frac{2\pi}{5} + i \sin \frac{2\pi}{5}$, we can, as Gauss suggested, approximate the sine and cosine values and determine which of the two roots corresponds to the principal root (which will tell us, among other things, $\cos \frac{2\pi}{5} = \frac{-1+\sqrt{5}}{4}$ and $\sin \frac{2\pi}{5} = \frac{1}{4}\sqrt{10 + 2\sqrt{5}}$). Since the $\sin \frac{2\pi}{5}$ is constructible, so is the regular pentagon, a fact known to the ancients: Euclid's construction appears as Proposition 11 of Book IV (though Ptolemy gave a much easier construction in the *Almagest*).

On the other hand, consider the regular heptagon. In the above, we found that the roots can be separated into two periods, $p + p^2 + p^4$ and $p^3 + p^5 + p^6$. Letting the sum of the roots in the set be the roots of a quadratic equation and reducing as before we obtain:

$$(y - (p + p^2 + p^4))(y - (p^3 + p^6 + p^5)) = y^2 + y + 2$$

with roots $y = \frac{-1\pm\sqrt{-7}}{2}$; if p is the principal root, then $p + p^2 + p^4$ has a positive imaginary component so $\frac{-1+\sqrt{-7}}{2} = p + p^2 + p^4$ and $\frac{-1-\sqrt{-7}}{2} = p^3 + p^5 + p^6$.

The next step would be letting p, p^2, and p^4 be the three roots of a cubic equation:

$$\begin{aligned}(z - p)(z - p^2)(z - p^4) &= z^3 - (p + p^2 + p^4)z^2 + (p^3 + p^6 + p^5)z - p^7 \\ &= z^3 - \left(\frac{-1 + \sqrt{-7}}{2}\right) z^2 + \left(\frac{-1 - \sqrt{-7}}{2}\right) z - 1\end{aligned}$$

While we can solve the cubic equation, we cannot do so by means of basic arithmetic operations and square roots alone; we must extract a cube root. Hence it would appear that the primitive seventh root of unity (and consequently the regular heptagon) is inconstructible.

The preceding example suggests the following: Suppose we wish to construct a regular n-gon, where n is prime. If $n - 1$ has any prime factors other than 2, then at some point in separating the roots, we will have to solve an equation of a degree higher than 2. Hence constructing a regular n-gon using this method requires $n = 2^k + 1$.

We can go a little further. If k has any odd factors, then $2^k + 1$ is composite; this follows because if $k = pq$ and q is odd, $x^{pq} + 1$ has a factor of $x^p + 1$. Thus a regular n-gon, where n is prime, *might* be constructible if n is a so-called Fermat prime, with $F_m = 2^{2^m} + 1$. The known Fermat primes are 3, 5, 17, 257, 65537; it is unknown if higher Fermat primes exist.

In any case, consider $n = 17$. The corresponding cyclotomic equation has 16 roots. Gauss split these into two sets of eight roots apiece; hence, a quadratic equation could be used to find the sum of eight of the roots. Each set of eight could in turn be split into two sets of four; again, a quadratic equation could be used to find the sum of four of the roots. Each set of four could be split into two sets of two, and the sum of two of the roots could be found. Finally the sets of two could be broken down into their individual roots, so a primitive 17th root of unity could be found. Since none of the equations has a degree higher than 2, the roots are constructible; hence a regular heptadecagon can be constructed using only compass and straightedge.

Wantzel

Gauss's method suggests but does not prove the constructibility of the 257- and 65,537-gons (we need Sylow's Theorem to guarantee constructibility); likewise, it suggests but does not prove the impossibility of constructing a regular heptagon.

The first proof of the impossibility of certain geometric constructions came from Pierre Wantzel (1814–1848) in **[9]** (1837). Wantzel began by considering a system of quadratic equations (which for brevity we will call a Wantzel System):

$$\begin{aligned} x_1^2 + Ax_1 + B &= 0 \\ x_2^2 + A_1x_2 + B_1 &= 0 \\ x_3^2 + A_2x_2 + B_2 &= 0 \\ &\vdots \\ x_n^2 + A_{n-1}x_2 + B_{n-1} &= 0 \end{aligned}$$

where A, B are rational functions of some given quantities; A_1, B_1 are rational functions of the given quantities together with x_1 (and hence the coefficients of the second equation are constructible numbers); A_2, B_2 are rational functions of the given quantities, together with x_1, x_2, and in general A_m, B_m are rational functions of the given quantities and the variables $x_1, x_2, \ldots, x_m$. Note that Gauss's method of showing the constructibility of a pentagon or heptadecagon made use of precisely such a system; in the case of the pentagon the Wantzel system is:

$$y^2 + y - 1 = 0$$
$$z^2 - yz + 1 = 0$$

More generally, every constructible number r corresponds to some Wantzel system.

Consider any of these equations $x_{m+1}^2 + A_m x_{m+1} + B_m = 0$. Remarkably, the rational functions A_m, B_m can always be reduced to a *linear* function of the form $A'_{m-1}x_m + B'_{m-1}$, where A'_{m-1} and B'_{m-1} are rational functions of the given quantities and the variables $x_1, x_2, \ldots, x_{m-1}$. This reduction can be performed in two steps. First, the preceding equation $x_m^2 + A_{m-1}x_m + B_{m-1} = 0$ can be used to eliminate the higher powers of x_m in the expression for A_m and B_m, reducing them to the form $\frac{C_m x_m + D_m}{E_m x_m + F_m}$. Then the numerator and denominator can be multiplied by a constant quantity to reduce the rational function to a linear one.

For example, suppose we have the system of equations:

$$x^2 - 5x + 2 = 0$$
$$y^2 + \left(\frac{x^3 + 3x + 1}{2x - 1}\right) y + \left(\frac{1}{x^2 + 7x + 5}\right) = 0$$

From the first equation we have $x^2 = 5x - 2$. Hence $x^3 = 5x^2 - 2x = 23x - 10$. Thus the second equation can be reduced to:

$$y^2 + \left(\frac{26x - 9}{2x - 1}\right) y + \left(\frac{1}{12x + 3}\right) = 0$$

How can we eliminate the rational functions? Consider the first rational function. Suppose we multiply numerator and denominator by some constant C so that

$$C(26x - 9) = (2x - 1)(\alpha x + \beta)$$

for some values of α, β; then the common factor of $2x - 1$ can be removed and the rational expression simplified to a linear one. Expanding gives us:

$$26Cx - 9C = 2\alpha x^2 + (2\beta - \alpha)x - \beta$$

We can make use of the substitution $x^2 = 5x - 2$ to eliminate the square term:

$$26Cx - 9C = (2\beta + 9\alpha)x - (\beta + 4\alpha)$$

Comparing coefficients gives us a system of 2 linear equations in 3 unknowns:

$$26C = 2\beta + 9\alpha \qquad 9C = \beta + 4\alpha$$

Because this system is underdetermined, we may express two of the variables in terms of the third. For example, one solution is $\alpha = 2$, $\beta = -23/4$, and $C = 1/4$; in other words $\frac{1}{4}(26x - 9) = (2x - 1)(2x - \frac{23}{4})$. Thus:

$$\frac{26x-9}{2x-1}=\frac{\frac{1}{4}}{\frac{1}{4}}\frac{26x-9}{2x-1}=\frac{(2x-1)(2x-23/4)}{\frac{1}{4}(2x-1)}=8x-23$$

In this way the final equation $x_n^2 + A_{n-1}x_n + B_{n-1} = 0$ can be converted into an equation where the coefficients A_{n-1} and B_{n-1} are linear functions of x_{n-1}.

Next, consider that x_{n-1} is one of the solutions to a quadratic equation. If we allow x_{n-1} to take on its two possible values, we obtain two different expressions for A_{n-1} and B_{n-1}, and consequently two different equations quadratic in x_n. Multiplying these two equations together will give us a fourth degree equation in x_n whose coefficients are functions of the given quantities and the variables $x_1, x_2, \ldots, x_{n-2}$. As before we can reduce these coefficients to linear functions of x_{n-2}; letting x_{n-2} take on its two possible values and multiplying the corresponding expressions will give us an eighth degree equation in x_n whose coefficients can be reduced to linear functions of x_{n-3}. Eventually we will end with an equation in x_n of degree 2^n whose coefficients are rational functions of the given quantities. This leads us to a preliminary theorem:

THEOREM. *Any Wantzel system of n equations corresponds to an equation of degree 2^n whose coefficients are rational functions of the given quantities; consequently, any constructible number is a root of an equation of degree 2^n whose coefficients are rational functions of the given quantities.*

For example, the Wantzel system corresponding to the construction of the pentagon was:

$$y^2 + y - 1 = 0$$
$$z^2 - yz + 1 = 0$$

Let the two roots of the first equation be $y = a$ and $y = b$. In the above we found these two roots, and used them to form a quadratic equation in z to find the principal fifth root of unity.

On the other hand, we can also write a single expression (which we will call the Wantzel polynomial) which contains all the roots. In this case, we can substitute the two roots $y = a$ and $y = b$ in to the left hand side of the second equation, then multiply the two expressions to obtain:

$$(z^2 - az + 1)(z^2 - bz + 1) = z^4 - (a+b)z^3 + (2+ab)z^2 - (a+b)z + 1$$

Since a and b are the two roots of $y^2 + y - 1$, we have $a + b = -1$ and $ab = -1$. Thus the equation

$$z^4 + z^3 + z^2 + z + 1 = 0$$

contains all solutions to the Wantzel system.

Next, suppose $x_n = r$ is a root of the Wantzel polynomial corresponding to a Wantzel system of n equations; further suppose that no Wantzel system of fewer than n equations exists with $x_n = r$ as a root. Wantzel then proved that no variable x_k could be expressed as a rational function of $x_1, x_2, \ldots, x_{k-1}$; equivalently, the quadratic equations are irreducible. This is because if one of the equations can be factored, then the preceding equation can be eliminated and we would obtain two Wantzel system of $n - 1$ equations, which would contain all the roots of the original system (and in particular r could be found by a Wantzel system of $n - 1$ equations). For example, consider the system:

$$x^2 - 3x - 7 = 0$$
$$y^2 - (4x - 1)y + 8x = 0$$
$$z^2 - (4y)z + (4y^2 - 1) = 0$$

and let $z = r$ be one of the roots. Note that the last equation factors, so we may write two separate Wantzel systems where the third equation differs, namely

$$x^2 - 3x - 7 = 0$$
$$y^2 - (4x - 1)y + 8x = 0$$
$$z - (2y + 1) = 0$$

and

$$x^2 - 3x - 7 = 0$$
$$y^2 - (4x - 1)y + 8x = 0$$
$$z - (2y - 1) = 0$$

where z can be expressed as a rational function of the preceding variables.

Consider the first system. Let the roots $y^2 - (4x - 1)y + 8x = 0$ be $y = a$ and $y = b$; letting y take on these two values in the third equation and multiplying the factors gives us the expression

$$(z - (2a + 1))\,(z - (2b + 1) = z^2 - (2a + 2a + 2)z + (4ab + 2a + 2b + 1)$$

But if the roots of $y^2 - (4x - 1)y + 8x = 0$ are $y = a$ and $y = b$, then $a + b = 4x - 1$, and $ab = 8x$; hence the second and third equation can be combined to form the single equation $z^2 - 8xz + (40x - 1) = 0$. Thus in place of the three equations, we have two equations:

$$x^2 - 3x - 7 = 0$$
$$z^2 - 8xz + (40x - 1) = 0$$

The reader can verify that the second Wantzel system would have $z^2 - (8x - 4)z + (24x - 4) = 0$ as its second equation. Thus in place a Wantzel system containing n equations, we would have two systems containing $n - 1$ equations, which between them contain all the roots z of the original system; hence $z = r$ would be the root of a Wantzel system containing $n - 1$ equations, which contradicts our original assumption.

Note that any solution x_n of the Wantzel polynomial $f(x)$ is a solution of $x_n^2 + A_{n-1}x_n + B_{n-1} = 0$, where A_{n-1}, B_{n-1} are found by substituting some set of solutions $\{x_1, x_2, \ldots, x_{n-1}\}$ to the equations of the Wantzel system. For example, the primitive fifth root of unity $z = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}$ of $z^4 + z^3 + z^2 + z + 1 = 0$ corresponds to a root of $z^2 - yz + 1 = 0$ where y is a solution to $y^2 + y - 1 = 0$.

Wantzel used this idea to prove that if another polynomial $F(x)$ had any root $x_n = a$ in common with $f(x)$, then it must have all roots in common; hence $f(x)$ is irreducible. Let $x_n = a$ be the root corresponding to the set $\{x_1, x_2, \ldots, x_{n-1}\}$, and let $F(x)$ be a polynomial with rational coefficients with $F(a) = 0$. As before we can reduce $F(x)$ to an expression of the form $A'_{n-1}x_n + B'_{n-1}$, where A'_{n-1}, B'_{n-1} are functions of the given quantities and the variables $x_1, x_2, \ldots, x_{n-1}$. Moreover, A'_{n-1} and B'_{n-1} must be equal to zero (since if they were not, x_n could be expressed as a rational function of $x_1, x_2, \ldots, x_{n-1}$); hence we have $A'_{n-1} = 0$ (and likewise, $B'_{n-1} = 0$). But A'_{n-1} can be reduced as before to a linear function of x'_{n-1}. Thus the equation $A'_{n-1} = 0$ gives us an equation of the form $A'_{n-2}x_{n-1} + B'_{n-2} = 0$, where A'_{n-2} and B'_{n-2} are functions of the given quantities and the variables $x_1, x_2, \ldots, x_{n-2}$.

As before A'_{n-2} and B'_{n-2} must be both equal to zero; from $A'_{n-2} = 0$ we can obtain an equation of the form $A'_{n-3}x_{n-2} + B'_{n-3} = 0$. Continuing in this fashion we will eventually arrive at an equation of the form $A'x_1 + B' = 0$, where A' and B' are functions of the given quantities only. Again, x_1 cannot be a rational function of the given quantities only, so A' and B' must both equal zero; since they contain no variables, they are identically zero. Thus the two roots of $x_1^2 + Ax_1 + B = 0$ satisfy $A'x_1 + B' = 0$.

Now consider the equation $A'_1x_2 + B'_1 = 0$. A'_1 and B'_1 have both been reduced to linear functions of x_1 that will equal zero for any value of x_1 that satisfies $x_1^2 + Ax_1 + B = 0$. Thus the two possible values of x_1 will make both A'_1 and B'_1 equal to zero; consequently the four possible values of x_2 will make $A'_1x_2 + B'_1 = 0$. In a like manner, the eight possible values of $x_3^2 + A_2x_3 + B_1 = 0$ will satisfy the equation $A'_2x_3 + B'_2 = 0$, and so on, and so the 2^n possible roots of $x_n^2 + A_{n-1}x_n + B_{n-1} = 0$ will satisfy $F(x)$. Hence if $F(x)$ shares any root with $f(x)$, it will share all the roots of $f(x)$.

For example, consider our system

$$y^2 + y - 1 = 0$$
$$z^2 - yz + 1 = 0$$

which corresponded to the single equation $z^4 + z^3 + z^2 + z + 1 = 0$. Let $z = z_1$ be the root corresponding to one of the roots $y = y_1$ of the first equation, and suppose there was another polynomial $F(z)$ with rational coefficients that also had $z = z_1$ as a root.

First, we can eliminate the higher powers of z in $F(z)$ by the equation $z^2 - yz + 1 = 0$. This allows us to write $F(z)$ as a polynomial in y and z of the form

$$(z^2 - yz + 1)f(y, z) + A_1z + B_1,$$

where A_1 and B_1 are functions of y and $f(y, z)$ is some polynomial in y and z. Since $z = z_1$ satisfies (by assumption) the equation $z^2 - yz + 1 = 0$ when $y = y_1$, then substituting in these values gives us $A_1z_1 + B_1$, which (since z_1 is a root of F) must equal zero. Since the system is minimal, z cannot be expressed as a rational function of y, so A_1 and B_1 must both equal zero when $y = y_1$.

Next take (for example) the expression A_1, which we can write as

$$(y^2 + y - 1)g(y) + A'y + B',$$

where A', B' are rational functions of the given quantities only. Since $y = y_1$ satisfies $y^2 + y - 1 = 0$, and (by the above) satisfies $A_1 = 0$, then $A'y_1 + B' = 0$. But y_1 (by assumption) cannot be written as a rational function of the given quantities; hence A' and B' are simultaneously equal to zero. Since they contain no variable terms at all, then A' and B' must be identically zero and $A_1 = (y^2 + y - 1)g(y)$. Hence any solution to $y^2 + y - 1 = 0$ will make $A_1 = 0$. The same reasoning applies to B_1.

Since $F(z)$ can be written as $(z^2 - yz + 1)f(y, z) + A_1z + B_1$, and $A_1 = 0$, $B_1 = 0$ when y is equal to either root of $y^2 - y + 1 = 0$, then any of the four roots of $z^4 + z^3 + z^2 + z + 1 = 0$ will satisfy $F(z) = 0$. Hence $F(z)$ must contain all the roots.

At last this gives us a necessary condition for constructibility:

WANTZEL'S THEOREM. *If r is a constructible number, it must be the root of an irreducible polynomial of degree 2^n.*

Equivalently, let r be the root of an irreducible polynomial $f(x)$. If the degree of f is not equal to 2^n, then r is not constructible. This proves the impossibility of duplicating the cube or trisecting an arbitrary angle. In the first case, $\sqrt[3]{2}$ is the root of $x^3 - 2 = 0$, which is irreducible but not of degree 2^n; the same reasoning proves that arbitrary nth roots cannot be found, unless n is a power of 2. Likewise trisecting an arbitrary angle requires finding a root of $l = 3x - 4x^3$, which will in general be irreducible and not of degree 2^n.

What about the cyclotomy problem? If n is prime, the corresponding cyclotomic equation is irreducible, but if n is not a Fermat prime, then the degree of this equation is not a power of 2 and so the regular n-gon will not be constructible. Thus it is impossible to construct regular polygons of 7, 11, 13, etc. sides using only compass and straightedge.

Wantzel's Theorem alone is insufficient to prove the impossibility of squaring the circle, though it does lay the groundwork for a proof. If $\sqrt{\pi}$ is a constructible number, it must be the root of an irreducible equation of degree 2^n. In 1882 Ferdinand Lindemann (1852–1939) proved that π is transcendental: hence no equation of any degree with rational coefficients can have π as a root. Consequently squaring the circle is impossible.

REFERENCES

1. A. Aaboe, *Episodes from the Early History of Mathematics*, MAA New Mathematical Library, 1964.
2. C. F. Gauss, *Disquisitiones Arithmeticae*, Yale University Press, 1966. Translated by A. A. Clarke.
3. T. L. Heath, *A History of Greek Mathematics*, Dover, 1981.
4. T. L. Heath, *The Thirteen Books of the Elements*, Dover, 1956.
5. J. L. Lagrange, *Œuvres*, Gauthiers-Villars, Paris, 1869.
6. ———, Réflexions sur la résolution algébrique des équations, *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres* (1770) 134–215; (1771) 138–253.
7. D. E. Smith, *The Geometry of Rene Descartes*, Dover, 1954.
8. A. T. Vandermonde, Mémoire sur la Résolution des Équations, *Histoire de l'Académie Royale des Sciences*, 1771, 365–416.
9. P. Wantzel, Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas. *Journal de Mathématiques Pures et Appliquées*, 1837, 366–372.

Polynomial Root Squeezing

MATTHEW BOELKINS
Grand Valley State University
Allendale, MI 49401
boelkinm@gvsu.edu

JUSTIN FROM
Central College
Pella, IA 50219

SAMUEL KOLINS
Cornell University
Ithaca, NY 14853
srk25@cornell.edu

Given a real polynomial with all its roots real, the Polynomial Root Dragging Theorem [**1**], [**2**] states that if one or more roots of the polynomial are moved to the right, then *all* of the critical numbers also move to the right (or possibly stay fixed, if a root is repeated) with none of the critical numbers moving as much as the root that is moved most. But what happens if some of the roots of the polynomial are dragged in opposing directions, either toward or away from each other?

Anderson's proof of the Root Dragging Theorem in [**2**] can be modified to show that for odd and even polynomials, if we drag some subset of the positive roots toward the origin, and simultaneously drag the corresponding negative roots toward the origin so that the updated polynomial remains odd or even, then all of the function's critical numbers move toward the origin (or stay fixed). For a polynomial that is not necessarily odd or even, the situation is more complicated. As shown in the example in Figure 1, it appears that when two consecutive roots are dragged toward each other (as indicated by the arrows), the critical numbers to the left of the first root move right, while the critical numbers to the right of the second root move left. In particular, for the critical numbers c_i of the original polynomial p and the critical numbers d_i of the updated polynomial q, we have that $c_1 < d_1$ and $c_2 < d_2$, while $c_4 > d_4$. It turns out that our observations in this example hold in general. In what follows we prove the Polynomial Root Squeezing Theorem, which shows that dragging certain pairs of roots toward

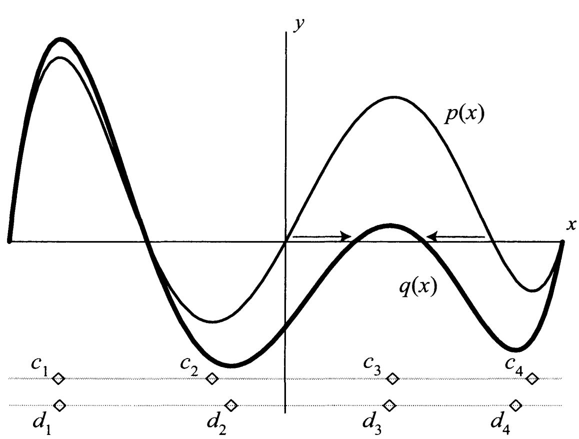


Figure 1 Two roots of the polynomial p have been squeezed toward each other to create the polynomial q.

one another in a uniform way causes the outer critical points of the function to move toward each other (or stay fixed, if a relevant root is repeated).

The *span* of a polynomial with all real zeros is the difference between its greatest and least roots. The Root Squeezing Theorem also provides insight into how the respective spans of a polynomial's derivatives depend on the location of the function's interior roots. We present an elementary argument for the fact [**4**] that for any degree n polynomial p with all real zeros and span $2b$, Span $p^{(k)} \geq$ Span $q^{(k)}$, where $q(x) = x^{n-2}(x^2 - b^2)$. Moreover, q is the unique monic polynomial with roots at $\pm b$ for which this minimum is attained.

For notational simplicity, we let $\mathbf{P}_{n,b}$ denote the set of all monic, degree n polynomials with all real roots $r_1 \leq r_2 \leq \cdots \leq r_n$ such that $r_1 = -b$ and $r_n = b$, where $b > 0$. By translations and scalings, results that hold for polynomials in $\mathbf{P}_{n,b}$ can be restated to apply for any polynomial whose zeros are all real. The critical numbers of a function $p \in \mathbf{P}_{n,b}$ will be denoted by $c_1 \leq c_2 \leq \cdots \leq c_{n-1}$, and we call r_i an *interior root* of p if $i \neq 1$ and $i \neq n$.

The root squeezing theorem

G. Peyser proved several results [**6**] that are very similar to the Polynomial Root Dragging Theorem. For example, removing the rightmost zero of a polynomial with all real zeros shifts the critical points of the resulting polynomial to the right. In addition, dragging the leftmost root of a polynomial to the right shifts the rightmost critical number to the right in the updated function. Our proof of the Root Squeezing Theorem uses analysis similar to Peyser's to demonstrate the impact of dragging two roots toward each other on certain critical points.

THEOREM 1. (THE POLYNOMIAL ROOT SQUEEZING THEOREM) *Let $p \in \mathbf{P}_{n,b}$ and say that p has critical numbers $c_1 \leq \cdots \leq c_{n-1}$. Let $r_j < r_k$ be any two interior roots of p and $d \in \mathbb{R}^+$ be such that*

$$d \leq \min\left\{ r_{j+1} - r_j, r_k - r_{k-1}, \frac{1}{2}(r_k - r_j) \right\}.$$

Let $\tilde{p}$ be the polynomial that results from squeezing r_j and r_k together by a distance $2d$. That is,

$$\tilde{p}(x) = (x - r_j - d)(x - r_k + d) \prod_{i \neq j,k} (x - r_i)$$

Denote the critical points of $\tilde{p}$ by $\tilde{c}_1 \leq \tilde{c}_2 \leq \ldots \leq \tilde{c}_{n-1}$. Then for $1 \leq i < j$ we have $\tilde{c}_i \geq c_i$, and for $k \leq i \leq n-1$ we have $\tilde{c}_i \leq c_i$.

Proof. Let $p \in \mathbf{P}_{n,b}$ and choose interior roots r_j and r_k so that $r_j < r_k$. Let c_i be a critical number of p such that $1 \leq i < j$ or $k \leq i \leq n-1$. We begin by considering the case where c_i lies at a repeated root of $p(x)$. If $r_i = c_i = r_{i+1}$ and neither r_i nor r_{i+1} are being shifted, then we have $\tilde{c}_i = r_i = r_{i+1} = c_i$. This is the only situation where c_i does not move in response to a pair of roots being squeezed. If $r_i = c_i = r_{i+1}$ and we shift r_{i+1} to the right, then by Rolle's Theorem $\tilde{c}_i > r_i = c_i$ as desired. Likewise, shifting r_i to the left shows $\tilde{c}_i < r_{i+1} = c_i$.

Therefore, the only remaining case is where $r_i < c_i < r_{i+1}$. Our goal is to compare c_i and $\tilde{c}_i$; to do so, we investigate the behavior of $\tilde{p}'$ at c_i and $\tilde{c}_i$. Define the function $q(x)$ so that $p(x) = (x - r_j)(x - r_k)q(x)$. Differentiating this expression gives

$$p'(x) = (x - r_j + x - r_k)q(x) + (x - r_j)(x - r_k)q'(x). \tag{1}$$

Similarly, we can differentiate $\tilde{p}(x) = (x - r_j - d)(x - r_k + d)q(x)$ to obtain

$$\tilde{p}'(x) = (x - r_j + x - r_k)q(x) + (x - r_j - d)(x - r_k + d)q'(x). \qquad (2)$$

Subtracting (1) from (2), simplifying, and evaluating the resulting expression at $x = c_i$ gives

$$\tilde{p}'(c_i) = d(r_k - r_j - d)q'(c_i). \qquad (3)$$

Since we assume that $(r_k - r_j) \geq 2d > d > 0$, the expression $d(r_k - r_j - d)$ must be positive. Therefore, (3) implies that $\tilde{p}'(c_i)$ and $q'(c_i)$ must have the same sign.

Now consider the case of a critical number that lies to the left of the two roots that were squeezed together. In particular, $1 \leq i < j$, so $c_i < r_j$. We examine the behavior of p and $\tilde{p}$ on the interval $r_i < x < r_{i+1}$. Obviously $p(x)$ is strictly positive or strictly negative on $r_i < x < r_{i+1}$. We will assume $p(x) < 0$ in what follows; the argument is similar if $p(x) > 0$.

Having assumed that $p(x) < 0$ over the interval $r_i < x < r_{i+1}$, we note specifically that $p(c_i) < 0$. Since $p(c_i) = (c_i - r_j)(c_i - r_k)q(c_i)$, we can also conclude that $q(c_i) < 0$. Using the fact that $p'(c_i) = 0$, a careful analysis of the signs of terms in (1) evaluated at $x = c_i$ reveals that $q'(c_i) < 0$. Since (3) implies that $\tilde{p}'(c_i)$ and $q'(c_i)$ have the same sign, we have $\tilde{p}'(c_i) < 0$.

Again using the assumption that $p(c_i) < 0$, a sign analysis of terms in the equation

$$p(x)(x - r_j - d)(x - r_k + d) = \tilde{p}(x)(x - r_j)(x - r_k)$$

evaluated at c_i implies that $\tilde{p}(c_i) < 0$. Letting $\tilde{r}_i$ denote the ith root of the updated polynomial $\tilde{p}$, we note that $r_i < c_i < r_{i+1}$ and $\tilde{r}_i = r_i < x < r_{i+1} \leq \tilde{r}_{i+1}$, so the fact that $\tilde{p}(c_i) < 0$ implies that $\tilde{p}(x) < 0$ on the entire interval $(\tilde{r}_i, \tilde{r}_{i+1})$. Therefore, the sign of $\tilde{p}'(x)$ must change once from negative to positive on this interval with the change occurring at the critical point $\tilde{c}_i$. Since we established earlier that $\tilde{p}'(c_i) < 0$, it follows that $c_i < \tilde{c}_i$, as desired.

A similar argument shows that $\tilde{c}_i \leq c_i$ when $k \leq i \leq n - 1$. ■

The Root Squeezing Theorem tells us that when we squeeze two interior roots toward each other, unless $r_1 = r_2 = \tilde{r}_2$ or $r_n = r_{n-1} = \tilde{r}_{n-1}$, the extreme left critical point $\tilde{c}_1$ lies to the right of the critical number c_1 in the original polynomial, and the extreme right critical point $\tilde{c}_{n-1}$ lies to the left of c_{n-1}. This will be important when we next consider the spans of a polynomial's derivatives. In addition, the Root Squeezing Theorem provides a measure of certain critical points' sensitivity to the particular root(s) being moved. For example, if we shift root r_j to the right d units (assuming d meets the conditions stated in Theorem 1), the Root Dragging Theorem tells us that *all* critical numbers in the updated function move to the right, with each moving less than d units. If we now choose some $r_k > r_j$ and move r_k to the *left* d units, the Root Dragging Theorem again implies that *all* of the critical numbers in the updated function have moved to the left. If instead of doing two sequential moves we apply the Root Squeezing Theorem and perform the two shifts simultaneously, we see that the critical numbers to the left of r_j have moved right, while the critical numbers to the right of r_k have moved left. This indicates that root r_j has more "pull" with respect to the critical numbers to its left than does the root r_k, and vice versa.

The span of a polynomial

The span of a polynomial with all real zeros is the difference between its least and greatest roots. In [7] R. Robinson proved that among all $p \in \mathbf{P}_{n,1}$, the function whose derivatives each have the maximum span must have form $(x+1)^k(x-1)^{n-k}$. Robinson's natural conjecture, that the maximum span of all the derivatives is achieved by the function(s) for which k and $n-k$ are closest to each other (equal when n is even, differing by 1 when n is odd), has been resolved in some cases, but remains an open question. Meir and Sharma [**3**] have proved several results on the maximum spans of a polynomial's derivatives for the situation where restrictions are placed on the first and second moments of the zeros. R. Pereira [**5**] has recently achieved some of these same results as a consequence of a more general theorem through a novel approach using the theory of majorizations.

Several authors [**4, 5, 8**] have independently proved the result that for any degree n polynomial p with all real zeros,

$$\sqrt{\frac{n-2}{n}}\,\text{Span } p \le \text{Span } p'. \tag{4}$$

Applying this inequality to consecutive derivatives of p reveals that

$$\text{Span } p^{(k)} \ge \sqrt{\frac{(n-k)(n-k-1)}{n(n-1)}}\,\text{Span } p. \tag{5}$$

Meir and Sharma [**4**] further note that this minimum is attained by the function $q(x) = x^{n-2}(x^2-b^2)$ for polynomials in $\mathbf{P}_{n,b}$.

We now use the Root Squeezing Theorem to show the natural way this function q arises and that it is the unique element in $\mathbf{P}_{n,b}$ whose derivatives have the minimum possible span. Finally, we show that a consequence of this approach is the inequality (5).

Two short lemmas

Two natural functions arise in the course of squeezing roots of a polynomial together. If we consider a given $p \in \mathbf{P}_{n,b}$ and imagine progressively squeezing all of its interior roots to a point $r \in [-b, b]$, we arrive at a function $p(x)$ of the form

$$p(x) = (x-r)^{n-2}(x^2-b^2).$$

Differentiating $p(x)$, we quickly see that the first and last zeros of p' lie at

$$x = \frac{2r \pm \sqrt{4r^2+4nb^2(n-2)}}{2n},$$

from which it follows that

$$\text{Span } p'(x) = \frac{\sqrt{4r^2+4nb^2(n-2)}}{n}. \tag{6}$$

Since b and n are fixed, we observe that Span $p'(x)$ is (uniquely) minimized when $r=0$. We state this formally as Lemma 1:

LEMMA 1. *Let $p(x)$ be a polynomial of form $p(x) = (x-r)^{n-2}(x^2-b^2)$ where $b>0$ and $-b \le r \le b$. Then the absolute minimum span of $p'(x)$ occurs only when $r=0$.*

From Lemma 1 we see that the function $p(x) = x^{n-2}(x^2-b^2)$ is of further importance. For this function, it follows from (6) that Span $p'(x) = 2b\frac{\sqrt{n^2-2n}}{n}$, so that the span of the derivative of p is an increasing function of b. Formally, this is our second lemma:

LEMMA 2. *If $p(x) = x^{n-2}(x^2 - b^2)$, then Span $p'(x)$ is directly proportional to b.*

From root squeezing to minimum span

We now use the Root Squeezing Theorem to provide an alternate proof that $q(x) = x^{n-2}(x^2 - b^2)$ is the unique polynomial whose derivatives have minimum span. We begin by showing this for the first derivative and then use induction to proceed to higher order derivatives.

THEOREM 2. *Let $q(x) = x^{n-2}(x^2 - b^2)$ for some $b > 0$ and let $p(x) \neq q(x)$ be any element of $\mathbf{P}_{n,b}$. Then Span $q'(x) <$ Span $p'(x)$.*

Proof. Let $p(x) = \prod_{i=1}^{n}(x - r_i)$ with $-b = r_1 \leq r_2 \leq \ldots \leq r_{n-1} \leq r_n = b$ and let a be the mean of the interior roots $\{r_2, r_3, \ldots, r_{n-1}\}$. Let j be the greatest integer such that $r_j < a$, let k be the smallest integer such that $r_k > a$, and let d be the minimum of $a - r_j$ and $r_k - a$. Shift the root r_j to the right d units and r_k to the left d units, and call the resulting polynomial $p_1(x)$. This process shifts r_j or r_k to $x = a$. From the Root Squeezing Theorem we know that Span $p_1'(x) \leq$ Span $p'(x)$. Note further that the mean of the interior roots of $p_1(x)$ remains at $x = a$.

By repeated application of this process on the resulting polynomial $p_1(x)$ we obtain the polynomial $\tilde{p}(x) = (x - a)^{n-2}(x^2 - b^2)$. Theorem 1 implies that Span $\tilde{p}'(x) \leq$ Span $p'(x)$; moreover, the discussion following the proof of Theorem 1 shows that equality holds if and only if $\tilde{p}(x) = p(x)$. Lemma 1 implies Span $q'(x) \leq$ Span $\tilde{p}'(x)$, with equality if and only if $a = 0$. Therefore, it follows Span $q'(x) \leq$ Span $p'(x)$, with equality only when $p(x)$ has all of its interior roots at $x = a = 0$, in which case $p(x) = q(x)$. ∎

Note particularly that the above argument shows that q is the unique minimizer among polynomials in $\mathbf{P}_{n,b}$. Next we show that $q(x)$ is in fact the unique polynomial whose higher order derivatives also have the least possible span.

THEOREM 3. *Let $q(x) = x^{n-2}(x^2 - b^2)$ and let $p(x) \in \mathbf{P}_{n,b}$ be such that $p(x) \neq q(x)$. Then Span $q^{(j)}(x) <$ Span $p^{(j)}(x)$ for all $1 \leq j \leq n - 2$.*

Proof. We will prove Theorem 3 by induction on the order of the derivative. By Theorem 2, Span $p'(x) <$ Span $q'(x)$, so the base case is true. Our inductive hypothesis is that Span $q^{(k)}(x) <$ Span $p^{(k)}(x)$ where $1 \leq k \leq n - 3$. We want to show Span $q^{(k+1)}(x) <$ Span $p^{(k+1)}(x)$.

We begin by noting that it is straightforward to show

$$q^{(k)}(x) = a_k(x^2 - e_k^2)x^{n-2-k}$$

for some real constants $e_k > 0$ and $a_k > 0$. Next we let $p^{(k)}(x) = t_k(x - r_1)\cdots(x - r_{n-k})$ where $r_1 \leq \cdots \leq r_{n-k}$ and $t_k > 0$ is a real constant, and perform a series of manipulations on the roots of $p^{(k)}(x)$. The constants a_k and t_k do not affect the position of the roots of the polynomials $q^{(k)}$ or $p^{(k)}$, so without loss of generality we set $a_k = t_k = 1$. Next we translate each root of $p^{(k)}(x)$ by $\frac{r_1 + r_{n-k}}{2}$ which creates the new polynomial $p_1(x) = \left(x + \frac{r_{n-k} - r_1}{2}\right)\cdots\left(x - \frac{r_{n-k} - r_1}{2}\right)$ of degree $n - k$. Letting $p_2(x) = \left(x^2 - \left(\frac{r_{n-k} - r_1}{2}\right)^2\right)x^{n-k-2}$, Theorem 2 implies that Span $p_2'(x) \leq$ Span $p_1'(x)$. It then follows that

$$\text{Span } p_2(x) = \text{Span } p_1(x) = \text{Span } p^{(k)}(x), \text{ and} \tag{7}$$

$$\text{Span } p_2'(x) \leq \text{Span } p_1'(x) = \text{Span } p^{(k+1)}(x). \tag{8}$$

From the definitions of q and p_2, (7), and the inductive hypothesis, we know

$$2e_k = \text{Span } q^{(k)}(x) < \text{Span } p_2(x) = r_{n-k} - r_1.$$

Then from Lemma 2, we know Span $q^{(k+1)}(x) <$ Span $p_2'(x)$. By (8) it follows that Span $q^{(k+1)}(x) <$ Span $p^{(k+1)}(x)$. ■

Having established that q is the unique function in $\mathbf{P}_{n,b}$ whose derivatives have minimum span, by finding the roots of $q^{(j)}(x)$ we can derive the lower bound on the span of the jth derivative of any polynomial. For $2 \leq j \leq n-2$ the jth derivative of $q(x)$ is

$$q^{(j)}(x) = 2\frac{(n-2)!}{(n-j)!}\left(jn - \frac{j(j+1)}{2}\right)x^{n-j} + \frac{(n-2)!}{(n-j-2)!}(x^2-1)x^{n-j-2}.$$

Setting $q^{(j)}(x) = 0$ and solving this equation for x gives the roots of $q^{(j)}(x)$ and leads to the following corollary, which is inequality (5).

COROLLARY. *If $p(x) \in \mathbf{P}_{n,b}$, then Span $p^{(j)}(x) \geq 2b\sqrt{\frac{(n-j)(n-j-1)}{n(n-1)}}$ where $1 \leq j \leq n-2$.*

Acknowledgment. We are grateful for the support of the National Science Foundation (grant DMS-0451254) which funds a Research Experience for Undergraduates program at Grand Valley State University.

REFERENCES

1. B. Anderson, Polynomial root dragging, *Amer. Math. Monthly* **100** (1993) 864–865.
2. B. Anderson, Where the inflection points of a polynomial may lie, this MAGAZINE **70** (1997) 32–39.
3. A. Meir and A. Sharma, Span of derivatives of polynomials, *Amer. Math. Monthly* **74** (1967) 527–531.
4. A. Meir and A. Sharma, On the zeros of derivatives of polynomials, *Canad. Math. Bull.* **11** (1968) 443–445.
5. R. Pereira, On the spans of polynomials and their derivatives, *J. Math. Anal. Appl.* **306** (2005) 425–431.
6. G. Peyser, On the roots of the derivative of a polynomial with real roots, *Amer. Math. Monthly* **74** (1967) 1102–1104.
7. R. Robinson, On the spans of derivatives of polynomials, *Amer. Math. Monthly* **71** (1964) 504–508.
8. P. Walker, Bounds for the separation of real zeros of polynomials, *J. Austral. Math. Soc. Ser. A* **59** (1995) 330–342.

NOTES

Paint It Black—A Combinatorial Yawp

ARTHUR T. BENJAMIN
Harvey Mudd College
Claremont, CA 91711
benjamin@hmc.edu

JENNIFER J. QUINN
Interdisciplinary Arts & Sciences
University of Washington, Tacoma
Tacoma, WA 98402
jjquinn@u.washington.edu

JAMES A. SELLERS
Pennsylvania State University
University Park, PA 16802
sellersj@math.psu.edu

MARK A. SHATTUCK
University of Tennessee
Knoxville, TN 37996
shattuck@math.utk.edu

Have you experienced a "mathematical yawp" lately? (Not sure you want to answer until you know what one is?) Well, the phrase "mathematical yawp" was coined by Francis Su in his James R. Leitzel Lecture at the 2006 MathFest. In essence, a mathematical yawp is one of those "light bulb" or "aha!" moments when a mathematician comes to an understanding of a topic so moving that it is accompanied by a yelp of joy or disbelief. By specialization, a *combinatorial yawp* is one of those moments achieved while counting.

Combinatorial proofs are appreciated for the elegance and/or simplicity of their arguments (see [**2**]). However, the true (and frequently underappreciated) beauty lies in their power to generalize results. Understanding the components of a mathematical identity in a concrete counting context provides the first clue for exploring natural extensions. Investigating and stretching the role of each parameter in turn, leads to different generalizations—ones that might not be connected without the combinatorial insight.

Our yawp occurred while exploring Problem #11220, proposed by David Beckwith, from the April 2006 issue of the *American Mathematical Monthly* [**1**], the innocuous-looking alternating binomial identity below.

IDENTITY 1. *For $n \geq 1$,*

$$\sum_{r=0}^{n}(-1)^r\binom{n}{r}\binom{2n-2r}{n-1} = 0.$$

Equipped with the ability to select subsets, to paint elements black, blue, or white, and to count, we will work through a novel proof of this identity and then explore numerous related results. What qualifies as a natural generalization is open to debate, but the greatest surprise is the sheer number of interesting generalizations to be explored.

To prove Identity 1, begin by understanding the unsigned quantity in the alternating sum, $\binom{n}{r}\binom{2n-2r}{n-1}$. Consider the set of n consecutive pairs, $\{\{1,2\},\{3,4\},\dots,\{2n-1,2n\}\}$. Given r, $0 \le r \le n$, select r of the **pairs** to paint black in $\binom{n}{r}$ ways. Of the remaining $2n-2r$ elements that have not yet been painted, select $n-1$ to paint blue. This can be done in $\binom{2n-2r}{n-1}$ ways. The remaining elements are then painted white. We call such a painted set a *configuration*. For example, when $n = 5$,

$$X = \{\{1, \underline{2}\}, \{\underline{3}, \underline{4}\}, \{\mathbf{5}, \mathbf{6}\}, \{\underline{7}, 8\}, \{\mathbf{9}, \mathbf{10}\}\}$$

is a configuration where black elements are bold, blue elements are underlined, and the remaining elements are white.

Now define two sets, denoted $\mathcal{E}$ and $\mathcal{O}$, that depend on the parameter r, the number of black pairs.

Set $\mathcal{E}$. All configurations with an even number of black pairs.

Set $\mathcal{O}$. All configurations with an odd number of black pairs.

Since a configuration from $\mathcal{E}$ contributes $+1$ to the summation while a configuration from $\mathcal{O}$ contributes -1, the left-hand side of Identity 1 is simply $|\text{Set } \mathcal{E}| - |\text{Set } \mathcal{O}|$. If we can show that $|\text{Set } \mathcal{E}| = |\text{Set } \mathcal{O}|$, then Identity 1 is proved. Our goal then is to find a bijection between $\mathcal{E}$ and $\mathcal{O}$.

Correspondence. Find the minimum integer j such that $1 \le j \le n$ and $\{2j-1, 2j\}$ contains no blue element, i.e., it is either a black pair or a white pair. Then toggle the color of this pair—if it is black, make it white and if it is white, make it black.

Since there are only $n-1$ blue elements (and n total pairs), every configuration has at least one pair containing no blue element. So j always exists and the correspondence is a bijection. Hence, $|\text{Set } \mathcal{E}| = |\text{Set } \mathcal{O}|$ and the proof is complete.

As an illustration, the previously considered configuration

$$X = \{\{1, \underline{2}\}, \{\underline{3}, \underline{4}\}, \{\mathbf{5}, \mathbf{6}\}, \{\underline{7}, 8\}, \{\mathbf{9}, \mathbf{10}\}\},$$

belongs to $\mathcal{E}$ since it contains $r = 2$ black pairs. By toggling the first blueless pair $\{5, 6\}$, X is matched with

$$X' = \{\{1, \underline{2}\}, \{\underline{3}, \underline{4}\}, \{5, 6\}, \{\underline{7}, 8\}, \{\mathbf{9}, \mathbf{10}\}\},$$

which belongs to $\mathcal{O}$, since it has $r = 1$ black pair.

At this point, many natural questions arise. Can we change the number of blue elements? What happens if we replace the *pairs* above by *k-sets*? Can we say something about partial sums? We will consider each of these questions in turn.

Changing the number of blue elements. If we paint fewer than $n-1$ elements blue in our proof above, the argument doesn't change. We are still guaranteed a blueless pair, so a toggle point exists. Letting m represent the number of blue elements to be painted, this gives

IDENTITY 2. *For* $0 \le m < n$,

$$\sum_{r=0}^{n} (-1)^r \binom{n}{r} \binom{2n-2r}{m} = 0.$$

What happens when m is larger than $n-1$? Well, the initial set-up is the same. Select r pairs to color black and m of the remaining elements to color blue. The sets $\mathcal{E}$ and $\mathcal{O}$ contain configurations with an even or odd number of black pairs. Again, toggle the color of the first blueless pair. Unfortunately, there are now unpaired configurations in our correspondence (so it is no longer a bijection). Since m is greater than or equal to n, we can no longer guarantee a toggle point exists. However, we know that the unpaired configurations have at least one blue element in every pair, so these configurations have zero black pairs and hence belong to $\mathcal{E}$.

For example, when $n = 5$ and $m = 7$, the configuration

$$X = \{\{\underline{1}, 2\}, \{\underline{3}, \underline{4}\}, \{\underline{5}, \underline{6}\}, \{\underline{7}, 8\}, \{9, \underline{10}\}\}$$

has no toggle point.

How many of these unpaired configurations are there? Such configurations have $m-n$ pairs where both elements are painted blue. So there are $\binom{n}{m-n}$ ways to select the blue pairs. Then, the other $n-(m-n) = 2n-m$ pairs have one blue element and one white element, and there are 2^{2n-m} ways to paint them. Thus, there are $\binom{n}{m-n}2^{2n-m}$ unpaired configurations, leading to our next generalization.

IDENTITY 3. *For* $n, m \ge 0$,

$$\sum_{r=0}^{n} (-1)^r \binom{n}{r} \binom{2n-2r}{m} = 2^{2n-m} \binom{n}{m-n}.$$

Note that this is a generalization of Identity 2 since $\binom{n}{m-n} = 0$ when $m < n$. To some, this would be enough for a yawp. But we press on for more!

From pairs to k-sets. Rather than creating n subsets by pairing consecutive elements of the set $\{1, 2, 3, \dots, 2n\}$, we ask what would happen if we group k consecutive elements from $\{1, 2, 3, \dots, kn\}$. By mimicking the argument for Identity 1, we can immediately generalize Identity 2 as follows.

IDENTITY 4. *For* $0 \le m < n$ *and* $k \ge 1$,

$$\sum_{r=0}^{n} (-1)^r \binom{n}{r} \binom{kn-kr}{m} = 0.$$

For example, when $n = 5$, $k = 3$, $m = 4$, the configuration

$$X = \{\{\underline{1}, 2, \underline{3}\}, \{\mathbf{4}, \mathbf{5}, \mathbf{6}\}, \{7, 8, 9\}, \{\underline{10}, \underline{11}, 12\}, \{\mathbf{13}, \mathbf{14}, \mathbf{15}\}\}$$

has $r = 2$ black 3-sets (and thus belongs to $\mathcal{E}$) and by toggling the first blueless 3-set, we get

$$X' = \{\{\underline{1}, 2, \underline{3}\}, \{4, 5, 6\}, \{7, 8, 9\}, \{\underline{10}, \underline{11}, 12\}, \{\mathbf{13}, \mathbf{14}, \mathbf{15}\}\}$$

(which belongs to $\mathcal{O}$).

Can we generalize Identity 4, allowing $m \ge n$ blue elements? Yes and no. We can formulate a general answer, but the alternating sum becomes a sum over integer partitions. Although it is not the nice answer we were hoping for, it still has some notable specializations.

In the general situation with $m \geq n$, unpaired objects are configurations with at least one blue element in every k-set. These objects necessarily belong to $\mathcal{E}$ since they have $r = 0$ black k-sets. For example, when $n = 5, k = 3, m = 8$, the configuration

$$\{\{\underline{1}, 2, 3\}, \{\underline{4}, \underline{5}, \underline{6}\}, \{\underline{7}, 8, 9\}, \{\underline{10}, 11, \underline{12}\}, \{13, \underline{14}, 15\}\}$$

has no blueless 3-set. We can count these by considering the distribution of m blue elements among the n different k-sets. Let x_i count the number of k-sets containing i blue elements ($1 \leq i \leq k$). In our example, $x_1 = 3, x_2 = 1, x_3 = 1$. The sum $\sum_{i=1}^{k} x_i$ counts the number of k-sets containing blue elements while the sum $\sum_{i=1}^{k} ix_i$ counts the number of blue elements. Only nonnegative integer solutions $(x_1, x_2, \ldots, x_n)$ to

$$\begin{cases} n = x_1 + x_2 + \cdots + x_k \\ m = x_1 + 2x_2 + \cdots + kx_k \end{cases}$$

contribute to the number of unpaired configurations. Since the number of ways to choose which x_i k-sets have i blue elements is the multinomial coefficient

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1!\, x_2! \cdots x_k!},$$

and a k-set with i blue elements can be painted $\binom{k}{i}$ ways, we get

IDENTITY 5. *For all $k, m, n \geq 1$,*

$$\sum_{r=0}^{n} (-1)^r \binom{n}{r} \binom{kn - kr}{m} = \sum_{(x_1, x_2, \ldots, x_k)} \binom{n}{x_1, x_2, \ldots, x_k} \prod_{i=1}^{k} \binom{k}{i}^{x_i},$$

where the sum on the right is taken over all simultaneous nonnegative integer solutions to $n = x_1 + x_2 + \cdots + x_k$ and $m = x_1 + 2x_2 \cdots + kx_k$.

Note that this is a generalization of Identity 4 since when m is less than n, the sum on the right is empty. Some special cases are worth mentioning because their right-hand sides reduce to simple one-term expressions:

- $m = n$

$$\sum_{r=0}^{n} (-1)^r \binom{n}{r} \binom{kn - kr}{n} = k^n.$$

- $m = n + 1$

$$\sum_{r=0}^{n} (-1)^r \binom{n}{r} \binom{kn - kr}{n+1} = nk^{n-1} \binom{k}{2}.$$

Partial sums. The final generalization considers what happens if we return to creating pairs from the set $\{1, 2, 3, \ldots, 2n\}$ and only consider the first s terms of the original sum. To make life easier, we restrict our attention to the situation where $m < n$ and consider

$$\sum_{r=0}^{s} (-1)^r \binom{n}{r} \binom{2n - 2r}{m}.$$

In this case, the development parallels Identity 2 except that only configurations with s or fewer black pairs are considered. To match configurations between $\mathcal{E}$ and $\mathcal{O}$, we toggle the color of the first blueless pair unless the configuration contains the maximum s black pairs and a white pair precedes them.

For example, when $n = 5$, $m = 2$, $s = 3$ the configuration

$$X = \{\{1, 2\}, \{\mathbf{3}, \mathbf{4}\}, \{\mathbf{5}, \mathbf{6}\}, \{\underline{7}, \underline{8}\}, \{\mathbf{9}, \mathbf{10}\}\},$$

is unmatched, since by toggling the first blueless set $\{1, 2\}$, we would wind up with four black pairs, exceeding our upper bound. We note that among the configurations with s black pairs and w white pairs, the fraction of those where a white pair comes before every black pair is $\frac{w}{w+s}$.

To count the number of unmatched objects, let b represent the number of blue pairs in a configuration. Since b blue pairs contain $2b$ blue elements, there must be $m - 2b$ pairs containing one blue and one white element (and since we have s black pairs, there are $n - b - (m - 2b) - s = n - m - s + b$ white pairs). So there are $2^{m-2b}\binom{n}{s,\ b,\ m-2b,\ n-m-s+b}$ configurations with s black pairs, b blue pairs, and a total of m blue elements. Of these, $\frac{n-m-s+b}{n-m+b}$ of the configurations have a white pair coming before all the black pairs. These unmatched configurations all belong to $\mathcal{E}$ or all belong to $\mathcal{O}$ depending on the parity of s. This yields the following identity:

IDENTITY 6. *For $0 \le m < n$ and $0 \le s \le n$,*

$$\sum_{r=0}^{s}(-1)^r\binom{n}{r}\binom{2n-2r}{m} = (-1)^s\sum_{b\ge 0}\frac{n-m-s+b}{n-m+b}2^{m-2b}\binom{n}{s,\ b,\ m-2b,\ n-m-s+b}.$$

Perhaps you don't find this solution satisfactory? Let's make one last restriction in hopes of finding a "nice" solution. Restrict the location of the black pairs to only occur in the first s positions. Then, for $1 \le m, n$, the alternating sum becomes

$$\sum_{r=0}^{s}(-1)^r\binom{s}{r}\binom{2n-2r}{m}.$$

The unsigned quantity in the alternating sum, $\binom{s}{r}\binom{2n-2r}{m}$, counts the ways to select r black pairs from $\{\{1, 2\}, \{3, 4\}, \dots, \{2s - 1, 2s\}\}$ and then paint m of the remaining uncolored elements from $\{1, 2, 3, \dots, 2n\}$ blue. We then use the same toggling argument as before:

Set $\mathcal{E}$. All configurations with an even number of black pairs.

Set $\mathcal{O}$. All configurations with an odd number of black pairs.

Correspondence. Find the minimum integer j such that $1 \le j \le s$ and $\{2j - 1, 2j\}$ contains no blue element, i.e., it is either a black pair or a white pair. Then toggle the color of the pair.

The solutions to this alternating sum depends on the size of m, the number of blue elements to be painted. If $m < s$, a toggle point always exists and our correspondence is a bijection, giving the following generalization of Identity 2.

IDENTITY 7. *For $0 \le m < s \le n$,*

$$\sum_{r=0}^{s}(-1)^r\binom{s}{r}\binom{2n-2r}{m} = 0.$$

If $m = s$, the unmatched configurations are those in which each of the first s pairs contains at least one blue element. (Unlike the previous situation, we don't have to worry about generating too many black pairs.) All 2^s of these unmatched configurations belong to $\mathcal{E}$, and we get

IDENTITY 8. *For* $0 \leq s \leq n$,

$$\sum_{r=0}^{s}(-1)^r\binom{s}{r}\binom{2n-2r}{s} = 2^s.$$

Lastly, if $m > s$, the unmatched configurations are again those in which each of the first s pairs contains at least one blue element. We convert the alternating sum into a positive sum by counting the configurations that are unmatched by the previous correspondence. Such unmatched configurations have at least one blue element among each of the first s pairs (and therefore have zero black elements). For $0 \leq w \leq s$, we claim that there are $\binom{s}{w}\binom{2n-s-w}{m-s}$ unmatched configurations where w of the first s pairs begin with a white element. To see this, note that once we choose which s pairs begin with a white element (which can be done $\binom{s}{w}$ ways) then those w pairs must end with a blue element and the remaining $s - w$ pairs must begin with a blue element. The remaining $m - s$ blue elements can be chosen among the unspecified $(s - w) + (2n - 2s) = 2n - s - w$ elements in $\binom{2n-s-w}{m-s}$ ways. Since all of the unmatched configurations belong to $\mathcal{E}$, we arrive at our final identity, which actually encapsulates Identities 7 and 8 too.

IDENTITY 9. *For all* $m, n, s \geq 0$,

$$\sum_{r=0}^{s}(-1)^r\binom{s}{r}\binom{2n-2r}{m} = \sum_{w=0}^{s}\binom{s}{w}\binom{2n-s-w}{m-s}.$$

So starting from a single alternating binomial identity, a concrete counting context, and a good correspondence, eight related identities were explored by manipulating the roles of the parameters (and sometimes introducing new ones). The resulting identities were often beautiful generalizations—though occasionally the results didn't quite qualify as "simple" or "nice." Regardless, the questions were worth asking, the answers worth exploring, and the connections worth making. We yawped. Did you?

Acknowledgments. This exploration records the questions and generalizations that actually occurred in Minicourse #5, *Combinatorially Thinking*, taught by Art Benjamin and Jenny Quinn at MathFest 2006, Knoxville. Mark Shattuck served as the informal TA. And while every member of the minicourse contributed to the discoveries listed here and many yawped out loud, we didn't think a publication with more than thirty authors was reasonable. Instead we thank everyone here: Karen Briggs, Sally Cockburn, Paul Coe, Joseph DeMaio, Anne Edlin, Robert Ellis, Tyler Evans, Leon Harkleroad, Masayuki Hino, Amy Hlavacek, Jon Johnson, John Konvalina, Mathew LaBar, Lidia Luquet, John Maceli, Steven Martin, Peter McNamara, Paul Olson, M.Tip Phaovibul, Richard Potter, Marc Renault, Michael Saclolo, Dave Schmidt, Jonathan Schneider, James Sellers, Mary Servatius, Therese Shelton, Douglas Smith, Paul Stanford, Eric Sundberg, Ronald Tiberio, James Uebelacker, Zhying Zheng, and Henry Zwick. Thanks also to the anonymous referees for their meticulous and thoughtful suggestions.

REFERENCES

1. David Beckwith, Problem #11220, *Amer. Math. Monthly* **113** (2006) 367.
2. A. T. Benjamin and J. J. Quinn, *Proofs that Really Count: The Art of Combinatorial Proof*, Mathematical Association of America, 2003.

Integration by Parts and Infinite Series

SHELBY J. KILMER
Missouri State University
Springfield, MO 65897
skilmer@missouristate.edu

We have all had students who, after having made a disastrous misstep at the beginning of a calculation, continued to grind away, oblivious to how complicated and unreasonable their result was becoming. Rarely does such perseverance pay off as well as it did for the second semester calculus student who submitted the following solution on her final exam.

$$\begin{aligned}\int xe^x\,dx &= \left(\frac{x^2}{2}e^x - \frac{x^3}{3!}e^x + \frac{x^4}{4!}e^x - \frac{x^5}{5!}e^x + \cdots\right) + C\\ &= -e^x + xe^x + e^x\left(1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + \frac{x^4}{4!} - \cdots\right) + C\\ &= xe^x - e^x + e^xe^{-x} + C\\ &= xe^x - e^x + C\end{aligned}$$

At first the professor grading the final thought he could give little credit for this work since the student had reversed the usual choices for f and g in applying the (tabular) integration by parts formula,

$$\int fg = \sum_{k=0}^{n}(-1)^k f^{(k)} g^{-(k+1)},$$

where $g^{-(k)}$ denotes the kth successive antiderivative and $f^{(n)}$ is constant.

However since the answer was correct he decided to give her work further consideration. Since $f^{(k)}$ was never constant, it was apparent that she had tacitly assumed the following series form for the solution.

$$\int fg = \sum_{k=0}^{\infty}(-1)^k f^{(k)} g^{-(k+1)} \tag{1}$$

In this calculation series (1) produced the correct solution, but was this a one-time shot-in-the-dark or could this method be used to find other valid series representations?

Throughout the remainder of the article we will refer to series (1) as an i.p. series, where i.p. is an abbreviation for integration by parts.

Let's first determine the i.p. series for the sine and cosine functions and then compare them with their Maclaurin series. We will address the convergence of (1) at the end of the article.

$$\begin{aligned}\sin x &= \int_0^x \cos t\,dt\\ &= x\cos x + \frac{x^2}{2}\sin x - \frac{x^3}{3!}\cos x - \frac{x^4}{4!}\sin x + \frac{x^5}{5!}\cos x + \frac{x^6}{6!}\sin x \cdots\\ &= \cos x\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}\cdots\right) + \sin x\left(\frac{x^2}{2} - \frac{x^4}{4!} + \frac{x^6}{6!} - \frac{x^8}{8!}\cdots\right)\end{aligned}$$

$$= \cos x \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \cdots\right)$$
$$- \sin x \left(1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \cdots\right) + \sin x,$$

hence if we let

$$\sigma_s = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} \qquad \text{and} \qquad \sigma_c = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!},$$

we obtain

$$\sigma_s \cos x - \sigma_c \sin x = 0. \tag{2}$$

On the other hand

$$1 - \cos x = \int_0^x \sin t \, dt$$
$$= x \sin x - \frac{x^2}{2} \cos x - \frac{x^3}{3!} \sin x + \frac{x^4}{4!} \cos x + \frac{x^5}{5!} \sin x - \frac{x^6}{6!} \cos x \cdots$$
$$= \sin x \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \cdots\right) + \cos x \left(-\frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \cdots\right)$$
$$= \sin x \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \cdots\right)$$
$$+ \cos x \left(1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \cdots\right) - \cos x$$

Thus with σ_s and σ_c as above, we have

$$\sigma_c \cos x + \sigma_s \sin x = 1. \tag{3}$$

Solving the simultaneous equations (2) and (3) yields

$$\sigma_s = \sin x \qquad \text{and} \qquad \sigma_c = \cos x$$

and shows that this method does indeed obtain valid series representations for $\sin x$ and $\cos x$, in fact their Maclaurin series.

There are several ways to find an i.p. series for π. The first we present uses the fact that π is the area of the unit circle. Recall that $(2k-1)!! = (2k-1)(2k-3)\cdots 5 \cdot 3 \cdot 1$. Then

$$\pi = 4 \int_0^1 \sqrt{1 - x^2}\, dx$$
$$= 4 \int_0^1 \sqrt{1 + x}\sqrt{1 - x}\, dx$$
$$= 4 \left[-\frac{2}{3}\sqrt{1 + x}(\sqrt{1 - x})^3 \right.$$
$$\left. + \sum_{k=1}^{\infty} (-1)^k \frac{(-1)^{k+1}(2k-3)!!}{2^k(\sqrt{1+x})^{2k-1}} \cdot \frac{(-1)^{k+1} 2^{k+1} (\sqrt{1-x})^{2k+3}}{(2k+3)!!} \right] \Bigg|_0^1$$

$$= 4\left[-\frac{2}{3}\sqrt{1+x}(\sqrt{1-x})^3 + 2\sum_{k=1}^{\infty}\frac{(-1)^k(\sqrt{1-x})^{2k+3}}{(2k+3)(2k+1)(2k-1)(\sqrt{1+x})^{2k-1}}\right]\Bigg|_0^1$$

$$= \frac{8}{3} + 8\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{(2k+3)(2k+1)(2k-1)}$$

$$= 8\cdot\sum_{k=0}^{\infty}\frac{(-1)^{k+1}}{(2k+3)(2k+1)(2k-1)}.$$

Partial fractions decomposition can also be used to derive this series.

Les Reid, the author's colleague at Missouri State University, derived the following elegant example of an i.p. series for π via the arctan function. Consider $\int \frac{dx}{1+x^2} = \arctan x$ and the partial fraction decomposition

$$\frac{1}{1+x^2} = \frac{1}{2i}\left(\frac{1}{x-i} - \frac{1}{x+i}\right).$$

The i.p. series method yields

$$\int \frac{dx}{x+a} = \frac{x}{x+a} + \frac{1}{2}\left(\frac{x}{x+a}\right)^2 + \frac{1}{3}\left(\frac{x}{x+a}\right)^3 + \frac{1}{4}\left(\frac{x}{x+a}\right)^4 + \cdots + C$$

$$= \sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{x}{x+a}\right)^k.$$

Therefore

$$\begin{aligned}
\pi &= 4\arctan 1 \\
&= 4\int_0^1 \frac{dx}{1+x^2} \\
&= \frac{2}{i}\int_0^1\left(\frac{1}{x-i} - \frac{1}{x+i}\right)dx \\
&= \frac{2}{i}\sum_{k=1}^{\infty}\frac{1}{k}\left[\left(\frac{1}{1-i}\right)^k - \left(\frac{1}{1+i}\right)^k\right]_0^1 \\
&= \frac{2}{i}\sum_{k=1}^{\infty}\frac{(1+i)^k - (1-i)^k}{2^k k} \\
&= 4\sum_{k=1}^{\infty}\frac{\sqrt{2}^k\left(e^{\frac{\pi k i}{4}} - e^{-\frac{\pi k i}{4}}\right)}{2^k k\cdot 2i} \\
&= 4\sum_{k=1}^{\infty}\frac{\sin\frac{\pi k}{4}}{2^{\frac{k}{2}}k}.
\end{aligned}$$

In some situations the i.p. series method can be used to sum a series. Here we calculate

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)(k+2)\cdots(k+n)}$$

for $n \geq 1$. This number will appear as a constant in the following equation.

$$\begin{aligned}\int x^{n-1} \ln x \, dx &= \frac{x^n}{n} \ln x + \sum_{k=1}^{\infty} (-1)^k \frac{x^{n+k}(n-1)!}{(n+k)!} \cdot \frac{(-1)^{k+1}(k-1)!}{x^k} + C \\ &= \frac{x^n}{n} \ln x - x^n (n-1)! \sum_{k=1}^{\infty} \frac{(k-1)!}{(n+k)!} + C \\ &= \frac{x^n}{n} \ln x - x^n (n-1)! \sum_{k=1}^{\infty} \frac{1}{k(k+1)(k+2)\cdots(k+n)} + C.\end{aligned}$$

Differentiating both sides eliminates the constant of integration.

$$x^{n-1} \ln x = x^{n-1} \ln x + \frac{x^{n-1}}{n} - n x^{n-1} (n-1)! \sum_{k=1}^{\infty} \frac{1}{k(k+1)(k+2)\cdots(k+n)}.$$

By simplifying this equation we obtain

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)(k+2)\cdots(k+n)} = \frac{1}{n!\,n}.$$

This result can also be obtained by telescoping.

Conditions for convergence of i.p. series are generally easily satisfied. By finitely many applications of integration by parts we obtain

$$\int_a^b f(t)g(t)\,dt = \sum_{k=0}^{n-1} (-1)^k f^{-(k+1)}(t) g^{(k)}(t) \Big|_a^b + (-1)^{n-1} \int_a^b f^{-(n)} g^{(n)}(t)\,dt, \tag{4}$$

hence in order to determine convergence of the i.p. series, it suffices to show the right hand integral in (4) tends to zero. However

$$\begin{aligned}\left| \int_a^b f^{-(n)} g^{(n)}(t) dt \right| &\leq \int_a^b \left| f^{-(n)} g^{(n)}(t) \right| dt \\ &\leq (b-a) \sup\{ |f^{-(n)}(t) g^{(n)}(t)| : a < t < b \}.\end{aligned}$$

Thus an i.p. series is convergent to the corresponding integral, provided

$$\sup\{ |f^{-(n)}(t) g^{(n)}(t)| : a < t < b \} \to 0. \tag{5}$$

As an example we verify the convergence of the second i.p. series for π presented above. Since

$$\pi = 4 \arctan 1 = \int_0^1 \frac{dx}{1+x^2} = \frac{2}{i} \int_0^1 \left(\frac{1}{x-i} - \frac{1}{x+i} \right) dx,$$

it suffices to show i.p. convergence for $\int_0^1 (x+a)^{-1}\,dx$, where $a = \pm i$. To that end, let $f(x) = 1$ and $g(x) = (x+a)^{-1}$.

Let n be a nonnegative integer. Then on $(0, 1]$, $\sqrt{\frac{x^2}{x^2+1}}$ and x^n are increasing functions, hence their composition is increasing and we have

$$\begin{aligned}
&\sup\{|f^{-(n)}(x)g^{(n)}(x)| : 0 < x < 1\} \\
&= \sup\left\{\left|\frac{x^n}{n!}\frac{n!}{(x+a)^{n+1}}\right| : 0 < x < 1\right\} \\
&= \sup\left\{\left(\sqrt{\frac{x^2}{x^2+1}}\right)^n \cdot \frac{1}{\sqrt{x^2+1}} : 0 < x < 1\right\} \\
&\leq \left(\frac{1}{\sqrt{2}}\right)^n \cdot 1 \to 0
\end{aligned}$$

Thus by (5) above we see this i.p. series indeed converges to π.

Since the convergence criteria for an i.p. series are so easily satisfied and because of the variety in their form, i.p. series have the potential for wide application. Due to the elementary nature of integration by parts and infinite series they offer new topics for the classroom and projects for advanced students. They give the working mathematician an elegant method for deriving series and the potential for discovering new ones.

What Fraction of a Soccer Ball Is Covered with Pentagons?

P. K. ARAVIND
Physics Department
Worcester Polytechnic Institute
Worcester, MA 01609
paravind@wpi.edu

The surface of many soccer balls is covered with pentagons and hexagons in such a way that one pentagon and two hexagons meet at each vertex, there being twelve pentagons and twenty hexagons altogether (see Fig. 1). The pentagons are generally set off in a different color to give the ball some contrast and make it easier to see. While watching the World Cup soccer matches last summer, I found myself wondering what fraction of the surface of a soccer ball is covered with pentagons. It is the purpose of this note to give the answer to this geometrical puzzle.

A rough answer to the puzzle can be obtained using Euclidean geometry if one assumes that the pentagons and hexagons on the ball are all planar. If l denotes the common edge length of the (planar) pentagons and hexagons, the area of a pentagon is $A_p = (5l^2/4)\cot(\pi/5)$ and that of a hexagon $A_h = 3\sqrt{3}l^2/2$. Letting $\phi = 2\cos(\pi/5) = (1+\sqrt{5})/2$ be the golden ratio, the fraction of the ball covered with pentagons can be worked out as

$$F = \frac{12A_p}{12A_p + 20A_h} = \frac{\phi}{\phi + \sqrt{48 - 12\phi^2}} \cong 0.28435. \qquad (1)$$

Figure 1 A common type of soccer ball, covered with 12 pentagons and 20 hexagons.

However the pentagons and hexagons are not planar, and so the accuracy of this estimate is open to question. It would be nice to have an exact answer to compare to the above result.

One way of getting the exact answer is to use a theorem of spherical geometry, according to which the area of a spherical triangle is equal to the product of its "angular excess" (i.e. the amount by which the sum of its angles, in radians, exceeds π) and the square of the radius of the sphere on which it lies [1]. This theorem can be used to calculate the area of a pentagon on a soccer ball as five times the area of one of the elemental triangles into which it is divided by the great circle arcs that join its center to its vertices. Let us take the radius of the soccer ball to be unity and denote by θ_p the vertex angle of a pentagon on it. Then the angles of the elemental triangle of which the pentagon is made up are $2\pi/5$, $\theta_p/2$, and $\theta_p/2$, and the area of this triangle is $2\pi/5 + \theta_p/2 + \theta_p/2 - \pi = \theta_p - 3\pi/5$, from which it follows that the area of the pentagon is $5(\theta_p - 3\pi/5)$. The fractional area occupied by the pentagons is therefore

$$F = \frac{12 \cdot 5 \cdot (\theta_p - 3\pi/5)}{4\pi} = \frac{15}{\pi}\left(\theta_p - \frac{3\pi}{5}\right). \tag{2}$$

The above formula suggests an empirical method of determining F, based on measuring the angle θ_p on a soccer ball. However this method proves to be unsatisfactory because $(\theta_p - 3\pi/5)$, the difference in the vertex angles of the spherical and planar pentagons, is only on the order of a few degrees and requires θ_p to be measured to a small fraction of a degree if F is to be calculated accurately via Eq. (2). Needless to say, most soccer balls are not put together with this end in mind!

An alternative approach to calculating F is based on a formula for the area of a spherical triangle due to Euler and Lagrange. Let $\overline{a}$, $\overline{b}$, and $\overline{c}$ be vectors from the center of a unit sphere to the vertices of a spherical triangle on it. We assume that the triangle is an *Euler triangle*, i.e., one in which no side or angle (both expressed in radian measure) exceeds π. Then the area, Ω, of this triangle is given by

$$\tan\left(\frac{\Omega}{2}\right) = \frac{|\overline{a} \cdot \overline{b} \times \overline{c}|}{1 + \overline{a} \cdot \overline{b} + \overline{b} \cdot \overline{c} + \overline{c} \cdot \overline{a}}. \tag{3}$$

A derivation of this formula, together with some of its history, can be found in Erikkson [2]. The reader can convince him(her)-self of the correctness of this formula in at least one special case by applying it to a spherical triangle all of whose angles are right angles (and for which the unit vectors $\overline{a}$, $\overline{b}$, and $\overline{c}$ are mutually orthogonal), for which it yields the expected result $\Omega = \pi/2$. Let (θ, ϕ) be the (polar, azimuthal) angles of a point on the surface of the soccer ball, and take the angular coordinates of the center of a pentagonal face and two of its adjacent vertices to be $(0, 0)$, $(\theta_1, 0)$, and $(\theta_1, 2\pi/5)$, respectively. The (unit) vectors from the center of the soccer ball to these vertices are then $\overline{a} = (0, 0, 1)$, $\overline{b} = (\sin\theta_1, 0, \cos\theta_1)$, and $\overline{c} = [\sin\theta_1 \cos(2\pi/5), \sin\theta_1 \sin(2\pi/5), \cos\theta_1]$. Substituting these into (3) allows one to calculate the area of an elemental triangle, from which one can get F. All that is needed to carry out this calculation is knowledge of the angle θ_1.

It is at this point that one needs to delve a little more deeply into the geometry of a soccer ball. A soccer ball is modeled on a truncated icosahedron, obtained by slicing off the corners of a regular icosahedron in such a way that each of its twelve vertices gets replaced by a regular pentagon and each of its twenty (triangular) faces by a regular hexagon. For this to happen, it is necessary that only the central third of each edge of the icosahedron be retained (as one of the edges of a resulting hexagonal face), with the thirds at either end being discarded to make way for the new pentagonal faces. If the edges of the truncated icosahedron are then projected on to its circumscribing sphere in such a way that each edge goes into a great circle arc on the sphere, the soccer ball pattern results. The reader who wishes to study the geometry of a truncated icosahedron in more detail can consult [**3**], which also gives instructions for building one. Wenninger [**4**] has a nice diagram showing how projecting a truncated icosahedron on to its circumscribing sphere leads to a pattern similar to that seen on the surface of a soccer ball.

It is known [**1**, Chapter 10] from the geometry of a regular icosahedron that the angle subtended by one of its edges at its center is $\theta_0 = \arctan(2) \cong 63.43°$. Taking two adjacent vertices of an icosahedron inscribed in a unit sphere to have coordinates $\overline{u}_1 = (0, 0, 1)$ and $\overline{u}_2 = (\sin\theta_0, 0, \cos\theta_0)$, one finds that the two vertices of the soccer ball lying on the joining edge are $\overline{v}_1 = \frac{2}{3}\overline{u}_1 + \frac{1}{3}\overline{u}_2$ and $\overline{v}_2 = \frac{1}{3}\overline{u}_1 + \frac{2}{3}\overline{u}_2$. The angle θ_1 can then be calculated as the angle between the vectors $\overline{u}_1$ and $\overline{v}_1$, and the common edge length of a pentagon or hexagon on the soccer ball, which we will denote θ_s, as the angle between the vectors $\overline{v}_1$ and $\overline{v}_2$. A simple calculation involving dot products shows that

$$\theta_1 = \arccos\left[\sqrt{\frac{8\phi + 17}{8\phi + 21}}\right] \cong 20.08° \quad \text{and} \quad \theta_s = \arccos\left(\frac{8\phi + 1}{10\phi - 1}\right) \cong 23.28°, \tag{4}$$

where we repeatedly used the relation $\phi^2 = \phi + 1$ to cast the above expressions in the simplest form possible. It is evident from this construction that $2\theta_1 + \theta_s = \theta_0$ (as is also evident numerically). With θ_1 in hand, we can calculate F from (3) in the manner indicated earlier and find that

$$F = \frac{30}{\pi} \arctan\left[\frac{\sin^2\theta_1 \sin(2\pi/5)}{(1 + \cos\theta_1)^2 + \sin^2\theta_1 \cos(2\pi/5)}\right] \cong 0.28177. \tag{5}$$

An alternative way of calculating F is to calculate the vertex angle θ_p of a pentagon on the soccer ball and then use it in (2). One can find θ_p from θ_1 and θ_s by using the cosine rule of spherical trigonometry, according to which the angle A opposite the side a of a spherical triangle with sides a, b, and c (in radian measure) is given by

$$\cos A = \frac{\cos a - \cos b \cos c}{\sin b \sin c}. \tag{6}$$

Applying this to an elemental triangle within a pentagonal face, with $A = \theta_p/2$, $a = b = \theta_1$, and $c = \theta_s$, gives

$$\theta_p = 2\arccos\left[\frac{\cos\theta_1 - \cos\theta_1\cos\theta_s}{\sin\theta_1\sin\theta_s}\right] \cong 111.38^\circ. \tag{7}$$

Using the radian measure of this angle in (2) gives $F = 0.28177$ to five decimal places, which agrees with (5) and gives us additional confidence in this result.

A comparison of (2) and (5) shows that the "planar approximation" used in getting (2) is remarkably good and gives a value just about 1 percent higher than the true value. The makers of soccer balls are evidently well aware of this close convergence, because they put the ball together out of planar pentagonal and hexagonal patches. After the patches are sewn together and the ball is inflated, the patches flex gently to accommodate themselves to the demands of spherical geometry.

Despite the near equality of (2) and (5), it is worth noting that the vertex angles of the spherical pentagons and hexagons on a soccer ball differ appreciably from those of the planar pentagon and hexagon. The vertex angles of the spherical pentagon and hexagon are $\theta_p = 111.38^\circ$ and $\theta_h = 124.31^\circ$ (the latter following from the fact that $\theta_p + 2\theta_h = 360^\circ$), and these differ noticeably from the angles of 108° (for a planar pentagon) and 120° (for a planar hexagon), showing that the differences between spherical and planar geometry are not completely masked in local measurements on a soccer ball.

The truncated icosahedron that underlies a soccer ball also serves as the framework for a molecule of C-60, or "buckyball," which has a carbon atom at each vertex of this polyhedron. It is interesting to contrast buckyball with diamond and graphite, the other two allotropes of carbon. In diamond, each carbon atom occurs at the center of a tetrahedral cage formed by four other carbon atoms, with the angle between neighboring C-C bonds being $\arccos(-1/3) = 109.47^\circ$. In graphite the carbon atoms are arranged in planar hexagonal sheets, with the angle between neighboring C-C bonds being 120°. Buckyball interpolates neatly between these other two allotropes in having bond angles of 108° and 120°.

REFERENCES

1. H. S. M. Coxeter, *Introduction to Geometry*, New York, Wiley, 1969, pp. 94–95.
2. F. Eriksson, "On the measure of solid angles," this MAGAZINE **63** (1990) 184.
3. H. M. Cundy and A. P. Rollett, *Mathematical Models*, New York, Oxford University Press, 1957, Chap. III.
4. M. J. Wenninger, *Spherical Models*, New York, Cambridge University Press, 1979, Chap. II, p. 33.

Euler's Triangle Inequality via Proofs Without Words

ROGER B. NELSEN
Lewis & Clark College
Portland, OR 97219
nelsen@lclark.edu

In November 1983, this MAGAZINE published a special issue [7] in tribute to Leonhard Euler (1707–1783) on the occasion of the 200th anniversary of his death. In addition to a number of excellent survey articles, that issue contained a glossary of terms, formulas, equations and theorems that bear Euler's name, the last one of which was the following:

EULER'S THEOREM FOR A TRIANGLE. *The distance d between the circumcenter and incenter of a triangle is given by* $d^2 = R(R - 2r)$*, where* R, r *are the circumradius and inradius, respectively.*

An immediate consequence of this theorem is $R \geq 2r$, which is often referred to as *Euler's triangle inequality*. In this Note (on the occasion of the 300th anniversary of Euler's birth) we use "proofs without words" to prove three simple lemmas that can be combined with the arithmetic mean-geometric mean inequality in order to prove Euler's triangle inequality with only simple algebra (and without reference to the theorem above). The proof is derived from one that appears in [3]. Coxeter [1] notes that although Euler published this inequality in 1767 [2], it had appeared earlier (1746) in a publication by William Chapple.

As we have just noted, the "inequality" in Euler's triangle inequality is derived from the *arithmetic mean-geometric mean inequality*: For any two positive numbers x and y, the arithmetic mean $(x+y)/2$ is at least as great as the geometric mean $\sqrt{xy}$. Hence for any three positive numbers x, y, and z, we have $x + y \geq 2\sqrt{xy}$, $y + z \geq 2\sqrt{yz}$, and $z + x \geq 2\sqrt{zx}$. Multiplying these three inequalities yields

$$(x+y)(y+z)(z+x) \geq 8xyz. \tag{1}$$

Now consider a triangle with side lengths a, b, and c as shown in Figure 1(a), and bisect each angle to locate the center of the inscribed circle. Extending an inradius (length r) to each side partitions the triangle into six smaller right triangles with side lengths as indicated in Figure 1(b). Noting that $x + y = c$, $y + z = a$, and $z + x = b$, (1) becomes

$$abc \geq 8xyz. \tag{2}$$

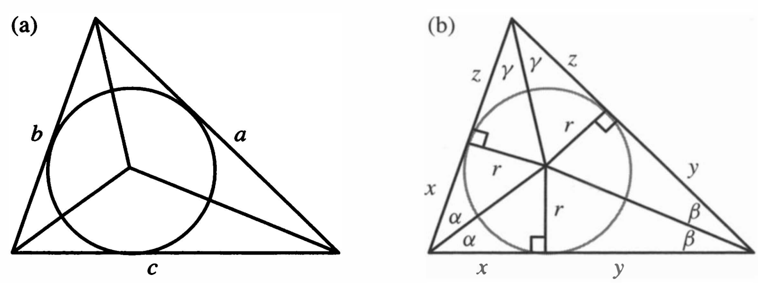
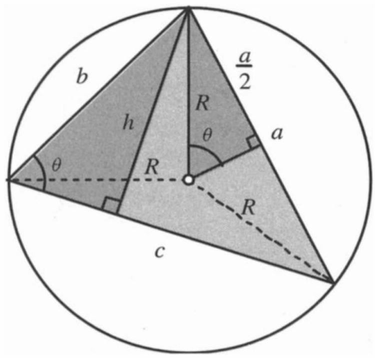


Figure 1

We now show that (2) is equivalent to $R \geq 2r$. To accomplish this, first we prove (wordlessly) three lemmas—which are of interest in their own right—from which Euler's triangle inequality readily follows. The proofs are elementary, employing nothing more sophisticated than similarity of triangles. The first expresses the area K of the triangle in terms of the three side lengths a, b, c and the circumradius R. The second, whose proof uses a rectangle composed of triangles similar to the right triangles in Figure 1(b), expresses the product xyz in terms of the inradius r and the sum $x + y + z$. The third gives the area K in terms of r and $x + y + z$.

LEMMA 1. $4KR = abc$.

Proof.



$$\frac{h}{b} = \frac{a/2}{R} \quad \Rightarrow \quad h = \frac{1}{2}\frac{ab}{R}$$

$$\therefore K = \frac{1}{2}hc = \frac{1}{4}\frac{abc}{R}$$

Figure 2 $4KR = abc$

LEMMA 2. $xyz = r^2(x + y + z)$.

Proof. Letting w denote $\sqrt{r^2 + x^2}$, we have

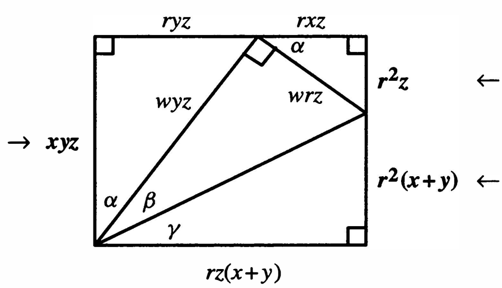


Figure 3 $xyz = r^2(x + y + z)$

LEMMA 3. $K = r(x + y + z)$.

Proof.

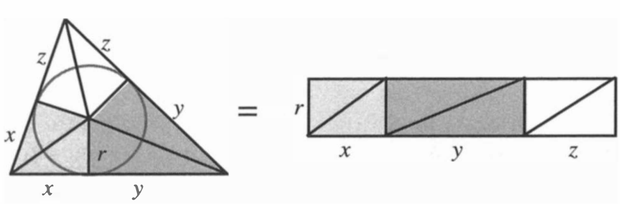


Figure 4 $K = r(x + y + z)$

We now prove

EULER'S TRIANGLE INEQUALITY. *In any triangle, the circumradius R and the inradius r satisfy $R \geq 2r$.*

Proof. Applying Lemma 1 to (2) yields $4KR \geq 8xyz$; invoking Lemma 2 then gives $4KR \geq 8r^2(x+y+z)$; and with Lemma 3 we have $4KR \geq 8Kr$; from which $R \geq 2r$ follows. ■

We conclude with a few comments about several results related to Euler's triangle inequality and the three lemmas used in our proof.

1. Euler's triangle inequality cannot be improved for general triangles, since $R = 2r$ if and only if the triangle is equilateral. However, for the class of right triangles, we have $R \geq (1+\sqrt{2})r$ with equality for isosceles right triangles. In fact, if one fixes one of the angles of the triangle, say α, then $R \sin\alpha \geq r\big(\tan(\alpha/2) + \sec(\alpha/2)\big)$. We leave the proofs of these inequalities as an exercise.
2. Since $2x = b + c - a$, $2y = c + a - b$, and $2z = a + b - c$, (1) can be written entirely in terms of a, b, and c as

$$abc \geq (a+b-c)(c+a-b)(b+c-a).$$

   This is known both as the *Lehmus inequality* [**1**] and *Padoa's inequality* [**5**], [**6**].
3. Lemmas 2 and 3 can be employed to produce a proof of *Heron's formula* for the area of the triangle: $K = \sqrt{s(s-a)(s-b)(s-c)}$, where s denotes the semiperimeter, and is given by $s = (a+b+c)/2 = x+y+z$. Since $s-a = x$, $s-b = y$, and $s-c = z$, the result in Lemma 2 can be written as $r^2 s = (s-a)(s-b)(s-c)$, or $(rs)^2 = s(s-a)(s-b)(s-c)$. But from Lemma 3 we have $rs = K$, from which Heron's formula for K now follows (for a wordless version of this proof and additional references, see [**4**]).
4. Dividing both sides of the result in Lemma 2 by r^3 yields $\frac{x}{r}\cdot\frac{y}{r}\cdot\frac{z}{r} = \frac{x}{r}+\frac{y}{r}+\frac{z}{r}$, which proves the following: If α, β, and γ are any three positive angles whose sum is $\pi/2$, then

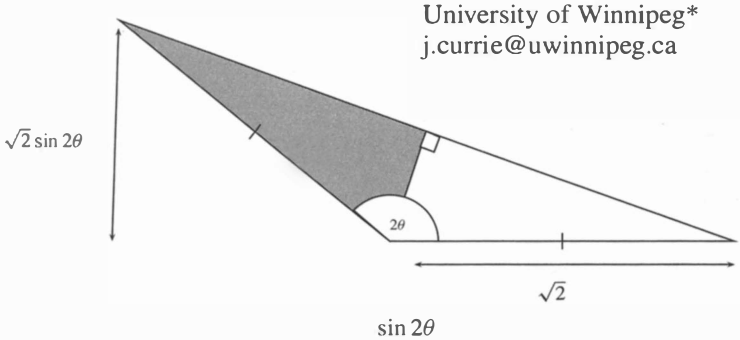
$$\cot\alpha\cot\beta\cot\gamma = \cot\alpha + \cot\beta + \cot\gamma.$$

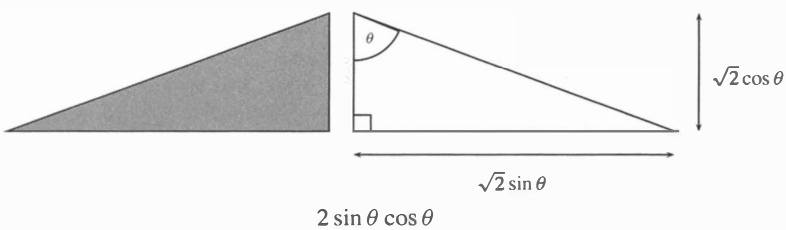
REFERENCES

1. H. S. M. Coxeter, "The Lehmus inequality," *Aequationes Mathematicae* **28** (1985) 1–12.
2. L. Euler, "Solutio facilis problematum quorundam geometricorum difficillimorum," *Novi Commentarii academiae scientiarum imperialis Petropolitanae* **11** (1767) 103–123. Reprinted in *Opera Omnia*, **I 26**, 139–157, 1953.
3. M. S. Klamkin, "On some geometric inequalities," *Mathematics Teacher* **60** (1967) 323–328.
4. R. B. Nelsen, "Heron's formula via proofs without words," *College Mathematics Journal* **32** (2001) 290–292.
5. R. B. Nelsen, "Proof without words: Padoa's inequality," this MAGAZINE **79** (2006) 53.
6. A. Padoa, "Una questione di minimo," *Periodico di Matematiche* **4** (1925) 80–85.
7. D. Schattschneider (editor), "A Tribute to Leonhard Euler 1707–1783," this MAGAZINE **56** (1983) 257–331.

Proof Without Words: Double Angle Formula via Area

James D. Currie
Department of Mathematics & Statistics
University of Winnipeg*
j.currie@uwinnipeg.ca





*The author is supported by an NSERC Discovery grant

PROBLEMS

Proposals

To be considered for publication, solutions should be received by July 1, 2007.

1786. *Proposed by Marian Tetiva, Bîrlad, Romania.*

Let $n \geq 2$ be a positive integer and let $O_n = \{1, 3, \ldots, 2n-1\}$ be the set of odd positive integers less than or equal to $2n-1$.

a. Prove that if m is a positive integer with $3 \leq m \leq n^2$ and $m \neq n^2 - 2$, then m can be written as a sum of distinct elements from O_n.

b. Prove that $n^2 - 2$ cannot be written as a sum of distinct elements of O_n.

1787. *Proposed by Ovidiu Bagdasar, Babes Bolyai University, Cluj Napoca, Romania.*

Let k and n be positive integers with $k \leq n$, and let $a_1, a_2, \ldots, a_n$ be nonnegative real numbers. Prove that

$$a_1 a_2 \cdots a_k + a_2 a_3 \cdots a_{k+1} + \cdots + a_{n-k+1} a_{n-k+2} \cdots a_n \leq \left(\frac{a_1 + a_2 + \cdots + a_n}{k} \right)^k.$$

1788. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let D be a nonempty compact set of real numbers, let $\{f_n\}$ be a sequence of real valued functions on D, and let f be a real valued function defined on D. Suppose that $\lim_{n\to\infty} f_n(x_n) = f(x)$ for any sequence $\{x_n\}$ in D with $x_n \to x \in D$.

a. Must it be the case that $f_n \to f$ uniformly on D?

b. Must it be the case that f is continuous on D?

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.

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

1789. *Proposed by Harris Kwong, SUNY Fredonia, Fredonia, NY.*

For nonzero real numbers $a_1, a_2, \ldots, a_n$, define $s = \sum_{k=1}^{n} \frac{1}{a_k}$ and

$$A = \begin{pmatrix} t + a_1 & t & \cdots & t \\ t & t + a_2 & \cdots & t \\ \vdots & \vdots & \ddots & \vdots \\ t & t & \cdots & t + a_n \end{pmatrix},$$

where t is a real number with $st \neq -1$. Find A^{-1} and $\det(A)$.

1790. *Proposed by Erwin Just (Emeritus), Bronx Community College of the City University of New York, Bronx, NY.*

Let R be a ring and assume that for each $x \in R$,

$$x + x^2 + x^3 + x^4 = x^{11} + x^{12} + x^{13} + x^{28}.$$

Prove that there is an integer $N > 1$ such that for each $x \in R$, we have $x = x^N$.

Quickies

Answers to the Quickies are on page 68.

Q977. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let $\{x_k\}_{k=0}^{\infty}$ be an infinite sequence of real numbers for which there exist positive constants a, b, and c, with $a + b + c < 1$, such that

$$|x_{n+1} - x_{m+1}| \leq a|x_n - x_{n+1}| + b|x_n - x_m| + c|x_m - x_{m+1}|,$$

for all nonnegative integers m and n. Prove that $\{x_k\}$ converges.

Q978. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let a, b, c be the side lengths of a triangle and let $x \geq 1$. Prove that

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

Solutions

Sets of the same cardinality **February 2007**

1761. *Proposed by Steve Butler, University of California San Diego, La Jolla, CA.*

For integer $n \geq 2$ define the sets

$$A(n) = \{(k, l) : 1 \leq k \leq l \leq n, k + l \leq n, \text{and} \gcd(k, l) = 1\}$$

$$B(n) = \{(k, l) : 1 \leq k \leq l \leq n, k + l > n, \text{and} \gcd(k, l) = 1\},$$

where $\gcd(k, l)$ denotes the greatest common divisor of the integers k and l. Prove that $A(n)$ and $B(n)$ have the same cardinality.

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

For $n \geq 2$, define $f : A(n) \to B(n)$ by

$$f(u, v) = (v, u + jv),$$

where $j = j(u, v)$ is the unique positive integer with $u + jv \leq n$ and $u + (j + 1)v > n$. For $(u, v) \in A(n)$, $\gcd(u, v) = 1$, so $\gcd(v, u + jv) = 1$. It then follows from the definition of j that $(v, u + jv) \in B(n)$. Thus, it suffices to show that f is one-to-one and onto $B(n)$.

- **one-to-one:** Suppose that $f(u, v) = (v, u + jv) = (q, p + iq) = f(p, q)$. Then $v = q$ and $u + jv = p + iq$. Assume, without loss of generality, that $p \leq u$. Then

$$u - p = (i - j)q. \tag{1}$$

  However, from $1 \leq p \leq q \leq n$ and $1 \leq u \leq q \leq n$, it follows that $0 \leq u - p \leq q - 1$. This contradicts (1) unless $i = j$. Thus, $(u, v) = (p, q)$, so f is one-to-one.
- **onto:** Given $(x, y) \in B(n)$ let $(u, v) = (y - j_0 x, x)$ where j_0 is the unique positive integer satisfying $1 \leq y - j_0 x \leq x$. Note that we then have $1 \leq u \leq v \leq n$, $u + v = x + (y - j_0 x) \leq y \leq n$, and $\gcd(u, v) = \gcd(x, y) = 1$, so $(u, v) \in A(n)$. In addition we have

$$u + j_0 v = y \leq n \qquad \text{and} \qquad u + (j_0 + 1)v = x + y > n.$$

  Thus $f(u, v) = (v, u + j_0 v) = (x, y)$. Hence, f maps onto $B(n)$.

II. *Solution by Robert L. Doucette, McNeese State University, Lake Charles, LA.*

The cardinality of the set $A(n) \cup B(n)$ is the number of relatively prime pairs in the array

$$\begin{array}{ccccc} (1,1) & (1,2) & \cdots & (1,n-1) & (1,n) \\ & (2,2) & \ldots & (2,n-1) & (2,n) \\ & & & & \vdots \\ & & & & (n,n) \end{array}$$

The number of pairs of relatively prime numbers in column m is $\phi(m)$, where ϕ denotes Euler's totient function. Hence

$$|A(n) \cup B(n)| = \sum_{m=1}^{n} \phi(m).$$

For $2 \leq m \leq 2n$, let d_m (the $m - 1$-st diagonal) be the set of ordered pairs (u, v) in the above array with $u + v = m$. We count the number of elements in $A(n)$ by counting, on diagonal d_m, $3 \leq m \leq n$, the relatively prime pairs $(k, m - k)$ with $1 \leq k \leq m - k$:

$$\begin{aligned} &|\{(k, m-k) \,:\, 1 \leq k \leq m-k,\ \gcd(k, m-k) = 1\}| \\ &\qquad = \frac{1}{2}\,|\{(k, m-k) \,:\, 1 \leq k \leq m-1,\ \gcd(k, m-k) = 1\}| \\ &\qquad = \frac{1}{2}\,|\{(k, m-k) \,:\, 1 \leq k \leq m-1,\ \gcd(k, m) = 1\}| \\ &\qquad = \frac{1}{2}\phi(m), \end{aligned}$$

where we have used the fact that $\gcd(k, m - k) = \gcd(k, m)$. Adding one for the pair $(1, 1)$, it follows that

$$|A(n)| = 1 + \frac{1}{2}\sum_{m=3}^{n} \phi(m) = \frac{1}{2}|A(n) \cup B(n)|.$$

Because $A(n)$ and $B(n)$ are disjoint, this completes the proof.

Also solved by Michael Andreoli, Michel Bataille (France), Jean Bogaert (Belgium), Robert Calcaterra, CMC 328, John Christopher, Commeca Problem Group, Jim Delany, Joe DeMaio and Andy Lightcap, Fejéntaláltuka Szeged Problem Solving Group (Hungary), Dmitry Fleishman, G.R.A.20 Problem Solving Group (Italy), Peter Gressis and Dennis Gressis, Russell Jay Hendel, Harris Kwong, Peter W. Lindstrom, Graham Lord, José H. Nieto (Venezuela), Northwestern University Math Problem Solving Group, Michael Reid, Nicholas C. Singer, Albert Stadler (Switzerland), Marian Tetiva (Romania), Paul Weisenhorn (Germany), and the proposer.

Modular Goldbach **February 2007**

1762. *Proposed by Erwin Just (Emeritus), Bronx Community College of the City University of New York, New York, NY.*

Let n be in integer with $n \geq 2$. Prove that for any even integer k, there exist odd primes p and q such that $p + q \equiv k \pmod{n}$.

Solution by Michel Bataille, Rouen, France.

Let $p_1, p_2, \ldots, p_r$ be the prime divisors of n. For $i = 1, 2, \ldots, r$, the prime p_i cannot divide both the odd integers $k - 1$ and $k + 1$ (otherwise p_i would be odd and would divide $k + 1 - (k - 1) = 2$.) Let $a_i \in \{k - 1, k + 1\}$ be such that p_i does not divide a_i. From the Chinese Remainder Theorem, there exists $a \in \mathbb{Z}$ satisfying $a \equiv a_i \pmod{p_i}$ for $i = 1, 2, \ldots, r$. For such an integer a, we have

$$a(k - a) \equiv a_i(k - a_i) \pmod{p_i}$$

and, since $a_i \not\equiv 0 \pmod{p_i}$ and $k - a_i = 1$ or -1, we see that $a(k - a) \not\equiv 0 \pmod{p_i}$ for $i = 1, 2, \ldots, r$. As a result, a and $b = k - a$ are both coprime to n.

Now, by Dirichlet's Theorem, there exist infinitely many primes in the arithmetic progressions $a + \lambda n$ and $b + \lambda n$ ($\lambda = 0, 1, 2, \ldots$). In particular, one can find odd primes p, q such that $p \equiv a \pmod{n}$ and $q \equiv b \pmod{n}$. This completes the proof since these primes p, q satisfy

$$p + q \equiv a + b = k \pmod{n}.$$

Also solved by Brian D. Beasley, John Christopher, Commeca Problem Group, Chip Curtis, Ron Dotzel, G.R.A.20 Problem Solving Group (Italy), Russell Jay Hendel, Peter W. Lindstrom, Jerry Metzger, Albert Stadler (Switzerland), Marian Tetiva (Romania), Doug Wilcox, and the proposer. There was one incorrect submission.

The volume of hull **February 2007**

1763. *Proposed by Joshua T. Wood and William P. Wardlaw, U. S. Naval Academy, Annapolis, MD.*

Let ℓ_1 and ℓ_2 be two lines in three space, let the distance between ℓ_1 and ℓ_2, measured along a mutual perpendicular to both lines, be d, and let θ be the angle determined by the direction vectors of ℓ_1 and ℓ_2. A line segment of length a lies on ℓ_1 and a line segment of length b lies on ℓ_2. Determine the volume of the convex hull of these two segments.

Solution by Jim Delany, Emeritus, California Polytechnic State University, San Luis Obispo, CA.

Without loss of generality we may assume that ℓ_1 is the x-axis and that the mutual perpendicular to ℓ_1 and ℓ_2 is the z-axis, chosen so that ℓ_2 intersects it at $(0, 0, d)$. Then one equation for ℓ_2 is $\mathbf{r}(t) = (t\cos\phi, t\sin\phi, d)$ where ϕ is either θ or $-\theta$, depending on the orientations of ℓ_1 and ℓ_2.

Suppose that the end points of the line segment on ℓ_1 are $P(a_1, 0, 0)$ and $Q(a_2, 0, 0)$ with $a_2 - a_1 = a$, and the end points of the segment on ℓ_2 are $R(b_1\cos\phi, b_1\sin\phi, d)$ and $S(b_2\cos\phi, b_2\sin\phi, d)$ where $b_2 - b_1 = b$. The convex hull of these four points is the tetrahedron $PQRS$. Its volume is

$$\frac{1}{6}|\overrightarrow{PQ}\cdot(\overrightarrow{PR}\times\overrightarrow{PS})| = \frac{1}{6}\,|(a_2-a_1)(b_2-b_1)d\sin\theta| = \frac{1}{6}abd|\sin\theta|.$$

Also solved by Michel Bataille (France), Jean Bogaert (Belgium), Herman Bubbert, Robert Calcaterra, Chip Curtis, Knut Dale (Norway), Euler's FOILers, Peter Gressis and Dennis Gressis, Kim McInturff, José H. Nieto (Venezuela), Paul Weisenhorn (Germany), and the proposers. There was one incorrect submission.

Euler-Mascheroni meets *e*. February 2007

1764. *Proposed by Ovidiu Furdui, student, Western Michigan University, Kalamazoo MI.*

For positive integer n, let $g_n = 1+\frac{1}{2}+\cdots+\frac{1}{n}-\ln n$. Prove that

$$\lim_{n\to\infty}\left(\frac{g_n^\gamma}{\gamma^{g_n}}\right)^{2n} = \frac{e}{\gamma},$$

where γ is the Euler-Mascheroni constant.

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

Euler's summation formula yields the well known estimate

$$g_n = 1+\frac{1}{2}+\cdots+\frac{1}{n}-\ln n = \gamma+\frac{1}{2n}+O\left(\frac{1}{n^2}\right).$$

(See *Concrete Mathematics* by Ronald Graham, Donald Knuth, and Oren Patashnik, 2nd edition, Addison Wesley, New York, 1994.) Thus $g_n = \gamma+\epsilon_n$, where $\epsilon_n = \frac{1}{2n}+O\left(\frac{1}{n^2}\right)$ and $2n\epsilon_n = 1+O\left(\frac{1}{n}\right)$. We then have

$$\left(\frac{g_n^\gamma}{\gamma^{g_n}}\right)^{2n} = \left(\frac{(\gamma+\epsilon_n)^\gamma}{\gamma^{\gamma+\epsilon_n}}\right)^{2n} = \frac{\left(1+\frac{2n\epsilon_n}{2\gamma n}\right)^{2\gamma n}}{\gamma^{2n\epsilon_n}} = \frac{\left(1+\frac{1+O\left(\frac{1}{n}\right)}{2\gamma n}\right)^{2\gamma n}}{\gamma^{1+O\left(\frac{1}{n}\right)}}.$$

Because

$$\lim_{n\to\infty}\left(1+\frac{1+O\left(\frac{1}{n}\right)}{2\gamma n}\right)^{2\gamma n} = \lim_{n\to\infty}\left(1+\frac{1}{2\gamma n}\right)^{2\gamma n} = e \quad\text{and}\quad \lim_{n\to\infty}\gamma^{1+O\left(\frac{1}{n}\right)} = \gamma,$$

the result follows.

Also solved by Michel Bataille (France), Michael S. Becker, Gerald E. Bilodeau, Jean Bogaert (Belgium), Paul Bracken, Brian Bradie, Erhard Braüne (Austria), Ghinea Catalin (Hungary), Hongwei Chen, Chip Curtis, Knut Dale (Norway), David Doster, Robert L. Doucette, Alex Fok (China), G.R.A.20 Problem Solving Group (Italy), Eugene A. Herman, Dan Jurca, Kee-Wai Lau (China), David Lovit, Paolo Perfetti (Italy), Zouk Mosbeh (Lebanon), John M. Sayer, Nicholas C. Singer, Albert Stadler (Switzerland), Marian Tetiva (Romania), Paul Weisenhorn (Germany), and the proposer.

Can you translate that? February 2007

1765. *Proposed by Eugene A. Herman, Grinnell College, Grinnell, IA.*

An object in 3-space is translated by a fixed vector **t** and then rotated using a rotation matrix whose axis of rotation has unit direction vector **a** and for which the angle of rotation in a plane perpendicular to **a** is $\theta = \frac{\pi}{n}$, where n is a positive integer. This translation-rotation move is repeated for a total of $2n$ times. When this is done, what are the position and orientation of the object relative to its initial position and orientation?

Solution by José H. Nieto, Universidad del Zulia, Maracaibo, Venezuela.

The object will be translated by the vector $2n(\mathbf{t}\cdot\mathbf{a})\mathbf{a}$ (i.e., $2n$ times the orthogonal projection of $\mathbf{t}$ on $\mathbf{a}$).

Let A denote the rotation matrix. Observe that A^{2n} is the identity. If $\mathbf{x}$ is the position vector of a point in the object, after the first translation-rotation it will be at position $A(\mathbf{x}+\mathbf{t}) = A\mathbf{x} + A\mathbf{t}$; after the second translation-rotation it will be at $A^2\mathbf{x} + A^2\mathbf{t} + A\mathbf{t}$; ... ; and after the $2n$th translation-rotation it will be at

$$A^{2n}\mathbf{x} + A^{2n}\mathbf{t} + A^{2n-1}\mathbf{t} + \cdots + A^2\mathbf{t} + A\mathbf{t} = \mathbf{x} + A^{2n-1}\mathbf{t} + \cdots + A^2\mathbf{t} + A\mathbf{t} + \mathbf{t}.$$

Now put $\mathbf{t} = \mathbf{u} + \mathbf{w}$, where $\mathbf{u} = (\mathbf{t}\cdot\mathbf{a})\mathbf{a}$ and $\mathbf{w} = \mathbf{t} - (\mathbf{t}\cdot\mathbf{a})\mathbf{a}$ is orthogonal to $\mathbf{a}$. Because $A\mathbf{a} = \mathbf{a}$ we have

$$A^{2n-1}\mathbf{u} + \cdots + A^2\mathbf{u} + A\mathbf{u} + \mathbf{u} = 2n\mathbf{u} = 2n(\mathbf{t}\cdot\mathbf{a})\mathbf{a}.$$

On the other hand $\mathbf{v} = A^{2n-1}\mathbf{w} + \cdots + A^2\mathbf{w} + A\mathbf{w} + \mathbf{w}$ must be $\mathbf{0}$, because it is orthogonal to the axis of rotation and $A\mathbf{v} = \mathbf{v}$. Therefore, after the $2n$-th translation-rotation, $\mathbf{x}$ will become

$$\begin{aligned}\mathbf{x} &+ (A^{2n-1}\mathbf{u} + \cdots + A^2\mathbf{u} + A\mathbf{u} + \mathbf{u}) + (A^{2n-1}\mathbf{w} + \cdots + A^2\mathbf{w} + A\mathbf{w} + \mathbf{w})\\ &= \mathbf{x} + 2n(\mathbf{t}\cdot\mathbf{a})\mathbf{a}.\end{aligned}$$

Also solved by Michel Bataille (France), Jean Bogaert (Belgium), Herman Bubbert, Robert Calcaterra, Chip Curtis, Jim Delany, Robert L. Doucette, G.R.A.20 Problem Solving Group (Italy), Jeffrey M. Groah, and the proposer. There were two incorrect submissions.

Answers

Solutions to the Quickies from page 64.

A977. Setting $m = n - 1$ in the given inequality, we obtain

$$|x_{n+1} - x_n| \le a|x_n - x_{n+1}| + b|x_n - x_{n-1}| + c|x_{n-1} - x_n|.$$

It follows that

$$|x_{n+1} - x_n| \le \left(\frac{b+c}{1-a}\right)|x_n - x_{n-1}|,$$

for all positive integers n. Because $0 < b + c < 1 - a$, we have $0 < \frac{b+c}{1-a} < 1$. Therefore $\{x_n\}$ is a contractive sequence, and hence converges.

A978. If $x = 1$ then the inequality is the well known triangle inequality. Thus we assume $x > 1$. Then

$$\frac{c^x}{a^x + b^x} \le \frac{(a+b)^x}{a^x + b^x} = \frac{\left(1 + \frac{b}{a}\right)^x}{1 + \left(\frac{b}{a}\right)^x}. \tag{1}$$

Let g be the function defined by

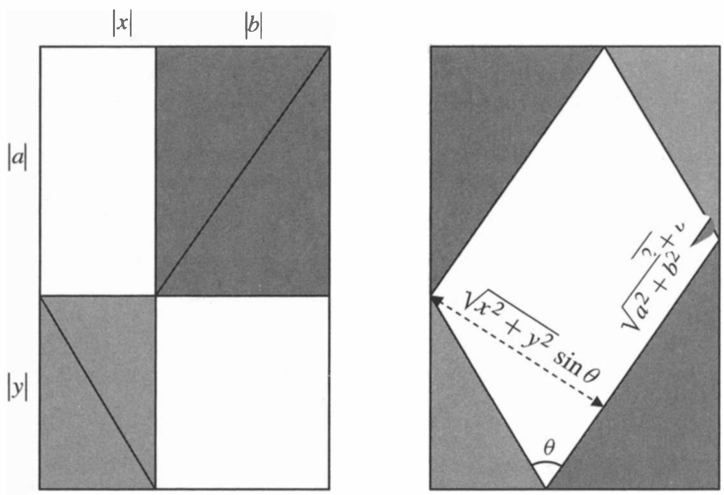
$$g(t) = \frac{(1+t)^x}{1+t^x}, \qquad t > 0.$$

An easy calculation shows that

$$g'(t) = \frac{x(t+1)^{x-1}(1-t^{x-1})}{(1+t^x)^2},$$

so the only critical point is at $t = 1$. Because $g(t) \to 1$ as $t \to 0$ and $t \to +\infty$ and $g(1) = 2^{x-1} > 1$, it follows that g assumes its maximum at $t = 1$. The desired inequality now follows from (1).

Proof Without Words: The Cauchy-Schwarz Inequality



$$|a|\,|x| + |b|\,|y| = \sqrt{a^2+b^2}\sqrt{x^2+y^2}\,\sin\theta$$

$$\Rightarrow \quad |\langle a, b\rangle \cdot \langle x, y\rangle| \le \|\langle a, b\rangle\|\,\|\langle x, y\rangle\|$$

REFERENCE

Roger B. Nelsen, *Proof Without Words*, Mathematical Association of America, Washington, D.C., 1993, p. 64.

SIDNEY H. KUNG
Cupertino, CA 95014
sidneykung@yahoo.com

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles, books, and other materials 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.

Netz, Reviel, and William Noel, *The Archimedes Codex: How a Medieval Prayer Book Is Revealing the True Genius of Antiquity's Greatest Scientist*, Da Capo Press, 2007; ix + 313 pp + 16 pp color photos, $27.50. ISBN 978-0-306-81580-5.

This book on the Archimedes Palimpsest is a collaboration between its leading interpreter (at Stanford) and its curator (at the Walters Museum in Baltimore). It is an absolutely fascinating tale of the rescue and significance of what Noel terms a "conservation disaster zone," with far more details on the provenance of the book than previously revealed. One can only regret its modern circumstances: theft from a monastery in Greece, acquisition by a collector who must have known that it had been stolen and who likely "enhanced" its marketability with forged illustrations obliterating some of the text, and sale to a family that let it molder in a basement for 75 years. Three pages, extant in 1906, now with forged illustrations painted over both the words of Archimedes and the medieval prayers over them, are still missing. It took three and a half years, with the utmost of care, just to separate the pages of the book. Its unidentified owner "Mr. B.," to whom we owe a great debt, continues to fund techniques to read its contents below the accumulated layers of ink and paint; the latest and most successful efforts have used the Stanford Linear Accelerator.

Deakin, Michael A.B., *Hypatia of Alexandria, Mathematician and Martyr*, Prometheus, 2007; 231 pp, $28. ISBN 978-1-59102-520-7.

This "first book-length biography to attempt an evaluation of Hypatia's mathematics" expands on an article by the author in the *American Mathematical Monthly* 101 (1994) 234–243. His task is not easy, since there are no extant works indisputably by her. Nevertheless, author Deakin sets out the historical, intellectual, and religious backgrounds to her life and attested achievements; relates the facts of her life and death; and tries to deduce what she may have written about mathematics. Appendices give mathematical background; a sketch of an earlier woman mathematician, Pandrosion, at Alexandria; and (quite usefully) new translations of the primary sources about Hypatia. (I take some delight that what Deakin terms "the best accessible summary of all" about Hypatia, by Ian Mueller, appeared in a book that I co-edited 20 years ago.)

Ruelle, David, *The Mathematician's Brain*, Princeton University Press, 2007; ix + 160 pp, $22.95. ISBN 978-0-691-12982-2.

In 23 essays, author Ruelle concentrates on an excellent presentation for the general reader of the "formal and structural aspects" of mathematics, including psychological aspects; but even though he is a mathematical physicist, he scarcely mentions applied mathematics. Most mathematicians will recognize their subject in his succinct portrait of it, even if they disagree with some of his opinions. Two essays have a narrower focus, on algebraic geometry and the fate of Alexander Grothendieck. Ruelle was at IHES (Institute des Hautes Études Scientifiques) in France with Grothendieck, and he laments the loss when Grothendieck "abandoned mathematics" or "was abandoned by it"; a few pages later, Ruelle calls "the disposal of Grothendieck" a "disgrace in the history of twentieth-century mathematics."

Diaconis, Persi, Susan Holmes, and Richard Montgomery, Dynamical bias in the coin toss, *SIAM Review* 49 (2) (2007) 211–235.

Want to have an edge in winning a coin toss? Bet on the side facing up before the flip! (If the coin is turned over after catching, bet on the down side.) You can expect to win about 51% of the time. The authors' inspiration was the ability of magicians to "flip" coins so that the toss appears normal but the coin never turns over, and their own construction of a coin-tossing machine that always produces the same result ("coin tossing is 'physics' not 'random'"). The authors model the physics of coin-tossing, prove theorems about the outcome, and estimate parameters of the model from filmed human tosses. The crux of their analysis is to take into account precession (change of axis of rotation) of the coin; they neglect air resistance (a flip lasts 1/6 second). What are probabilists to make of this iconoclasm, which appears to "behead" the coin toss as the prototypical random event? Apparently contrary to their analysis, however, they conclude that "The classical assumptions with probability 1/2 are pretty solid."

Segal, Mark, Chess, chance, and conspiracy, *Statistical Science* 22 (1007) (1) 98–108.

Former chess champion Bobby Fischer has claimed that the 1985 championship match Karpov vs. Kasparov was fixed, move by move. This paper focuses on a basis for that claim, the fact that in one game White made 18 consecutive moves of pieces on the light squares. The paper applies methods for analyzing the probabilities of runs in Bernoulli trials, including the non-identically-distributed case of varying success probabilities (via embedding into a Markov chain). The authors also compare the moves made with evaluations by computer chess programs and search game databases for similar runs. They conclude that the run in question was not so remarkable and end with the unnecessary and libelous suggestion that "perhaps" Fischer's own 19 consecutive wins en route to the championship match "was part of some conspiracy."

Larsen, Mogens Esrom, *Summa Summarum*, A K Peters, 2007; xii + 232 pp, $49. ISBN 978-1-56881-323-3.

It's a thin book but it "aims to provide ... a collection, of all known algebraic sums and a guide to find the sum you need." In other words, it is a compendium of binomial (and other) identities, classified in an unconventional way that avoids hypergeometric functions. Maybe if I look hard, I will find a result for a particular sum that I have been seeking for some years....

Almeida, D.F., G.G. Joseph, and J. Penzel, J. ("Aryabhata Group"), Transmission of the calculus from Kerala to Europe, in *Proceedings of the International Seminar and Colloquium on 1500 Years of* Aryabhateeyam (Kerala Sastra Sahitya Parishad, Kochi, India, 2003), edited by G.G. Joseph, 33–48. Joseph, George Gheverghese, Infinite series in Kerala: Background and motivation, *ibid.* Almeida, Dennis F., and George G. Joseph, Eurocentrism in the history of mathematics: The case of the Kerala school, *Race and Class* 45 (4) (2004) 45–59. Almeida, D.F., J.K. John, and A. Zadorozhnyy, Kerala mathematics: Its possible transmission to Europe and the consequential educational implications, *Journal of Natural Geometry* 20 (2001) 77–104. Raju, C.K., *Cultural Foundations of Mathematics: The Nature of Mathematical Proof and the Transmission of the Calculus from India to Europe in the 16th c. CE*, Longman, New Delhi, 2007. Bressoud, David, Was calculus invented in India?, *College Mathematics Journal* 33 (1) (2002) 2–13. Kerala school of astronomy and mathematics, Wikipedia, `http://en.wikipedia.org/wiki/Kerala_school_of_astronomy_and_mathematics`.

George G. Joseph is known for *The Crest of the Peacock: Non-European Roots of Mathematics* (2nd ed., Princeton University Press, 2000), which claims that histories of mathematics are eurocentrically biased. Born in Kerala, he no doubt has unusual interest in the achievements of Keralese mathematicians (14th to 16th centuries). They derived series expansions for sine, cosine, and arctangent, for computational rather than geometric purposes. Joseph, Almeida, and co-authors tantalize that Keralese discoveries *could* have been brought to the West by Jesuits or others. But they offer only wishful thinking, no evidence. Their papers, published out of the mainstream, are hard to find (write `g.g.joseph@exeter.ac.uk` for copies). I have not yet seen the book by Raju (who accuses Joseph and Almeida of plagiarizing his work, a claim that they reject). And "No," Bressoud begins and concludes, three series do not amount to calculus.

NEWS AND LETTERS

68th Annual William Lowell Putnam Mathematical Competition

Editor's Note: Additional solutions will be printed in the *Monthly* later in the year.

PROBLEMS

A1. Find all values of α for which the curves $y = \alpha x^2 + \alpha x + \frac{1}{24}$ and $x = \alpha y^2 + \alpha y + \frac{1}{24}$ are tangent to each other.

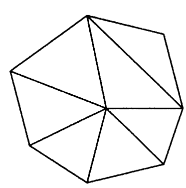
A2. Find the least possible area of a convex set in the plane that intersects both branches of the hyperbola $xy = 1$ and both branches of the hyperbola $xy = -1$. (A set S in the plane is called *convex* if for any two points in S the line segment connecting them is contained in S.)

A3. Let k be a positive integer. Suppose that the integers $1, 2, 3, \dots, 3k + 1$ are written down in random order. What is the probability that at no time during this process, the sum of the integers that have been written up to that time is a positive integer divisible by 3? Your answer should be in closed form, but may include factorials.

A4. A *repunit* is a positive integer whose digits in base 10 are all ones. Find all polynomials f with real coefficients such that if n is a repunit, then so is $f(n)$.

A5. Suppose that a finite group has exactly n elements of order p, where p is a prime. Prove that either $n = 0$ or p divides $n + 1$.

A6. A *triangulation* $\mathcal{T}$ of a polygon P is a finite collection of triangles whose union is P, and such that the intersection of any two triangles is either empty, or a shared vertex, or a shared side. Moreover, each side of P is a side of exactly one triangle in $\mathcal{T}$. Say that $\mathcal{T}$ is *admissible* if every internal vertex is shared by 6 or more triangles. For example



Prove that there is an integer M_n, depending only on n, such that any admissible triangulation of a polygon P with n sides has at most M_n triangles.

B1. Let f be a polynomial with positive integer coefficients. Prove that if n is a positive integer, then $f(n)$ divides $f(f(n) + 1)$ if and only if $n = 1$.

B2. Suppose that $f : [0, 1] \to \mathbb{R}$ has a continuous derivative and that $\int_0^1 f(x)\,dx = 0$. Prove that for every $\alpha \in (0, 1)$,

$$\left| \int_0^{\alpha} f(x)\,dx \right| \leq \frac{1}{8} \max_{0 \leq x \leq 1} |f'(x)|.$$

B3. Let $x_0 = 1$ and for $n \geq 0$, let $x_{n+1} = 3x_n + \lfloor x_n\sqrt{5}\rfloor$. In particular, $x_1 = 5$, $x_2 = 26$, $x_3 = 136$, $x_4 = 712$. Find a closed-form expression for x_{2007}. ($\lfloor a\rfloor$ means the largest integer $\leq a$.)

B4. Let n be a positive integer. Find the number of pairs P, Q of polynomials with real coefficients such that

$$P(X)^2 + Q(X)^2 = X^{2n} + 1$$

and $\deg P > \deg Q$.

B5. Let k be a positive integer. Prove that there exist polynomials $P_0(n), P_1(n), \ldots, P_{k-1}(n)$ (which may depend on k) such that for any integer n,

$$\left\lfloor \frac{n}{k} \right\rfloor^k = P_0(n) + P_1(n)\left\lfloor \frac{n}{k} \right\rfloor + \cdots + P_{k-1}(n)\left\lfloor \frac{n}{k} \right\rfloor^{k-1}.$$

B6. For each positive integer n, let $f(n)$ be the number of ways to make $n!$ cents using an unordered collection of coins, each worth $k!$ cents for some k, $1 \leq k \leq n$. Prove that for some constant C, independent of n,

$$n^{n^2/2-Cn}e^{-n^2/4} \leq f(n) \leq n^{n^2/2+Cn}e^{-n^2/4}.$$

SOLUTIONS

Solution to A1. (Based on a student solution) From the first curve we get $\frac{dy}{dx} = 2\alpha x + \alpha$ and from the second, $\frac{dx}{dy} = 2\alpha y + \alpha$. We conclude that at a point of tangency, $2\alpha x + \alpha = \dfrac{1}{2\alpha y + \alpha}$, or equivalently, $\alpha^2(2x+1)(2y+1) = 1$. The two curves are parabolas, one the reflection of the other across the line $y = x$, and we conclude that a point of tangency must lie on this line, so $\alpha^2(2x+1)^2 = 1$, $\alpha(2x+1) = \pm 1$, $x = \dfrac{\pm 1 - \alpha}{2\alpha}$. However, we can also solve $x = \alpha x^2 + \alpha x + 1/24$ to find $x = \frac{1-\alpha\pm\sqrt{\alpha^2-\frac{13}{6}\alpha+1}}{2\alpha}$. Hence, if $x = \dfrac{1-\alpha}{2\alpha}$, $\sqrt{\alpha^2 - \frac{13}{6}\alpha + 1} = 0$. If $x = \dfrac{-1-\alpha}{2\alpha}$, $\sqrt{\alpha^2 - \frac{13}{6}\alpha + 1} = 2$. Solving in both cases gives $\alpha = \frac{2}{3}, \frac{3}{2}, \frac{13\pm\sqrt{601}}{12}$.

Solution to A2. The convex set must contain a quadrilateral Q with a vertex on each branch of the hyperbolas, so it suffices to minimize the area of the quadrilateral Q. Let A, B, C, D be the vertices of Q on the respective branches of the hyperbola in quadrants 1, 2, 3, 4, respectively. Because side AD of the quadrilateral intersects the x-axis to the right of the origin O, and side BC does to the left, O must be inside the quadrilateral. The area of triangle AOD, $A = (a, 1/a)$, $D = (d, -1/d)$, is

$$\frac{1}{2}\det\begin{pmatrix} d & a \\ -1/d & 1/a \end{pmatrix} = \frac{1}{2}\left(\frac{d}{a} + \frac{a}{d}\right) \geq 1$$

(since $x + 1/x \geq 2$ for all positive x). The quadrilateral region can be divided into four such triangular regions, so the area of the quadrilateral is at least 4. But the square with vertices $(\pm 1, \pm 1)$ has area 4, so we're done.

Solution to A3. The number of ways to write down $1, 2, 3, \ldots, 3k+1$ in random order is $(3k+1)!$, so we want to count the number of ways in which none of the "partial sums" is divisible by 3. First, consider the integers modulo $3: 1, 2, 0, 1, 2, 0, \ldots, 1, 2, 0, 1$. To write these with none of the partial sums divisible by 3, we must start with a 1 or a 2. After that, we can include or omit 0's at will without affecting whether any of the partial sums are divisible by 3, so suppose we omit all 0's. The remaining sequence of 1's and 2's must then be of the form

$$1, 1, 2, 1, 2, 1, 2, \ldots$$

or

$$2, 2, 1, 2, 1, 2, 1, \ldots$$

(once you start, the rest of the sequence is forced by the condition that no partial sum is divisible by 3). However, a sequence of the form 2, 2, 1, 2, 1, 2, 1, ... has one more 2 than 1, and we need to have one more 1 than 2. So the *only* possibility for our sequence modulo 3, once the 0's are omitted, is 1, 1, 2, 1, 2, 1, 2, ... There are $2k + 1$ numbers in this sequence, and the k 0's can be returned to the sequence arbitrarily except at the beginning. So the number of ways to form the complete sequence modulo 3 equals the number of ways to distribute the k identical 0's over $2k + 1$ boxes (the "slots" after the 1's and 2's), which by a standard "stars and bars" argument is $\binom{3k}{k}$. Once this is done, there are $k!$ ways to replace the k 0's in the sequence modulo 3 by the actual integers $3, 6, \ldots, 3k$. Also, there are $k!$ ways to "reconstitute" the 2's and $(k+1)!$ ways for the 1's. So the answer is

$$\frac{1}{(3k+1)!}\binom{3k}{k}k!\,k!\,(k+1)! = \frac{k+1}{3k+1}\cdot\frac{k!\,k!}{(2k)!} = \frac{k+1}{3k+1}\cdot\binom{2k}{k}^{-1}.$$

Solution to A4. Clearly, any constant polynomial whose value is a repunit will do. We'll show that the nonconstant polynomials with the prescribed property are precisely those of the form $f(X) = \dfrac{(9X+1)^d\cdot 10^\ell - 1}{9}$ for positive integer constants d and ℓ.

Let f be such a polynomial. From the hypothesis, there exists a sequence $(a_n)_{n\ge 1}$ of positive integers such that $f\left(\dfrac{10^n-1}{9}\right) = \dfrac{10^{a_n}-1}{9}$. Let $\deg f = d \ge 1$. Then there is a nonzero number A such that $f(x) \sim Ax^d$ as $x \to \infty$. Therefore $f\left(\dfrac{10^n-1}{9}\right) \sim \dfrac{A}{9^d}\cdot 10^{nd}$. Thus, $10^{a_n} \sim \dfrac{A}{9^{d-1}}\cdot 10^{nd}$. This shows that the sequence $(a_n - nd)_{n\ge 1}$ converges to a limit ℓ such that $A = 9^{d-1}\cdot 10^\ell$. Because this sequence consists of integers, it eventually becomes equal to the constant sequence (ℓ). Thus, from a certain point we have $f\left(\dfrac{10^n-1}{9}\right) = \dfrac{10^{nd+\ell}-1}{9}$. If $x_n = \dfrac{10^n-1}{9}$, we deduce that the polynomial equation $f(x) = \dfrac{(9x+1)^d\cdot 10^\ell - 1}{9}$ has infinitely many solutions x_n, so $f(X) = \dfrac{(9X+1)^d\cdot 10^\ell - 1}{9}$. It is clear that all such polynomials satisfy the conditions of the problem, so we are done.

Solution to A5. Let G be the group, let A be the set of elements of order p, and let g be an element of A. Then the elements of A other than g and its powers can be partitioned into parts of size p as follows: If $h \in A$ commutes with g, then its part is the set of elements of the form $g^k h$. If it does not commute with g, then its part is the set of elements of the form $g^k h g^{-k}$. Since g has $p - 1$ nontrivial powers, the statement follows.

Solution to A6. We claim that the sequence (M_n) such that

$$M_3 = 1, \qquad M_n = M_{n-1} + \frac{n}{3} + 1$$

will do. By Euler's formula for a polygonal tiling on a planar graph, $V - E + F = 1$. We can say that each face and each vertex has Euler number 1, and each edge has Euler number -1, and the total for all elements is therefore 1. We can then redistribute the Euler numbers by donating $\frac{1}{3}$ from each triangle to each of its 3 vertices, and donating $-\frac{1}{2}$ from each edge to each of its 2 vertices. The vertices then come in three kinds: positive, negative, and zero, and their total is still 1. The admissibility condition says that interior vertices are nonpositive.

Therefore the total adjusted Euler number of the boundary vertices is at least 1. The adjusted Euler number of one boundary vertex is $\frac{1}{2} - \frac{t}{6}$ if it is met by t triangles. If $t = 1$, then we can remove the one triangle that meets the vertex and reduce the number of sides of the polygon by 1, which confirms the claim. If there are no boundary vertices with $t = 1$, then there must be at least 6 with $t = 2$; indeed, there must be 6 pairs without any negative boundary vertices with $t > 3$ between them. At least one such pair is k sides apart with $k \leq \frac{n}{6}$, and only boundary vertices with $t = 3$ between them. Precisely $2k + 1$ triangles touch these k sides, and we can remove these triangles to again obtain a polygon with one fewer total sides.

Solution to B1. First observe that for integers m and n,

$$m - n \text{ divides } f(m) - f(n). \tag{1}$$

Assume that $f(n)$ divides $f(f(n)+1)$. Then, by (1), $f(n)$ divides $f(f(n)+1) - f(1)$, so $f(n)$ divides $f(1)$.

The converse is false if f is a constant polynomial, so assume that f is non-constant. Because f is a polynomial with positive integer coefficients, $f(n) > f(1)$ if $n > 1$, so, from above, $f(n)$ does not divide $f(f(n)+1))$. However, if $n = 1$, then (1) implies that $f(1)$ divides $f(f(1)+1) - f(1)$, so $f(1)$ divides $f(f(1)+1)$.

Solution to B2. Since the extreme values of $\int_0^\alpha f(x)\,dx$ (as a function of α) occur at values of α where

$$\frac{d}{d\alpha}\left(\int_0^\alpha f(x)\,dx\right) = f(\alpha) = 0,$$

we may assume that $f(\alpha) = 0$. Let $M = \max_{0\leq x\leq 1} |f'(x)|$. Then by the Mean Value Theorem, for $0 \leq x \leq \alpha$,

$$|f(x)| = |f(x) - f(\alpha)| \leq M|x - \alpha| = M(\alpha - x),$$

so

$$\left|\int_0^\alpha f(x)\,dx\right| \leq \int_0^\alpha |f(x)|\,dx \leq \int_0^\alpha M(\alpha - x)\,dx = \frac{1}{2}M\alpha^2.$$

Thus, if $\alpha \leq 1/2$, we have $\left|\int_0^\alpha f(x)\,dx\right| \leq \frac{1}{8}M$, and we are done.

If $\alpha > 1/2$, note that

$$\int_\alpha^1 f(x)\,dx = \int_0^1 f(x)\,dx - \int_0^\alpha f(x)\,dx = -\int_0^\alpha f(x)\,dx,$$

so

$$\left|\int_0^\alpha f(x)\,dx\right| \leq \int_\alpha^1 |f(x)|\,dx \leq \int_\alpha^1 M(x - \alpha)\,dx = \frac{1}{2}M(1-\alpha)^2 < \frac{1}{8}M,$$

and we are done.

Solution to B3. By factoring the first few terms, we see that

$$x_1 = 5 = 1\cdot 5, \qquad x_2 = 26 = 2\cdot 13, \qquad x_3 = 136 = 4\cdot 34, \qquad x_4 = 8\cdot 89,$$

which leads us to conjecture that $x_n = 2^{n-1}\cdot F_{2n+3}$ for $n \geq 1$. Here, F_m is the mth Fibonacci number, which is given by Binet's formula (or by solving the Fibonacci recurrence):

$$F_m = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^m - \left(\frac{1-\sqrt{5}}{2}\right)^m\right).$$

We will prove $x_n = 2^{n-1}\cdot F_{2n+3}$ by induction. If it is true for n, then

$$x_{n+1} = 3x_n + \lfloor x_n\sqrt{5}\rfloor = \lfloor (3+\sqrt{5})x_n\rfloor = \left\lfloor \left(\frac{1+\sqrt{5}}{2}\right)^2 x_n\cdot 2\right\rfloor$$

$$= \left\lfloor \left(\frac{1+\sqrt{5}}{2}\right)^2 \cdot \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2}\right)^{2n+3} - \left(\frac{1-\sqrt{5}}{2}\right)^{2n+3} \right) \cdot 2^n \right\rfloor$$

$$= \left\lfloor \left(F_{2n+5} + \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^{2n+3} \left(\left(\frac{1-\sqrt{5}}{2}\right)^2 - \left(\frac{1+\sqrt{5}}{2}\right)^2 \right) \right) \cdot 2^n \right\rfloor$$

$$= F_{2n+5} \cdot 2^n + \left\lfloor -\left(\frac{1-\sqrt{5}}{2}\right)^{2n+3} \cdot 2^n \right\rfloor.$$

But $-\left(\frac{1-\sqrt{5}}{2}\right)^{2n+3} \cdot 2^n = \frac{\sqrt{5}-1}{4} \cdot (3-\sqrt{5})^{n+1}$ is between 0 and 1, so $x_{n+1} = 2^n \cdot F_{2n+5}$ and the induction is done.

In particular, $x_{2007} = \frac{2^{2006}}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2}\right)^{4017} - \left(\frac{1-\sqrt{5}}{2}\right)^{4017} \right).$

Solution to B4. The expression is equivalent to a factorization

$$X^{2n} + 1 = \big(P(X) + i\,Q(X)\big)\big(P(X) - i\,Q(X)\big),$$

where the leading coefficient of $P(X) + i\,Q(X)$ is either 1 or -1. We can suppose it is 1, so that $P(X) + i\,Q(X)$ is monic, and multiply the answer by 2.

The roots of $X^{2n} + 1$ all have multiplicity 1 and occur as complex conjugate pairs. For each such pair, one root is a root of $P(X) + i\,Q(X)$ and the other is a root of $P(X) - i\,Q(X)$. The choices of $P(X) + i\,Q(X)$ amount to n binary choices, plus the choice of overall sign. Thus there are 2^{n+1} solutions.

Solution to B5. The question is equivalent to showing that $\left(\frac{n}{k} - \left\lfloor \frac{n}{k} \right\rfloor\right)^k$ is a linear combination of $\left(\frac{n}{k} - \left\lfloor \frac{n}{k} \right\rfloor\right)^j$ for $0 \le j \le k-1$. It's equivalent because you can expand all of the binomials and collect powers of $\left\lfloor \frac{n}{k} \right\rfloor$. Such a linear combination is plausible because all of the functions involved are periodic with period k. So, we look for coefficients A_i such that

$$A_0 + A_1\left(\frac{i}{k}\right) + A_2\left(\frac{i}{k}\right)^2 + \cdots + A_{k-1}\left(\frac{i}{k}\right)^{k-1} = \left(\frac{i}{k}\right)^k$$

for $0 \le i \le k-1$. The matrix of coefficients for this system of equations is the Vandermonde matrix $V_{ij} = \left(\frac{i}{k}\right)^j$, which is well-known to be nonsingular. Therefore, the system has a solution, and we are done.

Solution to B6. It is clear that $f(n)$ is just the number of nonnegative integer solutions of the equation $a_1 \cdot 1! + a_2 \cdot 2! + \cdots + a_n \cdot n! = n!$, which is the same as the number of solutions in nonnegative integers of the inequality $a_2 \cdot 2! + a_3 \cdot 3! + \cdots + a_{n-1} \cdot (n-1)! + a_n \cdot n! \le n!$. For any such solution different from $(0, 0, \ldots, 0, n!)$ we have $a_n = 0$ and we will consider the hypercube $H(a_2, a_3, \ldots, a_{n-1}) = [a_2, a_2 + 1) \times [a_3, a_3 + 1) \times \cdots \times [a_{n-1}, a_{n-1} + 1)$. It is clear that these hypercubes are disjoint for distinct $(a_2, \ldots, a_{n-1})$. So the number of solutions of the inequality is the total volume of these hypercubes. Now, observe that any such hypercube is included in the set of points $(x_2, \ldots, x_{n-1})$ with $x_i \ge 0$ and $\sum_{i=2}^{n-1}(x_i - 1)x! < n!$. Also, the union of these cubes covers the region consisting of those points $(x_2, \ldots, x_{n-1})$ with $x_i \ge 0$ and $\sum_{i=2}^{n-1} x_i \cdot i! \le n!$. Indeed, take a point $(x_2, \ldots, x_{n-1})$ in this region. Then $(\lfloor x_2 \rfloor, \ldots, \lfloor x_{n-1} \rfloor)$ is a solution of the inequality and the point belongs to the corresponding hypercube. Now, more generally, let us consider the region $R(a_1, a_2, \ldots, a_n; A)$ defined by the inequalities $x_i \ge 0$ and

$a_1x_1 + a_2x_2 + \cdots + a_nx_n \le A$. Its volume is

$$\begin{aligned}
\mathrm{Vol}(R(a_1,\ldots,a_n;A)) &= \int_{x_i\ge 0,\, a_1x_1+\cdots+a_nx_n\le A} dx_1\,dx_2\ldots dx_n \\
&= \int_{0\le x_n\le \frac{A}{a_n}} \int_{x_1,\ldots,x_{n-1}\ge 0,\, a_1x_1+\cdots+a_{n-1}x_{n-1}\le A-a_nx_n} dx_1\ldots dx_n \\
&= \int_0^{\frac{A}{a_n}} \mathrm{Vol}(R(a_1,\ldots,a_{n-1};A-a_nx_n))\,dx_n \\
&= \mathrm{Vol}(R(a_1,\ldots,a_{n-1};1))\cdot\int_0^{\frac{A}{a_n}} (A-a_nx_n)^{n-1}dx_n \\
&= \frac{A^n}{na_n}\cdot \mathrm{Vol}(R(a_1,\ldots,a_{n-1};1)).
\end{aligned}$$

This implies, by induction, that $\mathrm{Vol}(R(a_1,a_2,\ldots,a_n;A)) = \dfrac{A^n}{n!\cdot a_1a_2\cdots a_n}$. Thus, because the sum of the volumes of the hypercubes is between the volume of $R(2!,3!,\ldots,(n-1)!;n!)$ and $R(2!,3!,\ldots,(n-1)!;2!+3!+\cdots+(n-1)!+n!)$, by counting the solution $(0,0,\ldots,0,n!)$, we deduce that the number of solutions satisfies

$$1+\frac{(n!)^{n-2}}{(n-2)!\,2!\,3!\cdots(n-1)!} \le f(n) \le 1+\frac{\big(n!+2!+3!+\cdots+(n-1)!\big)^{n-2}}{(n-2)!\,2!\,3!\cdots(n-1)!}.$$

Let

$$\begin{aligned}
u_n &= \ln\left(\frac{(n!)^{n-2}}{(n-2)!\,2!\,3!\cdots(n-1)!}\right) = (n-2)\ln(n!) - \ln(n-2)! - \sum_{k=1}^{n-1}\ln k! \\
&= (n-2)\ln n! - \ln(n-2)! - \sum_{k=1}^{n-1}(n-k)\ln k = \mathrm{O}(\ln n!) + \sum_{k=1}^{n-1} k\ln k.
\end{aligned}$$

An easy in integral estimation gives $\sum_{k=1}^{n-1} k\ln k = \dfrac{n^2\ln n}{2} + \mathrm{O}(n\ln n)$. Thus, $u_n = \frac{1}{2}n^2\ln n + \mathrm{O}(n\ln n)$ and because $n! + 2! + \cdots + (n-1)! < 3n!$, it follows that

$$\ln f(n) = u_n + \mathrm{O}(\ln n!) = \frac{n^2}{2}\cdot\ln n + \mathrm{O}(n\ln n).$$

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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