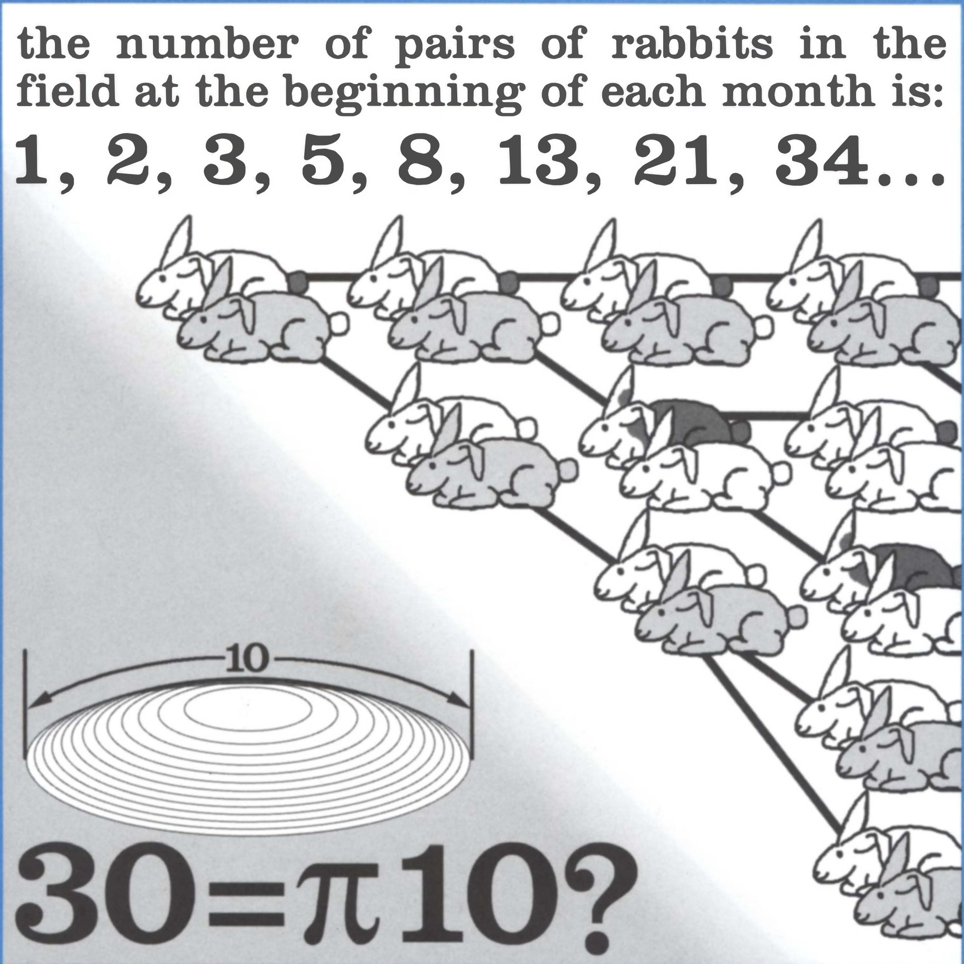
Vol. 76, No. 3, June 2003

MATHEMATICS MAGAZINE



- The Fibonacci Numbers—Exposed
- The Fibonacci Numbers—Exposed More Discretely
- Gaspard Monge and the Monge Point of the Tetrahedron

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

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

Cover image: Bunnies Hop toward a Molten Sea, by Jason Challas. A biblical story describes a "molten sea" in a way that seems to imply that π is 3. According to Anderson, Stumpf, and Miller (see p. 225), the given measurements would be accurate if the surface of the "sea" bulged upward in a spherical shape. Fibonacci's rabbits are eager to dive in. Jason Challas lectures on computer art (assuring his students that π is *not* three) at Santa Clara University.

AUTHORS

Dan Kalman received his Ph.D. from the University of Wisconsin in 1980. Before joining the mathematics faculty at American University in 1993, he worked for eight years in the aerospace industry in Southern California. During that period, two events occurred that were crucial to the evolution of "Fibonacci Numbers Exposed": he met Robert Mena on a visit to California State University, Long Beach, and he learned about the Pythagorean-Fibonacci connection on a visit to California State University, Northridge. He delights in puns and word play of all kinds, and is an avid fan of Douglas Adams, J. R. R. Tolkien, and Gilbert and Sullivan.

Robert Mena joined the faculty at California State University, Long Beach in 1988 after 15 years on the faculty at the University of Wyoming, and graduate school at the University of Houston. It was while being a "rotator" at the National Science Foundation in Arlington, Virginia during the academic year 2000–2001 that he had the pleasure of reconnecting with Dan. Coincidentally, it was after a talk by H. S. Wilf at a conference at Northridge that he reconnected with the Fibonacci sequence. He is an enthusiastic solver of quotes and other sorts of acrostics. Even after 30 years of teaching, he still purports to enjoy the company of his students and the thrill of teaching mathematics.

Arthur T. Benjamin and **Jennifer J. Quinn** have coauthored more than a dozen papers together and have just completed a book for the MAA, *Proofs That Really Count: The Art of Combinatorial Proof*. Benjamin is Professor and Chair of the mathematics department at Harvey Mudd College. Quinn is Associate Professor and Chair of the mathematics department at Occidental College. Both have been awarded the Distinguished Teaching Award from the Southern California Section of the MAA, with Benjamin winning the MAA's Haimo Prize in 2000. They have been selected as the next editors of *Math Horizons* magazine, and look forward to your submissions.

Robert Alan Crabbs was born and raised in Nassau County, Long Island, New York. He obtained a B.A. in liberal arts with a minor in mathematics from Hofstra University in Hempstead, New York in 1979. In 1994 he received an M.A. in mathematics education from the University of Central Florida. There, in 1989, he was introduced to the Monge point in an analytic geometry course given by Professor Howard Eves. "Gaspard Monge and the Monge Point of the Tetrahedron" is the result of Professor Eves' superlative teaching, and of his friendship.

Vol. 76, No. 3, June 2003

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 Dave Riska (*driska@maa.org*), Advertising Manager, the Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

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

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

Printed in the United States of America

The Fibonacci Numbers—Exposed

DAN KALMAN
American University
Washington, D.C. 20016
kalman@american.edu

ROBERT MENA
California State University
Long Beach, CA 90840
rmena@csulb.edu

Among numerical sequences, the Fibonacci numbers F_n have achieved a kind of celebrity status. Indeed, Koshy gushingly refers to them as one of the "two shining stars in the vast array of integer sequences" [**16**, p. xi]. The second of Koshy's "shining stars" is the *Lucas* numbers, a close relative of the Fibonacci numbers, about which we will say more below. The Fibonacci numbers are famous for possessing wonderful and amazing properties. Some are well known. For example, the sums and differences of Fibonacci numbers are Fibonacci numbers, and the ratios of Fibonacci numbers converge to the golden mean. Others are less familiar. Did you know that any four consecutive Fibonacci numbers can be combined to form a Pythagorean triple? Or how about this: The greatest common divisor of two Fibonacci numbers is another Fibonacci number. More precisely, the gcd of F_n and F_m is F_k, where k is the gcd of n and m.

With such fabulous properties, it is no wonder that the Fibonacci numbers stand out as a kind of super sequence. But what if it is not such a special sequence after all? What if it is only a rather pedestrian sample from an entire race of super sequences? In this case, the home world is the planet of two-term recurrences. As we shall show, its inhabitants are all just about as amazing as the Fibonacci sequence.

The purpose of this paper is to demonstrate that many of the properties of the Fibonacci numbers can be stated and proved for a much more general class of sequences, namely, second-order recurrences. We shall begin by reviewing a selection of the properties that made Fibonacci numbers famous. Then there will be a survey of second-order recurrences, as well as general tools for studying these recurrences. A number of the properties of the Fibonacci numbers will be seen to arise simply and naturally as the tools are presented. Finally, we will see that Fibonacci connections to Pythagorean triples and the gcd function also generalize in a natural way.

Famous Fibonacci properties

The Fibonacci numbers F_n are the terms of the sequence $0, 1, 1, 2, 3, 5, \ldots$ wherein each term is the sum of the two preceding terms, and we get things started with 0 and 1 as F_0 and F_1. You cannot go very far in the lore of Fibonacci numbers without encountering the companion sequence of Lucas numbers L_n, which follows the same recursive pattern as the Fibonacci numbers, but begins with $L_0 = 2$ and $L_1 = 1$. The first several Lucas numbers are therefore 2, 1, 3, 4, 7.

Regarding the origins of the subject, Koshy has this to say:

> The sequence was given its name in May of 1876 by the outstanding French mathematician François Edouard Anatole Lucas, who had originally called it "the series of Lamé," after the French mathematician Gabriel Lamé [**16**, p. 5].

Although Lucas contributed a great deal to the study of the Fibonacci numbers, he was by no means alone, as a perusal of Dickson [**4**, Chapter XVII] reveals. In fact, just about all the results presented here were first published in the nineteenth century. In particular, in his foundational paper [**17**], Lucas, himself, investigated the generalizations that interest us. These are sequences A_n defined by a recursive rule $A_{n+2} = aA_{n+1} + bA_n$ where a and b are fixed constants. We refer to such a sequence as a *two-term recurrence*.

The popular lore of the Fibonacci numbers seems not to include these generalizations, however. As a case in point, Koshy [**16**] has devoted nearly 700 pages to the properties of Fibonacci and Lucas numbers, with scarcely a mention of general two-term recurrences. Similar, but less encyclopedic sources are Hoggatt [**9**], Honsberger [**11**, Chapter 8], and Vajda [**21**]. There has been a bit more attention paid to so-called *generalized Fibonacci numbers*, A_n, which satisfy the same recursive formula $A_{n+2} = A_{n+1} + A_n$, but starting with arbitrary initial values A_0 and A_1, particularly by Horadam (see for example Horadam [**12**], Walton and Horadam [**22**], as well as Koshy [**16**, Chapter 7]). Horadam also investigated the same sort of sequences we consider, but he focused on different aspects from those presented here [**14, 15**]. In [**14**] he includes our Examples 3 and 7, with an attribution to Lucas's 1891 *Théorie des Nombres*. With Shannon, Horadam also studied Pythagorean triples, and their paper [**20**] goes far beyond the connection to Fibonacci numbers considered here. Among more recent references, Bressoud [**3**, chapter 12] discusses the application of generalized Fibonacci sequences to primality testing, while Hilton and Pedersen [**8**] present some of the same results that we do. However, none of these references share our general point of emphasis, that in many cases, properties commonly perceived as unique to the Fibonacci numbers, are actually shared by large classes of sequences.

It would be impossible to make this point here in regard to all known Fibonacci properties, as Koshy's tome attests. We content ourselves with a small sample, listed below. We have included page references from Koshy [**16**].

Sum of squares $\sum_1^n F_i^2 = F_n F_{n+1}$. (Page 77.)

Lucas-Fibonacci connection $L_{n+1} = F_{n+2} + F_n$. (Page 80.)

Binet formulas The Fibonacci and Lucas numbers are given by

$$F_n = \frac{\alpha^n - \beta^n}{\alpha - \beta} \quad \text{and} \quad L_n = \alpha^n + \beta^n,$$

where

$$\alpha = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \beta = \frac{1-\sqrt{5}}{2}.$$

(Page 79.)

Asymptotic behavior $F_{n+1}/F_n \to \alpha$ as $n \to \infty$. (Page 122.)

Running sum $\sum_1^n F_i = F_{n+2} - 1$. (Page 69.)

Matrix form We present a slightly permuted form of what generally appears in the literature. Our version is

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^n = \begin{bmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{bmatrix}.$$

(Page 363.)

Cassini's formula $F_{n-1}F_{n+1} - F_n^2 = (-1)^n$. (Page 74)

Convolution property $F_n = F_m F_{n-m+1} + F_{m-1} F_{n-m}$. (Page 88, formula 6.)

Pythagorean triples If w, x, y, z are four consecutive Fibonacci numbers, then $(wz, 2xy, yz - wx)$ is a Pythagorean triple. That is, $(wz)^2 + (2xy)^2 = (yz - wx)^2$. (Page 91, formula 88.)

Greatest common divisor $\gcd(F_m, F_n) = F_{\gcd(m,n)}$. (Page 198.)

This is, as mentioned, just a sample of amazing properties of the Fibonacci and Lucas numbers. But they all generalize in a natural way to classes of two-term recurrences. In fact, several of the proofs arise quite simply as part of a general development of the recurrences. We proceed to that topic next.

Generalized Fibonacci and Lucas numbers

Let a and b be any real numbers. Define a sequence A_n as follows. Choose initial values A_0 and A_1. All succeeding terms are determined by

$$A_{n+2} = aA_{n+1} + bA_n. \tag{1}$$

For fixed a and b, we denote by $\mathcal{R}(a, b)$ the set of all such sequences. To avoid a trivial case, we will assume that $b \neq 0$.

In $\mathcal{R}(a, b)$, we define two distinguished elements. The first, F, has initial terms 0 and 1. In $\mathcal{R}(1, 1)$, F is thus the Fibonacci sequence. In the more general case, we refer to F as the (a, b)-Fibonacci sequence. Where no confusion will result, we will suppress the dependence on a and b. Thus, in every $\mathcal{R}(a, b)$, there is an element F that begins with 0 and 1, and this is the Fibonacci sequence for $\mathcal{R}(a, b)$.

Although F is the primordial sequence in $\mathcal{R}(a, b)$, there is another sequence L that is of considerable interest. It starts with $L_0 = 2$ and $L_1 = a$. As will soon be clear, L plays the same role in $\mathcal{R}(a, b)$ as the Lucas numbers play in $\mathcal{R}(1, 1)$. Accordingly, we refer to L as the (a, b)-Lucas sequence. For the most part, there will be only one a and b under consideration, and it will be clear from context which $\mathcal{R}(a, b)$ is the home for any particular mention of F or L. In the rare cases where some ambiguity might occur, we will use $F^{(a,b)}$ and $L^{(a,b)}$ to indicate the F and L sequences in $\mathcal{R}(a, b)$.

In the literature, what we are calling F and L have frequently been referred to as *Lucas sequences* (see Bressoud [**3**, chapter 12] and Weisstein [**23**, p. 1113]) and denoted by U and V, the notation adopted by Lucas in 1878 [**17**]. We prefer to use F and L to emphasize the idea that there are Fibonacci and Lucas sequences in each $\mathcal{R}(a, b)$, and that these sequences share many properties with the traditional F and L. In contrast, it has sometimes been the custom to attach the name *Lucas* to the L sequence for a particular $\mathcal{R}(a, b)$. For example, in $\mathcal{R}(2, 1)$, the elements of F have been referred to as *Pell numbers* and the elements of L as *Pell-Lucas* numbers [**23**, p. 1334].

Examples Of course, the most familiar example is $\mathcal{R}(1, 1)$, in which F and L are the famous Fibonacci and Lucas number sequences. But there are several other choices of a and b that lead to familiar examples.

Example 1: $\mathcal{R}(11, -10)$. The Fibonacci sequence in this family is $F = 0, 1, 11, 111, 1111, \ldots$ the sequence of repunits, and $L = 2, 11, 101, 1001, 10001, \ldots$. The initial 2, which at first seems out of place, can be viewed as the result of putting two 1s in the same position.

Example 2: $\mathcal{R}(2, -1)$. Here F is the sequence of whole numbers $0, 1, 2, 3, 4, \ldots$, and L is the constant sequence $2, 2, 2, \ldots$. More generally, $\mathcal{R}(2, -1)$ consists of all the arithmetic progressions.

Example 3: $\mathcal{R}(3,-2)$. $F = 0, 1, 3, 7, 15, 31, \ldots$ is the Mersenne sequence, and $L = 2, 3, 5, 9, 17, 33, \ldots$ is the Fermat sequence. These are just powers of 2 plus or minus 1.

Example 4: $\mathcal{R}(1,-1)$. $F = 0, 1, 1, 0, -1, -1, 0, 1, 1, \ldots$ and $L = 2, 1, -1, -2, -1, 1, 2, 1, -1, \ldots$. Both sequences repeat with period 6, as do all the elements of $\mathcal{R}(1,-1)$.

Example 5: $\mathcal{R}(3,-1)$. $F = 0, 1, 3, 8, 21, \ldots$ and $L = 2, 3, 7, 18, \ldots$. Do you recognize these? They are the *even-numbered* Fibonacci and Lucas numbers.

Example 6: $\mathcal{R}(4,1)$. $F = 0, 1, 4, 17, 72, \ldots$ and $L = 2, 4, 18, 76, \ldots$. Here, L gives every third Lucas number, while F gives 1/2 of every third Fibonacci number.

Example 7: $\mathcal{R}(2,1)$. $F = 0, 1, 2, 5, 12, 29, 70, \ldots$ and $L = 2, 2, 6, 14, 34, 82, \ldots$. These are the Pell sequences, mentioned earlier. In particular, for any n, $(x, y) = (F_{2n} + F_{2n-1}, F_{2n})$ gives a solution to Pell's Equation $x^2 - 2y^2 = 1$. This extends to the more general Pell equation, $x^2 - dy^2 = 1$, when $d = k^2 + 1$. Then, using the F sequence in $\mathcal{R}(2k, 1)$, we obtain solutions of the form $(x, y) = (kF_{2n} + F_{2n-1}, F_{2n})$. Actually, equations of this type first appeared in the Archimedean cattle problem, and were considered by the Indian mathematicians Brahmagupta and Bhaskara [**2**, p. 221]. Reportedly, Pell never worked on the equations that today bear his name. Instead, according to Weisstein [**23**], "while Fermat deserves the credit for being the first [European] to extensively study the equation, the erroneous attribution to Pell was perpetrated by none other than Euler."

Coincidentally, the even terms F_{2n} in $\mathcal{R}(a, 1)$ also appear in another generalized Fibonacci result, related to an identity discussed elsewhere in this issue of the MAGAZINE [**6**]. The original identity for normal Fibonacci numbers is

$$\arctan\left(\frac{1}{F_{2n}}\right) = \arctan\left(\frac{1}{F_{2n+1}}\right) + \arctan\left(\frac{1}{F_{2n+2}}\right).$$

For $F \in \mathcal{R}(a, 1)$ the corresponding result is

$$\arctan\left(\frac{1}{F_{2n}}\right) = \arctan\left(\frac{a}{F_{2n+1}}\right) + \arctan\left(\frac{1}{F_{2n+2}}\right).$$

The wonderful world of two-term recurrences

The Fibonacci and Lucas sequences are elements of $\mathcal{R}(1, 1)$, and many of their properties follow immediately from the recursive rule that each term is the sum of the two preceding terms. Similarly, it is often easy to establish corresponding properties for elements of $\mathcal{R}(a, b)$ directly from the fundamental identity (1). For example, in $\mathcal{R}(1, 1)$, the *Sum of Squares* identity is

$$F_1^2 + F_2^2 + \cdots + F_n^2 = F_n F_{n+1}.$$

The generalization of this to $\mathcal{R}(a, b)$ is

$$b^n F_0^2 + b^{n-1} F_1^2 + \cdots + bF_{n-1}^2 + F_n^2 = \frac{F_n F_{n+1}}{a}. \tag{2}$$

This can be proved quite easily using (1) and induction.

Many of the other famous properties can likewise be established by induction. But to provide more insight about these properties, we will develop some analytic methods, organized loosely into three general contexts. First, we can think of $\mathcal{R}(a, b)$ as a subset of $\mathbb{R}^\infty$, the real vector space of real sequences, and use the machinery of difference operators. Second, by deriving Binet formulas for elements of $\mathcal{R}(a, b)$, we obtain explicit representations as linear combinations of geometric progressions. Finally, there is a natural matrix formulation which is tremendously useful. We explore each of these contexts in turn.

Difference operators We will typically represent elements of $\mathbb{R}^\infty$ with uppercase roman letters, in the form

$$A = A_0, A_1, A_2, \ldots.$$

There are three fundamental linear operators on $\mathbb{R}^\infty$ to consider. The first is the *left-shift*, Λ. For any real sequence $A = A_0, A_1, A_2, \ldots$, the shifted sequence is $\Lambda A = A_1, A_2, A_3, \ldots$.

This shift operator is a kind of discrete differential operator. Recurrences like (1) are also called difference equations. Expressed in terms of Λ, (1) becomes

$$(\Lambda^2 - a\Lambda - b)A = 0.$$

This is analogous to expressing a differential equation in terms of the differential operator, and there is a theory of difference equations that perfectly mirrors the theory of differential equations. Here, we have in mind linear constant coefficient differential and difference equations.

As one fruit of this parallel theory, we see at once that $\Lambda^2 - a\Lambda - b$ is a linear operator on $\mathbb{R}^\infty$, and that $\mathcal{R}(a, b)$ is its null space. This shows that $\mathcal{R}(a, b)$ is a subspace of $\mathbb{R}^\infty$. We will discuss another aspect of the parallel theories of difference and differential equation in the succeeding section on Binet formulas.

Note that any polynomial in Λ is a linear operator on $\mathbb{R}^\infty$, and that all of these operators commute. For example, the *forward difference* operator Δ, defined by $(\Delta A)_k = A_{k+1} - A_k$, is given by $\Delta = \Lambda - 1$. Similarly, consider the k-term sum, Σ_k, defined by $(\Sigma_k A)_n = A_n + A_{n+1} + \cdots + A_{n+k-1}$. To illustrate, $\Sigma_2(A)$ is the sequence $A_0 + A_1, A_1 + A_2, A_2 + A_3, \ldots$. These sum operators can also be viewed as polynomials in Λ: $\Sigma_k = 1 + \Lambda + \Lambda^2 + \cdots + \Lambda^{k-1}$.

Because these operators commute with Λ, they are operators on $\mathcal{R}(a, b)$, as well. In general, if Ψ is an operator that commutes with Λ, we observe that Ψ also commutes with $\Lambda^2 - a\Lambda - b$. Thus, if $A \in \mathcal{R}(a, b)$, then $(\Lambda^2 - a\Lambda - b)\Psi A = \Psi(\Lambda^2 - a\Lambda - b)A = \Psi 0 = 0$. This shows that $\Psi A \in \mathcal{R}(a, b)$. In particular, $\mathcal{R}(a, b)$ is closed under differences and k-term sums.

This brings us to the second fundamental operator, the *cumulative sum* Σ. It is defined as follows: $\Sigma(A) = A_0, A_0 + A_1, A_0 + A_1 + A_2, \ldots$. This is not expressible in terms of Λ, nor does it commute with Λ, in general. However, there is a simple relation connecting the two operators:

$$\Delta\Sigma = \Lambda. \tag{3}$$

This is a sort of discrete version of the fundamental theorem of calculus. In the opposite order, we have

$$(\Sigma\Delta A)_n = A_{n+1} - A_0,$$

a discrete version of the other form of the fundamental theorem. It is noteworthy that Leibniz worked with these sum and difference operators as a young student, and later identified this work as his inspiration for calculus (Edwards [**5**, p. 234]).

The final fundamental operator is the *k-skip*, Ω_k, which selects every kth element of a sequence. That is, $\Omega_k(A) = A_0, A_k, A_{2k}, A_{3k}, \ldots$. By combining these operators with powers of Λ, we can sample the terms of a sequence according to any arithmetic progression. For example,

$$\Omega_5 \Lambda^3 A = A_3, A_8, A_{13}, \ldots .$$

Using the context of operators and the linear space $\mathcal{R}(a, b)$, we can derive useful results. First, it is apparent that once A_0 and A_1 are chosen, all remaining terms are completely determined by (1). This shows that $\mathcal{R}(a, b)$ is a two-dimensional space. Indeed, there is a natural basis $\{E, F\}$ where E has starting values 1 and 0, and F, with starting values 0 and 1, is the (a, b)-Fibonacci sequence. Thus

$$E = 1, 0, b, ab, a^2b + b^2, \ldots$$

$$F = 0, 1, a, a^2 + b, a^3 + 2ab, \ldots .$$

Clearly, $A = A_0E + A_1F$ for all $A \in \mathcal{R}(a, b)$. Note further that $\Lambda E = bF$, so that we can easily express any A just using F:

$$A_n = bA_0F_{n-1} + A_1F_n \tag{4}$$

As an element of $\mathcal{R}(a, b)$, L can thus be expressed in terms of F. From (4), we have

$$L_n = 2bF_{n-1} + aF_n.$$

But the fundamental recursion (1) then leads to

$$L_n = bF_{n-1} + F_{n+1}. \tag{5}$$

This is the analog of the Lucas-Fibonacci connection stated above.

Recall that the difference and the k-term sum operators all preserve $\mathcal{R}(a, b)$. Thus, ΔF and $\Sigma_k F$ are elements of $\mathcal{R}(a, b)$ and can be expressed in terms of F using (4). The case for Σ is a more interesting application of operator methods. The question is this: If $A \in \mathcal{R}(a, b)$, what can we say about ΣA?

As a preliminary step, notice that a sequence is constant if and only if it is annihilated by the difference operator Δ. Now, suppose that $A \in \mathcal{R}(a, b)$. That means $(\Lambda^2 - a\Lambda - b)A = 0$, and so too

$$\Lambda(\Lambda^2 - a\Lambda - b)A = 0.$$

Now commute Λ with the other operator, and use (3) to obtain

$$(\Lambda^2 - a\Lambda - b)\Delta\Sigma A = 0.$$

Finally, since Δ and Λ commute, pull Δ all the way to the front to obtain

$$\Delta(\Lambda^2 - a\Lambda - b)\Sigma A = 0.$$

This shows that while $(\Lambda^2 - a\Lambda - b)\Sigma A$ may not be 0 (indicating $\Sigma A \notin \mathcal{R}(a, b)$), at worst it is constant. Now it turns out that there are two cases. If $a + b \neq 1$, it can be shown that ΣA differs from an element of $\mathcal{R}(a, b)$ by a constant. That tells us at once that there is an identity of the form

$$(\Sigma A)_n = c_0 F_n + c_1 F_{n-1} + c_2,$$

which corresponds to the running sum property for Fibonacci numbers. We will defer the determination of the constants c_i to the section on Binet formulas.

Here is the verification that ΣA differs from an element of $\mathcal{R}(a, b)$ by a constant when $a + b \neq 1$. We know that $(\Lambda^2 - a\Lambda - b)\Sigma A$ is a constant c_1. Suppose that we can find another constant, c, such that $(\Lambda^2 - a\Lambda - b)c = c_1$. Then we would have $(\Lambda^2 - a\Lambda - b)(\Sigma A - c) = 0$, hence $\Sigma A - c \in \mathcal{R}(a, b)$. It is an exercise to show c can be found exactly when $a + b \neq 1$.

When $a + b = 1$ we have the second case. A little experimentation with (1) will show you that in this case $\mathcal{R}(a, b)$ includes all the constant sequences. The best way to analyze this situation is to develop some general methods for solving difference equations. We do that next.

Binet formulas We mentioned earlier that there is a perfect analogy between linear constant coefficient difference and differential equations. In the differential equation case, a special role is played by the exponential functions, $e^{\lambda t}$, which are eigenvectors for the differential operator: $De^{\lambda t} = \lambda \cdot e^{\lambda t}$. For difference equations, the analogous role is played by the geometric progressions, $A_n = \lambda^n$. These are eigenvectors for the left-shift operator: $\Lambda\lambda^n = \lambda \cdot \lambda^n$. Both differential and difference equations can be formulated in terms of polynomials in the fundamental operator, Λ or D, respectively. These are in fact characteristic polynomials—the roots λ are eigenvalues and correspond to eigenvector solutions to the differential or difference equation. Moreover, except in the case of repeated roots, this leads to a basis for the space of all solutions.

We can see how this all works in detail in the case of $\mathcal{R}(a, b)$, which is viewed as the null space of $p(\Lambda) = \Lambda^2 - a\Lambda - b$. When is the geometric progression $A_n = \lambda^n$ in this null space? We demand that $A_{n+2} - aA_{n+1} - bA_n = 0$, so the condition is

$$\lambda^{n+2} - a\lambda^{n+1} - b\lambda^n = 0.$$

Excluding the case $\lambda = 0$, which produces only the trivial solution, this leads to $p(\lambda) = 0$ as a necessary and sufficient condition for $\lambda^n \in \mathcal{R}(a, b)$. Note also that the roots of p are related to the coefficients in the usual way. Thus, if the roots are λ and μ, then

$$\lambda + \mu = a \tag{6}$$

$$\lambda\mu = -b. \tag{7}$$

Now if λ and μ are distinct, then λ^n and μ^n are independent solutions to the difference equation. And since we already know that the null space is two dimensional, that makes $\{\lambda^n, \mu^n\}$ a basis. In this case, $\mathcal{R}(a, b)$ is characterized as the set of linear combinations of these two geometric progressions. In particular, for $A \in \mathcal{R}(a, b)$, we can express A in the form

$$A_n = c_\lambda \lambda^n + c_\mu \mu^n. \tag{8}$$

The constants c_λ and c_μ are determined by the initial conditions

$$A_0 = c_\lambda + c_\mu$$

$$A_1 = c_\lambda \lambda + c_\mu \mu.$$

We are assuming λ and μ are distinct, so this system has the solution

$$c_\lambda = \frac{A_1 - \mu A_0}{\lambda - \mu}$$
$$c_\mu = \frac{\lambda A_0 - A_1}{\lambda - \mu}.$$

Now let us apply these to the special cases of F and L. For F, the initial values are 0 and 1, so $c_\lambda = 1/(\lambda - \mu)$ and $c_\mu = -1/(\lambda - \mu)$. For L the initial terms are 2 and $a = \lambda + \mu$. This gives $c_\lambda = c_\mu = 1$. Thus,

$$F_n = \frac{\lambda^n - \mu^n}{\lambda - \mu} \tag{9}$$
$$L_n = \lambda^n + \mu^n. \tag{10}$$

These are the *Binet Formulas* for $\mathcal{R}(a, b)$.

When $\lambda = \mu$, the fundamental solutions of the difference equation are $A_n = \lambda^n$ and $B_n = n\lambda^n$. Most of the results for $\mathcal{R}(a, b)$ have natural extensions to this case. For example, in the case of repeated root λ, the Binet formulas become

$$F_n = n\lambda^{n-1}$$
$$L_n = 2\lambda^n.$$

Extensions of this sort are generally quite tractable, and we will not typically go into the details. Accordingly, we will assume from now on that p has distinct roots, or equivalently, that $a^2 + 4b \neq 0$.

Another special case of interest occurs when one root is 1. In this case, the geometric progression 1^n is constant, and $\mathcal{R}(a, b)$ contains all the constant sequences. As we saw earlier, the condition for this is $a + b = 1$. Now the Binet representation gives a new way of thinking about this result. It is an exercise to verify that $a + b = 1$ if and only if 1 is a root of p.

If both roots equal 1, the fundamental solutions are $A_n = 1$ and $B_n = n$. This shows that $\mathcal{R}(2, -1)$ consists of all the arithmetic progressions, confirming our earlier observation for Example 2.

Let us revisit the other examples presented earlier, and consider the Binet formulas for each.

Example 0: $\mathcal{R}(1, 1)$. For the normal Fibonacci and Lucas numbers, $p(t) = t^2 - t - 1$, and the roots are α and β as defined earlier. The general version of the Binet formulas reduce to the familiar form upon substitution of α and β for λ and μ.

Example 1: $\mathcal{R}(11, -10)$. Here, with $p(t) = t^2 - 11t + 10$, the roots are 10 and 1. In this case the Binet formulas simply tell us what is already apparent: $F_n = (10^n - 1)/9$ and $L_n = 10^n + 1$.

Example 3: $\mathcal{R}(3, -2)$. In this example, $p(t) = t^2 - 3t + 2$, with roots 2 and 1. The Binet formulas confirm the pattern we saw earlier: $F_n = 2^n - 1$ and $L_n = 2^n + 1$.

Example 4: $\mathcal{R}(1, -1)$. Now $p(t) = t^2 - t + 1$. Note that $p(t)(t + 1) = t^3 + 1$, so that roots of p are cube roots of -1 and hence, sixth roots of 1. This explains the periodic nature of F and L. Indeed, since $\lambda^6 = \mu^6 = 1$ in this case, every element of $\mathcal{R}(1, -1)$ has period 6 as well.

Example 5: $\mathcal{R}(3, -1)$. The roots in this example are α^2 and β^2. The Binet formulas involve only even powers of α and β, hence the appearance of the even Fibonacci and Lucas numbers.

Example 6: $\mathcal{R}(4, 1)$. This example is similar to the previous one, except that the roots are α^3 and β^3.

Example 7: $\mathcal{R}(2, 1)$. For this example $p(t) = t^2 - 2t - 1$, so the roots are $1 \pm \sqrt{2}$. The Binet formulas give

$$F_n = \frac{(1+\sqrt{2})^n - (1-\sqrt{2})^n}{2\sqrt{2}} \quad \text{and} \quad L_n = (1+\sqrt{2})^n + (1-\sqrt{2})^n.$$

Characterizing $\mathcal{R}(a, b)$ in terms of geometric progressions has immediate applications. For example, consider the ratio of successive terms of a sequence in $\mathcal{R}(a, b)$. Using (8), we have

$$\frac{A_{n+1}}{A_n} = \frac{c_\lambda \lambda^{n+1} + c_\mu \mu^{n+1}}{c_\lambda \lambda^n + c_\mu \mu^n}.$$

Now assume that $|\lambda| > |\mu|$, and divide out λ^n:

$$\frac{A_{n+1}}{A_n} = \frac{c_\lambda \lambda + c_\mu \mu(\mu^n/\lambda^n)}{c_\lambda + c_\mu(\mu^n/\lambda^n)}.$$

Since $(\mu/\lambda)^n$ will go to 0 as n goes to infinity, we conclude

$$\frac{A_{n+1}}{A_n} \to \lambda \quad \text{as} \quad n \to \infty.$$

In words, the ratio of successive terms of a sequence in $\mathcal{R}(a, b)$ always tends toward the dominant eigenvalue as n goes to infinity. That is the general version of the asymptotic behavior we observed for Fibonacci numbers.

As a second example, if $A_n = c_\lambda \lambda^n + c_\mu \mu^n$, then $\Omega_k A_n = c_\lambda \lambda^{kn} + c_\mu \mu^{kn}$. This is a linear combination of two geometric progressions as well, with eigenvalues λ^k and μ^k. Consequently, $\Omega_k A \in \mathcal{R}(a', b')$ for some a' and b'. Now using the relationship between roots and coefficients again, we deduce that $a' = \lambda^k + \mu^k$, and by (10) that gives $a' = L_k^{(a,b)}$. Similarly, we find $b' = -(\lambda\mu)^k = -(-b)^k$. Thus,

$$\Omega_k : \mathcal{R}(a, b) \to \mathcal{R}(L_k^{(a,b)}, -(-b)^k). \tag{11}$$

We can extend this slightly. If $A \in \mathcal{R}(a, b)$, then so is $\Lambda^d A$ for any positive integer d. Thus, $\Omega_k \Lambda^d A \in \mathcal{R}(a', b')$. In other words, when $A \in \mathcal{R}(a, b)$, the sequence $B_n = A_{kn+d}$ is in $\mathcal{R}(a', b')$. This corresponds to sampling A at the terms of an arithmetic progression.

In the particular case of F and L, we can use the preceding results to determine the effect of Ω_k explicitly. For notational simplicity, we will again denote by a' and b' the values $L_k^{(a,b)}$ and $-(-b)^k$, respectively. We know that $\Omega_k F^{(a,b)} \in \mathcal{R}(a', b')$, and begins with the terms 0 and $F_k^{(a,b)}$. This is necessarily a multiple of $F^{(a',b')}$, and in particular, gives

$$\Omega_k F^{(a,b)} = F_k^{(a,b)} \cdot F^{(a',b')}. \tag{12}$$

Similarly, $\Omega_k L^{(a,b)}$ begins with 2 and $L_k^{(a,b)}$. But remember that the latter of these is exactly $a' = L_1^{(a',b')}$. Thus,

$$\Omega_k L^{(a,b)} = L^{(a',b')}. \tag{13}$$

Of course, this last equation is easily deduced directly from the Binet formula for $L^{(a,b)}$, as well. The observations in Examples 5 and 6 are easily verified using (12) and (13).

For one more example, let us return to the analysis of ΣA for $A \in \mathcal{R}(a, b)$. Again using the expression $A_n = c_\lambda \lambda^n + c_\mu \mu^n$ we find the terms of ΣA as

$$\Sigma A_n = c_\lambda \frac{\lambda^{n+1} - 1}{\lambda - 1} + c_\mu \frac{\mu^{n+1} - 1}{\mu - 1}.$$

Evidently, this is invalid if either λ or μ equals 1. So, as before, we exclude that possibility by assuming $a + b \neq 1$.

Under this assumption, we found earlier that ΣA must differ from an element of $\mathcal{R}(a, b)$ by a constant. Now we can easily determine the value of that constant. Rearranging the preceding equation produces

$$\Sigma A_n = \frac{c_\lambda \lambda}{\lambda - 1}\lambda^n + \frac{c_\mu \mu}{\mu - 1}\mu^n - \left(\frac{c_\lambda}{\lambda - 1} + \frac{c_\mu}{\mu - 1}\right).$$

This clearly reveals ΣA as the sum of an element of $\mathcal{R}(a, b)$ with the constant $C = -(c_\lambda/(\lambda - 1) + c_\mu/(\mu - 1))$.

In general, the use of this formula requires expressing A in terms of λ and μ. But in the special case of F, we can express the formula in terms of a and b. Recall that when $A = F$, $c_\lambda = 1/(\lambda - \mu)$ and $c_\mu = -1/(\lambda - \mu)$. Substituting these in the earlier formula for C, leads to

$$\begin{aligned} C &= -\frac{1}{\lambda - \mu}\left(\frac{1}{\lambda - 1} - \frac{1}{\mu - 1}\right) \\ &= -\frac{1}{\lambda - \mu}\frac{\mu - \lambda}{(\lambda - 1)(\mu - 1)} = \frac{1}{(\lambda\mu - \lambda - \mu + 1)}. \end{aligned}$$

Once again using (6) and (7), this yields

$$C = \frac{1}{1 - a - b}. \tag{14}$$

As an example, let us consider ΣF for $\mathcal{R}(2, 3)$. In the table below, the first several terms of F and ΣF are listed.

n	0	1	2	3	4	5
F_n	0	1	2	7	20	61
ΣF_n	0	1	3	10	30	91

In this example, we have $C = 1/(1 - 2 - 3) = -1/4$. Accordingly, adding 1/4 to each term of ΣF should produce an element of $\mathcal{R}(2, 3)$. Carrying out this step produces

$$\Sigma F + \frac{1}{4} = \frac{1}{4}(1, 5, 13, 41, 121, 365, \ldots).$$

As expected, this is an element of $\mathcal{R}(2, 3)$.

Applying a similar analysis in the general case (with the assumption $a + b \neq 1$) leads to the identity

$$\Sigma F_n = \frac{1}{a+b-1}(F_{n+1} + bF_n - 1).$$

This reduces to the running sum property for Fibonacci numbers when $a = b = 1$. A similar analysis applies in the case $a + b = 1$. We leave the details to the reader.

In the derivation of the Binet formulas above, a key role was played by the eigenvectors and eigenvalues of the shift operator. It is therefore not surprising that there is a natural matrix formulation of these ideas. That topic is the third general context for tool development.

Matrix formulation Using the natural basis $\{E, F\}$ for $\mathcal{R}(a, b)$, we can represent Λ by a matrix M. We already have seen that $\Lambda E = bF$, so the first column of M has entries 0 and b. Applying the shift to F produces $(1, a, \ldots) = E + aF$. This identifies the second column entries of M as 1 and a, so

$$M = \begin{bmatrix} 0 & 1 \\ b & a \end{bmatrix}. \tag{15}$$

Now if $A \in \mathcal{R}(a, b)$, then relative to the natural basis it is represented by $[A] = [A_0 \;\; A_1]^T$. Similarly, the basis representation of $\Lambda^n A$ is $[A_n \;\; A_{n+1}]^T$. On the other hand, we can find the same result by applying M n times to $[A]$. Thus, we obtain

$$\begin{bmatrix} 0 & 1 \\ b & a \end{bmatrix}^n \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} = \begin{bmatrix} A_n \\ A_{n+1} \end{bmatrix}. \tag{16}$$

Premultiplying by $[1 \;\; 0]$ then gives

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ b & a \end{bmatrix}^n \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} = A_n. \tag{17}$$

This gives a matrix representation for A_n.

Note that in general, the ith column of a matrix M can be expressed as the product $M\mathbf{e}_i$, where $\mathbf{e}_i$ is the ith standard basis element. But here, the standard basis elements are representations $[E]$ and $[F]$. In particular, $M^n[E] = [E_n \;\; E_{n+1}]^T$ and $M^n[F] = [F_n \;\; F_{n+1}]^T$. This gives us the columns of M^n, and therefore

$$M^n = \begin{bmatrix} E_n & F_n \\ E_{n+1} & F_{n+1} \end{bmatrix}.$$

Then, using $\Lambda E = bF$, we have

$$\begin{bmatrix} 0 & 1 \\ b & a \end{bmatrix}^n = \begin{bmatrix} bF_{n-1} & F_n \\ bF_n & F_{n+1} \end{bmatrix}. \tag{18}$$

This is the general version of the matrix form for Fibonacci numbers.

The matrix form leads immediately to two other properties. First, taking the determinant of both sides of (18), we obtain

$$bF_{n-1}F_{n+1} - bF_n^2 = (-b)^n.$$

Simplifying,

$$F_{n-1}F_{n+1} - F_n^2 = (-1)^n b^{n-1},$$

the general version of Cassini's formula.

Second, start with $M^n = M^m M^{n-m}$, expressed explicitly in the form

$$\begin{bmatrix} bF_{n-1} & F_n \\ bF_n & F_{n+1} \end{bmatrix} = \begin{bmatrix} bF_{m-1} & F_m \\ bF_m & F_{m+1} \end{bmatrix} \begin{bmatrix} bF_{n-m-1} & F_{n-m} \\ bF_{n-m} & F_{n-m+1} \end{bmatrix}.$$

By inspection, we read off the 1, 2 entry of both sides, obtaining

$$F_n = F_m F_{n-m+1} + bF_{m-1}F_{n-m}, \tag{19}$$

generalizing the *Convolution Property* for regular Fibonacci numbers. As a special case, replace n with $2n+1$ and m with $n+1$, producing

$$F_{2n+1} = F_{n+1}^2 + bF_n^2. \tag{20}$$

This equation will be applied in the discussion of Pythagorean triples.

This concludes our development of general tools. Along the way, we have found natural extensions of all but two of our famous Fibonacci properties. These extensions are all simple and direct consequences of the basic ideas in three general contexts: difference operators, Binet formulas, and matrix methods. Establishing analogs for the remaining two properties is just a bit more involved, and we focus on them in the next section.

The last two properties

Pythagorean triples In a way, the connection with Pythagorean triples is trivial. The well-known parameterization $(x^2 - y^2, 2xy, x^2 + y^2)$ expresses primitive Pythagorean triples in terms of quadratic polynomials in two variables. The construction using Fibonacci numbers is similar. To make this clearer, note that if w, x, y, z are four consecutive Fibonacci numbers, then we may replace w with $y - x$ and z with $y + x$. With these substitutions, the Fibonacci parameterization given earlier for Pythagorean triples becomes

$$(wz, 2xy, yz - wx) = \big((y-x)(y+x), 2yx, y(y+x) - x(y-x)\big).$$

Since we can reduce the parameterization to a quadratic combination of two parameters in this way, the ability to express Pythagorean triples loses something of its mystery. In fact, if w, x, y, z are four consecutive terms of any sequence in $\mathcal{R}(a, b)$, we may regard x and y as essentially arbitrary, and so use them to define a Pythagorean triple $(x^2 - y^2, 2xy, x^2 + y^2)$. Thus, we can construct a Pythagorean triple using *just two* consecutive terms of a Fibonacci-like sequence.

Is that cheating? It depends on what combinations of the sequence elements are considered legitimate. The Fibonacci numbers have been used to parameterize Pythagorean triples in a variety of forms. The version given above, $(wz, 2xy, yz - wx)$, appears in Koshy [**16**] with a 1968 attribution to Umansky and Tallman. Here and below we use consecutive letters of the alphabet rather than the original subscript formulation, as a notational convenience. Much earlier, Raine [**19**] gave it this way: $(wz, 2xy, t)$, where, if w is F_n then t is F_{2n+3}. Boulger [**1**] extended Raine's results and observed that the triple can also be expressed $(wz, 2xy, x^2 + y^2)$. Horadam [**13**] reported it in the form $(xw, 2yz, 2yz + x^2)$. These combinations use a variety of different quadratic monomials, including both yz and x^2. So, if those are permitted, why not simply use the classical $(x^2 - y^2, 2xy, x^2 + y^2)$ and be done with it? The more complicated parameterizations we have cited then seem to be merely exercises in complexification.

In light of these remarks, it should be no surprise that the Fibonacci parameterization of Pythagorean triples can be generalized to $\mathcal{R}(a, b)$. For example, Shannon and Horadam [**20**] give the following version: $((a/b^2)xw, 2Pz(Pz - x), x^2 + 2Pz(Pz - x))$ where $P = (a^2 - -b)/2b^2$.

Using a modified version of the diophantine equation, we can get closer to the simplicity of Raine's formulation. For $\mathcal{R}(a, b)$ we replace the Pythagorean identity with

$$X^2 + bY^2 = Z^2 \tag{21}$$

and observe that the parameterization

$$(X, Y, Z) = (v^2 - bu^2, 2uv, v^2 + bu^2)$$

always produces solutions to (21). Now, if w, x, y, and z are four consecutive terms of $A \in \mathcal{R}(a, b)$, then we can express the first and last as

$$w = \frac{1}{b}(y - ax)$$
$$z = bx + ay.$$

Define constants $c = b/a$ and $d = c - a$. Then a calculation verifies that

$$(X, Y, Z) = (cwz - dxy, 2xy, xz + bwy) \tag{22}$$

is a solution to (21). In fact, with $u = x$ and $v = y$, it is exactly the parameterization given above.

In the special case that $x = F_n^{(a,b)}$, we can also express (22) in the form

$$(X, Y, Z) = (cwz - dxy, 2xy, t)$$

where $t = F_{2n+1}$. This version, which generalizes the Raine result, follows from (20). Note, also, that when $a = b = 1$, (22) becomes $(wz, 2xy, xz + wy)$, which is another variant on the Fibonacci parameterization of Pythagorean triples.

Greatest common divisor The Fibonacci properties considered so far make sense for real sequences in $\mathcal{R}(a, b)$. Now, however, we will consider divisibility properties that apply to integer sequences. Accordingly, we henceforth assume that a and b are integers, and restrict our attention to sequences $A \in \mathcal{R}(a, b)$ for which the initial terms A_0 and A_1 are integers, as well. Evidently, this implies A is an integer sequence. In order to generalize the gcd property, we must make one additional assumption: that a and b are relatively prime. Then we can prove in $\mathcal{R}(a, b)$, that the gcd of F_m and F_n is F_k, where k is the gcd of m and n. The proof has two parts: We show that F_k is a divisor of both F_m and F_n, and that F_m/F_k and F_n/F_k are relatively prime. The first of these follows immediately from an observation about the skip operator already presented. The second part depends on several additional observations.

OBSERVATION 1. *F_k is a divisor of F_{nk} for all $n > 0$.*

Proof. We have already noted that $\Omega_k F = F_k \cdot F^{(a',b')}$ so every element of $\Omega_k F = F_0, F_k, F_{2k}, \ldots,$ is divisible by F_k.

OBSERVATION 2. *F_n and b are relatively prime for all $n \geq 0$.*

Proof. Suppose p is prime divisor of b. Since a and b are relatively prime, p is not a divisor of a. Modulo p, the fundamental recursion (1) becomes $F_{n+2} \equiv aF_{n+1}$, so $F_n \equiv F_1 a^{n-1}$ for $n \geq 1$. This shows that $F_n \not\equiv 0$, since p is not a divisor of a.

OBSERVATION 3. *If $A \in \mathcal{R}(a,b)$, and if p is a common prime divisor of A_k and A_{k+1}, but is not a divisor of b, then p is a divisor of A_n for all $n \geq 0$.*

Proof. If $k > 0$, $A_{k+1} = aA_k + bA_{k-1}$, so p is a divisor of A_{k-1}. By induction, p divides both A_0 and A_1, and therefore A_n for all $n \geq 0$.

OBSERVATION 4. *If positive integers h and k are relatively prime, then so are F_h and F_k.*

Proof. If p is a prime divisor of F_h and F_k, then by Observation 2, p is not a divisor of b. Since h and k are relatively prime, there exist integers r and s such that $rh + sk = 1$. Clearly r and s must differ in sign. Without loss of generality, we assume that $r < 0$, and define $t = -r$. Thus, $sk - th = 1$. Now by Observation 1, F_{sk} is divisible by F_k, and hence by p. Similarly, F_{th} is divisible by F_h, and hence, also by p. But F_{th} and F_{sk} are consecutive terms of F, so by Observation 3, p is a divisor of all F_n. That is a contradiction, and shows that F_h and F_k can have no common prime divisor.

OBSERVATION 5. *If $a' = L_k^{(a,b)}$ and $b' = -(-b)^k$, then a' and b' are relatively prime.*

Proof. Suppose, to the contrary, that p is a common prime divisor of a' and b'. Then clearly p is a divisor of b, and also a divisor of $L_k^{(a,b)}$, which equals $bF_{k-1}^{(a,b)} + F_{k+1}^{(a,b)}$ by (5). This makes p a divisor of $F_{k+1}^{(a,b)}$, which contradicts Observation 2.

With these observations, we now can prove the

THEOREM. *The* gcd *of F_m and F_n is F_k, where k is the* gcd *of m and n.*

Proof. Let $s = m/k$ and $t = n/k$, and observe that s and t are relatively prime. We consider $A = \Omega_k F = F_0, F_k, F_{2k}, \ldots$. As discussed earlier, A can also be expressed as $F_k \cdot F^{(a',b')}$ where $a' = L_k$ and $b' = -(-b)^k$. Moreover, by Observation 5, a' and b' are relatively prime. As in Observation 1, we see at once that every A_j is a multiple of F_k, so in particular, F_k is a divisor of $A_s = F_{ks} = F_m$ and $A_t = F_{kt} = F_n$. On the other hand, $F_m/F_k = F_s^{(a',b')}$ and $F_n/F_k = F_t^{(a',b')}$, are relatively prime by Observation 4. Thus, F_k is the gcd of F_m and F_n. ∎

Several remarks about this result are in order. First, in Michael [**18**], the corresponding result is established for the traditional Fibonacci numbers. That proof depends on the $\mathcal{R}(1,1)$ instances of (19), Observation 1, and Observation 3, and extends to a proof for $\mathcal{R}(a,b)$ in a natural way.

Second, Holzsager [**10**] has described an easy construction of other sequences A_n for which $\gcd(A_n, A_m) = A_{\gcd(m,n)}$. First, for the primes p_k, define $A_{p_k} = q_k$ where the q_k are relatively prime. Then, extend A to the rest of the integers multiplicatively. That is, if $n = \prod p_i^{e_i}$ then $A_n = \prod q_i^{e_i}$. Such a sequence defines a mapping on the positive integers that carries the prime factorization of any subscript into a corresponding factorization involving the qs. This mapping apparently will commute with the gcd. By Observation 4, the terms $F_{p_k}^{(a,b)}$ are relatively prime, but since $F_2 = a$ and $F_4 = a^3 + 2ab$, the F mapping is not generally multiplicative. Thus, Holzsager's construction does not lead to examples of the form $F^{(a,b)}$.

Finally, we note that there is similar result for the (a,b)-Lucas numbers, which we omit in the interest of brevity. Both that result and the preceding Theorem also appear in Hilton and Pedersen [**8**]. Also, the general gcd result for F was known to Lucas, and we may conjecture that he knew the result for L as well.

In particular and in general

We have tried to show in this paper that much of the mystique of the Fibonacci numbers is misplaced. Rather than viewing F as a unique sequence with an amazing host of algebraic, combinatorial, and number theoretic properties, we ought to recognize that it is simply one example of a large class of sequences with such properties. In so arguing, we have implicitly highlighted the tension within mathematics between the particular and the general. Both have their attractions and pitfalls. On the one hand, by focusing too narrowly on a specific amazing example, we may lose sight of more general principles at work. But there is a countervailing risk that generalization may add nothing new to our understanding, and result in meaningless abstraction.

In the case at hand, the role of the skip operator should be emphasized. The proof of the gcd result, in particular, was simplified by the observation that the skip maps one $\mathcal{R}(a, b)$ to another. This observation offers a new, simple insight about the terms of Fibonacci sequences—an insight impossible to formulate without adopting the general framework of two-term recurrences.

It is not our goal here to malign the Fibonacci numbers. They constitute a fascinating example, rich with opportunities for discovery and exploration. But how much more fascinating it is that an entire world of such sequences exists. This world of the super sequences should not be overlooked.

REFERENCES

1. William Boulger, Pythagoras Meets Fibonacci, *Mathematics Teacher* **82** (1989), 277–782.
2. Carl B. Boyer, *A History of Mathematics*, 2nd ed., Revised by Uta C. Merzbach, Wiley, New York, 1991.
3. David M. Bressoud, *Factorization and Primality Testing*, Springer, New York, 1989.
4. Leonard E. Dickson, *History of the Theory of Numbers*, vol. 1: Divisibility and Primality, Chelsea, New York, 1952.
5. C. H. Edwards, Jr., *The Historical Development of the Calculus*, Springer, New York, 1979.
6. Ko Hayashi, Fibonacci numbers and the arctangent function, this MAGAZINE, **76**:3 (2003), 214–215.
7. Peter Hilton, Derek Holton, and Jean Pedersen, *Mathematical Reflections: In a Room with Many Mirrors*, Springer, New York, 1997.
8. Peter Hilton and Jean Pedersen, Fibonacci and Lucas numbers in teaching and research, *Journées Mathématiques & Informatique* **3** (1991–1992), 36–57.
9. Verner E. Hoggatt, Jr., *Fibonacci and Lucas Numbers*, Fibonacci Association, Santa Clara, CA, 1969.
10. Richard Holzsager, private correspondence, July, 2000.
11. Ross Honsberger, *Mathematical Gems III*, Mathematical Association of America, Washington, DC, 1985.
12. A. F. Horadam, A generalized Fibonacci sequence, *American Mathematical Monthly* **68** (1961), 455–459.
13. A. F. Horadam, Fibonacci number triples, *Amer. Math. Monthly* **68** (1961), 751–753.
14. A. F. Horadam, Generating functions for powers of a certain generalised sequences of numbers, *Duke Math. J.* **32** (1965), 437–446.
15. A. F. Horadam, Jacobsthal and Pell curves, *Fibonacci Quart.* **26** (1988), 77–83.
16. Thomas Koshy, *Fibonacci and Lucas Numbers with Applications*, Wiley, New York, 2001.
17. Edouard Lucas, Théorie des fonctions numériques simplement périodiques, *Amer. J. Math.* **1** (1878), 184–240, 289–321.
18. G. Michael, A new proof for an old property, *Fibonacci Quart.* **2** (1964), 57–58.
19. Charles W. Raine, Pythagorean triangles from the Fibonacci series, *Scripta Mathematica* 14 (1948), 164.
20. A. G. Shannon and A. F. Horadam, Generalized Fibonacci number triples, *Amer. Math. Monthly* **80** (1973), 187–190.
21. S. Vajda, *Fibonacci & Lucas Numbers, and the Golden Section, Theory and Applications,* Ellis Horwood Ltd., Chichester, 1989.
22. J. E. Walton and A. F. Horadam, Some further identities for the generalized Fibonacci sequence $\{H_n\}$, *Fibonacci Quart.* **12** (1974), 272–280.
23. Eric W. Weisstein, *CRC Concise Encyclopedia of Mathematics,* CRC Press, Boca Raton, FL, 1999. Also available on-line at `http://mathworld.wolfram.com`.

The Fibonacci Numbers—Exposed More Discretely

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

JENNIFER J. QUINN
Occidental College
Los Angeles, CA 90041
jquinn@oxy.edu*

In the previous article, Kalman and Mena [**5**] propose that Fibonacci and Lucas sequences, despite the mathematical favoritism shown them for their abundant patterns, are nothing more than ordinary members of a class of super sequences. Their arguments are beautiful and convinced us to present the same material from a more *discrete* perspective. Indeed, we will present a simple combinatorial context encompassing nearly all of the properties discussed in [**5**].

As in the Kalman-Mena article, we generalize Fibonacci and Lucas numbers: Given nonnegative integers a and b, the *generalized Fibonacci sequence* is

$$F_0 = 0, \quad F_1 = 1, \quad \text{and for } n \geq 2, \quad F_n = aF_{n-1} + bF_{n-2}. \tag{1}$$

The *generalized Lucas sequence* is

$$L_0 = 2, \quad L_1 = a, \quad \text{and for } n \geq 2, \quad L_n = aL_{n-1} + bL_{n-2}.$$

When $a = b = 1$, these are the celebrity Fibonacci and Lucas sequences. For now, we will assume that a and b are nonnegative integers. But at the end of the article, we will see how our methods can be extended to noninteger values of a and b.

Kalman and Mena prove the following generalized Fibonacci identities

$$F_n = F_m F_{n-m+1} + bF_{m-1}F_{n-m} \tag{2}$$

$$(a + b - 1)\sum_{i=1}^{n} F_i = F_{n+1} + bF_n - 1 \tag{3}$$

$$a\left(b^n F_0^2 + b^{n-1}F_1^2 + \cdots + bF_{n-1}^2 + F_n^2\right) = F_n F_{n+1} \tag{4}$$

$$F_{n-1}F_{n+1} - F_n^2 = (-1)^n b^{n-1} \tag{5}$$

$$\gcd(F_n, F_m) = F_{\gcd(n,m)} \tag{6}$$

$$L_n = aF_n + 2bF_{n-1} \tag{7}$$

$$L_n = F_{n+1} + bF_{n-1} \tag{8}$$

using the tool of difference operators acting on the real vector space of real sequences. In this paper, we offer a purely combinatorial approach to achieve the same results. We hope that examining these identities from different perspectives, the reader can more fully appreciate the unity of mathematics.

*Editor's Note: Readers interested in clever counting arguments will enjoy reading the authors' upcoming book, *Proofs That Really Count: The Art of Combinatorial Proof*, published by the MAA.

Fibonacci numbers—The combinatorial way

There are many combinatorial interpretations for Fibonacci and Lucas numbers [**3**]. We choose to generalize the "square and domino tiling" interpretation here. We show that the classic Fibonacci and Lucas identities naturally generalize to the (a, b) recurrences simply by adding a splash of color.

For nonnegative integers a, b, and n, let f_n count the number of ways to tile a $1 \times n$ board with 1×1 colored squares and 1×2 colored dominoes, where there are a color choices for squares and b color choices for dominoes. We call these objects *colored n-tilings*. For example, $f_1 = a$ since a length 1 board must be covered by a colored square; $f_2 = a^2 + b$ since a board of length 2 can be covered with two colored squares or one colored domino. Similarly, $f_3 = a^3 + 2ab$ since a board of length 3 can be covered by 3 colored squares or a colored square and a colored domino in one of 2 orders. We let $f_0 = 1$ count the empty board. Then for $n \geq 2$, f_n satisfies the generalized Fibonacci recurrence

$$f_n = af_{n-1} + bf_{n-2},$$

since a board of length n either ends in a colored square preceded by a colored $(n-1)$-tiling (tiled in af_{n-1} ways) or a colored domino preceded by a colored $(n-2)$-tiling (tiled in bf_{n-2} ways.) Since $f_0 = 1 = F_1$ and $f_1 = a = F_2$, we see that for all $n \geq 0$, $f_n = F_{n+1}$. After defining $f_{-1} = 0$, we now have a combinatorial definition for the generalized Fibonacci numbers.

THEOREM 1. *For $n \geq 0$, $F_n = f_{n-1}$ counts the number of colored $(n-1)$-tilings (of a $1 \times (n-1)$ board) with squares and dominoes where there are a colors for squares and b colors for dominoes.*

Using Theorem 1, equations (2) through (6) can be derived and appreciated combinatorially. In most of these, our combinatorial proof will simply ask a question and answer it two different ways.

For instance, if we apply Theorem 1 to equation (2) and reindex by replacing n by $n + 1$ and m by $m + 1$, we obtain

IDENTITY 1. *For $0 \leq m \leq n$,*

$$f_n = f_m f_{n-m} + bf_{m-1} f_{n-m-1}.$$

Question: How many ways can a board of length n be tiled with colored squares and dominoes?

Answer 1: By Theorem 1, there are f_n colored n-tilings.

Answer 2: Here we count how many colored n-tilings are *breakable* at the m-th cell and how many are not. To be breakable, our tiling consists of a colored m-tiling followed by a colored $(n-m)$-tiling, and there are $f_m f_{n-m}$ such tilings. To be *unbreakable* at the m-th cell, our tiling consists of a colored $(m-1)$-tiling followed by a colored domino on cells m and $m+1$, followed by a colored $(n-m-1)$-tiling; there are $bf_{m-1} f_{n-m-1}$ such tilings. Altogether, there are $f_m f_{n-m} + bf_{m-1} f_{n-m-1}$ colored n-tilings.

Since our logic was impeccable for both answers, they must be the same. The advantage of this proof is that it makes the identity *memorable* and visualizable. See FIGURE 1 for an illustration of the last proof.

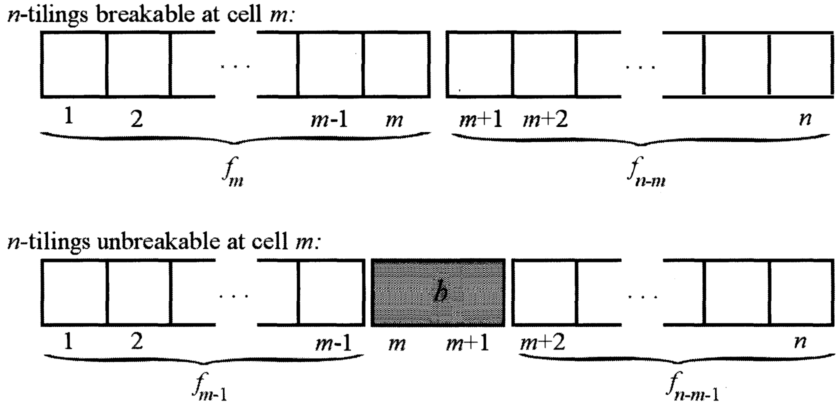


Figure 1 A colored n-tiling is either breakable or unbreakable at cell m

Equation (3) can be rewritten as the following identity.

IDENTITY 2. *For* $n \geq 0$,

$$f_n - 1 = (a-1)f_{n-1} + (a+b-1)[f_0 + f_1 + \cdots + f_{n-2}].$$

Question: How many colored n-tilings exist, excluding the tiling consisting of all white squares?

Answer 1: By definition, $f_n - 1$. (Notice how our question and answer become shorter with experience!)

Answer 2: Here we partition our tilings according to the last tile that is not a white square. Suppose the last tile that is not a white square begins on cell k. If $k = n$, that tile is a square and there are $a - 1$ choices for its color. There are f_{n-1} colored tilings that can precede it for a total of $(a-1)f_{n-1}$ tilings ending in a nonwhite square. If $1 \leq k \leq n-1$, the tile covering cell k can be a nonwhite square or a domino covering cells k and $k+1$. There are $a + b - 1$ ways to pick this tile and the previous cells can be tiled f_{k-1} ways. Altogether, there are $(a-1)f_{n-1} + \sum_{k=1}^{n-1}(a+b-1)f_{k-1}$ colored n-tilings, as desired.

Notice how easily the argument generalizes if we partition according to the last tile that is not a square of color 1 or 2 or... or c. Then the same reasoning gives us for any $1 \leq c \leq a$,

$$f_n - c^n = (a-c)f_{n-1} + \big((a-c)c + b\big)[f_0c^{n-2} + f_1c^{n-3} + \cdots + f_{n-2}]. \quad (9)$$

Likewise, by partitioning according to the last tile that is not a black domino, we get a slightly different identity, depending on whether the length of the tiling is odd or even:

$$f_{2n+1} = a(f_0 + f_2 + \cdots + f_{2n}) + (b-1)(f_1 + f_3 + \cdots + f_{2n-1}),$$
$$f_{2n} - 1 = a(f_1 + f_3 + \cdots + f_{2n-1}) + (b-1)(f_0 + f_2 + \cdots + f_{2n-2}).$$

After applying Theorem 1 to equation (4) and reindexing ($n \to n+1$) we have

IDENTITY 3. *For* $n \geq 0$,

$$a\sum_{k=0}^{n} f_k^2 b^{n-k} = f_n f_{n+1}.$$

Question: In how many ways can we create a colored n-tiling and a colored $(n+1)$-tiling?

Answer 1: $f_n f_{n+1}$.

Answer 2: For this answer, we ask for $0 \le k \le n$, how many colored tiling pairs exist where cell k is the last cell for which both tilings are breakable? (Equivalently, this counts the tiling pairs where the last square occurs on cell $k+1$ in exactly one tiling.) We claim this can be done $af_k^2 b^{n-k}$ ways, since to construct such a tiling pair, cells 1 through k of the tiling pair can be tiled f_k^2 ways, the colored square on cell $k+1$ can be chosen a ways (it is in the n-tiling if and only if $n-k$ is odd), and the remaining $2n-2k$ cells are covered with $n-k$ colored dominoes in b^{n-k} ways. See FIGURE 2. Altogether, there are $a\sum_{k=0}^{n} f_k^2 b^{n-k}$ tilings, as desired.

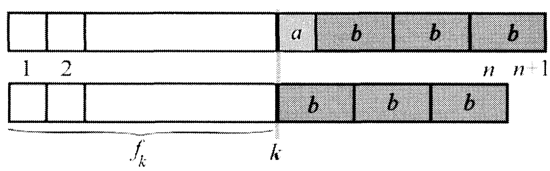


Figure 2 A tiling pair where the last mutually breakable cell occurs at cell k

The next identity uses a slightly different strategy. We hope that the reader does not *find fault* with our argument.

Consider the two colored 10-tilings offset as in FIGURE 3. The first one tiles cells 1 through 10; the second one tiles cells 2 through 11. We say that there is a *fault* at cell i, $2 \le i \le 10$, if both tilings are breakable at cell i. We say there is a fault at cell 1 if the first tiling is breakable at cell 1. Put another way, the pair of tilings has a fault at cell i for $1 \le i \le 10$ if neither tiling has a domino covering cells i and $i+1$. The pair of tilings in FIGURE 3 has faults at cells 1, 2, 5, and 7. We define the *tail* of a tiling to be the tiles that occur after the last fault. Observe that if we swap the tails of FIGURE 3 we obtain the 11-tiling and the 9-tiling in FIGURE 4 and it has the same faults.

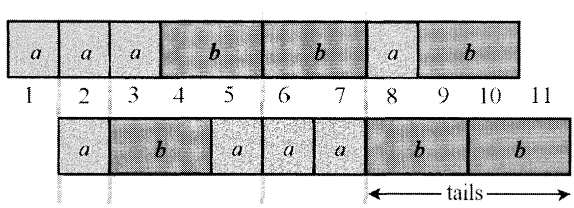


Figure 3 Two 10-tilings with their faults (indicated with gray lines) and tails

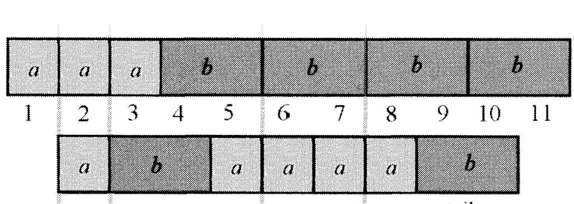


Figure 4 After tail-swapping, we have an 11-tiling and a 9-tiling with exactly the same faults

Tail swapping is the basis for the identity below, based on (5). At first glance, it may appear unsuitable for combinatorial proof due to the presence of the $(-1)^n$ term. Nonetheless, we will see that this term is merely the error term of an *almost* one-to-one correspondence between two sets whose sizes are easily counted. We use a different format for this combinatorial proof.

IDENTITY 4. $f_n^2 = f_{n+1}f_{n-1} + (-1)^n b^n$

Set 1: Tilings of two colored n-boards (a *top* board and a *bottom* board). By definition, this set has size f_n^2.

Set 2: Tilings of a colored $(n+1)$-board and a colored $(n-1)$-board. This set has size $f_{n+1}f_{n-1}$.

Correspondence: First, suppose n is odd. Then the top and bottom board must each have at least one square. Notice that a square in cell i ensures that a fault must occur at cell i or cell $i-1$. Swapping the tails of the two n-tilings produces an $(n+1)$-tiling and an $(n-1)$-tiling with the same tails. This produces a 1-to-1 correspondence between all pairs of n-tilings and all tiling pairs of sizes $n+1$ and $n-1$ that have faults. Is it possible for a tiling pair with sizes $n+1$ and $n-1$ to be fault free? Yes, with all colored dominoes in *staggered formation* as in FIGURE 5, which can occur b^n ways. Thus, when n is odd, $f_n^2 = f_{n+1}f_{n-1} - b^n$.

Similarly, when n is even, tail swapping creates a 1-to-1 correspondence between faulty tiling pairs. The only fault-free tiling pair is the all domino tiling of FIGURE 6. Hence, $f_n^2 = f_{n+1}f_{n-1} + b^n$. Considering the odd and even case together produces our identity.

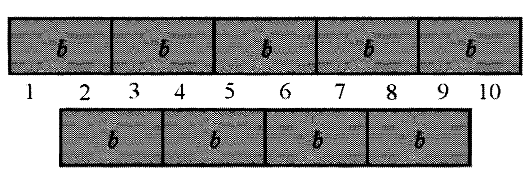


Figure 5 When n is odd, the only fault-free tiling pairs consist of all dominoes

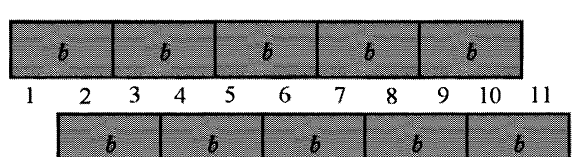


Figure 6 When n is even, the only fault-free tiling pairs consist of all dominoes

We conclude this section with a combinatorial proof of what we believe to be the most beautiful Fibonacci fact of all.

THEOREM 2. *For generalized Fibonacci numbers defined by (1) with relatively prime integers a and b,*

$$\gcd(F_n, F_m) = F_{\gcd(n,m)}. \qquad (10)$$

We will need to work a little harder to prove this theorem combinatorially, but it can be done. Fortuitously, we have already combinatorially derived the identities needed to prove the following lemma.

LEMMA 1. *For generalized Fibonacci numbers defined by (1) with relatively prime integers a and b and for all $m \geq 1$, F_m and bF_{m-1} are relatively prime.*

Proof. First we claim that F_m is relatively prime to b. By conditioning on the location of the last colored domino (if any exist), equation (9) says (after letting $c = a$ and reindexing),

$$F_m = a^{m-1} + b\sum_{j=1}^{m-2} a^{j-1} F_{m-1-j}.$$

Consequently, if $d > 1$ is a divisor of F_m and b, then d must also divide a^{m-1}, which is impossible since a and b are relatively prime.

Next we claim that F_m and F_{m-1} are relatively prime. This follows from equation (5) since if $d > 1$ divides F_m and F_{m-1}, then d must divide b^{m-1}. But this is impossible since F_m and b are relatively prime.

Thus since $\gcd(F_m, b) = 1$ and $\gcd(F_m, F_{m-1}) = 1$, then $\gcd(F_m, bF_{m-1}) = 1$, as desired. ∎

To prove Theorem 2, we exploit Euclid's algorithm for computing greatest common divisors: If $n = qm + r$ where $0 \leq r < m$, then

$$\gcd(n, m) = \gcd(m, r).$$

Since the second component gets smaller at each iteration, the algorithm eventually reaches $\gcd(g, 0) = g$, where g is the greatest common divisor of n and m. The identity below shows one way that F_n can be expressed in terms of F_m and F_r. It may look formidable at first but makes perfect sense when viewed combinatorially.

IDENTITY 5. *If $n = qm + r$, where $0 \leq r < m$, then*

$$F_n = (bF_{m-1})^q F_r + F_m \sum_{j=1}^{q} (bF_{m-1})^{j-1} F_{(q-j)m+r+1}.$$

Question: How many colored $(qm + r - 1)$-tilings exist?

Answer 1: $f_{qm+r-1} = F_{qm+r} = F_n$.

Answer 2: First we count all such colored tilings that are unbreakable at every cell of the form $jm - 1$, where $1 \leq j \leq q$. Such a tiling must have a colored domino starting on cell $m - 1, 2m - 1, \ldots, qm - 1$, which can be chosen b^q ways. Before each of these dominoes is an arbitrary $(m - 2)$-tiling that can each be chosen f_{m-2} ways. Finally, cells $qm + 1, qm + 2, \ldots, qm + r - 1$ can be tiled f_{r-1} ways. See FIGURE 7. Consequently, the number of colored

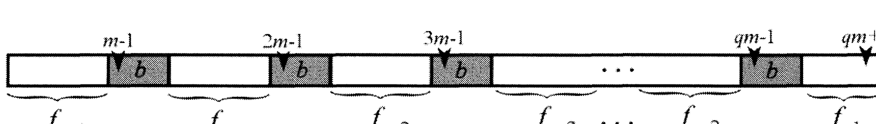


Figure 7 There are $(bF_{m-1})^q F_r$ colored $(qm + r - 1)$-tilings with no breaks at any cells of the form $jm - 1$ where $1 \leq j \leq q$

tilings with no $jm-1$ breaks is $b^q(f_{m-2})^q f_{r-1} = (bF_{m-1})^q F_r$. Next, we partition the remaining colored tilings according to the first breakable cell of the form $jm-1$, $1 \le j \le q$. By similar reasoning as before, this can be done $(bF_{m-1})^{j-1} F_m F_{(q-j)m+r+1}$ ways. (See FIGURE 8.) Altogether, the number of colored tilings is $(bF_{m-1})^q F_r + F_m \sum_{j=1}^{q} (bF_{m-1})^{j-1} F_{(q-j)m+r+1}$.

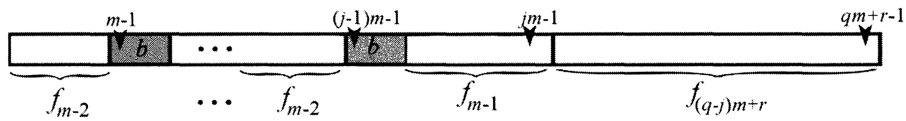


Figure 8 There are $(bF_{m-1})^{j-1} F_m F_{(q-j)m+r+1}$ colored $(qm+r-1)$-tilings that are breakable at cell $jm-1$, but not at cells of the form $im-1$ where $1 \le i < j$

The previous identity explicitly shows that F_n is an integer combination of F_m and F_r. Consequently, d is a common divisor of F_n and F_m if and only if d divides F_m and $(bF_{m-1})^q F_r$. But by Lemma 1, since F_m is relatively prime to bF_{m-1}, d must be a common divisor of F_m and F_r. Thus F_n and F_m have the same common divisors (and hence the same gcd) as F_m and F_r. In other words,

COROLLARY 1. *If $n = qm + r$, where $0 \le r < m$, then*

$$\gcd(F_n, F_m) = \gcd(F_m, F_r).$$

But wait!! This corollary is the same as Euclid's algorithm, but with $F's$ inserted everywhere. This proves Theorem 2 by following the same steps as Euclid's algorithm. The $\gcd(F_n, F_m)$ will eventually reduce to $\gcd(F_g, F_0) = (F_g, 0) = F_g$, where g is the greatest common divisor of m and n.

Lucas numbers—the combinatorial way

Generalized Lucas numbers are nothing more than generalized Fibonacci numbers running in circles. Specifically, for nonnegative integers a, b, and n, let ℓ_n count the number of ways to tile a circular $1 \times n$ board with slightly curved colored squares and dominoes, where there are a colors for squares and b colors for dominoes. Circular tilings of length n will be called *n-bracelets*. For example, when $a = b = 1$, $\ell_4 = 7$, as illustrated in FIGURE 9. In general, $\ell_4 = a^4 + 4a^2b + 2b^2$.

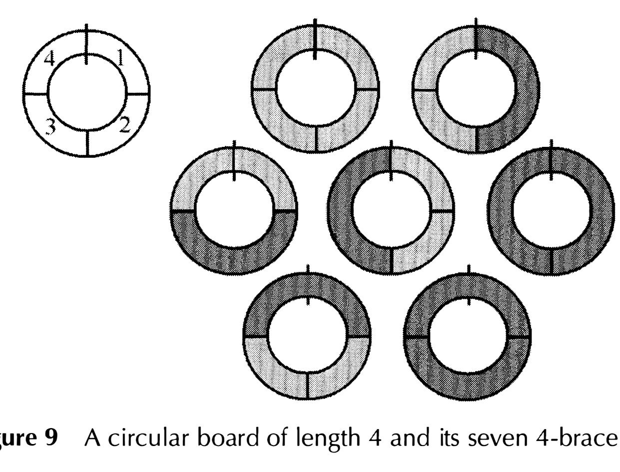


Figure 9 A circular board of length 4 and its seven 4-bracelets

From the definition of ℓ_n it follows that $\ell_n \ge f_n$ since an n-bracelet can have a domino covering cells n and 1; such a bracelet is called *out-of-phase*. Otherwise, there is a break between cells n and 1, and the bracelet is called *in-phase*. The first 5 bracelets in FIGURE 9 are in-phase and the last 2 are out-of-phase. Notice $\ell_1 = a$ and $\ell_2 = a^2 + 2b$ since a circular board of length 2 can be covered with two squares, an in-phase domino, or an out-of-phase domino. We define $\ell_0 = 2$ to allow 2 empty bracelets, one in-phase and one out-of-phase. In general for $n \geq 2$, we have

$$\ell_n = a\ell_{n-1} + b\ell_{n-2}$$

because an n-bracelet can be created from an $(n-1)$-bracelet by inserting a square to the left of the first tile or from an $(n-2)$-bracelet by inserting a domino to the left of the first tile. The *first tile* is the one covering cell 1 and it determines the phase of the bracelet; it may be a square, a domino covering cells 1 and 2, or a domino covering cells n and 1.

Since $\ell_0 = 2 = L_0$ and $\ell_1 = a = L_1$, we see that for all $n \geq 0$, $\ell_n = L_n$. This becomes our combinatorial definition for the generalized Lucas numbers.

THEOREM 3. *For all $n \geq 0$, $L_n = \ell_n$ counts the number of n-bracelets created with colored squares and dominoes where there are a colors for squares and b colors for dominoes.*

Now that we know how to combinatorially think of Lucas numbers, generalized identities are a piece of cake. Equation (7), which we rewrite as

$$L_n = af_{n-1} + 2bf_{n-2},$$

reflects the fact that an n-bracelet can begin with a square (af_{n-1} ways), an in-phase domino (bf_{n-2} ways), or an out-of-phase domino (bf_{n-2} ways). Likewise, equation (8), rewritten as

$$L_n = f_n + bf_{n-2},$$

conditions on whether or not an n-bracelet is in-phase (f_n ways) or out-of-phase (bf_{n-2} ways.)

You might even think these identities are too easy, so we include a couple more generalized Lucas identities for you to ponder along with visual hints. For more details see [**4**].

$$f_{n-1}L_n = f_{2n-1} \qquad \text{See FIGURE 10.}$$
$$L_n^2 = L_{2n} + 2 \cdot (-b)^n \qquad \text{See FIGURE 11.}$$

Further generalizations and applications

Up until now, all of our proofs have depended on the fact that the recurrence coefficients a and b were nonnegative integers, even though most generalized Fibonacci identities remain true when a and b are negative or irrational or even complex numbers. Additionally, our sequences have had very specific initial conditions ($F_0 = 0$, $F_1 = 1$, $L_0 = 2$, $L_1 = a$), yet many identities can be extended to handle arbitrary ones. This section illustrates how combinatorial arguments can still be used to overcome these apparent obstacles.

Arbitrary initial conditions Let a, b, A_0, and A_1 be nonnegative integers and consider the sequence A_n defined by the recurrence, for $n \geq 2$, $A_n = aA_{n-1} + bA_{n-2}$. As

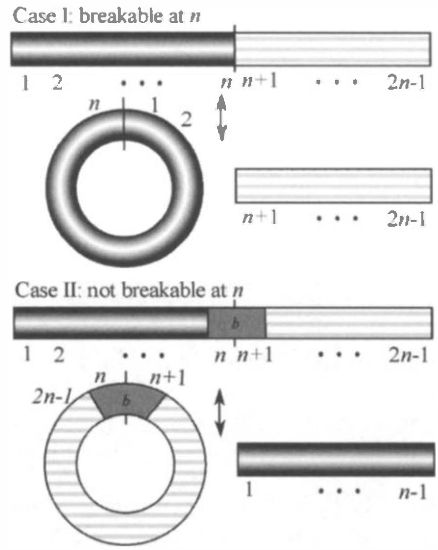


Figure 10 Picture for $f_{n-1}L_n = f_{2n-1}$

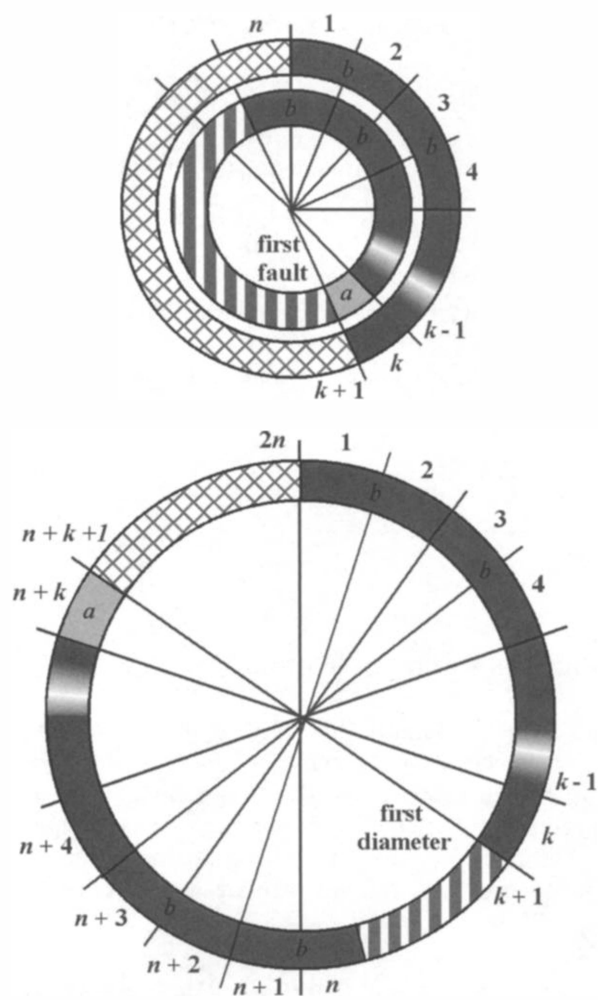


Figure 11 Picture for $L_n^2 = L_{2n} + 2 \cdot (-b)^n$ when n is even

described in [**1**] and Chapter 3 of [**4**], the *initial conditions* A_0 and A_1 determine the number of choices for the *initial tile*. Just like F_n, A_n counts the number of colored n-tilings where *except for the first tile* there are a colors for squares and b colors for dominoes. For the first tile, we allow A_1 choices for a square and bA_0 choices for a domino. So as not to be confused with the situation using ideal initial conditions, we assign the first tile a *phase* instead of a color.

For example, when $A_0 = 1$ and $A_1 = a$, the ideal initial conditions, we have a choices for the phase of an initial square and b choices for the phase of an initial domino. Since *all* squares have a choices and *all* dominoes have b choices, it follows that $A_n = f_n$. When $A_0 = 0$ and $A_1 = 1$, A_n counts the number of colored n-tilings that begin with an "uncolored" square; hence $A_n = f_{n-1} = F_n$. When $A_0 = 2$ and $A_1 = a$, A_n counts the number of colored n-tilings that begin with a square in one of a phases or a domino in one of $2b$ phases. This is equivalent to a colored n-bracelet since there are an equal number of square phases as colors and twice as many domino phases as colors (representing whether the initial domino is in-phase or out-of-phase.) Thus when $A_0 = 2$ and $A_1 = a$, we have $A_n = L_n$.

In general, there are $A_1 f_{n-1}$ colored tilings that begin with a phased square and $bA_0 f_{n-2}$ colored tilings that being with a phased domino. Hence we obtain the following identity from Kalman and Mena [**5**]:

$$A_n = bA_0 F_{n-1} + A_1 F_n. \tag{11}$$

Arbitrary recurrence coefficients Rather than assigning a discrete number of colors for each tile, we can assign weights. Squares have weight a and dominoes have weight b except for the initial tile, which has weight A_1 as a square and weight bA_0 as a domino. Here a, b, A_0, and A_1 do not have to be nonnegative integers, but can be chosen from the set of complex numbers (or from any commutative ring). We define the *weight of an n-tiling* to be the product of the weights of its individual tiles. For example, the 7-tiling "square-domino-domino-square-square" has weight a^3b^2 with ideal initial conditions and has weight $A_1a^2b^2$ with arbitrary initial conditions. Inductively one can prove that for $n \geq 1$, A_n is the sum of the weights of all weighted n-tilings, which we call the *total weight* of an n-board.

If X is an m-tiling of weight w_X and Y is an n-tiling of weight w_Y, then X and Y can be glued together to create an $(m + n)$-tiling of weight $w_X w_Y$. If an m-board can be tiled s different ways and has total weight $A_m = w_1 + w_2 + \cdots + w_s$ and an n-board can be tiled t ways with total weight $A_n = x_1 + x_2 + \cdots + x_t$, then the sum of the weights of all weighted $(m + n)$-tilings breakable at cell m is

$$\sum_{i=1}^{s}\sum_{j=1}^{t} w_i x_j = (w_1 + w_2 + \cdots + w_s)(x_1 + x_2 + \cdots + x_t) = A_m A_n.$$

Now we are prepared to revisit some of our previous identities using the weighted approach. For Identity 1, we find the total weights of an n-board in two different ways. On the one hand, since the initial conditions are ideal, the total weight is $A_n = f_n$. On the other hand, the total weight is comprised of the total weight of those tilings that are breakable at cell m ($f_m f_{n-m}$) plus the total weight of those tilings that are unbreakable at cell m ($f_{m-1} b f_{n-m-1}$). Identities 2, 3, and 5 can be argued in similar fashion.

For Identity 4, we define the weight of a tiling pair to be the product of the weights of all the tiles, and define the total weight as before. Next we observe that tail swapping preserves the weight of the tiling pair since no tiles are created or destroyed in the process. Consequently, the total weight of the set of *faulty* tiling pairs (X, Y) where

X and Y are n-tilings equals the total weight of the faulty tiling pairs (X', Y'), where X' is an $(n+1)$-tiling and Y' is an $(n-1)$-tiling. The fault-free tiling pair, for the even and odd case, will consist of n dominoes and therefore have weight b^n. Hence identity 4 remains true even when a and b are complex numbers.

Kalman and Mena **[5]** prove Binet's formulas for Fibonacci numbers

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right], \tag{12}$$

and for more general sequences.

These can also be proved combinatorially **[2]**. Binet's formula follows from considering a random tiling of an infinitely long strip with cells $1, 2, 3, \ldots$, where squares and dominoes are randomly and independently inserted from left to right. The probability of inserting a square is $1/\phi$ and the probability of inserting a domino is $1/\phi^2$, where $\phi = (1+\sqrt{5})/2$. (Conveniently, $1/\phi + 1/\phi^2 = 1$.) By computing the probability of being breakable at cell $n-1$ in two different ways, Binet's formula instantly appears. This approach can be extended to generalized Fibonacci numbers and beyond, as described in **[1]**.

Finally, we observe that the Pythagorean Identity presented in **[5]** for traditional Fibonacci numbers, which can be written as

$$(f_{n-1}f_{n+2})^2 + (2f_n f_{n+1})^2 = f_{2n+2}^2$$

can also be proved combinatorially. For details, see **[4]**.

We hope that this paper illustrates that Fibonacci and Lucas sequences are members of a very special class of sequences satisfying beautiful properties, namely sequences defined by second order recurrence relations. But why stop there? Combinatorial interpretations can be given to sequences that satisfy higher-order recurrences. That is, if we define $a_j = 0$ for $j < 0$ and $a_0 = 1$, then for $n \geq 1$, $a_n = c_1 a_{n-1} + \cdots + c_k a_{n-k}$ counts the number of ways to tile a board of length n with colored tiles of length at most k, where each tile of length i has c_i choices of color. Again, this interpretation can be extended to handle complex values of c_i and arbitrary initial conditions. See Chapter 3 of **[4]**. Of course, the identities tend to be prettier for the two-term recurrences, and are usually prettiest for the traditional Fibonacci and Lucas numbers.

Acknowledgment. We thank Jeremy Rouse, Dan Kalman, and Robert Mena for their inspiration.

REFERENCES

1. A. T. Benjamin, C. R. H. Hanusa, and F. E. Su, Linear recurrences and Markov chains, *Utilitas Mathematica* **64** (2003), 3–17.
2. A. T. Benjamin, G. M. Levin, K. Mahlburg, and J. J. Quinn, Random approaches to Fibonacci identities, *Amer. Math. Monthly* **107**:6 (2000), 511–516.
3. A. T. Benjamin and J. J. Quinn, Recounting Fibonacci and Lucas Identities, *College Math. J.* **30**:5 (1999), 359–366.
4. ———, *Proofs That Really Count: The Art of Combinatorial Proof*, Mathematical Association of America, Washington, D.C., 2003.
5. D. Kalman and R. Mena, The Fibonacci Numbers—Exposed, this MAGAZINE **76**:3 (2003), 167–181.
6. T. Koshy, *Fibonacci and Lucas Numbers with Applications*, John Wiley and Sons, NY, 2001.
7. S. Vajda, *Fibonacci & Lucas Numbers, and the Golden Section: Theory and Applications,* John Wiley and Sons, New York, 1989.

Gaspard Monge and the Monge Point of the Tetrahedron

ROBERT ALAN CRABBS
P. O. Box 509
Loughman, FL 33858-0509

Monge and his contributions to mathematics

Gaspard Monge (1746–1818) was a man of extraordinary talent. Despite humble origins, he founded one new branch of mathematics, made major early contributions to a second, and became a close friend of Napoleon Bonaparte (1769–1821).

Monge was born on May 9 or 10, 1746, in Beaune, France [**15**, p. 9]. His father, a peddler and later a storekeeper, valued education and saw to the education of his three sons. By age 14, Monge made his exceptional ability evident by independently constructing a fire engine. At the Collège de la Trinité in Lyon, he so impressed his teachers that they invited him to teach physics at age 16 or 17. In the summer of 1764, at age 18, Monge returned to Beaune. There, having devised his own plans of observation and constructed his own surveying instruments, he created with remarkable skill and care a large-scale map of his hometown. (The original is still at the Beaune library.) Word of this map reached the prestigious École Royale du Génie de Mézières (Mézières Royal School of Engineering), and a high-ranking officer there offered Monge a position as a draftsman. Although Monge was unaware that he could not become a student officer because he was a commoner by birth, his decision to accept the offer turned out to be a good one.

Monge was politically active, and held several government posts. From 1783 to 1789 he was an examiner of naval cadets, and from August or September 1792 until April 1793 he was Minister of the Navy, a position made difficult by the troubles and failures of the French navy, which made Monge a target of criticism. Shortly after resigning as Minister of the Navy, Monge began supervising armaments factories and writing instruction manuals for the workers. Monge supported the French Revolution, but in the turmoil that developed, many people had unjust accusations leveled against them and were executed. Monge himself was sometimes in danger, and at one point, after being denounced by the porter at his lodgings, he left Paris. In 1796, Napoleon wrote to Monge to offer him his friendship and a position. The two had met when Monge was Minister of the Navy and Napoleon, then a little-known artillery officer, had been impressed by how Monge had treated him. Monge was sent to Italy to obtain artworks for France, and during 1798 and 1799 he accompanied Napoleon on his Egyptian campaign. After establishing the Consulate in 1799, Napoleon named Monge a senator for life.

Monge's strong sense of justice and equality, his honesty, and his kindness were evident throughout his political career. As an examiner of naval cadets Monge rejected outright unqualified sons of aristocrats. He spoke frankly with Napoleon, and may have exerted a moderating influence on him [**2**, pp. 190, 196, 204]. When Napoleon returned from exile in Elba, for instance, Monge successfully counseled against taking excessive vengeance. Sadly for Monge, he was stripped of his honors after Napoleon's final fall from power in 1815. He died in Paris on July 28, 1818.

Monge held several teaching positions, beginning with his teaching duties in physics at age 16 or 17. In 1769, at age 22, Monge became a mathematics professor at the École Royale du Génie de Mézières; by 1772 he was teaching physics as well. After his election to the Académie des Sciences in 1780, Monge began to spend long periods in Paris. Doing so allowed him to teach hydrodynamics at the Louvre (this position had been created by Anne Robert Turgot (1727–1781), a statesman and economist [**15**, p. 23]), but it also forced him to resign his post at Mézières in 1784. During 1794–1795 Monge taught at the École Normale de l'an III, and starting in 1795, and again after returning from Egypt, he taught at the École Polytechnique. Declining health finally forced him to give up teaching in 1809.

Monge was an influential educator for several reasons. One is his close association with the École Polytechnique. Monge successfully argued for a single engineering school rather than several specialized schools, helped to develop the curriculum, talked with professors, advised administrators, and supervised the opening of the school. In 1797, Monge was appointed Director. With Monge, Laplace, and Lagrange among its first faculty, the École Polytechnique was influential from the beginning, and it remains France's best-known technical school.

A second reason for Monge's influence is that he was an exceptional and inspiring teacher, well liked and respected by his students. Monge was, according to Boyer, "perhaps the most influential mathematics teacher since the days of Euclid." [**4**, p. 468] Among his students who made mathematical contributions of their own were Lazare Carnot (1753–1823), Charles Brianchon (1785–1864), Jean Victor Poncelet (1788–1867), Charles Dupin (1784–1873), J. B. Meusnier (1754–1793), E. L. Malus (1775–1812), and O. Rodrigues (1794–1851).

Monge's major mathematical contributions were in the areas of descriptive geometry, differential geometry, and analytic geometry. His work in descriptive geometry began in his early days at Mézières, when he was assigned to produce a plan for a fortress that would hide and protect its defenders from enemy attack. At the time such problems were solved by long computations, but Monge used a geometrical method to solve the problem so quickly that his superior officer at first refused to believe Monge's results. (For details on the problem and on Monge's solution, see Taton's account [**15**, pp. 12–14].) When the solution was found to be correct, it was made a military secret and Monge was assigned to teach the method. (He was not allowed to teach the method publicly until 1794, at the École Normale de l'an III.)

In 1802, Monge and Jean-Nicolas-Pierre Hachette (1769–1834) published *Application d'algèbre à la géométrie*, in the *Journal de l'École Polytechnique*. (This work appeared again in 1805 and 1807.) It summarized Monge's lectures in solid analytic geometry. Boyer remarks: "The notation, phraseology, and methods are virtually the same as those to be found in any textbook of today. The definitive form of analytic geometry finally had been achieved, more than a century and a half after Descartes and Fermat had laid the foundations." [**3**, p. 220]

Monge published two papers [**11, 12**] entitled "Sur la pyramide triangulaire," in 1809 and 1811. In the remainder of the present paper, we focus mainly on the content of these two papers. A detailed account of all of Monge's work on the tetrahedron is given by Taton [**15**, pp. 241–246].

The Monge point of the tetrahedron

Two edges of a tetrahedron are called *opposite edges* if they have no common vertex. A tetrahedron $ABCD$ has three pairs of opposite edges; one pair is AB and CD. With this convention, we can define six *Monge planes*, one for each edge:

DEFINITION. *A* Monge plane *is perpendicular to one edge of a tetrahedron and contains the midpoint of the opposite edge.*

The following theorem guarantees the existence of a *Monge point*, where the Monge planes meet.

THEOREM 1. *The six Monge planes of a tetrahedron are concurrent.*

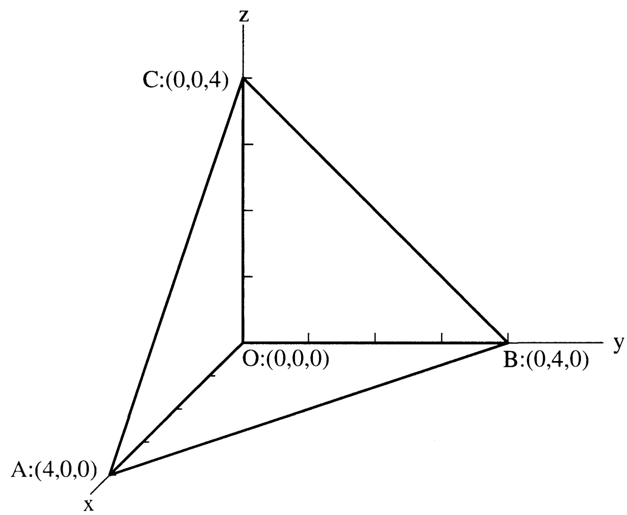


Figure 1 A tetrahedron with vertices at the points $O(0, 0, 0)$, $A(4, 0, 0)$, $B(0, 4, 0)$, and $C(0, 0, 4)$. The six Monge planes $x = 0$, $y = 0$, $z = 0$, $x = y$, $y = z$, and $x = z$ intersect at the origin.

Example: FIGURE 1 shows the tetrahedron with vertices at $O(0, 0, 0)$, $A(4, 0, 0)$, $B(0, 4, 0)$, and $C(0, 0, 4)$. The six Monge planes of this tetrahedron are the three coordinate planes and the perpendicular bisectors of edges AB, BC, and CA, that is, the planes $x = 0$, $y = 0$, $z = 0$, $x = y$, $y = z$, and $z = x$. The origin lies on all six of these planes and is therefore the Monge point. Notice also that three Monge planes are perpendicular to any given face of the tetrahedron and therefore intersect in a line perpendicular to the face. For example, the Monge planes $x = 0$, $y = 0$, and $x = y$ are perpendicular to face AOB and intersect in the z-axis. The three Monge planes perpendicular to face ABC intersect in the line $x = y = z$.

The following proof differs only in details from the one given by Monge [**12**, pp. 263–265]. A similar proof is given by Thompson [**16**].

Proof. In tetrahedron $ABCD$, construct a triangle $B'C'D'$ by connecting the midpoints of the three edges issuing from vertex A (as in FIGURE 2). Each of the three Monge planes perpendicular to an edge of face BCD cuts the triangle $B'C'D'$ in one of its altitudes, for each contains a vertex of this triangle and cuts its opposite side perpendicularly. Therefore, these Monge planes intersect in the perpendicular from the orthocenter H' of triangle $B'C'D'$ to face BCD. Let us call this line the *Monge normal* of face BCD. By the same reasoning, each face of the tetrahedron has a Monge normal associated with it.

The Monge normals of faces ABC and ACD intersect at a point M, for they are not parallel and they both lie in the Monge plane perpendicular to edge AC. We can now be certain that M lies on every Monge plane except possibly the one perpendicular to edge BD. Since M lies on the Monge planes perpendicular to edges AB and AD, it lies on their line of intersection. But their line of intersection is the Monge normal of face

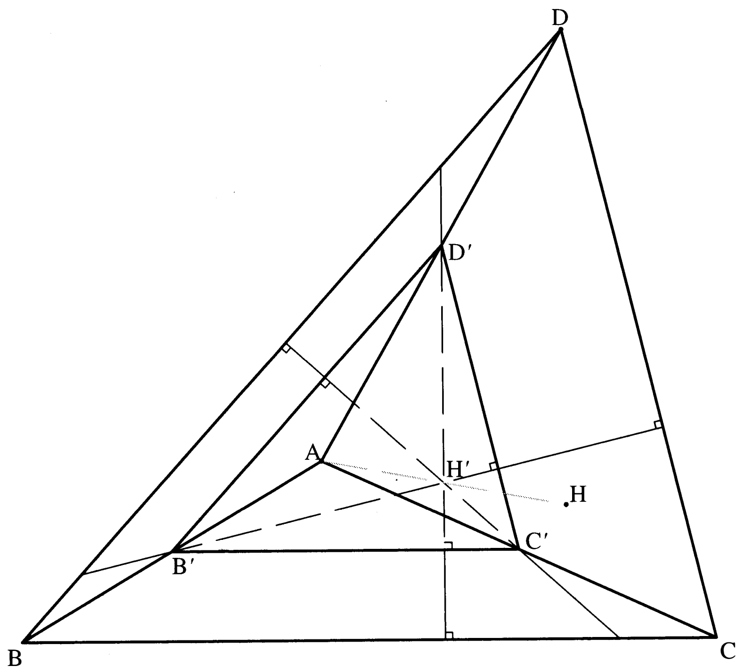


Figure 2 A tetrahedron *ABCD* with a midpoint triangle *B′C′D′* parallel to face *BCD*. Face *BCD* lies in the plane of the paper and the view is from directly above the tetrahedron. Intersections involving the three Monge planes perpendicular to the edges of face *BCD* are marked as follows: thin solid lines indicate the intersections of the Monge planes with that portion of the surface of the tetrahedron lying between face *BCD* and triangle *B′C′D′*; dashed lines indicate the intersections with *B′C′D′*.

ABD, which lies in the Monge plane perpendicular to *BD*. Hence, *M* lies on all six Monge planes, and the theorem is established. ∎

The Monge point, the circumcenter, and the centroid

Monge actually discovered and proved more than Theorem 1. Before stating his more general result, we will describe relationships among the bimedians and the centroid; relate the centroid to the center of mass and the center of gravity; define the circumcenter; and give an example.

DEFINITION. *A* bimedian *is a line segment connecting the midpoints of opposite edges of a tetrahedron.*

LEMMA. *The three bimedians of a tetrahedron bisect each other.*

We give analytic and synthetic proofs. The first proof is very simple; Monge himself gave the second **[11**, p. 2].

Analytic proof. Write the four vertices as $A(a_1, a_2, a_3)$, $B(b_1, b_2, b_3)$, $C(c_1, c_2, c_3)$, and $D(d_1, d_2, d_3)$. Straightforward uses of the midpoint formula (three for each bimedian) show that the midpoint of each bimedian is the point $((a_1 + b_1 + c_1 + d_1)/4$, $(a_2 + b_2 + c_2 + d_2)/4, (a_3 + b_3 + c_3 + d_3)/4)$. ∎

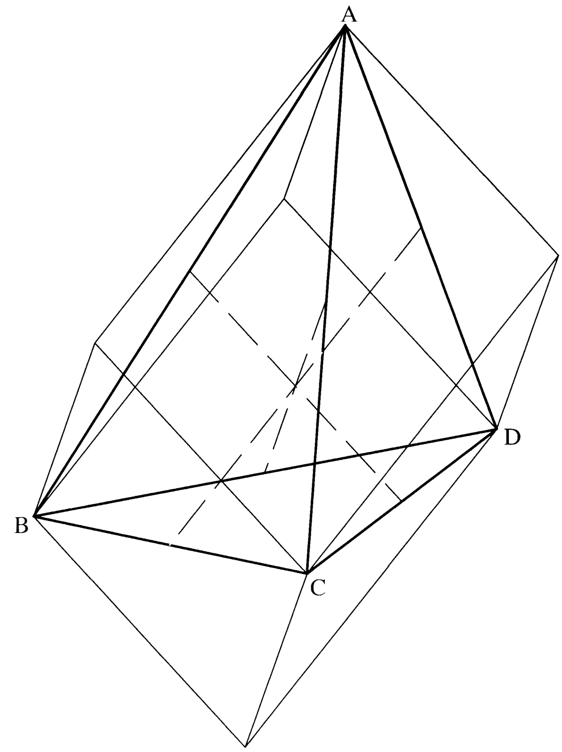


Figure 3 Tetrahedron *ABCD* (thick solid lines), its circumscribed parallelepiped (thin solid lines), and its three bimedians (dashed lines). The bimedians are the axes of the parallelepiped, so they bisect each other at the center of the parallelepiped.

Synthetic proof. Through each edge of the tetrahedron, construct a plane that is parallel to the opposite edge. (To construct the desired plane through, say, edge AB, construct a line through AB that is parallel to edge CD.) This process produces a parallelepiped that circumscribes the tetrahedron. (See FIGURE 3.) Each edge of the tetrahedron is a diagonal of a face of the parallelepiped; opposite edges of the tetrahedron are diagonals of opposite faces of the parallelepiped. Hence, the three bimedians of the tetrahedron are the axes of the parallelepiped (the segments connecting the centers of opposite faces). Therefore, the bimedians bisect each other at the center of the parallelepiped, and the proof is complete. ■

Before stating the next theorem, we note that the centroid is sometimes referred to as the center of mass or the center of gravity. Monge himself [**11, 12**] uses the term "center of gravity." The centroid of a solid coincides with its center of mass if the solid has uniform density. (The centroid of a solid is also known as its center of volume.) The center of gravity of a solid coincides with its center of mass if the solid is subjected to a uniform gravitational field. Hence, for the case of an ideal tetrahedron that is both of uniform density and subjected to a uniform gravitational field, the center of mass, center of gravity, and centroid all coincide.

Polya discusses the center of gravity (centroid) of the tetrahedron in an intuitive manner [**14**, pp. 38–45].

THEOREM 2. *The centroid of a tetrahedron lies at the common midpoint of its three bimedians.*

This theorem plays a crucial role in what follows. Monge [**11**] gave two complete proofs of it, and also stated a previously-known theorem from which he said the result could be easily derived. (The theorem asserts that $d_G = (d_A + d_B + d_C + d_D)/4$, where the variables are the signed distances of the centroid G and the vertices from an arbitrary plane, the distances being positive for points on one side of the plane and negative for points on the other.) Following are some key ideas from Monge's first proof.

Let us consider a tetrahedron as an infinite collection of line segments, all parallel to one edge of the tetrahedron. A *median plane* of a tetrahedron contains one edge and the midpoint of the opposite edge. If a median plane bisects the edge that the line segments are parallel to, it will bisect all of the line segments that constitute the tetrahedron. (See FIGURE 4.) It follows that the centers of mass of all of the line segments—and therefore also the center of mass of the entire tetrahedron—lie on the median plane. Since we may consider line segments parallel to any edge of the tetrahedron, we conclude that the centroid lies on all six median planes. But median planes that contain opposite edges intersect in a bimedian. Hence, the centroid of the tetrahedron lies on each of the three bimedians. Since, by the Lemma, the bimedians bisect each other, the centroid is their common midpoint.

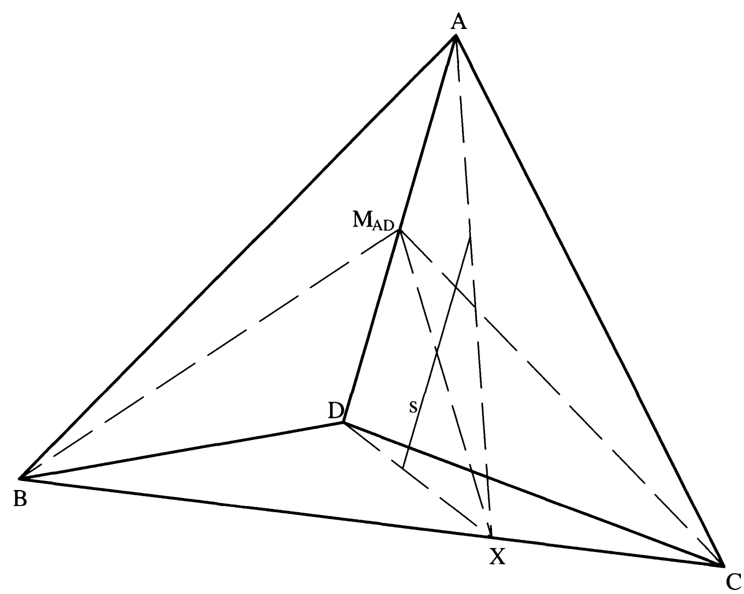


Figure 4 An arbitrary segment s that is parallel to edge AD and that has its endpoints in faces ABC and BCD. Triangle ADX is the intersection of the tetrahedron with the plane of edge AD and segment s. Median plane BCM_{AD} bisects segment s because it contains XM_{AD}, which is a median of triangle ADX and bisects s.

The centroid of a tetrahedron can be defined as the intersection of its bimedians or as the intersection of its four medians (the line segments from a vertex to the centroid of the opposite face). Monge related these two definitions [**11**, p. 4]. If two median planes contain opposite edges of a tetrahedron, they will intersect in a bimedian; if median planes contain edges issuing from the same vertex of a tetrahedron, they will intersect in a median. To see this consider the three median planes ABM_{CD}, ACM_{BD}, and ADM_{BC}. Since BM_{CD}, CM_{BD}, and DM_{BC} are the medians of face BCD, each of these median planes contains the centroid G_{BCD} of face BCD. They therefore intersect in the median AG_{BCD} of the tetrahedron.

The *circumcenter* of a tetrahedron is the center of the sphere defined by the four vertices of the tetrahedron. The circumcenter is equidistant from the four vertices and lies at the intersection of the perpendicular bisectors of the six edges of the tetrahedron.

For the tetrahedron in FIGURE 1 the centroid is (1, 1, 1) (by Theorem 2) and the circumcenter is (2, 2, 2). Since the Monge point is at the origin, it is seen to be symmetric to the circumcenter with respect to the centroid. A similar property holds for every tetrahedron:

MONGE'S THEOREM. *The six Monge planes of a tetrahedron are concurrent at the reflection of the circumcenter in the centroid.*

An analytic proof is straightforward, provided the origin is placed at the circumcenter of the tetrahedron. The coordinates of the reflection of the circumcenter in the centroid, M, are then double the coordinates of the centroid (which were given in the analytic proof of the Lemma). One can show that M satisfies the equation of any of the Monge planes by substituting the coordinates of M into the equation, simplifying, and using the fact that the vertices of the tetrahedron are equidistant from the origin. Eves gives a complete analytic proof along these lines [**7**, p. 149]. Forder gives a somewhat different analytic proof [**10**, p. 471].

The following synthetic proof appears in Altshiller-Court's book [**1**, p. 76] (an excellent source of information on the geometry of the tetrahedron).

Proof. Consider the perpendicular bisector of any edge of a tetrahedron and the Monge plane perpendicular to the same edge. These two planes are parallel. A bimedian crosses between them, for the perpendicular bisector contains the midpoint of one edge and the Monge plane contains the midpoint of the opposite edge. The centroid of the tetrahedron lies midway between the planes, for the centroid is the midpoint of the bimedian (Theorem 2). Hence, the reflection in the centroid of any point in one plane lies in the other plane. In particular, since the circumcenter lies in the perpendicular bisector, its reflection in the centroid is a point M lying in the Monge plane. By this reasoning, M lies on all six Monge planes and is the Monge point. ∎

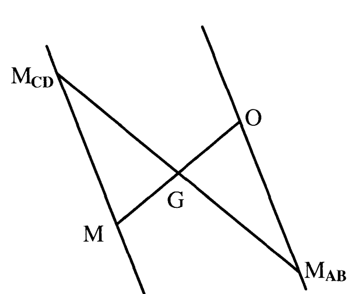


Figure 5 The plane of the circumcenter O, the centroid G, the Monge point M, and the midpoints M_{AB} and M_{CD} of edges AB and CD of tetrahedron $ABCD$. The perpendicular bisector of edge AB intersects the plane of the figure in $M_{AB}O$; the Monge plane perpendicular to AB intersects it at $M_{CD}M$.

FIGURE 5 depicts the plane of the circumcenter, the centroid, the Monge point, and one of the three bimedians.

Except for a brief introduction and a final short paragraph, Monge devotes "Sur la pyramide triangulaire" [**12**] entirely to a long two-part proof of this theorem. He begins the final paragraph with the assertion that the entire proof can be greatly simplified, but he gives us only key points of the simplified proof.

The first part of the long proof, which establishes the existence of the Monge point, we gave as our proof of Theorem 1. The second part establishes that the Monge point is the reflection of the circumcenter in the centroid, and it is not very elegant. We omit this part, but remark that it involves the construction of three new vertices related to a tetrahedron T. In the final paragraph Monge extends the construction from the previous proof by adding a fourth new vertex to form a conjugate or twin tetrahedron T'. He tells us that the twin tetrahedra have a common centroid, that their circumcenters are symmetric with respect to the centroid, and that the Monge point of T is the circumcenter of T'. A citation to a paper in which Monge defines twin tetrahedra is given.

Why would Monge have given a long proof when he had a much shorter and more elegant proof, and why would he have given the long proof in detail and simply outlined the short one? Perhaps he wanted to stimulate the reader's interest by setting up the simplification, to entice him into trying to fill in the details. Here we may see a hint of Monge's gift for teaching.

Using the information Monge gives us [**12**, p. 266] and a complete proof using twin tetrahedra [**1**, p. 76], we speculate on how Monge might have filled in the details. In FIGURE 3 construct the twin T' of tetrahedron $T = ABCD$ by connecting the four vertices of the parallelepiped that do not belong to T. Label the vertices of T' so that the diagonals of the parallelepiped are AA', BB', CC', and DD'. (See FIGURE 6.) Now, consider the Monge plane perpendicular to AB and containing M_{CD}. This Monge plane is perpendicular to $A'B'$ because AB and $A'B'$ are parallel. It contains $M_{A'B'}$ because M_{CD} and $M_{A'B'}$ coincide, both being midpoints of diagonals of the same parallelogram. Hence, the Monge plane perpendicular to edge AB of T is the perpendicular bisector of edge $A'B'$ of T'. Applying the same reasoning to the other Monge planes of T, we establish the key idea of the proof: The six Monge planes of T are the perpendicular bisectors of T'. Since the perpendicular bisectors of T' intersect at the circumcenter O' of T', we have $M = O'$, and the existence of the Monge point is established. Now T and T' are clearly symmetric with respect to the center of the parallelepiped. Therefore, by FIGURE 3 and Theorem 2, T and T' are symmetric with respect to the centroid G of T, whence the circumcenters O and O' are reflections in G. Substituting M for O' establishes the symmetry of M and O with respect to G and completes the proof.

In the proof just given we guessed that Monge would have used the circumscribing parallelepiped. This seems plausible because Monge defined twin tetrahedra in terms of the circumscribing parallelepiped in another paper [**15**, p. 242], and a figure such as FIGURE 6 may help one see the theorem. However, use of the circumscribing parallelepiped can be avoided in the above paragraph with the following changes. (1) Define T' as the reflection of T in G. (2) Replace the sentence, "It [the Monge plane perpendicular to AB and containing M_{CD}] contains $M_{A'B'}$ because M_{CD} and $M_{A'B'}$ coincide, both being midpoints of diagonals of the same parallelogram" with, "It contains $M_{A'B'}$ because M_{CD} and $M_{A'B'}$ coincide, both being the reflection of M_{AB} in G, $M_{A'B'}$ because T and T' are reflections in G and M_{CD} because of Theorem 2."

Notice that in each of the three synthetic proofs of Monge's theorem we have discussed (single tetrahedron and twin tetrahedra with and without the circumscribing parallelepiped) Theorem 2 plays an important role.

For a different approach to proving Monge's theorem see Thompson's proof [**16**].

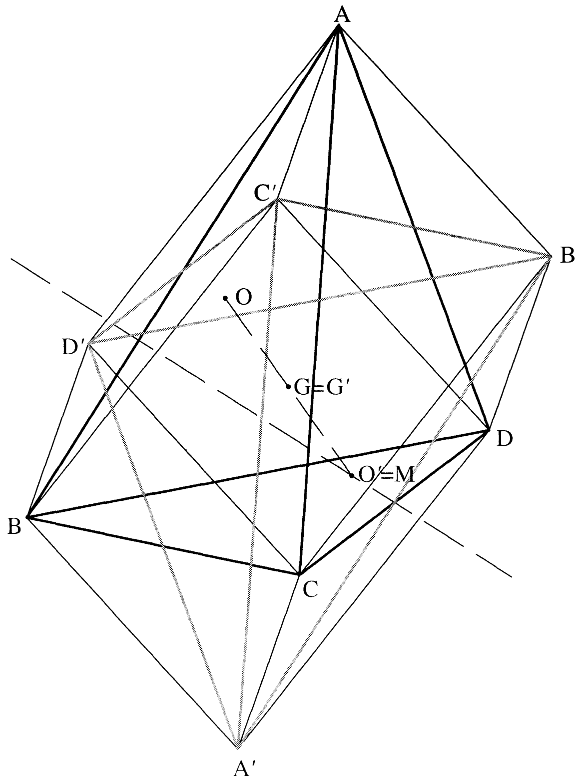


Figure 6 Twin tetrahedra *ABCD* and *A′B′C′D′* (thick black and gray solid lines, respectively) and their circumscribed parallelepiped (thin solid lines). Face *ABD* is in the plane of the paper. The Monge plane perpendicular to edge *AB* and containing the midpoint of edge *CD* is seen edge on (dashed line). This plane is also the perpendicular bisector of edge *A′B′*.

The rectangular tetrahedron: a special case

Monge's theorem invites comparison between the triangle and the tetrahedron; Monge stresses this analogy [**12**]. In a plane triangle, the circumcenter, the centroid, and the orthocenter lie on a line called the *Euler line*. The centroid lies between the other two points, twice as far from the orthocenter as from the circumcenter. (Two elegant proofs of this theorem are given in books by Dörrie [**6**, p. 141] and Eves [**9**, p. 109].) By analogy, the line containing the circumcenter, the centroid, and the Monge point of a tetrahedron is called the Euler line of the tetrahedron.

This analogy raises natural questions: Must a tetrahedron have an orthocenter, that is, a point at which its four altitudes meet? If an orthocenter exits, how is it related to the Monge point? The following definition and theorem provide answers.

DEFINITION. *A* rectangular, *or* orthocentric, tetrahedron *is one in which opposite edges have perpendicular directions.*

THEOREM 3. *A tetrahedron has an orthocenter if and only if it is rectangular. In a rectangular tetrahedron, the orthocenter is the Monge point.*

Proof. Suppose that tetrahedron *ABCD* has an orthocenter H_T. Since plane ABH_T contains the altitudes to faces *BCD* and *ACD*, it is perpendicular to both of these faces and to their intersection at edge CD. Hence, edges AB and CD have perpendicular directions. By the same reasoning, edges AC and BD and AD and BC have perpendicular directions and tetrahedron $ABCD$ is rectangular.

Assume that the tetrahedron is rectangular. Since opposite edges have perpendicular directions, each Monge plane must contain not only a midpoint of an edge but an entire edge of the tetrahedron. Hence, if three Monge planes are perpendicular to the edges of any one face of the tetrahedron, they contain the opposite vertex, and the Monge normal and the altitude associated with the face coincide. It now follows that the four altitudes intersect at the Monge point. This completes the proof. ∎

Altshiller-Court [**1**, p. 71] gives a proof of the concurrency of the four altitudes that does not assume knowledge of the Monge point. FIGURE 1 provides an example of a rectangular tetrahedron. The four altitudes do indeed coincide with the Monge normals and meet in the Monge point O.

Mannheim's theorem

A second set of planes, named for A. Mannheim (1831–1906), also intersect at the Monge point.

DEFINITION. *A* Mannheim plane *contains the altitude to and the orthocenter of a face of a tetrahedron.*

MANNHEIM'S THEOREM. *The four Mannheim planes of a tetrahedron intersect at the Monge point.*

Proof. In FIGURE 2, let the orthocenters of face BCD and triangle $B'C'D'$ be H and H', respectively. Note that triangle $B'C'D'$ is a dilation of face BCD with center A and ratio $1/2$; thus, H' is the midpoint of AH. Consider the Mannheim plane that contains H and the altitude from vertex A, that is, the Mannheim plane that is perpendicular to face BCD. The line AH lies in this plane, whence H' lies in it, whence the Monge normal of face BCD lies in it (by our proof of Theorem 1), whence the Monge point lies in it. By the same reasoning, the Monge point lies in the other three Mannheim planes, and the proof is complete. ∎

Two additional proofs of this theorem appear in Altshiller-Court's book [**1**, p. 78].

Acknowledgments. I thank Professor Howard Eves for his invaluable help in bringing this paper to fruition. He first inspired my interest in the Monge point and he gave me both encouragement and practical help.

I also thank Mr. Gary Krupa, who, using French sources, answered my questions with patience, concern for detail, and genuine interest. The biographical information given here is more complete and accurate because of his efforts.

REFERENCES

1. Nathan Altshiller-Court, *Modern Pure Solid Geometry*, 2nd ed., Chelsea Publishing Co., New York, 1964.
2. Eric Temple Bell, *Men of Mathematics*, Simon & Schuster, New York, 1986.
3. Carl B. Boyer, *History of Analytic Geometry*, The Scripta Mathematica Studies Nos. 6 and 7, Scripta Mathematica, Yeshiva University, New York, 1956.
4. ———, *A History of Mathematics*, 2nd ed., revised by Uta C. Merzbach, John Wiley & Sons, New York, 1991.
5. Julian Lowell Coolidge, *A History of Geometrical Methods*, Oxford University Press, Oxford, 1947.
6. Heinrich Dörrie, *100 Great Problems of Elementary Mathematics: Their History and Solution*, Dover Publications Inc., New York, 1965, translated by David Antin.

7. Howard Eves, *The Analytical Method in Geometry*, unpublished text used at the University of Central Florida in 1989.
8. ———, *An Introduction to the History of Mathematics*, 6th ed., Saunders College Publishing, Philadelphia, 1990.
9. ———, *A Survey of Geometry*, revised edition, Allyn and Bacon, Boston, 1972.
10. H. G. Forder, A Theorem in Coolidge's "Circle and Sphere," Math. Gaz. **15** (1930–1931), 470–471.
11. Gaspard Monge, Sur la pyramide triangulaire, in *Correspondance sur l'École Impériale Polytechnique*, edited by J. Hachette, 1813–1816, Vol. 2, pp. 1–6.
12. ———, Sur la pyramide triangulaire, in *Correspondance sur l'École Impériale Polytechnique*, edited by J. Hachette, 1813–1816, Vol 2, pp. 263–266.
13. J. J. O'Connor and E. F. Robertson, Gaspard Monge, The MacTutor History of Mathematics Archive, School of Mathematics and Statistics, University of St. Andrews, Scotland, `http://www-groups.dcs.st-andrews.ac.uk/~history/Mathematicians/Monge.html`.
14. George Polya, *How to Solve It*, 2nd ed., Princeton University Press, Princeton, NJ, 1973.
15. René Taton, *L'oeuvre scientifique de Monge*, Presses Universitaires de France, Paris, 1951.
16. H. F. Thompson, A Geometrical Proof of a Theorem connected with the Tetrahedron, Proc. Edinb. Math. Soc. **17** (1908–1909), 51–53.

Math Bite: A Novel Proof of the Infinitude of Primes, Revisited

DANIEL CASS
GERALD WILDENBERG
St. John Fisher College
Rochester NY 14618
cass@sjfc.edu
wildenbe@sjfc.edu

We rephrase a proof of the infinitude of primes by Fürstenberg [**1**]. The original used arithmetic progressions as the basis for a topology on the integers. Our approach avoids the language of topology.

Recall: For A a subset of the integers, the *characteristic function* of A has value 1 for x in A and 0 otherwise. A *periodic* set of integers is one whose characteristic function is periodic.

Observe that if S and T are periodic sets with periods S and T respectively, then $S \cup T$ is periodic with a period dividing $\operatorname{lcm}(s,t)$ and that this is easily extended to all finite unions. And observe that if S is periodic, then the complement of S is periodic.

THEOREM. *There are infinitely many prime integers.*

Proof. For each prime p, let $S_p = \{n \cdot p : n \in \mathbb{Z}\}$. Define $\mathcal{S}$ to be the union of all the sets S_p, each of which is periodic. If this union is taken over a finite set, then $\mathcal{S}$ is periodic and then so is its complement. But the complement of $\mathcal{S}$ is {-1,1}, which being finite is not periodic. Hence the number of primes is infinite.

As a pedagogic coda, we note that filling in the details of our proof would be a nice exercise in a course serving as a bridge to upper division mathematics.

REFERENCES

1. H. Fürstenberg, On the infinitude of primes, *Amer. Math. Monthly* **62** (1955), 353.

NOTES

Periodic Points of the Open-Tent Function

DANRUN HUANG
DANIEL SCULLY
Saint Cloud State University
St. Cloud, MN 56301-4498
dhuang@stcloudstate.edu
djscully@stcloudstate.edu

Given a function $f : S \to S$, it is of great interest in the field of dynamical systems to figure out which points in the set S are eventually sent back to themselves through repeated applications of f. More precisely, people like to know which points x and which positive integers n have the property that $f^n(x) = x$, where f^n denotes the nth iteration of f. Such a point x is called a *periodic point*, and the smallest such n is called the *prime period* of x. (Note that this does not require that n be a prime number.)

According to a famous theorem by Li and Yorke [**8**], for a continuous function f on a line or a closed interval S, if f has a point of prime period 3, then f has a point of prime period n for every n. This amazing result turns out to be a special case of the even more amazing Sarkovskii theorem [**4**, Ch. 11].

To construct a simple example of a continuous function with a point of prime period 3 on the unit interval, we choose $0 \to 1/2 \to 1 \to 0$ as our 3-cycle and connect the points (0, 1/2), (1/2, 1), and (1, 0) by a piecewise-linear function

$$f(x) = \begin{cases} x + 1/2 & \text{if } 0 \le x < 1/2, \\ 2 - 2x & \text{if } 1/2 \le x \le 1. \end{cases}$$

We call f the *open-tent function*. Its graph is given in FIGURE 1.

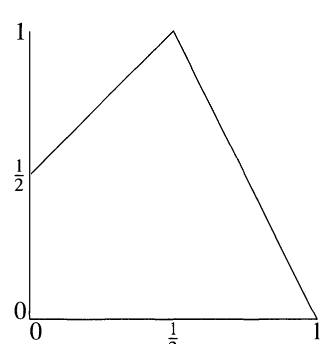


Figure 1 The open-tent function

The open-tent function f is well known and is used as a simple example to illustrate that prime period 3 implies chaos [**1**, p. 248; **3**, p. 135; **6**]. However, knowing the existence of points with various periods and actually finding them are two different matters. In a nice note, David Sprows [**10**] uses binary expansions to construct for the open-tent function f a point of prime period n for each positive integer n. However, with this method, information about the orbits of f is far from clear. The method we present in this Note gives a far more explicit picture.

The *orbit* of x is the set $\{x, f(x), f^2(x), f^3(x), \ldots\}$. For instance, the orbit of $2/3$ consists of a single point, while the orbit of $1/3$ includes a single additional point, $5/6$. Clearly, the orbit is finite if x is periodic.

Our key for giving a precise description of the orbits is through a method called *encoding* and *decoding*. One might view a binary expansion for any $x \in [0, 1]$ as an identification number, or "ID," for x. Our strategy is to provide each $x \in [0, 1]$ with a different infinite sequence of zeros and ones as its new ID; this is called *encoding*. With this new ID (encoding), we can know the orbital information of the open-tent function f much better. For example, we can tell *how many* points of period n there are and how we can locate *all* of them. We can also locate many other points with interesting orbital features, such as a point that stays obediently in the interval $[1/2, 1]$ for every single iteration of f, escaping exactly once on the one-millionth time.

The open-tent function is an example of a *dynamical system* (S, f), a set S together with a function f from S back to itself. The open-tent example, where the set is $[0, 1]$ would be written $([0, 1], f)$. The key idea of this note is as follows: First, we use a "digital (or symbolic) model" to encode the open-tent function system $([0, 1], f)$, namely, the well-known symbolic dynamical system (G, σ), called the *golden-mean shift*. Then we investigate the encoded orbital information of $([0, 1], f)$ in (G, σ) which is much easier to handle digitally. Finally, we decode the information obtained from (G, σ) back to the system $([0, 1], f)$ in the same way that a CD player decodes its digital codes back into music.

This approach connects many interesting topics in undergraduate mathematics, such as the golden mean, Fibonacci and Lucas numbers, directed graphs, matrices, binary expansions, and coding. Our technique is standard in the field of dynamical systems [**1, 4, 5, 7, 9**], but we provide a rigorous and complete coding algorithm for the open-tent function, including the coding for the numbers of the form $j/2^n$ (the boundary points that arise upon repeatedly bisecting the unit interval), which has previously been unavailable to students.

Unlike the *tent function* (obtained by replacing $x + 1/2$ by $2x$ for $0 \le x < 1/2$ in the definition of f), which is mentioned in almost every dynamical system text and utilizes all 0-1 sequences for its coding, the open-tent function gives us an elementary yet nontrivial example of coding in terms of a proper subset of the set of all 0-1 sequences, as well as a simple yet rich application of symbolic dynamics—a fast-growing branch of modern mathematics [**7, 9**].

The golden-mean shift A *symbolic dynamical system* (X, σ) of the kind considered in this note consists of a set X of infinite sequences of symbols and a shift function σ that knocks off the first term of each sequence. As an example, let $\{0, 1\}$ be the symbol set, and let $X = \{0, 1\}^\infty$ be the set of all infinite 0-1 sequences of the form $x = c_0c_1c_2\cdots$, where $c_i = 0$ or 1. Define the shift function $\sigma : X \to X$ by $\sigma(x) = c_1c_2c_3\cdots$. For the open-tent function, we let G denote the subset of $\{0, 1\}^\infty$ that consists of the sequences in which adjacent zeros are forbidden. The set G together with the shift function σ defined above is called the *golden-mean shift*.

A directed graph H associated with G is shown in FIGURE 2. The vertices of H are the two symbols 0 and 1. The directed edges on H give the rule indicating which symbol can follow another in the sequences of G. Since there is no edge on H from 0 to itself, adjacent zeros are forbidden in the sequences of G. It is easy to see that the elements of G represent all infinite walks on H that start at either of the two vertices

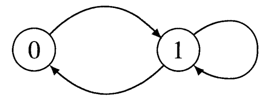


Figure 2 Directed graph H of golden-mean shift

and continue forever. The symbols in the sequence indicate the vertices visited during the walk in the order they are visited. The directed graph H can be recorded by the integer matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

as follows. Let the (i, j) entry, $A(i, j)$, of A be the number of edges from vertex i to vertex j. The matrix A is called the *adjacency matrix* of H. Note that the eigenvalues of A are the golden means

$$\frac{1 \pm \sqrt{5}}{2}.$$

Inductively, we have

$$A^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, A^1 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, A^2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \ldots,$$
$$A^n = \begin{pmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{pmatrix},$$

where $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$, are the *Fibonacci numbers*, treated extensively elsewhere in this issue. It is well known that the nth power of an adjacency matrix counts the walks of length n on its graph. In fact, $A^n(i, j)$ counts the walks on H of length n from vertex i to vertex j, and the trace of A^n, $\mathrm{tr}(A^n) = A^n(1, 1) + A^n(2, 2)$, equals the number of closed walks of length n on H. The sequence $\{\mathrm{tr}(A^n)\}_{n=0}^{\infty}$ also satisfies the Fibonacci recurrence relation since $\mathrm{tr}(A^n) = F_{n-1} + F_{n+1}$ with $\mathrm{tr}(A^0) = 2 = L_0$ and $\mathrm{tr}(A^1) = 1 = L_1$. Therefore, $\{\mathrm{tr}(A^n)\}_{n=0}^{\infty}$ is the sequence of famous *Lucas numbers* L_n.

We use the notation $(c_0c_1 \cdots c_{n-1})^{\infty}$ to indicate the sequence in G or $\{0, 1\}^{\infty}$ formed by concatenating infinitely many copies of $c_0c_1 \cdots c_{n-1}$. Hence,

$$\sigma((c_0c_1 \cdots c_{n-1})^{\infty}) = (c_1c_2 \cdots c_{n-1}c_0)^{\infty} \quad \text{and}$$
$$\sigma^n((c_0c_1 \cdots c_{n-1})^{\infty}) = (c_0c_1 \cdots c_{n-1})^{\infty},$$

so $(c_0c_1 \cdots c_{n-1})^{\infty}$ has period n under σ. In G, then, we have only one fixed point $1^{\infty} = 11\cdots$ since $\sigma(1^{\infty}) = 1^{\infty}$ and 0^{∞} is not in G. We have two points $(01)^{\infty}$ and $(10)^{\infty}$ with prime period 2, since $\sigma((01)^{\infty}) = (10)^{\infty}$ and $\sigma((10)^{\infty}) = (01)^{\infty}$. The element $1^{\infty} = (11)^{\infty}$ also has period 2 though its prime period is 1.

The one-to-one correspondence between the elements of G and the infinite walks on H implies

$$\begin{pmatrix} \text{the number} \\ \text{of period-}n \\ \text{points in } G \end{pmatrix} = \begin{pmatrix} \text{the number of} \\ n\text{-step closed} \\ \text{walks in } H \end{pmatrix} = \mathrm{tr}(A^n) = L_n. \tag{1}$$

Encoding and decoding The link between a general dynamical system and its symbolic dynamical system is realized by the encoding and decoding processes. Through them we show that there is a one-to-one correspondence between the period n points of the open-tent function in $[0, 1]$ and those of the golden-mean shift in G. We begin our encoding by making a partition of the unit interval, $I_0 = [0, 1/2)$ and $I_1 = [1/2, 1]$. We then encode every point $x \in [0, 1]$ as an infinite sequence as follows:

$$E(x) = c_0c_1c_2\cdots, \text{ where } c_k = 0 \text{ if } f^k(x) \in I_0 \text{ and } c_k = 1 \text{ if } f^k(x) \in I_1.$$

The sequence $E(x) = c_0c_1c_2\cdots$ is called the *encoding* of x (or the *itinerary* of x). It is the new ID of x. For the open-tent function f, we see that $f(I_0) \subseteq I_1$, so every 0 in the encoding of a number must be followed by a 1, that is, adjacent zeros are forbidden, and $E([0, 1]) \subseteq G$. As an example, since $f^0(0) = 0 \in I_0$, $f(0) = 1/2 \in I_1$, and $f^2(0) = 1 \in I_1$, the first three digits in the encoding of 0 are 011. Since f sends 1 back to 0, the sequence repeats. So, $E(0) = (011)^\infty$. Similarly, $E(1/2) = (110)^\infty$ and $E(1) = (101)^\infty$.

Though a brute force encoding is always possible, decoding is not as straightforward. To make a more general analysis possible, we move back and forth between x and $E(x)$ through the binary expansion of x. Things are complicated a bit by the fact that some rational numbers have two distinct binary expansions. For example, in binary $1/2 = 0.1\bar{0} = 0.0\bar{1}$ just as in decimal $1/2 = 0.5\bar{0} = 0.04\bar{9}$. We must proceed carefully.

Rational numbers of the form $j/2^n$ are called *dyadic numbers*. The dyadic numbers in $(0, 1]$ are exactly the rational numbers in the unit interval that have two distinct binary expansions. If $x \in (0, 1]$ is dyadic, then there exist nonnegative integers j and n such that in lowest terms

$$x = \frac{j}{2^n} = 0.x_1x_2\ldots x_{n-1}1\bar{0} = 0.x_1x_2\ldots x_{n-1}0\bar{1}.$$

Before presenting technical coding formulas, let us see some heuristic descriptions. Because the second piece of f has a slope of -2, an interval in I_1 is stretched by f to double its length, and its orientation is reversed (if $x < y$, then $f(x) > f(y)$), while the first piece of f simply slides an interval in I_0 to the right $1/2$ unit into I_1. The nth iteration of f is a piecewise function that is linear on dyadic intervals of the form $(p/2^n, (p+1)/2^n)$. For the dyadic numbers, we must make the proper choice of binary expansion. A function ψ is defined in (2) to serve this purpose.

Let us consider a generic case where $x \in [0, 1]$ is not dyadic. Suppose $x = 0.x_1x_2x_3\ldots$ is its binary expansion and $E(x) = c_0c_1c_2\cdots$ is its encoding. If $x_1 = 0$, then $x \in I_0$ and $f(x) \in I_1$, so we can determine that c_0c_1 must be 01. Similarly, if $x_1 = 1$, then we can determine that $c_0 = 1$. This is our first step of encoding through the binary expansion. Note that in the first case ($x_1 = 0$), f is applied once to determine the first two symbols in the encoding. In the second case ($x_1 = 1$), f is not applied at all and only the first symbol of the encoding was determined. In both cases, the encoding step ends with a symbol 1. That is, when the orbit enters I_1. To summarize:

$$x_1 = 0 \Rightarrow c_0c_1 = 01,$$
$$x_1 = 1 \Rightarrow c_0 = 1.$$

Having used the first digit of the binary expansion, we ignore it and focus on the second for step 2, because this digit determines the next entry in the encoding, whether it is c_1 or c_2. Suppose $x = 0.*x_2\ldots$. If $x_2 = 0$, then $x \in (0, 1/4)$ or $(1/2, 3/4)$. If the former, then we know that f moved $(0, 1/4)$ onto $(1/2, 3/4)$ without an orientation reversal in step 1. The next iteration of f sends $(1/2, 3/4)$ to I_1, so the next symbol in the coding is 1 and there is a total of one orientation reversal. Similarly, if $x_2 = 1$, then the next two symbols in the coding are 01 with one orientation reversal, and again, the step ends with the orbit entering I_1. Thus,

$$\begin{aligned} x_2 = 0 &\Rightarrow \text{ the next symbol in the encoding is } 1, \\ x_2 = 1 &\Rightarrow \text{ the next two symbols in the encoding are } 01. \end{aligned}$$

The third step deals with x_3. It produces another orientation reversal, so the orientation is the same as in step 1. Thus,

$$\begin{aligned} x_3 = 0 &\Rightarrow \text{ the next two symbols in the encoding are } 01, \\ x_3 = 1 &\Rightarrow \text{ the next symbol in the encoding is } 1. \end{aligned}$$

The process continues as above through the binary expansion of x. The encoding rules alternate, using x_2 as the template for x_n if n is even, and x_3 if n is odd. The argument, including the subtle handling of the dyadic numbers, is in the proof of Theorem 1. A casual reader could skip the proof.

We now develop technical algorithms for encoding and decoding. The expansions presented are binary. Define $\psi : [0, 1] \to \{0, 1\}^\infty$ by

$$\psi(x) = \begin{cases} x_1 x_2 \cdots & \text{if } x = 0.x_1 x_2 \ldots \text{ and is not dyadic, or } 0, \\ x_1 x_2 \cdots x_{n-1} 10^\infty & \text{if } x = 0.x_1 x_2 \ldots x_{n-1} 1\bar{0} \text{ and } n \text{ is odd,} \\ x_1 x_2 \cdots x_{n-1} 01^\infty & \text{if } x = 0.x_1 x_2 \ldots x_{n-1} 0\bar{1} \text{ and } n \text{ is even.} \end{cases} \tag{2}$$

We call $\psi(x)$ the *proper binary expansion* for x. Obviously, ψ is one-to-one and has a left inverse $\phi : \{0, 1\}^\infty \to [0, 1]$ defined by

$$\phi(z_1 z_2 z_3 \cdots) = 0.z_1 z_2 z_3 \ldots = \sum_{k=1}^{\infty} \frac{z_k}{2^k}$$

with $\phi \circ \psi = \mathrm{Id}_{[0,1]}$. It is also clear that ϕ is onto and almost one-to-one—except that it maps two binary expansion sequences to each nonzero dyadic number.

The next function, B, is a bijection between $\{0, 1\}^\infty$ and G. This function and its inverse are at the heart of the encoding and decoding processes, since they provide the correspondence between the proper binary expansion of a point and its encoding. Define $B : \{0, 1\}^\infty \to G$ by $B(z_1 z_2 z_3 \cdots) = y_1 y_2 y_3 \cdots$, where

$$y_n = \begin{cases} 01 & \text{if } n \text{ is odd and } z_n = 0, \\ 1 & \text{if } n \text{ is odd and } z_n = 1, \\ 1 & \text{if } n \text{ is even and } z_n = 0, \\ 01 & \text{if } n \text{ is even and } z_n = 1. \end{cases} \tag{3}$$

For example,

$$\begin{aligned} B(0^\infty) &= B(0000\cdots) = 011011\cdots = (011)^\infty \quad \text{and} \\ B(01^\infty) &= B(0111\cdots) = 0(101)^\infty. \end{aligned}$$

It is easy to see that B is a bijection with the inverse given by $B^{-1}(y_1 y_2 y_3 \cdots) = z_1 z_2 z_3 \cdots$ where $y_1 y_2 y_3 \cdots \in G$, y_n equals 01 or 1, and

$$z_n = \begin{cases} 0 & \text{if } n \text{ is odd and } y_n = 01, \\ 1 & \text{if } n \text{ is odd and } y_n = 1, \\ 1 & \text{if } n \text{ is even and } y_n = 01, \\ 0 & \text{if } n \text{ is even and } y_n = 1. \end{cases} \tag{4}$$

Define the *decoder* $D : G \to [0, 1]$ of E by $D = \phi \circ B^{-1}$. We have the following:

THEOREM 1. *$E = B \circ \psi$, $D \circ E = Id_{[0,1]}$, and $E \circ D|_{E([0,1])} = Id_{E([0,1])}$. In particular, the encoder E is one-to-one and the decoder D is onto.*

Proof. We first show that $E = B \circ \psi$. Suppose $x \in [0, 1]$ is not dyadic and $\psi(x) = x_1x_2x_3\cdots$. In addition, let $E(x) = c_0c_1c_2\cdots = y_1y_2y_3\cdots$, where c_k is 0 or 1 depending on whether $f^k(x)$ is in I_0 or I_1 and y_n is 01 or 1. Suppose further that $B \circ \psi(x) = y_1'y_2'y_3'\cdots$, where y_n' is 01 or 1. We show that $y_n = y_n'$ for all n by induction. If $x_1 = 0$, then $x = f^0(x) \in I_0$ and $f^1(x) \in I_1$, so $c_0 = 0$, $c_1 = 1$, and $y_1 = 01$. On the other hand, by the definition of ψ and (3), $x_1 = 0$ implies $y_1' = 01$. Similarly, if $x_1 = 1$, then $x \in I_1$, so $y_1 = 1$ and $y_1' = 1$. In either case, $y_1 = y_1'$.

Now assume $y_i = y_i'$ for $i = 1, \ldots, n-1$. Since x is not dyadic, there exists an integer j such that $0 \le j \le 2^{n-1} - 1$ and $x \in (j/2^{n-1}, (j+1)/2^{n-1})$, which lies entirely in I_0 or I_1. Each application of f sends such a dyadic interval to another dyadic interval. If the left branch of f is applied, its width remains the same, but if the right branch is applied, the width doubles and the endpoints of the image can both be written with the denominator 2^{n-2}. Such intervals still fall entirely in I_0 or I_1 until they are stretched to a width of 1. Therefore, upon the $(n-1)$st visit of x to I_1, this interval has been stretched to $(1/2, 1)$. If x is in the left half of $(j/2^{n-1}, (j+1)/2^{n-1})$, then $x_n = 0$. If n is even, then upon the $(n-1)$st visit to I_1 the left half of $(j/2^{n-1}, (j+1)/2^{n-1})$ has been stretched to $(1/2, 3/4)$. The next application of f sends the iteration of x already in $(1/2, 3/4)$ to I_1, so $y_n = 1$. But, n even and $x_n = 0$ implies $y_n' = 1$ by (3). Likewise, n odd implies $y_n = 01$ and $y_n' = 01$. In either case $y_n = y_n'$. The parallel argument shows that if x is in the right half of $(j/2^{n-1}, (j+1)/2^{n-1})$, then $x_n = 1$ and $y_n = y_n'$. Therefore, we proved that $y_n = y_n'$. for all n if x is not dyadic.

Suppose $x \in [0, 1]$ is dyadic. It is easy to check that $E(x) = B \circ \psi(x)$ for $x = 0, 1/2$, or 1. If x is dyadic and different from those three, then $x = j/2^n$ in lowest terms with $n \ge 2$, and $x = j/2^n$ is the midpoint of an interval of the form $(p/2^{n-1}, (p+1)/2^{n-1})$ that falls entirely in I_0 or I_1. Let $E(x) = c_0c_1c_2\cdots = y_1y_2y_3\cdots$ as before, and suppose that $B \circ \psi(x) = y_1'y_2'y_3'\cdots$. Since the midpoint of $(p/2^{n-1}, (p+1)/2^{n-1})$ is an element of $(p/2^{n-1}, (p+1)/2^{n-1})$, an argument that parallels the nondyadic case shows that $y_k = y_k'$, but only for $k = 1, 2, \ldots n-1$. Upon the $(n-1)$st visit to I_1, the interval $(p/2^{n-1}, (p+1)/2^{n-1})$ is stretched onto $(1/2, 1)$ and $x = j/2^n$ is mapped to $3/4$. Let q equal the number of applications of f required to produce $n-1$ visits by $(p/2^{n-1}, (p+1)/2^{n-1})$ to I_1, then $c_0c_1c_2\cdots c_q = y_1y_2y_3\cdots y_{n-1}$ with $c_q = 1$. Another application of f sends x to $1/2$, so $c_{q+1} = 1$ and $y_n = 1$. Since $E(1/2) = (110)^\infty$, $E(x) = c_0c_1c_2\cdots c_{q-1}1(110)^\infty = y_1y_2y_3\cdots y_{n-1}(110)^\infty = y_1y_2y_3\cdots y_{n-1}1(101)^\infty$. Thus, if n is odd, $B \circ \psi(x) = B(x_1x_2x_3\cdots x_{n-1}10^\infty) = y_1'y_2'y_3'\cdots y_{n-1}'1(101)^\infty = E(x)$. Similarly, if n is even $B \circ \psi(x) = B(x_1x_2x_3\cdots x_{n-1}01^\infty) = y_1'y_2'y_3'\cdots y_{n-1}'1(101)^\infty = E(x)$.

By construction of the following maps

$$[0,1] \underset{\phi}{\overset{\psi}{\rightleftarrows}} \{0,1\}^\infty \underset{B^{-1}}{\overset{B}{\rightleftarrows}} G,$$

we get $D \circ E = (\phi \circ B^{-1}) \circ (B \circ \psi) = \phi \circ \psi = \mathrm{Id}_{[0,1]}$. In particular, E is one-to-one and D is onto. Moreover, for any $y \in E([0, 1])$, there is an $x \in [0, 1]$ such that $E(x) = y$, so $E \circ D(y) = E(D(E(x))) = E(x) = y$, thus $E \circ D|_{E([0,1])} = \mathrm{Id}_{E([0,1])}$. This ends the proof. ∎

Theorem 1 states that the encoder and the decoder are not inverses of each other, but almost. Hence, we cannot identify the open-tent function with its symbolic model the golden-mean shift. However, many important dynamical features like periodic points are still in one-to-one correspondence between the two systems.

Using (4) and $D(y) = \sum_{i=1}^{\infty} z_i/2^i$, we can decode any 0-1 sequence in G into a number in [0,1]. The first example is straightforward, but there are a few subtleties of decoding as demonstrated in Examples 2 & 3.

EXAMPLE 1. *To decode the element* $(11011)^\infty$, *use (4) directly to obtain* $D[(11011)^\infty] = 0.\overline{1000} = 2^3/(2^4-1) = 8/15$. *So,* $8/15$ *is a point of period* 5.

EXAMPLE 2. *A careless decoding may suggest that* $D[(1101)^\infty] = 0.\overline{100} = 4/7$, *but the odd number of ones in the string* 1101 *tells us that the even-odd parity is switched in the second appearance of* 1101 *in the infinite string* $(1101)^\infty$. *In this case we must list the repeating string twice to get an even number of* 1s. *Thus,* $D[(11011101)^\infty] = 0.\overline{100011} = 5/9$, *and* $5/9$ *has period* 4.

EXAMPLE 3. *What do we do with that final* 0 *when decoding* $(11110)^\infty$? *Simply note that* $(11110)^\infty = 1(11101)^\infty$, *so* $D[(11110)^\infty] = D[1(11101)^\infty] = 0.1\overline{0100} = 19/30$.

EXAMPLE 4. *What is the point* $x \in [0,1]$ *such that* $f^n(x) \geq 1/2$ *for all n except when n equals one million? We get the answer by decoding an element of G with the right properties:*

$$x = D(1^{1,000,000}01^\infty) = \phi \circ B^{-1}[(11)^{500,000}(01)(11)^\infty] = \phi[(10)^{500,000}0(01)^\infty]$$

$$= \left(\sum_{k=0}^{\infty} \frac{1}{2^{2k+1}}\right) - \frac{1}{2^{10^6+1}} = \frac{2}{3} - \frac{1}{2^{10^6+1}} = \frac{2^{10^6+2}-1}{3 \cdot 2^{10^6+1}}.$$

The encoder $E : [0,1] \to G$ is not onto. Let us find exactly which elements of G fall outside $E([0,1])$. Suppose $w \in G$, but $w \notin E([0,1])$. Let $x = D(w) = \phi \circ B^{-1}(w)$. Both $B^{-1}(w)$ and $\psi(x)$ are binary expansion sequences of x by the definitions of ϕ and ψ respectively. But $B^{-1}(w) \neq \psi(x)$, for otherwise $w = B(B^{-1}(w)) = B(\psi(x)) = E(x) \in E([0,1])$. A contradiction. Since ϕ is almost a bijection except for the dual representation of the dyadic numbers,

> *the elements of G that fall outside* $E([0,1])$ *are precisely those elements of G that correspond through B to the improper binary expansions of the dyadic numbers.*

It is easy to spot these elements. If $y \in G$ begins with a 1, then $\sigma^{-1}(y) = \{0y, 1y\}$, while if y begins with a 0, then $\sigma^{-1}(y) = \{1y\}$. In FIGURE 3 we have an infinite directed graph for $(101)^\infty$ and its preimages. It shows the complete genealogy of the ambiguous sequences in G corresponding to the dyadic numbers. An arrow from w to y indicates $\sigma(w) = y$.

Since the dyadic numbers are the preimages of 1 under f^n for various n and $E(1) = (101)^\infty$, the elements of $E([0,1])$ that are preimages of $(101)^\infty$ under σ for various n decode to the dyadic numbers. Notice, however, that $\sigma^{-1}[(101)^\infty] = \{(110)^\infty, 0(101)^\infty\}$ even though $f^{-1}(1) = \{1/2\}$. The element $0(101)^\infty \notin E([0,1])$ and the whole right-hand branch of the directed graph in FIGURE 3 that passes through $0(101)^\infty$ lies outside of $E([0,1])$.

> *The necessary and sufficient conditions for* $y \in G$ *being in* $E([0,1])$ *are that either y does not have the repeating block* $(101)^\infty$, *or, if it does, then it has a* 1 *just before its repeating block* $(101)^\infty$.

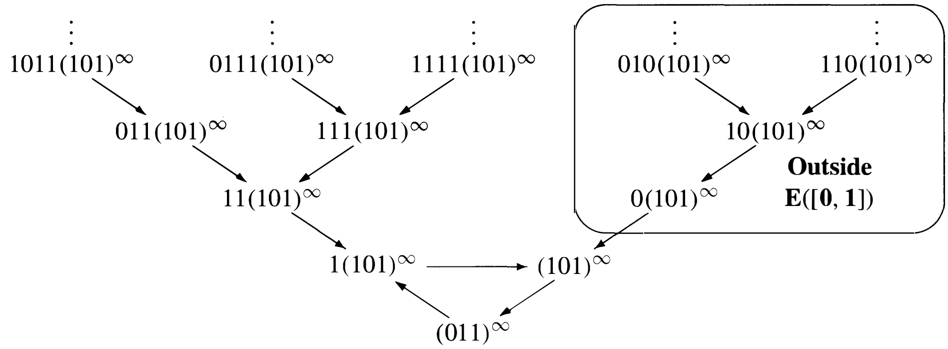


Figure 3 Preimages of $(101)^\infty$ under σ^n

None of the points in FIGURE 3 except the bottom three are periodic though all of their orbits eventually enter a periodic cycle. Such points are called *eventually periodic points*. Now we can describe all the periodic points and eventually periodic points of the open-tent function as our main result.

Periodic points and eventually periodic points

THEOREM 2. *Let f be the open-tent function on $[0, 1]$, and let (G, σ) be the golden-mean shift.*

1. *There is a period-preserving bijection between the set of all periodic points in $([0, 1], f)$ and those in (G, σ). In particular, we can locate all periodic points of f precisely.*
2. *The number of period-n points of f equals the Lucas number L_n.*
3. *The only 3-cycle of f consists of the three dyadic numbers $0 \xrightarrow{f} 1/2 \xrightarrow{f} 1 \xrightarrow{f} 0$. All other dyadic numbers in $[0, 1]$ are eventually period-three points with their orbits eventually ending with the 3-cycle above. The converse is also true.*
4. *A number $x \in [0, 1]$ is a periodic point of f if and only if either*
   - *(a) x is a rational number that can be written as a fraction with an odd denominator (this includes 0 and 1), or*
   - *(b) x is a rational number that can be written in the form $1/2 + j/(2k)$ for some nonnegative integer j and odd positive integer k (this includes $1/2$ and 1).*
5. *A number $x \in [0, 1]$ is a periodic or eventually periodic point of f if and only if x is rational.*

Proof. (1,2). By design, if $E(x) = c_0c_1c_2\cdots$, then $E(f(x)) = c_1c_2c_3\cdots$. Hence, $E \circ f = \sigma \circ E$. This along with the fact that E is one-to-one (Theorem 1) implies that $f^n(x) = x$ if and only if $\sigma^n(E(x)) = E(x)$. Thus, x has period n under f if and only if $E(x)$ has period n under σ. Since none of the elements of G that fall outside $E([0, 1])$ are periodic (FIGURE 3), E serves as a bijection between the period-n points of $([0, 1], f)$ and those of (G, σ). So, the number of period-n points of $[0, 1]$ under f equals the Lucas number L_n by (1). To prove *(3)*, observe that the only prime-period-three elements in G are $(011)^\infty$, $(110)^\infty$, and $(101)^\infty$. They decode to the only prime-period-three elements 0, $1/2$, and 1 respectively in $[0, 1]$. The proof of Theorem 1 implies the rest. We leave the proofs of *(4)* and *(5)* as exercises in the application of decoding. ∎

TABLE 1 lists the period-n points of (G, σ) and $([0, 1], f)$ for $n = 1, 2, 3, 4$.

TABLE 1: Periodic points of the open-tent function

Prime Period	$E(x)$	x
1	1^∞	2/3
2	$(01)^\infty$, $(10)^\infty$	1/3, 5/6
3	$(011)^\infty$, $(110)^\infty$, $(101)^\infty$	0, 1/2, 1
4	$(0111)^\infty$, $(1110)^\infty$, $(1101)^\infty$, $(1011)^\infty$,	2/9, 13/18, 5/9, 8/9

Let q_n denote the number of points in [0, 1] having prime period n under f. If $k < n$ and k divides n, then the q_k elements of [0, 1] with prime period k are counted in L_n along with the q_n elements with prime period n. Thus,

$$q_n = L_n - \sum_{k|n, k<n} q_k.$$

With the help of a computer, we calculate some values of q_n in TABLE 2.

TABLE 2: Number of period-n points

n	No. of Period n Pts. L_n	No. of Prime Period n Pts. q_n
1	1	1
2	3	2
4	7	4
5	11	10
10	123	110
20	15,127	15,000
25	167,761	167,750
50	28,143,753,123	28,143,585,250
100	792,070,839,848,373,253,127	792,070,839,820,228,485,000

We should be aware of the limitations of a computer for such a seemingly simple process as calculating the iterations of f at some point x. Try using a spreadsheet or mathematical software that uses floating-point arithmetic to investigate this; you will find that *all orbits* of f end with the 3-cycle $0 \to 1/2 \to 1 \to 0$. Why? The computer uses a finite binary expansion to represent the seed number. In doing so, it has rounded the seed to a dyadic number. By Theorem 2, the orbits of all dyadic points end in that 3-cycle. This phenomenon is quite unique to the open-tent function. It is no longer true if we just move the top of the tent a bit higher or lower! Interested readers may study the orbit diagram (by *Maple*) in FIGURE 4 of the following family of functions with the parameter c:

$$f_c(x) = \begin{cases} cx + 1/2 & x < 1/2 \\ (1+c)(1-x) & x \geq 1/2 \end{cases}, \quad \text{for } -1 \leq c \leq \frac{1+\sqrt{5}}{2}.$$

The orbit diagram plots the parameter c with a gap of 0.02 on the horizontal axis versus the eventual orbit of the critical point 1/2 under f_c on the vertical axis. A different family that contains the open-tent function is discussed by Bassein [**2**].

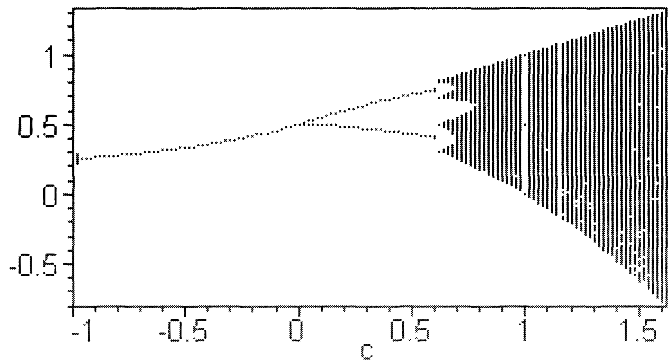


Figure 4 Orbit diagram of $x = 1/2$ under a modified tent function f_c

FIGURE 4 presents a familiar picture of transition to chaos through period-doubling bifurcations [**4**, Ch. 8]. For $-1 < c < 0$, the orbit of $x = 1/2$ tends to an attracting fixed point. At $c = 0$, the family has a period-doubling bifurcation where the attracting fixed point turns into a repelling fixed point and gives birth to an attracting 2-cycle. For $c > 0$, the orbit of $1/2$ tends to an attracting 2-cycle until $c \approx 0.617$ when another period-doubling bifurcation happens that gives birth to an attracting 4-cycle. The dark region of the diagram shows that the orbits of $x = 1/2$ under corresponding f_c are trapped in one or more vertical intervals, jumping back and forth chaotically. When $c = 1$, f_c is the open-tent function, and the orbit of $1/2$ is represented by the three dots that appear on the vertical line $c = 1$. The reason we can see these three dots is not because it is an attracting 3-cycle, but because all numbers are rounded by computer to dyadic numbers that eventually enter the 3-cycle $0 \to 1/2 \to 1 \to 0$ (see FIGURE 3). The open-tent function is unique in this family $\{f_c\}$. For c just off from 1, the orbit of $1/2$ under f_c is chaotic. When $c > (1 + \sqrt{5})/2$, the orbit of $1/2$ escapes, so we see the golden mean one more time to end the orbit diagram! We still do not know how to locate all the periodic points of f_c for all $c \neq 1$ as we do for the open-tent function.

Acknowledgment. The authors are grateful to the referees for valuable remarks and suggestions.

REFERENCES

1. K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos—An Introduction to Dynamical Systems*, Springer, New York, 1996.
2. Susan Bassein, The dynamics of a family of one-dimensional maps, *Amer. Math. Monthly* **105** (1998), 118–130.
3. G. Chang and T. W. Sederberg, *Over and Over Again*, New Math Library, MAA, 1997.
4. R. L. Devaney, *A First Course in Chaotic Dynamical Systems—Theory and Experiment*, Addison-Wesley, 1992.
5. D. Gulick, *Encounters with Chaos*, McGraw-Hill, 1992.
6. X-C. Huang, From intermediate value theorem to chaos, this MAGAZINE **62** (1992), 91–103.
7. B. P. Kitchens, *Symbolic Dynamics, One-sided, Two-sided and Countable State Markov Shifts*, Springer, Berlin-Heidelberg, 1998.
8. T. Li and J. A. Yorke, Period three implies chaos, *Amer. Math. Monthly* **82** (1975), 985–992.
9. D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge Univ. Pr. UK, 1995.
10. D. Sprows, Digitally determined periodic points, this MAGAZINE **71** (1998), 304–305.

Fibonacci Numbers and the Arctangent Function

KO HAYASHI
The King's Academy
Sunnyvale, CA 94085

This note provides several geometric illustrations of three identities involving the arctangent function and the reciprocals of Fibonacci numbers. The Fibonacci numbers are defined by $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$, for $n > 1$. The following identities link the Fibonacci numbers to the arctangent function. Only the first is evident in the literature [**1, 2, 3**].

$$\arctan\left(\frac{1}{F_{2i}}\right) = \arctan\left(\frac{1}{F_{2i+1}}\right) + \arctan\left(\frac{1}{F_{2i+2}}\right) \tag{1}$$

$$\arctan\left(\frac{2}{F_{2i+2}}\right) = \arctan\left(\frac{1}{F_{2i+1}}\right) + \arctan\left(\frac{1}{F_{2i+4}}\right) \tag{2}$$

$$\arctan\left(\frac{1}{F_{2i}}\right) = \arctan\left(\frac{2}{F_{2i+2}}\right) + \arctan\left(\frac{1}{F_{2i+3}}\right) \tag{3}$$

Identities (1)–(3) can be proven formally using Cassini's identity [**1**, p. 127]

$$F_{k+1}^2 = F_k F_{k+2} + (-1)^k$$

and the addition formula for the tangent function. Interested readers are invited to do so.

The following six diagrams illustrate special cases of equations (1)–(3). FIGURE 1, a representation of Euler's famous formula for π [**4, 5**], illustrates (1) for $i = 1$. One can see that $\angle ABD$ plus $\angle DBC$ is equal to $\angle ABC$.

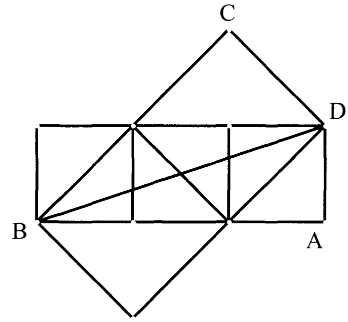


Figure 1 $\frac{\pi}{4} = \arctan(1) = \arctan(\frac{1}{2}) + \arctan(\frac{1}{3})$

FIGURE 2 illustrates (1) for $i = 2$, using the larger squares to form the arctangent of $1/5$ and the smaller squares being used to form the arctangents of $1/3$ and of $1/8$.

The two diagrams in FIGURE 3 illustrate (2) for the values $i = 1$ and $i = 2$.

The diagrams in FIGURE 4 illustrate equation (3) for the values $i = 1$ and $i = 2$.

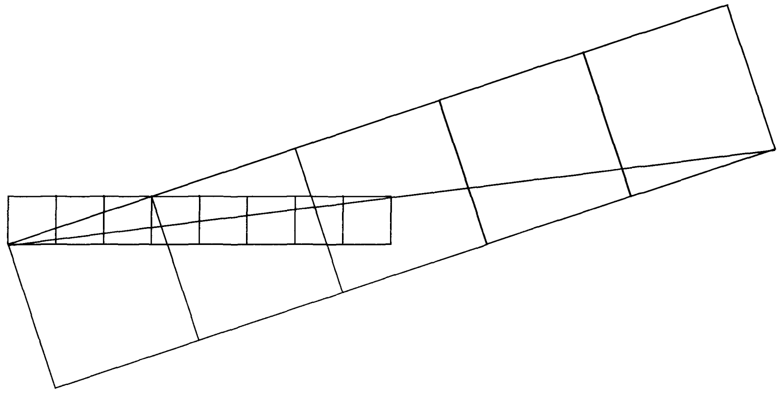


Figure 2 $\arctan(\frac{1}{3}) = \arctan(\frac{1}{5}) + \arctan(\frac{1}{8})$

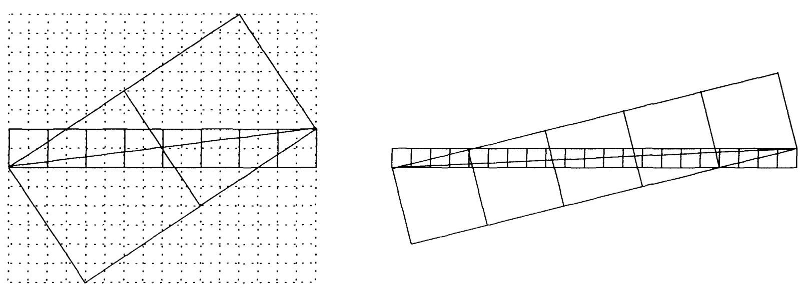


Figure 3 $\arctan(\frac{2}{3}) = \arctan(\frac{1}{2}) + \arctan(\frac{1}{8})$; $\arctan(\frac{1}{4}) = \arctan(\frac{1}{5}) + \arctan(\frac{1}{21})$

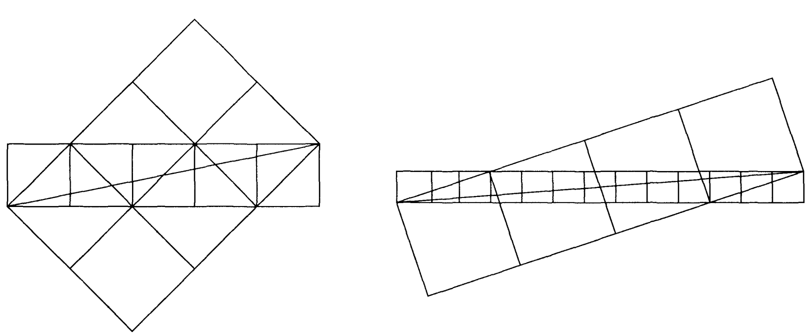


Figure 4 $\arctan(1) = \arctan(\frac{2}{3}) + \arctan(\frac{1}{5})$; $\arctan(\frac{1}{3}) = \arctan(\frac{1}{4}) + \arctan(\frac{1}{13}))$

Acknowledgments. The author would like to thank Professor Paul Garrett for reviewing the mathematics and to thank Ching-Yi Wang for his formatting of the manuscript.

REFERENCES

1. Robert M. Young, *Excursions in Calculus: An Interplay of the Continuous and the Discrete*, Dolciani Mathematical Expositions, MAA, Washington, DC, 1992, p. 136.
2. Marjorie Bicknell and Vernon E. Hoggatt Jr., *A Primer for the Fibonacci Numbers*, The Fibonacci Association, San Jose, 1972, pp. 49–50.
3. Problem 3801, *Amer. Math. Monthly* **45** (1938), 636.
4. Martin Gardner, *Mathematical Circus*, Spectrum Series, MAA, Washington, DC, 1992, p. 125.
5. Edward Kitchen, Dörrie tiles and related miniatures, this MAGAZINE **67** (1994), 128–130.

Hypercubes and Pascal's Triangle: A Tale of Two Proofs

MICHAEL L. GARGANO
JOSEPH F. MALERBA
Pace University
NY, NY 10038

MARTY LEWINTER
Purchase College
Purchase, NY 10577

The entries of the nth row of Pascal's triangle consists of the combinatorial numbers

$$\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \ldots, \binom{n}{n-2}, \binom{n}{n-1}, \binom{n}{n}, \quad \text{where } \binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

These numbers are called the binomial coefficients, because they satisfy the binomial theorem:

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

Upon setting $x = 1$, we obtain

$$2^n = \sum_{k=0}^{n} \binom{n}{k}. \tag{2}$$

Differentiating both sides of (1) with respect to x, we have

$$n(1+x)^{n-1} = \sum_{k=1}^{n} k\binom{n}{k} x^{k-1} = \binom{n}{1} + 2\binom{n}{2}x + 3\binom{n}{3}x^2 + \cdots + n\binom{n}{n}x^{n-1}. \tag{3}$$

Setting $x = 1$, we finally obtain the well-known identity [**4**, p. 11],

$$n2^{n-1} = \binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n}. \tag{4}$$

This last identity can also be proven without calculus. For a typical short proof, see Rosen [**9**, Section 4.3, Exercise 51] or Buckley and Lewinter [**3**, Section 1.4, Exercise 9].

We shall prove identity (4) using graph theory. In contrast to the previously mentioned proofs, which suggest that (4) is an algebraic accident, our approach here will count a combinatorial object in two different ways, thereby yielding insight into *why* the identity is true. The hypercube, Q_n, is an important graph, with applications in computer science [**1**]–[**3**], [**5**]–[**8**]. Its vertex set is given by $V(Q_n) = \{(x_1, x_2, \ldots, x_n) \mid x_i = 0 \text{ or } 1; i = 1, 2, \ldots, n\}$, i.e., each vertex is labeled by a binary n-dimensional vector. It follows that $|V(Q_n)| = 2^n$. Vertices $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are adjacent if and only if $\sum_{i=1}^{n} |x_i - y_i| = 1$, from which it follows that Q_n is n-regular. Since the degree sum is $n2^n$, we find that Q_n has $n2^{n-1}$

edges, that is, $|E(Q_n)| = n2^{n-1}$. The distance between vertices x and y is given by $\sum_{i=1}^{n} |x_i - y_i|$, that is, the number of place disagreements in their binary vectors.

Calling the vertex $(0, 0, \ldots, 0)$ the ***origin***, define the ith ***distance set*** D_i, as the set of vertices whose distance from the origin is i. Then for each $i = 0, 1, 2, \ldots, n$, we have $D_i = \{(x_1, x_2, \ldots, x_n) \mid \sum_{i=1}^{n} x_i = i\}$, that is, D_i consists of those vertices with exactly i 1s in their binary n-vectors. Moreover, we have $|D_i| = \binom{n}{i}$. The fact that the D_is partition $V(Q_n)$ demonstrates Equation (2) rather nicely.

Now observe that the induced subgraph on any D_i contains no edges, since all of the binary vectors of the vertices in D_i contain the same number of 1s. (If two vertices are adjacent, the number of 1s in their binary vectors must differ by exactly one.) Furthermore, if $|i - j| \geq 2$, then if $x \in D_i$ and $y \in D_j$, it follows that x and y are nonadjacent, that is, $xy \notin E(Q_n)$. Then all edges are of the form uv, where $u \in D_i$ and $v \in D_{i+1}$, for $i = 0, 1, 2, \ldots, n-1$. Since each vertex in D_{i+1} has $i+1$ 1s in its binary vector, it is adjacent to exactly $i+1$ vertices in D_i. (These vertices are obtained by replacing one 1 by 0 in the binary vector of the chosen vertex in D_{i+1}.) This implies that the number of edges with endpoints in both D_i and D_{i+1} is $(i+1)|D_{i+1}| = (i+1)\binom{n}{i+1}$. It follows that the total number of edges in Q_n is given by $\sum_{i=0}^{n-1} (i+1)\binom{n}{i+1}$. Finally, since $|E(Q_n)| = n2^{n-1}$, we are done with the proof of (4).

Acknowledgment. The authors thank the referees for their helpful suggestions.

REFERENCES

1. D. Aulicino, F. Harary, and M. Lewinter, Embedding k-dimensional meshes in hypercubes, *GTN of NY* **XXXVII** (1999), 51–53.
2. D. Aulicino and M. Lewinter, K-equidistant sets of Q_n, *GTN of NY* **XXXVI** (1999), 17–19.
3. Buckley and Lewinter, *A Friendly Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, 2003.
4. D. Burton, *Elementary Number Theory*, 4th ed., McGraw-Hill, New York, NY, 1997.
5. F. Harary and M. Lewinter, Spanning subgraphs of a hypercube III: Meshes, *Intern. J. Computer Math.* **25** (1988), 1–4.
6. F. Harary and M. Lewinter, Spanning subgraphs of a hypercube VI: Survey and unsolved problems, in *Graph Theory, Combinatorics, and Applications*, edited by Y.Alavi et al., J. Wiley, New York, NY, 1991.
7. F. Harary and M. Lewinter, Spanning subgraphs of a hypercube IV: Rooted trees, *Mathl. Comput. Modelling* **17** (1993), 85–88.
8. M.Lewinter and W.Widulski, Equipartition sets of hypercubes, *J. Comb. Info. & System Sci.* **16** (1991), 19–24.
9. Rosen, *Discrete Math and its Applications*, 4th ed. McGraw-Hill, New York, 1999.

A Derivation of Taylor's Formula with Integral Remainder

DIMITRI KOUNTOUROGIANNIS
69 Dover street #2
Somerville, MA 02144
dimitrik@alum.mit.edu

PAUL LOYA
Binghamton University
Binghamton, NY 13902-6000
paul@math.binghamton.edu

Taylor's formula with integral remainder is usually derived using integration by parts [**4, 5**], or sometimes by differentiating with respect to a parameter [**1, 2**]. According to M. Spivak [7, p. 390], integration by parts is applied in a "rather tricky way" to derive Taylor's formula, using a substitution that "one might discover after sufficiently many similar but futile manipulations". In this MAGAZINE, Lampret [3] derived both Taylor's formula and the Euler-Maclaurin summation formula using a rather heroic application of integration by parts.

We derive the remainder formula in a way that avoids tricks and heroics. The key step is changing the order of integration in multiple integrals, a topic that many students in an analysis class will benefit from reviewing. This derivation has almost certainly been found many times before [6], however, most people seem to be unaware of it.

The Taylor formula Suppose that a function $f(x)$ and all its derivatives up to $n+1$ are continuous on the real line. Then Taylor's formula for $f(x)$ about 0 is

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + R(x), \tag{1}$$

where the remainder, $R(x)$, is given by

$$R(x) = \frac{1}{n!}\int_0^x (x-u)^n f^{(n+1)}(u)\,du.$$

Our derivation is based on the following simple idea: Try to reconstruct f by integrating $f^{(n+1)}$, $n+1$ times. This approach is suggested by the case $n=0$, when (1) is merely the fundamental theorem of calculus. For notational simplicity, we prove (1) for only $n=2$; however, the general case is similar. Thus, consider

$$\widetilde{R}(x) := \int_0^x \int_0^w \int_0^v f^{(3)}(u)\,du\,dv\,dw. \tag{2}$$

Now let's evaluate this integral in two ways. The first way is by direct integration using the fundamental theorem of calculus three times:

$$\widetilde{R}(x) = f(x) - f(0) - xf'(0) - \frac{x^2}{2!}f''(0). \tag{3}$$

The second way to integrate (2) is by interchanging the order of integration:

$$\int_0^w \int_0^v f^{(3)}(u)\,du\,dv = \int_0^w \int_u^w f^{(3)}(u)\,dv\,du = \int_0^w (w-u)f^{(3)}(u)\,du.$$

Interchanging the order of integration again gives

$$\begin{aligned}\int_0^x \left\{\int_0^w \int_0^v f^{(3)}(u)\,du\,dv\right\} dw &= \int_0^x \left\{\int_0^w (w-u)f^{(3)}(u)\,du\right\} dw \\ &= \int_0^x \int_u^x (w-u)f^{(3)}(u)\,dw\,du \\ &= \frac{1}{2}\int_0^x (x-u)^2 f^{(3)}(u)\,du. \qquad (4)\end{aligned}$$

Equating (3) and (4) yields the Taylor formula (1) for $n=2$.

Acknowledgments. The second author was supported in part by a Ford Foundation Fellowship.

REFERENCES

1. Richard Courant and Fritz John, *Introduction to Calculus and Analysis, Vol. I*, Springer-Verlag, New York, 1989.
2. Allen Devinatz, *Advanced Calculus*, Holt, Rinehart and Winston, Inc., New York, 1968.
3. Vito Lampret, The Euler-Maclaurin and Taylor formulas: Twin, elementary derivations, this MAGAZINE, **74**:2 (2001), 109–122.
4. Serge Lang, *A First Course in Calculus*, 5th ed., Addison-Wesley Pub. Co., Reading, Mass., 1964.
5. Angus E. Taylor and W. Robert Mann, *Advanced Calculus*, 3rd ed., John Wiley & Sons Inc., New York, 1983.
6. Oved Shisha, On Taylor's theorem, *J. Res. Nat. Bur. Standards Sect. B*, **72B** (1968), 5–6.
7. Michael Spivak, *Calculus*, 2nd ed., Publish or Perish, Berkeley, 1980.

A Theorem Involving the Denominators of Bernoulli Numbers

PANTELIS DAMIANOU
University of Cyprus
Nicosia 16578
Cyprus

PETER SCHUMER
Middlebury College
Middlebury, VT 05753
schumer@middlebury.edu

The Swiss mathematician, Jakob Bernoulli (1654–1705), successfully sought a general method for summing the first n kth powers for arbitrary positive integers n and k. Let us define

$$S_k(n) = \sum_{j=1}^{n} j^k = 1^k + 2^k + \cdots + n^k.$$

Define the average of the first n kth powers by

$$\mu_k(n) = \frac{S_k(n)}{n}.$$

We pose and answer the following natural question: For which values of n and k is $\mu_k(n)$ an integer? Our answer, although it does involve the denominators of Bernoulli numbers, which undergraduates may not have seen, relies primarily upon elementary divisibility arguments.

Background In his *Ars Conjectandi*, published posthumously in 1713 and dedicated primarily to the theory of probability, Bernoulli presented a recursive solution for $S_k(n)$. It states that for $k \geq 1$,

$$(n+1)^{k+1} = (n+1) + \sum_{j=1}^{k} \binom{k+1}{j} S_j(n),$$

where the binomial coefficients are defined as usual:

$$\binom{k+1}{j} = \frac{(k+1)!}{j!\,(k+1-j)!}.$$

Furthermore, if we define what are now called the *Bernoulli numbers* by

$$B_0 = 1 \quad \text{and} \quad (k+1)B_k = -\sum_{j=0}^{k-1} \binom{k+1}{j} B_j \quad \text{for } k \geq 1,$$

then for $k \geq 1$, the sums $S_k(n)$ satisfy:

$$(k+1)S_k(n) = \sum_{j=0}^{k} \binom{k+1}{j} B_j (n+1)^{k+1-j}.$$

The Bernoulli numbers are the rational coefficients of the linear terms of the $(k+1)$st degree polynomials, $S_k(n-1)$. For example,

$$\begin{aligned} S_0(n-1) &= 1n - 1, \\ S_1(n-1) &= \frac{1}{2}n^2 - \frac{1}{2}n, \\ S_2(n-1) &= \frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n, \\ S_3(n-1) &= \frac{1}{4}n^4 - \frac{1}{2}n^3 + \frac{1}{4}n^2 + 0n, \quad \text{and} \\ S_4(n-1) &= \frac{1}{5}n^5 - \frac{1}{2}n^4 - \frac{4}{15}n^3 - \frac{1}{30}n. \end{aligned}$$

It follows that $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$, $B_3 = 0$, and $B_4 = -1/30$. In fact, $B_{2k+1} = 0$ for all $k \geq 1$. More compactly, we can define the Bernoulli numbers by the following power series:

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} \frac{B_k x^k}{k!}.$$

For even $k \geq 2$, we write $B_k = N_k/D_k$, where N_k and D_k are relatively prime and $D_k \geq 1$. The numerators N_k have played a significant role in number theory due largely to their connection with Fermat's Last Theorem. A prime p is a *regular* prime if p does not divide any of the numbers $N_2, N_4, \ldots, N_{p-3}$. (The only irregular primes less than 100 are 37, 59, and 67.) In 1850, Ernst Kummer proved that Fermat's Last Theorem is true for every exponent that is a regular prime. Of course, as of 1995 Andrew Wiles has proved Fermat's Last Theorem in toto.

The denominators D_k have played a less significant role in mathematics even though they can be clearly described. The 1840 Von Staudt-Clausen Theorem states that for k even, D_k is the product of all primes p with $(p-1) \mid k$. An interesting consequence is that D_k is square-free for all k. The theorem was proven independently (and nearly simultaneously) by the two mathematicians.

Examples We begin by considering a few examples, deriving results directly using congruence relations.

- $k = 1$: We have $\mu_1(n) = (n+1)/2$. Hence, $\mu_1(n) \in \mathbb{Z}$ if and only if n is odd. This is an exceptional case due to the fact that $B_1 \neq 0$.
- $k = 2$: In this case, $\mu_2(n) = (n+1)(2n+1)/6$. We claim that $\mu_2(n) \in \mathbb{Z}$ if and only if n is not divisible by 2 or 3. First, suppose that n is not divisible by 2

or 3. Clearly, $(n+1)(2n+1)$ is even. If $n \equiv 1 \pmod 3$, then $3 \mid (2n+1)$ and if $n \equiv 2 \pmod 3$, then $3 \mid (n+1)$. In any event, $6 \mid (n+1)(2n+1)$ and so $\mu_2(n) \in \mathbb{Z}$. Second, suppose that n is divisible by either 2 or 3. If $2 \mid n$, then $(n+1)(2n+1)$ is odd and hence is not divisible by 6. If $3 \mid n$, then $n = 3k$ for appropriate integer k, and $(n+1)(2n+1) = (3k+1)(6k+1) = 18k^2 + 9k + 1$, a number not divisible by 3 (nor by 6).

- $k = 3$: We have $\mu_3(n) = n(n+1)^2/4$. We claim that $\mu_3(n) \in \mathbb{Z}$ as long as n is not congruent to 2 modulo 4. If n is congruent to 0, 1, or 3 modulo 4, then $4 \mid n(n+1)^2$. However, if $n \equiv 2 \pmod 4$, then $n(n+1)^2 \equiv 2 \pmod 4$, and so 4 does not divide $n(n+1)^2$.
- $k = 4$: In this case, $\mu_4(n) = (n+1)(2n+1)(3n^2+3n-1)/30$. We claim that $\mu_4(n) \in \mathbb{Z}$ if and only if n is not divisible by 2, 3, or 5. Suppose that n is relatively prime to 30 (equivalently not divisible by 2, 3, or 5). Then $n+1$ is even and $(n+1)(2n+1)$ is divisible by 3. Furthermore, $(n+1)(2n+1)(3n^2+3n-1) = 6n^4 + 15n^3 + 10n^2 - 1 \equiv n^4 - 1 \pmod 5$. But by Fermat's Little Theorem, $n^4 - 1 \equiv 0 \pmod 5$ and so $5 \mid (n+1)(2n+1)(3n^2+3n-1)$. Hence, $\mu_4(n) \in \mathbb{Z}$ in this case as well.

  In the other direction, if $2 \mid n$, then $(n+1)(2n+1)(3n^2+3n-1)$ is odd and not divisible by 30. If $3 \mid n$, then $(n+1)(2n+1)(3n^2+3n-1) \equiv -1 \pmod 3$ and so is not divisible by 30. Finally, if $5 \mid n$, then $(n+1)(2n+1)(3n^2+3n-1) \equiv -1 \pmod 5$ and so is not divisible by 30.

These examples hint that the situation is very different for odd and even values of n. We develop our main theorem in two sections. Only the even case involves the Bernoulli numbers. In both parts, we use the easily noted fact that $\mu_k(n)$ is an integer if and only if $S_k(n) \equiv 0 \pmod n$.

An "odd" theorem

THEOREM 1. *For odd numbers $k \geq 3$, $\mu_k(n)$ is an integer if and only if $n \not\equiv 2 \pmod 4$.*

Proof. Suppose k is odd and $k \geq 3$. Since $(n-a)^k \equiv -a^k \pmod n$ for all a, we can pair up the terms of $S_k(n)$ from the outside in.

(a) If n is odd, then

$$\begin{aligned} S_k(n) &= [1^k + (n-1)^k] + [2^k + (n-2)^k] + \cdots \\ &\quad + \left[\left(\frac{n-1}{2}\right)^k + \left(\frac{n+1}{2}\right)^k\right] + n^k \\ &\equiv (1^k - 1^k) + (2^k - 2^k) + \cdots + 0 = 0 \pmod n. \end{aligned}$$

(b) If n is even, then there are two subcases depending on whether or not n is divisible by 4.

   (i) If $n \equiv 0 \pmod 4$, then

$$\begin{aligned} S_k(n) &= [1^k + (n-1)^k] + [2^k + (n-2)^k] + \cdots \\ &\quad + \left[\left(\frac{n}{2} - 1\right)^k + \left(\frac{n}{2} + 1\right)^k\right] + \left(\frac{n}{2}\right)^k + n^k \\ &\equiv 0 \pmod n \quad \text{since } k > 1 \text{ and } \frac{n}{2} \text{ is even.} \end{aligned}$$

(ii) If $n \equiv 2 (\bmod 4)$, then

$$\begin{aligned} S_k(n) &= [1^k + (n-1)^k] + [2^k + (n-2)^k] + \cdots \\ &\quad + \left[\left(\frac{n}{2}-1\right)^k + \left(\frac{n}{2}+1\right)^k\right] + \left(\frac{n}{2}\right)^k + n^k \\ &\equiv \left(\frac{n}{2}\right)^k (\bmod n). \end{aligned}$$

But $n/2$ is odd and so $(n/2)^k$ is odd. Since n is even, $(n/2)^k$ is not congruent to $0 (\bmod n)$. ■

An "even" more interesting theorem

THEOREM 2. *For even numbers $k \geq 2$, $\mu_k(n)$ is an integer if and only if n is relatively prime to D_k.*

Proof. The Von Staudt-Clausen theorem [**1**, Theorem 118] states that the kth Bernoulli denominator can be written as a product of primes as follows:

$$D_k = \prod_{p \text{ prime and } p-1|k} p.$$

To prove our result it must be shown that $S_k(n) \equiv 0 (\bmod n)$ if and only if for every prime p dividing n, that $p \nmid D_k$. By Von Staudt-Clausen it suffices to establish that

$$S_k(n) \equiv 0 (\bmod n) \text{ iff for every prime } p \text{ that divides } n,\ (p-1) \nmid k. \tag{1}$$

For the sake of completeness, we state and prove the following easily established result [**1**, Theorem 119]:

LEMMA 1. *For any prime p,*

$$\begin{aligned} \sum_{m=1}^{p} m^k &\equiv -1 (\bmod p) \textit{ if } (p-1) \mid k \\ &\equiv 0 (\bmod p) \textit{ if } (p-1) \nmid k. \end{aligned} \tag{2}$$

Proof of Lemma 1. If $(p-1) \mid k$, then $k = (p-1)r$ for some integer r. Hence, for $m < p$, $m^k = (m^{p-1})^r \equiv 1 (\bmod p)$ by Fermat's Little Theorem. It follows that $\sum_{m=1}^{p} m^k \equiv p - 1 \equiv -1 (\bmod p)$.

If $(p-1) \nmid k$, then let g be a primitive root of p. It follows that the set $\{g, 2g, \ldots, (p-1)g\}$ is identical to the set $\{1, 2, \ldots, p-1\}$ of reduced residues modulo p. Hence $\sum_{m=1}^{p-1} (mg)^k \equiv \sum_{m=1}^{p-1} m^k (\bmod p)$, and so $(g^k - 1) \sum_{m=1}^{p-1} m^k \equiv 0 (\bmod p)$.

But g^k is not congruent to $1 (\bmod p)$ since g is a primitive root mod p. Thus $\sum_{m=1}^{p} m^k \equiv 0 (\bmod p)$. This establishes Lemma 1. ■

Returning to the proof of our main result, it is convenient to first assume that *n is square-free*.

We establish (1):

($\Leftarrow$) Suppose that for all p dividing n that $(p-1) \nmid k$. Choose a prime $p \mid n$. By (2),

$$\sum_{m=1}^{p} m^k \equiv 0 (\bmod p).$$

Similarly,

$$\sum_{m=rp+1}^{(r+1)p} m^k \equiv 0 (\text{mod}\, p) \quad \text{for } 0 \le r \le \frac{n}{p} - 1.$$

Hence $S_k(n) = \sum_{m=1}^{n} m^k = \sum_{r=0}^{\frac{n}{p}-1} \sum_{m=rp+1}^{(r+1)p} m^k \equiv 0(\text{mod}\, p)$. But p arbitrary and n square-free implies that $S_k(n) \equiv 0(\text{mod}\, n)$.

($\Rightarrow$) We prove the contrapositive. Suppose there exists a prime $p \mid n$ such that $(p-1) \mid k$. By (2)

$$\sum_{m=1}^{p} m^k \equiv -1(\text{mod}\, p).$$

Similarly,

$$\sum_{m=rp+1}^{(r+1)p} m^k \equiv -1(\text{mod}\, p) \quad \text{for } 0 \le r \le \frac{n}{p} - 1.$$

Hence $S_k(n) \equiv -n/p(\text{mod}\, p)$, which is not congruent to $0(\text{mod}\, p)$ since p and n/p are relatively prime. Thus $S_k(n)$ is not congruent to $0(\text{mod}\, n)$.

Now suppose that n *is not square-free.*

($\Leftarrow$) Suppose that for all p dividing n that $(p-1) \nmid k$. If there is a prime p exactly dividing n (that is, $p \mid n$, but p^2 does not divide n), then as in the square-free case, $S_k(n) \equiv 0(\text{mod}\, p)$.

Now let p be a prime with $p^a \parallel n$ with $a \ge 2$. (The notation $p^a \parallel n$ means that $p^a \mid n$ and $p^{a+1} \nmid n$.)

LEMMA 2. *Let p be a prime with $(p-1) \nmid k$. Then*

$$1^k + 2^k + \cdots + (p^a)^k \equiv 0(\text{mod}\, p^a).$$

Proof of Lemma 2 (Induction on a). If $a = 1$, then

$$1^k + 2^k + \cdots + p^k \equiv 0(\text{mod}\, p) \text{ by (2).}$$

Assume that the lemma holds for $a-1$, namely that

$$1^k + 2^k + \cdots + (p^{a-1})^k \equiv 0(\text{mod}\, p^{a-1}).$$

Now consider $S_k(p^a) = \sum_{r=0}^{p-1} \sum_{j=1}^{p^{a-1}} (rp^{a-1} + j)^k$. The binomial theorem implies that

$$(rp^{a-1} + j)^k = \sum_{i=0}^{k} \binom{k}{i} r^i p^{(a-1)i} j^{k-i}.$$

Hence

$$S_k(p^a) = \sum_{r=0}^{p-1} \sum_{j=1}^{p^{a-1}} \sum_{i=0}^{k} \binom{k}{i} r^i p^{(a-1)i} j^{k-i}. \tag{3}$$

For $i \ge 2$, $p^{(a-1)i} \equiv 0(\text{mod}\, p^a)$ and so all terms of (3) with $i \ge 2$ are congruent to $0(\text{mod}\, p^a)$.

For $i = 0$, $\sum_{r=0}^{p-1} \sum_{j=1}^{p^{a-1}} j^k = p \cdot S_k(p^{a-1})$.
But $S_k(p^{a-1}) \equiv 0 (\text{mod}\, p^{a-1})$ by our inductive hypothesis. Hence

$$\sum_{r=0}^{p-1} \sum_{j=1}^{p^{a-1}} j^k \equiv 0 (\text{mod}\, p^a).$$

For $i = 1$,

$$\sum_{r=0}^{p-1} \sum_{j=1}^{p^{a-1}} krp^{a-1} j^{k-1} = \sum_{r=0}^{p-1} krp^{a-1} \cdot S_{k-1}(p^{a-1})$$
$$= kS_{k-1}(p^{a-1}) \cdot p^{a-1} \cdot \frac{(p-1)p}{2}.$$

But $S_{k-1}(p^{a-1}) \in \mathbb{Z}$ and $2 \mid (p-1)$. Thus

$$\sum_{r=0}^{p-1} \sum_{j=1}^{p^{a-1}} krp^{a-1} j^{k-1} \equiv 0 (\text{mod}\, p^a).$$

Therefore, $S_k(p^a) \equiv 0 (\text{mod}\, p^a)$ and Lemma 2 is proven. ■

In a manner analogous to Lemma 2, it follows that

$$\sum_{m=rp^a+1}^{(r+1)p^a} m^k \equiv 0 (\text{mod}\, p^a) \quad \text{for } 0 \le r \le \frac{n}{p^a} - 1.$$

Hence $S_k(n) \equiv 0 (\text{mod}\, p^a)$ for any $p \mid n$ with $p^a \parallel n$ and $a \ge 1$. It follows that $S_k(n) \equiv 0 (\text{mod}\, n)$.

($\Rightarrow$) A slight modification of the square-free proof works here, as follows.

On the one hand, if there exists a prime $p \parallel n$ such that $(p-1) \mid k$, then by (2), $\sum_{m=1}^{p} m^k \equiv -1 (\text{mod}\, p)$. Hence $S_k(n) \equiv -n/p (\text{mod}\, p)$, which is not congruent to $0 (\text{mod}\, p)$ since $p \parallel n$. Thus $n \nmid S_k(n)$ as in the square-free case.

On the other hand, suppose there exists a prime p with $p^a \parallel n$ with $a \ge 2$ and $(p-1) \mid k$. By (2), $\sum_{m=1}^{p} m^k \equiv -1 (\text{mod}\, p)$. Thus $\sum_{m=1}^{p} m^k \equiv (rp-1) (\text{mod}\, p^a)$ for some r with $1 \le r \le p^{a-1}$. But then $S_k(n) \equiv n/p(rp-1) \equiv -n/p (\text{mod}\, p^a)$. Hence $S_k(n)$ is not congruent to $0 (\text{mod}\, p^a)$ and so $n \nmid S_k(n)$.

This completes the proof of part (2) and establishes the theorem. ■

REFERENCES

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

Proof Without Words: Viviani's Theorem with Vectors

The sum of the distances from a point P in an equilateral to the three sides of the triangle is independent of the position of P (and so equal to the altitude of the triangle.)

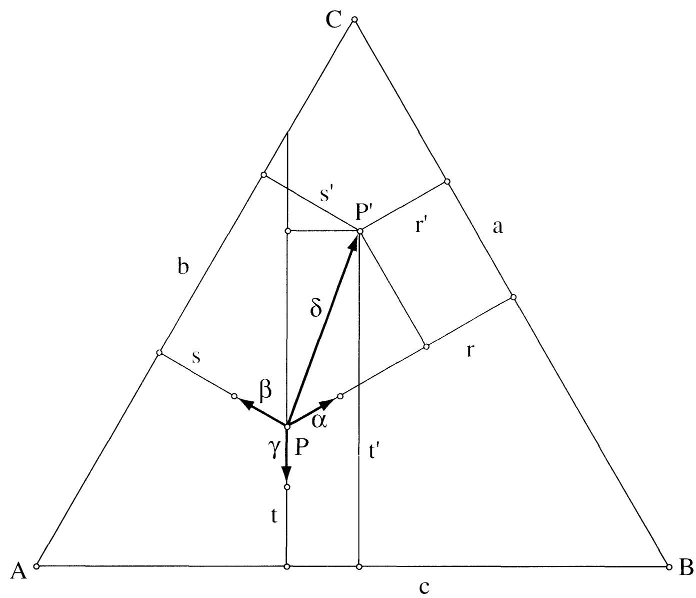
$$|\alpha| = |\beta| = |\gamma|$$

$$\alpha + \beta + \gamma = 0$$

$$\alpha \cdot \delta + \beta \cdot \delta + \gamma \cdot \delta = 0$$

$$(r - r') + (s - s') + (t - t') = 0$$

$$r + s + t = r' + s' + t'$$



—HANS SAMELSON
STANFORD, CA 94305

Let π be 3

ROBERT N. ANDERSEN
JUSTIN STUMPF
JULIE TILLER
University of Wisconsin–Eau Claire
Eau Claire, WI 54702

And he (Hirum of Tyre) made a molten sea, ten cubits from the one brim to the other: it was round all about, and its height was five cubits: and a line of thirty cubits did compass it round about. 1 Kings 7:23

The literal interpretation of this Biblical passage is that Hirum constructed a hemispherical basin that had a diameter of 10 cubits (a cubit is approximately one half of a meter) and had a circumference three times that value. This ratio of the circumference of a circle to its diameter apparently contradicts results of the works of Archimedes who established that the ratio of the circumference of any circle to its diameter is between $3\frac{10}{71}$ and $3\frac{1}{7}$. Since Archimedes provided a convincing argument for his values we are inclined to accept them as true and regard the numbers given to us in First Kings as approximate values whose error is due to rounding off. In this paper we do not write this off to a rounding error, but rather identify a setting where 3 is the correct value.

Archimedes' results were obtained in Euclidean geometry. Using alternate geometries we will establish that the ratio of the circumference of a circle to its diameter may take on a continuum of values, including three. In the next section we will discuss how Archimedes first determined his values. Then we will show how the ratio varies in spherical geometry. Finally, we discuss the possible values in hyperbolic geometry.

The results of Archimedes The computation of π has a long history [**2**]. Archimedes [**1**] first considered a regular polygon as inscribed within a circle and then as circumscribed about a circle, and thus was able to compute a lower approximation and an upper approximation for the ratio of the circumference of a circle to its diameter. He observed these ratios up to a polygon with 96 sides, and thus was able to conclude that the ratio of the circumference of any circle to its diameter is between $3\frac{10}{71}$ and $3\frac{1}{7}$. Using Archimedes' technique and modern trigonometry we can compute even better approximations for this ratio.

Let us assume that a circle is divided into 360 degrees, and consider regular polygons having n sides (in our diagrams $n = 6$), where the length of each of the sides is 1 unit.

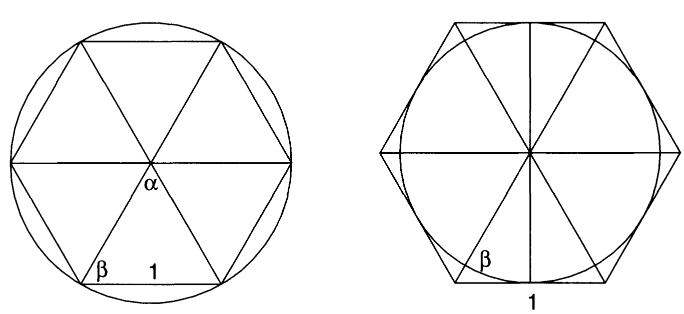


Figure 1 A circle circumscribed and inscribed by a regular polygon

From the easy observation that the measure of the angle α is $360°/n$, we find that the measure of the angle β is $90°(n-2)/n$. Now using the law of sines, we can see that the diameter of the circumscribed circle would be

$$2 \cdot \frac{\sin(90° \cdot (n-2)/n)}{\sin(360°/n)}.$$

The diameter of the inscribed circle is easily calculated to be

$$\tan(90° \cdot (n-2)/n).$$

We see that the ratio of the circumference of the circle to its diameter lies between ratio of the perimeter of the inscribed polygon to the circle's diameter and the ratio of the perimeter of the circumscribed polygon to the circle's diameter. For the remainder of the paper we will follow the universally accepted convention of using π as the ratio of the circumference of a circle to its diameter in Euclidean geometry. Computing the ratio of the perimeter of each polygon to the diameter of the circle, we derive the inequality

$$n \cdot \frac{\sin(360°/n)}{2 \cdot \sin(90°(n-2)/n)} < \pi < n \cdot \cot(90°(n-2)/n).$$

We can now take the limits as n approaches infinity. The results yield what we already know, that each ratio approaches the same value, which is approximately 3.141592654.

The ratio π in spherical geometry From First Kings the ratio of the circumference to the diameter is easily computed to be 3. For the sake of academic curiosity let us assume that the ratio of the circumference to the diameter of the "molten sea" was indeed supposed to be 3. How can this be? An answer lies in the geometry under consideration. Archimedes proved his results in Euclidean geometry, the geometry of plane or flat surfaces. But suppose we were to use spherical geometry. Can we then produce circles in which the ratios of the circumferences to the diameters are indeed exactly 3?

In Euclidean geometry, the surface upon which measurements are made is the plane. In spherical geometry, the universal surface upon which measurements are made is the sphere. To perform calculations on the sphere we imagine the sphere embedded in three-dimensional Euclidean space. We may then use the results of Euclidean geometry to compute measurements on the sphere.

In the spherical setting, the ratio of circumference to diameter need not be π; in fact, it is easy to see how to produce a circle whose circumference-to-diameter ratio is 2. Simply choose the circle to be a great arc (equator), then the diameter of the circle (the shortest path between two antipodal points on the circle) is half the length of the great arc. So the ratio of the great arc to $1/2$ of a great arc is $1/(1/2) = 2$.

Since we now know that on a sphere the ratio of a circumference of a circle to its diameter need not be π, we turn our attention to all possible values for this ratio.

Let a circle with circumference C and diameter $2r$ lie on a sphere with radius R. Let α be the center of the circle, β the center of the sphere, γ a point on the circle, and δ the center of the circle as viewed in Euclidean geometry. Call ρ the Euclidean radius of the circle in the planar cross-section shown, and θ the angle $\angle\alpha\beta\gamma$, as in FIGURE 2.

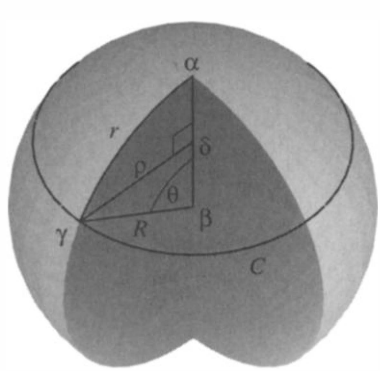


Figure 2 A circle on a sphere

We have $r = R\theta$, $\sin\theta = \rho/R$, and $2\rho\pi = C$, so that $\rho = R\sin\theta$ and $\theta = r/R$. Thus

$$C = 2\pi R \sin\frac{r}{R} = 2\pi\frac{r}{\theta}\sin\theta.$$

Let Π denote the ratio of the circumference of a circle to its diameter, which we represent as a function of θ. Since $2r$ is the diameter of the circle on the sphere, we can compute

$$\Pi(\theta) = \frac{C(\theta)}{2r} = \pi\frac{\sin\theta}{\theta}.$$

We now can compute $\Pi(\pi/2) = 2$, as we noted earlier. We can also compute $\Pi(\pi/6) = 3$, which is the ratio given in First Kings. We also note that the limiting value as the angle θ approaches 0 is

$$\lim_{\theta\to 0}\Pi(\theta) = \pi\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = \pi.$$

We may allow θ to grow larger than $\pi/2$ and define the diameter of the circle to be twice the radius, which is the length of the arc from a point on the circle to the center, which we might as well call the North Pole. The graph of $\Pi(\theta) = \pi\sin(\theta)/\theta$ will represent all possible values for Π on the sphere.

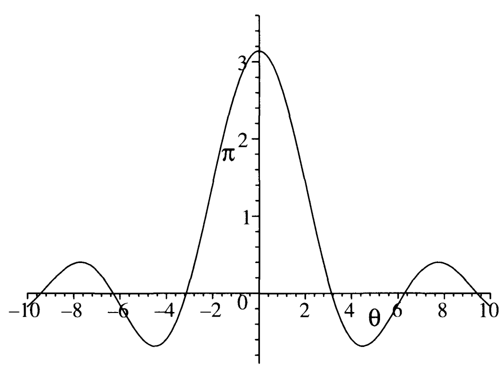


Figure 3 The graph of $\Pi(\theta) = \pi(\sin(\theta))/\theta$

As the terminus of angle θ passes the South Pole we will measure the circumference in the opposite direction, hence yielding the negative values for Π.

Also, we see that $\lim_{\theta\to\pm\infty}\pi(\sin\theta)/\theta = 0$. As θ grows larger the radius is wrapped once around the sphere for every multiple of 2π, but the absolute value of the circumference will never grow larger than the length of the equator.

The graph of our function suggests that a minimum value for Π occurs near ± 4.5. The standard calculus technique of setting the first derivative of $\Pi(\theta) = \pi(\sin\theta)/\theta$ equal to 0 yields the following results:

$$\Pi'(\theta) = \pi\frac{\theta\cos\theta - \sin\theta}{\theta^2} = 0 \Rightarrow \tan\theta = \theta,$$

so that $\theta \approx \pm 4.493409458$. We thus compute the minimum value to be approximately

$$\Pi(4.493409458) = \pi\frac{\sin 4.493409458}{4.493409458} = -.6824595706.$$

Thus we may conclude that for spherical geometry the ratio of the circumference of a circle to its diameter ranges over the values

$$-.6824595706 \le \Pi < \pi \approx 3.141592654.$$

If we return to the values given to us in First Kings we may compute the size of the sphere on which the measurements of the molten sea are made. Using the equation $C = 2\pi R \sin r/R$, we solve for R when $C = 30$ and $r = 5$:

$$30 = 2\pi R \sin \frac{5}{R} \quad \text{or} \quad \frac{15}{\pi} = R \sin \frac{5}{R}.$$

We have no closed form solution for this equation, but the function

$$F(R) = R \sin\left(\frac{5}{R}\right) - \frac{15}{\pi}$$

is a continuously differentiable function for $R > 0$, so Newton's method would produce an accurate approximation. This takes only a few seconds using computer programs such as *Maple* or *Mathematica*, and we find that $R \approx 9.549296586$ cubits. Thus a circle at latitude 60° on a sphere of radius 9.549296586 cubits will have a circumference of 30 cubits and a diameter of 10 cubits.

The ratio π in hyperbolic geometry Spherical geometry is not the only alternative to the Euclidean plane. Any smooth two-dimensional surfaces in $\mathbb{R}^3$ might do just as well. Any such surface has an intrinsic measurement of curvature, which gives us an idea of how curved the surface is at each point. The curvature, or more precisely the Gaussian curvature, is computed as the product of two other quantities called the *principal curvatures* at a point. These principal curvatures are the maximum and minimum curvatures of the collection of one-dimensional arcs through that point. For a circle the curvature is $1/R$ where $|R|$ is the radius. We comment that R may be positive or negative depending as to whether we make the measurement from a vantage point inside the circle or outside the circle. Since the curvature of every arc on a sphere through any given point is $1/R$ where $|R|$ is the radius of the sphere, we have the curvature of the sphere to be the constant $K = 1/R^2$, which is always positive.

It is possible for surfaces to have negative curvature. The saddle point of a hyperbolic paraboloid is such an example. Since the surface is curving in a concave fashion in one direction and a convex fashion in the other direction, the maximal principal curvature will be positive and the minimal principal curvature will be negative. The Gaussian curvature at the saddle point is thus the product of a negative value and a positive value, which must be negative. O'Neill [3] is one standard reference.

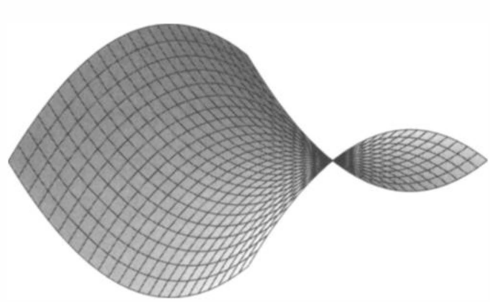


Figure 4 The hyperbolic paraboloid

We may now consider a space where at every point the principal curvatures are $1/R$ and $-1/R$, and hence the Gaussian curvature is the negative value $K = -1/R^2 = 1/(iR)^2$. We may consider this space to be a pseudo-sphere with an imaginary radius. The geometry on this space is known as hyperbolic geometry.

We can replace R with iR in our development of the function $\Pi(\theta)$ to get the corresponding formulas for hyperbolic geometry.

$$C = 2\pi i R \sin\frac{r}{iR} \quad \text{where } \frac{r}{R} = \theta.$$

Thus, we get a formula reminiscent of the spherical ratios,

$$C = 2\pi i r \frac{\sin\frac{\theta}{i}}{\theta} = 2\pi i r \frac{\sin(-i\theta)}{\theta}.$$

Apply the identity $\sin\theta = (e^{i\theta} - e^{-i\theta})/2i$, to give us

$$C = 2\pi i r \frac{e^{-i^2\theta} - e^{i^2\theta}}{2i\theta} = 2\pi r \frac{e^{\theta} - e^{-\theta}}{2\theta} \equiv 2\pi r \frac{\sinh\theta}{\theta}.$$

Thus

$$\Pi(\theta) = \frac{C}{2r} = \pi\frac{\sinh\theta}{\theta}.$$

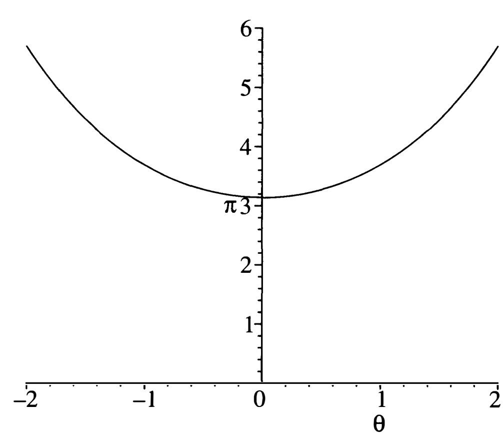


Figure 5 The graph of $\Pi(\theta) = \pi(\sinh\theta)/\theta$

The graph of $\Pi(\theta) = \pi(\sinh\theta)/\theta$ in FIGURE 5 reveals that Π takes on values greater than π in hyperbolic geometry. Easy limit computations produce

$$\lim_{\theta\to 0} \pi\frac{\sinh\theta}{\theta} = \pi \quad \text{and} \quad \lim_{\theta\to\pm\infty} \pi\frac{\sinh\theta}{\theta} = +\infty.$$

So we may conclude that, in hyperbolic geometry, Π takes on all values greater than π.

Conclusion The ancient Hebrews were certainly unaware of alternate geometries and were more concerned with the spiritual aspects of their lives than mathematical precision. Since it is highly unlikely that they would choose a sphere of approximately 9 meters in diameter on which to make their measurements, we can rationally conclude the discrepancies between First Kings and Archimedes is due to a very coarse approximation. But it is entertaining to realize that these measurements can be made exact by using the appropriate geometry.

Archimedes did not know the formal limit concept we use today, but he most surely knew the intuitive concept. Today the exact value of π is known to be the limit of the sequence produced by Archimedes. It is interesting to note that π is the limit of the ratio of circumferences of circles to their diameters in both the spherical and hyperbolic geometries. But this should not be surprising, since the limit is taken as the central angle approaches 0. If we imagine that the diameter of the circle is held constant, then the radius of the sphere or pseudo-sphere must approach infinity and the curvature approaches 0. Thus the Euclidean plane can be thought of as a sphere or pseudo-sphere with curvature 0.

Using our three geometries, π can be assigned any positive real value that you want, and even some negative values. We find it compelling to ponder the possibility of a geometry that would allow all negative values.

REFERENCES

1. Archimedes, Measurement of a circle, *The Works of Archimedes with Introductory Chapters by J. L. Heath*, Dover Publications, New York, 1912.
2. D. Castellanos, The ubiquitous π, this MAGAZINE, **61** (1988), 67–98 & 148–163.
3. Barrett O'Neill, *Elementary Differential Geometry*, Academic Press, San Diego, CA, 1966.

(*continued from page 247*)

circle. Thus, no points on the curved parts of the convex hull belong to any of the arcs whose lengths we are summing. Because the curved parts of the convex hull can be translated together to form a unit circle, their total length is 2π. Thus, our bound is improved to $2(n-1)\pi$. Combining this with our previous information, we have

$$\sum_{1\le i<j\le n} \frac{8}{O_iO_j} < 2(n-1)\pi.$$

Dividing by 8 yields the desired inequality.

PROBLEMS

Proposals

To be considered for publication, solutions should be received by November 1, 2003.

1672. *Proposed by Árpád Bényi, University of Kansas, Lawrence, KS, and Mircea Martin, Baker University, Baldwin City, KS.*

Let f and g be odd functions that are analytic in a neighborhood of 0. Given that $f'(0) = g'(0) \neq 0$, $f^{(3)}(0) = g^{(3)}(0) \neq 0$, and $0 \neq f^{(5)}(0) \neq g^{(5)}(0) \neq 0$, evaluate

$$\lim_{x\to 0} \frac{f(x) - g(x)}{f^{-1}(x) - g^{-1}(x)},$$

where h^{-1} denotes the inverse of the function h.

1673. *Proposed by P. Ivady, Budapest, Hungary.*

Prove that for $0 < x < \pi$,

$$\frac{\sin^3 x}{x^3} < \left(\frac{\pi^2 - x^2}{\pi^2 + x^2}\right)^2.$$

1674. *Proposed by H. A. Shah Ali, Tehran, Iran.*

Given that $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n$, and $x_{n+1} = x_1$, prove that

$$\sum_{k=1}^{n} \frac{x_k - x_{k+1}}{1 + x_k x_{k+1}} \geq 0.$$

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

Solutions should be written in a style appropriate for this MAGAZINE. Each solution should begin on a separate sheet.

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

1675. *Proposed by Michael Woltermann, Washington and Jefferson College, Washington, PA.*

Let ABC be a triangle, and let $\widehat{AB}$, $\widehat{BC}$, $\widehat{CA}$, respectively, be the arcs of its circumcircle subtended by sides AB, BC, CA. (The arcs are defined so that any two of the three arcs intersect in just one point.) Let X, Y, Z, respectively, be the midpoints of $\widehat{AB}$, $\widehat{BC}$, $\widehat{CA}$, and let X', Y', Z', respectively, be the reflections of X, Y, Z in sides AB, BC, CA. Triangle $X'Y'Z'$ is called the Fuhrman triangle of triangle ABC, and the circumcenter F of triangle $X'Y'Z'$ is the Fuhrman point of ABC. Let I and N be the incenter and nine point center, respectively, of triangle ABC. Prove that N is the midpoint of segment IF.

1676. *Proposed by Erwin Just (Emeritus), Bronx Community College of the City University of New York, Bronx, NY.*

Find all pairs of integers m and n such that

$$2m \equiv -1 (\operatorname{mod} n) \quad \text{and} \quad n^2 \equiv -2 (\operatorname{mod} m).$$

Quickies

Answers to the Quickies are on page 238.

Q931. *Proposed by Nick MacKinnon, Winchester College, Winchester, England.*

Six pupils disrupted my lesson on Pascal's triangle, so for a punishment I gave each one a different prime number and told them to work out its tenth power by hand. "That's too easy," said Blaise, whose prime was 3. "I'll work out the product of everybody else's answer instead." Then she read out the forty-one digit answer and left. What was the answer?

Q932. *Proposed by Robert Gregorac, Iowa State University, Ames, IA, and Murray S. Klamkin, University of Alberta, Edmonton, AB, Canada.*

Let A be a point on a circle of center O and radius a, and let P be a point on the extension of $\overline{OA}$ through A. A secant line from P intersects the circle in points Q and Q'. Given a fixed position of P, determine the maximum area of triangle AQQ'.

Solutions

Sums of Primes and Squares **June 2002**

1648. *Proposed by Erwin Just (Emeritus), Bronx Community College of the City University of New York, Bronx, NY.*

Prove that there exist an infinite number of integers, none of which is expressible as the sum of a prime and a perfect square.

The following solution was submitted by many readers.

We show that for each positive integer n, the number $(3n + 5)^2$ cannot be written as the sum of a square and a prime. Indeed if $(3n + 5)^2 = m^2 + p$ for some positive integers m and p, then

$$p = (3n - m + 5)(3n + m + 5).$$

If $3n - m + 5 > 1$, then p is not prime. If $3n - m + 5 = 1$, then $p = 6n + 9$ and again p is not prime.

Solved by Jack Abad, Bela Bajnok, Roy Barbara (Lebanon), Michel Bataille (France), Brian D. Beasley, D. Bednarchak, Ton Boerkoel, Jean Bogaert (Belgium), Pierre Bornsztein (France), Marc Brodie, Doug Cashing, John Christopher, Charles R. Diminnie, Daniele Donini (Italy), Russell Euler and Jawad Sadek, Fejentalaltuka Szeged Problem Solving Group (Hungary), Charles M. Fleming, Ovidiu Furdui, William Gasarch, Marty Getz and Dixon Jones, Julien Grivaux (France), Jerrold W. Grossman, Douglas Iannucci (Virgin Islands), Khudija Jamil, John H. Jaroma, D. Kipp Johnson, Lenny Jones, Murray S. Klamkin (Canada), Ken Korbin, Victor Y. Kutsenok, Elias Lampakis (Greece), Joe Langsam, Peter W. Lindstrom, S. C. Locke, Francesco Marino (Italy), Reiner Martin, Millsaps Problem Solving Group, Gary Raduns, Alex Rand, Robert C. Rhoades, Rolf Richberg (Germany), John P. Robertson, James S. Robertson, Elianna Ruppin, Harry Sedinger, Achilleas Sinefakopoulos, Richard M. Smith, Albert Stadler (Switzerland), Steven Steinsaltz, Dave Trautman, Daniel G. Treat, Jim Vandergriff, Edward Wang, Doug Wilcock, Dean Witter III, Japheth Wood, Li Zhou, David Zhu, and the proposer. There were two solutions with no name.

Rational Bisectors **June 2002**

1649. *Proposed by K. R. S. Sastry, Bangalore, India.*

Prove that if a right triangle has all sides of integral length, then it has at most one angle bisector of integral length.

Solution by Julien Grivaux, student, Université Pierre et Marie Curie, Paris, France.

We prove that there is at most one angle bisector of rational length. Let ABC be a triangle with integer side lengths $BC = a$, $CA = b$, and $AB = c$. Let ℓ_a, ℓ_b, and ℓ_c be the lengths of the angle bisectors from A, B, and C, respectively. It is well known that

$$\ell_a = \frac{2}{b+c}\sqrt{s(s-a)bc},$$

where $s = (a+b+c)/2$ is the semiperimeter of ABC. Similar expressions hold for ℓ_b and ℓ_c. Hence

$$\begin{aligned}\ell_a\ell_b\ell_c &= \frac{8sabc}{(b+c)(c+a)(a+b)}\sqrt{s(s-a)(s-b)(s-c)} \\ &= \frac{8sabc[ABC]}{(b+c)(c+a)(a+b)}, \qquad (*)\end{aligned}$$

where $[ABC]$ denotes the area of ABC. Now assume that ABC is a right triangle with right angle C. Then $[ABC] = ab/2$ is rational, and it follows from $(*)$ that $\ell_a\ell_b\ell_c$ is rational. Let H denote the intersection point of AB with the angle bisector from C. Then

$$[ABC] = [ACH] + [BCH] = \frac{1}{2}b\ell_c\sin(45°) + \frac{1}{2}a\ell_c\sin(45°) = \frac{\ell_c(a+b)}{2\sqrt{2}},$$

and it follows that ℓ_c is irrational. Thus $\ell_a\ell_b$ must be irrational, so ℓ_a and ℓ_b cannot both be rational. This completes the proof.

Note. Many readers noted that $a = 2kmn$, $b = k(m^2 - n^2)$ and $c = k(m^2 + n^2)$ where k, m, n are positive integers and m and n are relatively prime. Then

$$\ell_a = \frac{m^2-n^2}{m}\sqrt{m^2+n^2}, \quad \ell_b = \frac{2\sqrt{2}mn}{m+n}\sqrt{m^2+n^2}, \quad \text{and} \quad \ell_c = \frac{2\sqrt{2}mn(m+n)}{m-n},$$

so ℓ_c is irrational and at most one of ℓ_b and ℓ_c is rational.

Also solved by John Atkins and Herb Bailey, Roy Barbara (Lebanon), Michel Bataille (France), J. C. Binz (Switzerland), Jean Bogaert (Belgium), Pierre Bornsztein (France), Marc Brodie, John Christopher, Chip Curtis, M. N. Deshpande (India), Daniele Donini (Italy), Fejentalaltuka Szeged Problem Solving Group (Hungary), *Ovidiu Furdui, Marty Getz and Dixon Jones, Brian D. Ginsberg, John F. Goehl, D. Kipp Johnson, Ken Korbin, Victor Y. Kutsenok, Elias Lampakis (Greece), Peter W. Lindstrom, Francesco Marino (Italy), Rolf Richberg (Germany), Ralph Rush, Raul A. Simon (Chile), Helen Skala, Albert Stadler (Switzerland), H. T. Wang, Li Zhou, David Zhu, and the proposer.*

Cutting a Polygon into Rhombi **June 2002**

1650. *M. N. Deshpande, Nagpur, India*

Let $R(\theta)$ denote the rhombus with unit side and and a vertex angle of θ, and let $n \geq 2$ be a positive integer. Prove that a regular $4n$-gon of unit side can be tiled with collection of $n(2n-1)$ rhombi consisting of n copies of $R(\pi/2)$ and $2n$ copies of each of $R(\pi k/2n)$, $1 \leq k \leq n-1$.

Solution by Marty Getz and Dixon Jones, University of Alaska, Fairbanks, AK.

Arrange $2n-1$ copies of $R(\pi/2n)$ with vertices meeting at a point P, and so that all of the rhombi lie on or inside of an angle of measure $\pi(2n-1)/2n$ with vertex at P. Note that if two adjacent rhombi share an edge with vertices P and Q, then the other edges with vertex Q form an angle of measure $2\pi/2n$, and $2n-2$ such angles are formed. Add a second tier (as measured from P) of rhombi by nesting a copy of $R(2\pi/2n)$ in each of these angles. Continue this process, adding a third tier of $2n-3$ rhombi, a fourth tier of $2n-4$ rhombi, and so forth, until the $(2n-1)$st tier, consisting of one rhombus, is placed. The accompanying figure shows the resulting tiling for the case $n = 5$.

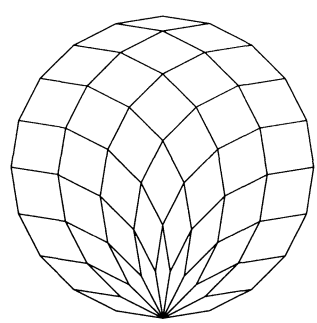


Figure 1 Tiling a regular 20-gon

We now show that the tiling described is indeed a tiling of a regular $4n$-gon and that the rhombi tiling it are the desired ones. As observed above, the first and second tiers consists of $2n-1$ copies of $R(\pi/2n)$ and $2n-2$ copies of $R(2\pi/2n)$, respectively. Now assume that $k\pi/2n$ and $(k+1)\pi/2n$ are the vertex angles (oriented towards P) for the kth and $(k+1)$st tiers, respectively. Then the vertex angle for the $(k+2)$nd tier is

$$2\pi - 2\left(\pi - \frac{(k+1)\pi}{2n}\right) - \frac{k\pi}{2n} = \frac{(k+2)\pi}{2n},$$

and $2n-(k+2)$ such angles are formed by adjacent pairs of rhombi in the $(k+1)$st tier. Thus, $2n-(k+2)$ copies of $R((k+2)\pi/2n)$ fit in these angles to make the $(k+2)$nd tier. Because $R(k\pi/2n) = R((2n-k)\pi/2n)$, there are $2n$ copies of each $R(k\pi/2n)$, $1 \leq k \leq n-1$, and n copies of $R(\pi/2)$ in this construction. The first tier contributes four outer edges to the tiling, and each subsequent tier contributes two outer edges. Thus the resulting figure is a $4n$-gon with each side of unit length. Because each pair of adjacent edges meets at an angle of

$$\frac{k\pi}{2n} + \left(\pi - \frac{(k+1)\pi}{2n}\right) = \frac{(2n-1)\pi}{2n},$$

it follows that the tiled $4n$-gon is regular.

Also solved by Daniele Donini (Italy), Fejentalaltuka Szeged Problem Solving Group (Hungary), Julien Grivaux (France), Elias Lampakis (Greece), Li Zhou, and the proposer.

Bounds on the Gamma Function **June 2002**

1651. *Proposed by Juan-Bosco Romero Márquez, Universidad de Valladolid, Valladolid, Spain*

Prove that for $x \geq 2$,

$$\left(\frac{x}{e}\right)^{x-1} \leq \Gamma(x) \leq \left(\frac{x}{2}\right)^{x-1},$$

where Γ is the gamma function.

Solution by Michel Bataille, Rouen, France.

Let $f(x) = (x-1)\ln(x/2) - \ln\Gamma(x)$ and $g(x) = \ln\Gamma(x) + (x-1) - (x-1)\ln x$. The inequality on the right will follow from (i) $f(x) \geq 0$, and the inequality on the left from (ii) $g(x) \geq 0$. We establish both of these inequalities.

It is well known that for $z \neq 0, -1, -2, \ldots$,

$$\Gamma(z) = \frac{e^{-\gamma z}}{z} \prod_{k=1}^{\infty} \left(\frac{k}{k+z}\right) e^{z/k},$$

where γ is Euler's constant. From this expression it is easy to show that

$$\frac{\Gamma'(x)}{\Gamma(x)} = -\gamma - \frac{1}{x} + \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{x+k}\right) \quad \text{and} \quad \frac{d}{dx}\frac{\Gamma'(x)}{\Gamma(x)} = \frac{1}{x^2} + \sum_{k=1}^{\infty} \frac{1}{(x+k)^2}, \qquad (*)$$

for $x > 0$.

(i) By $(*)$ we have

$$f'(x) = \ln\left(\frac{x}{2}\right) + \frac{x-1}{x} - \frac{\Gamma'(x)}{\Gamma(x)} \quad \text{and} \quad f''(x) = \frac{1}{x} - \sum_{k=1}^{\infty} \frac{1}{(x+k)^2}.$$

Because

$$\sum_{k=1}^{\infty} \frac{1}{(x+k)^2} \leq \int_0^{\infty} \frac{1}{(x+t)^2}\, dt = \frac{1}{x},$$

it follows that $f''(x) \geq 0$ and f' is increasing for $x > 0$. Next note that

$$f'(2) = \frac{1}{2} - \frac{\Gamma'(2)}{\Gamma(2)} = \frac{1}{2} - \left(-\gamma - \frac{1}{2} + \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{2+k}\right)\right) = \gamma - \frac{1}{2} > 0.$$

Thus $f'(x) \geq f'(2) > 0$ for $x \geq 2$, so f is increasing on $[2, \infty)$. The conclusion $f(x) \geq 0$ on this interval follows because $f(2) = -\ln\Gamma(2) = -\ln 1 = 0$.

(ii) By (∗),

$$g'(x) = \frac{\Gamma'(x)}{\Gamma(x)} + \frac{1}{x} - \ln x = -\gamma - \ln x + \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{x+k}\right).$$

For positive integer n, let $H_n = \sum_{k=1}^{n} \frac{1}{k}$. Because

$$\sum_{k=1}^{n} \frac{1}{x+k} \leq \int_0^n \frac{1}{x+t}\,dt = \ln(x+n) - \ln x,$$

we have

$$-\gamma + H_n - \ln n - \ln\left(1 + \frac{x}{n}\right) \leq -\gamma - \ln x + \sum_{k=1}^{n}\left(\frac{1}{k} - \frac{1}{x+k}\right).$$

Letting $n \to \infty$ and recalling that $\lim_{n\to\infty}(H_n - \gamma - \ln n) = 0$, we obtain $0 \leq g'(x)$. It follows that $g(x) \geq g(2) = 1 - \ln 2 > 0$. This completes the proof of (ii).

Also solved by Mihaly Bencze (Romania), Paul Bracken (Canada), Daniele Donini (Italy), Ovidiu Furdui, Julien Grivaux (France), Murray S. Klamkin (Canada), Elias Lampakis (Greece), Peter W. Lindstrom, Daniel A. Morales (Venezuela), Luis Moreno, Rolf Richberg (Germany), Li Zhou, and the proposer.

Some Geometric Inequalities **June 2002**

1652. *Proposed by Razvan A. Satnoianu, Oxford University, Oxford, United Kingdom.*

In triangle ABC, let r denote the radius of the inscribed circle, R the radius of the circumscribed circle, and p the semiperimeter. Prove the following inequalities, and show that in each case the constant on the right is the best possible:

(a) $\dfrac{R}{p} + \dfrac{p}{R} \geq 2.$

(b) $\dfrac{r}{p} + \dfrac{p}{r} \geq \dfrac{28\sqrt{3}}{9}.$

(c) $\dfrac{r}{p} + \dfrac{p}{r} \geq \dfrac{56}{31}\left(\dfrac{R}{p} + \dfrac{p}{R}\right).$

Solution by Achilleas Sinefakopoulos (student), Cornell University, Ithaca, NY.

It is well known that inequality (a) holds for all positive real numbers R and p. To show that the constant on the right is best possible, we need only show that there is a triangle with $R = p$. To this end, note that by the Intermediate Value Theorem there is a real number $\theta \in (0, \pi/2)$ with $\cos\theta(\sin\theta + 1) = \frac{1}{2}$. Consider the triangle ABC with $A = 2\theta$ and $B = C = \pi/2 - \theta$. Then by the law of sines

$$\frac{p}{R} = \sin A + \sin B + \sin C = \sin 2\theta + 2\cos\theta = 2\cos\theta(\sin\theta + 1) = 1,$$

showing that equality is possible.

Now let a, b, c be the lengths of the sides of ABC, and recall that

$$ab + bc + ca = p^2 + r^2 + 4Rr \quad \text{and} \quad abc = 4Rrp.$$

Thus

$$p^2 + r^2 + 4Rr = ab + bc + ca = \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c}\right) abc \geq \frac{9abc}{a+b+c} = \frac{9abc}{2p} = 18Rr,$$

so $p^2 + r^2 \geq 14Rr$. Hence

$$\frac{r}{p} + \frac{p}{r} = \frac{r^2 + p^2}{pr} \geq \frac{14Rr}{pr} = \frac{14R}{p}.$$

Thus, to prove inequalities (b) and (c), it suffices to show that

$$\frac{R}{p} \geq \frac{2\sqrt{3}}{9} \quad \text{and} \quad \frac{R}{p} \geq \frac{4}{31}\left(\frac{R}{p} + \frac{p}{R}\right),$$

respectively. However, both of these are equivalent to the well-known inequality $2p \leq 3\sqrt{3}R$, which follows from

$$\frac{p}{R} = \sin A + \sin B + \sin C \leq 3 \sin\left(\frac{\pi}{3}\right) = \frac{3\sqrt{3}}{2}.$$

When ABC is equilateral, the above inequalities become equalities, showing that the constants on the right of (b) and (c) are best possible.

Also solved by Herb Bailey, Roy Barbara (Lebanon), Michel Bataille (France), Jean Bogaert (Belgium), Pierre Bornsztein (France), Chip Curtis, Daniele Donini (Italy), Ovidiu Furdui, Julien Grivaux (France), Murray S. Klamkin (Canada), Ken Korbin, Elias Lampakis (Greece), Rolf Richberg (Germany), Raul A. Simon (Chile), Li Zhou, and the proposer.

Answers

Solutions to the Quickies from page 233.

A931. If the number Blaise recited was x^{10}, then $10^{40} \leq x^{10} < 10^{41}$, so $10^4 \leq x < 10^{4.1}$. The smallest possible value for x is $2 \cdot 5 \cdot 7 \cdot 11 \cdot 13 = 10010$, which falls in the desired range. The next smallest possibility is $2 \cdot 5 \cdot 7 \cdot 11 \cdot 17 = 13090$, which is greater than $10^{4.1}$. Therefore, Blaise worked out $(2 \cdot 5 \cdot 7 \cdot 11 \cdot 13)^{10}$. Our problem is to describe how she did this.

First observe that $7 \cdot 11 \cdot 13 = 1001$, and

$$1001^{10} = (1000 + 1)^{10} = 1000^{10} + 10 \cdot 1000^9 + 45 \cdot 1000^8 \cdots + 10 \cdot 1000 + 1,$$

with the coefficients 1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1 coming from the tenth row of Pascal's triangle, presumably still on the board from the lesson. Because none of these coefficients has more than three digits, Blaise can simply read out

$$1001^{10} = 001\,010\,045\,120\,210\,252\,210\,120\,045\,010\,001,$$

and append ten 0s to the end.

A932. Let $c = AP$ as in FIGURE 2. The altitudes H of $T = \triangle Q'OQ$ and H_1 of $T_1 = \triangle Q'AQ$ from O and A to PQ are parallel, so by similar triangles,

$$\frac{H_1}{H} = \frac{c}{a + c} = k.$$

It follows that the area of T_1 is k times the area of T because they have the same base QQ'. Thus T and T_1 will both have maximum area for the same secant PQ. It is clear that the maximum area for T is $a^2/2$ and is attained when $\angle Q'OQ = \pi/2$. The corresponding area of T_1 is $k(a^2/2) = a^2c/(2a + 2c)$.

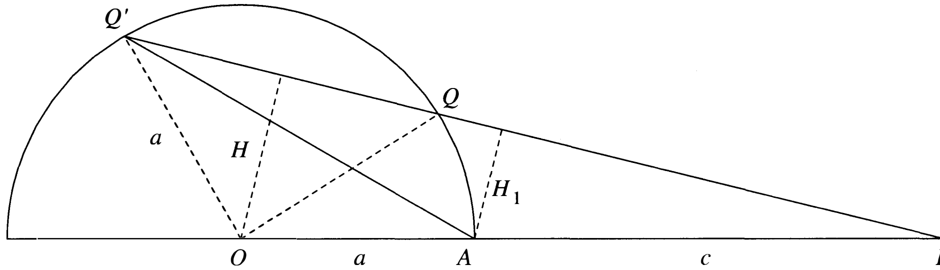


Figure 2

60 Years Ago in the MAGAZINE

Editorial comment by S. T. Sanders, Vol. **18**, No. 1, (Sept.–Oct., 1943):

Post-War Planning in Mathematics

Many teachers are nervously concerned over what may be the post-war status of school mathematics. The enormous expansion of the technical applications of the science under pressure of war has brought about a world-wide strengthening of mathematics in the school curriculum. Can this current academic primacy of mathematics be made permanent? Such is the question raised by those keenly mindful of the scant attention paid to this subject by the less recent curriculum makers.

A careful study of the matter should not discount the fact that in respect to mathematics, the war has served only to bring about greatly multiplied *uses* of mathematics a large proportion of which were already in existence. For, even in pre-war times there had been for many years a steadily growing public emphasis upon *applied* mathematics, rather than upon the logical or cultural aspects of the science.

In the light of this definite trend, a trend not rooted in any war, it could well be that the post-war school effort should first be directed to discovering the mathematical aids or needs of all the major peace-time industrial enterprises.

When they [the Delians] consulted him [Plato] on the problem set them by the Oracle, namely that of duplicating the cube, he replied, "It must be supposed, not that the god specially wished this problem solved, but that he would have the Greeks desist from war and wickedness and cultivate the Muses, so that, their passions being assuaged by philosophy and mathematics, they might live in innocent and mutually helpful intercourse with one another."

—from the Preface to Sir Thomas Heath's
A History of Greek Mathematics,
written during World War I.

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.

Jonsson, Bengt, Klassiskt matteproblem kan ha fått en lösnin [Classic math problem may have been solved], *Svenska Dagbladet* (23 February 2003), `http://www.svd.se/dynamiskt/Vetenskap/did_4943112.asp`.

Castro, Carlos, and Jorge Mahecha, Final steps toward a proof of the Riemann hypothesis (5 November 2002), `www.arxiv.org/format/hep-th/0208221`.

du Sautoy, Marcus, *The Music of the Primes: Searching to Solve the Greatest Mystery in Mathematics*, HarperCollins, 2003; 352 pp, $24.95. ISBN: 0–06621070–4.

Derbyshire, John, *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem*, Joseph Henry Press, 2003; $27.95. ISBN 0–30908549–7.

Sabbagh, Karl, *The Riemann Hypothesis: The Greatest Unsolved Problem in Mathematics*, Farrar Straus & Giroux, 2003; 304 pp, $25 (P). ISBN 0–37425007–3.

Maybe the Riemann Hypothesis (RH) has been proved, by Carlos Castro (Clark Atlanta University) and Jorge Mahecha (University of Antioquia, Colombia)! They follow a strategy of M. Pitkänen "based on the orthogonality relations between eigenfunctions of a non-Hermitian operator used in super-conformal transformations." Whether or not their proof stands up, the announcement could not come at a better time—no fewer than three books are just out to explain RH to a popular audience, one of them at this writing ranked 169th at Amazon. Moreover, an enormous bonanza of popular books about mathematics is in print. Apart from the difficulty of deciding which to recommend, I wonder whom these books reach—who buys and reads them? Mathematicians such as yourself? College libraries? School libraries? Public libraries, where the books languish next to 50-year-old arithmetic-made-simple books? Does the general public buy or borrow them?

Robinson, Sara, Russian reports he has solved a celebrated math problem, *New York Times* (15 April 2003) F3; `http:/www.nytimes.com/2003/04/15/science/15MATH.html`. Brodie, Josh, Perelman explains proof to famous math mystery, *Daily Princetonian* (17 April 2003), `http://www.dailyprincetonian.com/archives/2003/04/17/news/7979.shtml`. Johnson, George, A world of doughnuts and spheres, *New York Times* (20 April 2003) Section 4, p. 5; `http://www.nytimes.com/2003/04/20/weekinreview/20JOHN.html`.

Perelman, Grisha, The entropy formula for the Ricci flow and its geometric applications (12 November 2002), `www.arxiv.org/format/math.DG/0211159`; Ricci flow with surgery on three-manifolds (11 March 2003) `www.arxiv.org/format/math.DG/0303109`.

Maybe the Poincaré Conjecture (PC) has been proved, too, by Grigori Perelman (Steklov Institute)! In fact, Perelman's aim and claim is to have proved the more general geometrization conjecture for closed three-manifolds: William Thurston (UC–Davis, then at Princeton) had conjectured in the 1970s that there are only eight topologically distinct shapes for closed three-manifolds. PC itself says only that the three-sphere is the only bounded three-manifold with no holes. Perelman's proof uses a technique called the Ricci flow, invented by R.S. Hamilton (Columbia University); Perelman implements a program of Hamilton's for proving the geometrization conjecture.

Chui, Glennda, A good proof despite math goof, *San Jose Mercury News* (7 May 2003), `http://www.bayarea.com/mld/mercurynews/news/local/5805921.htm` . Small gaps between consecutive primes: Recent work of D. Goldston and C. Yildirim. `http://www.aimath.org/goldston_tech/` . On the error in Goldston and Yildirim's "Small gaps between consecutive primes," `http://aimath.org/primegaps/residueerror/` .

In early April, it looked as if there had been substantial progress toward the twin prime conjecture, by Daniel A. Goldston (San Jose State University) and Cem Y. Yildirim (Bogaziçi University, Istanbul)—but in late April, a gap emerged. The twin prime conjecture is that infinitely many pairs of primes differ by 2. On average, the spacing between primes near n is $\ln n$. Infinitely many pairs of primes are closer than one-half the average spacing; Goldston and Yildirim sought to extend that result to any fraction.

Havil, Julian, *Gamma: Exploring Euler's Constant*, Princeton University Press, 2003; xxiii + 266 pp, $29.95. ISBN 0–691–09983–9.

Well, in the past few years there have been popular books devoted to e, π, i, and 0; but I never imagined that there would be a book devoted solely to γ, the limit of the "discrepancy" involved in approximating a partial sum of the harmonic series by the natural logarithm:

$$\gamma = \lim_{n\to\infty}(H_n - \ln n) = \lim_{n\to\infty}\left(\frac{1}{1} + \frac{1}{2} + \cdots \frac{1}{n} - \ln n\right) \approx 0.577.$$

Actually, unlike some of those books on other constants, this book is not for a popular audience, since it is chock full of mathematical expressions involving sums, limits, and integrals, and it goes far beyond γ to explore the gamma function, the prime number theorem, and the Riemann Hypothesis (how could a book coming out this year be complete otherwise?). A background of a semester or two of calculus is really required.

Gottman, John M., James D. Murray, Catherine C. Swanson, Rebecca Tyson, Kristin R. Swanson, *The Mathematics of Marriage: Dynamic Nonlinear Models*, MIT Press, 2003; xvii + 402 pp, $42.95. ISBN 0–262–07226–2. Ellenberg, Jordan, Love by the numbers, *Slate* (16 April 2003), `http://slate.msn.com/id/2081484/` . Glenn, David, Every unhappy family has its own bilinear influence function, *Chronicle of Higher Education* (25 April 2003) A14; `http://chronicle.com/free/v49/i33/33a01401.htm` .

This book sets out toward the ambitious goal of providing "the foundation for a scientific theory of marital relations." The authors, who include a psychologist, a mathematician, and their students, set out to provide an underlying theory to explain their ability to predict the fate of a marriage (with 90% accuracy) from observing a single 15-minute interaction of a couple. They briefly review research on marriage, followed by 100 pp of introduction to calculus, differential equations, and catastrophe theory, before getting down to the modeling. The model itself is couched as a pair of difference equations that involve two threshold and four influence parameters for each spouse; those parameters can be estimated from scoring the 15-minute interaction of a couple. The authors' analysis focuses on steady states of the model and their stability. They go on to assess the value of the model; apply it to newlyweds, homosexual relationships, and parent-baby interaction; and use it to evaluate forms of couples therapy.

Fusaro, B.A., and P.C. Kenschaft (eds.), *Environmental Mathematics in the Classroom*, MAA, 2003; viii + 255 pp, $49.95 (P) (members, $35.95). ISBN 0–88385–714–6.

This book has 14 chapters that treat environmental problems of all kinds (with real data) using only high-school mathematics; each chapter includes exercises and references. A book such as this poses a new paradigm for a mathematics course to satisfy a distribution or quantitative requirement: Treat environmental problems, focus on mathematical modeling, apply mathematics that is mostly familiar, and introduce new mathematics on a "just in time" basis. Alumni of such a course would have no doubt about the usefulness and importance of mathematics in addressing vital problems.

NEWS AND LETTERS

43rd International Mathematical Olympiad

Glasgow, Scotland, United Kingdom

July 24 and 25, 2002

edited by Titu Andreescu and Zuming Feng

2002 Olympiad Results

The top twelve students on the 2002 USAMO were (in alphabetical order):

Steve Byrnes	West Roxbury, MA
Michael Hamburg	South Bend, IN
Neil Herriot	Palo Alto, CA
Daniel Kane	Madison, WI
Anders Kaseorg	Charlotte, NC
Ricky Liu	Newton, MA
Tiankai Liu	Saratoga, CA
Po-Ru Loh	Madison, WI
Alison Miller	Niskayuna, NY
Gregory Price	Falls Church, VA
Tong-ke Xue	Chandler, AZ
Inna Zakharevich	Palo Alto, CA

Daniel Kane, Ricky Liu, Tiankai Liu, Po-Ru Loh, and Inna Zakharevich, all with perfect scores, tied for first on the USAMO. They shared college scholarships of $30,000 provided by the Akamai Foundation. The Clay Mathematics Institute (CMI) award, for a solution of outstanding elegance, and carrying a $1,000 cash prize, was presented to Michael Hamburg, for the second year in a row, for his solution to USAMO Problem 6.

The USA team members were chosen according to their combined performance on the 31st annual USAMO and the Team Selection Test that took place at the Mathematics Olympiad Summer Program (MOSP) held at the University of Nebraska-Lincoln, June 18–July 13, 2002. Members of the USA team at the 2002 IMO (Glasgow, United Kingdom) were Daniel Kane, Anders Kaseorg, Ricky Liu, Tiankai Liu, Po-Ru Loh, and Tong-ke Xue. Titu Andreescu (Director of the American Mathematics Competitions) and Zuming Feng (Phillips Exeter Academy) served as team leader and deputy leader, respectively. The team was also accompanied by Reid Barton (Massachusetts Institute of Technology), Steven Dunbar (University of Nebraska-Lincoln) and Zvezdelina Stankova (Mills College), as the observers of the team leader and deputy leader.

At the 2002 IMO, gold medals were awarded to students scoring between 29 and 42 points (there were three perfect papers on this very difficult exam), silver medals to students scoring between 23 and 28 points, and bronze medals to students scoring between 14 and 22 points. Loh's 36 tied for fourth place overall. The team's individual performances were as follows:

Kane	West HS	GOLD Medallist
Kaseorg	Home-schooled	SILVER Medallist
R. Liu	Newton South HS	GOLD Medallist
T. Liu	Phillips Exeter Academy	GOLD Medallist
Loh	James Madison Memorial HS	GOLD Medallist
Xue	Hamilton HS	Honorable Mention

In terms of total score (out of a maximum of 252), the highest ranking of the 84 participating teams were as follows:

China	212	Taiwan	161
Russia	204	Romania	157
USA	171	India	156
Bulgaria	167	Germany	144
Vietnam	166	Iran	143
Korea	163	Canada	142

Note: For interested readers, the editors recommend the *USA and International Mathematical Olympiads 2002*. There many of the problems are presented together with a collection of remarkable solutions developed by the examination committees, contestants, and experts, during or after the contests.

Problems

1. Let n be a positive integer. Let T be the set of points (x, y) in the plane where x and y are nonnegative integers and $x + y < n$. Each point of T is colored red or blue. If a point (x, y) is red, then so are all points (x', y') of T with both $x' \leq x$ and $y' \leq y$. Define an X-set to be a set of n blue points having distinct x-coordinates, and a Y-set to be a set of n blue points having distinct y-coordinates. Prove that the number of X-sets is equal to the number of Y-sets.
2. Let BC be a diameter of circle ω with center O. Let A be a point of circle ω such that $0 < \angle AOB < 120°$. Let D be the midpoint of arc $\widehat{AB}$ not containing C. Line ℓ passes through O and is parallel to line AD. Line ℓ intersects line AC at J. The perpendicular bisector of segment OA intersects circle ω at E and F. Prove that J is the incenter of triangle CEF.
3. Find all pairs of integers $m, n \geq 3$ such that there exist infinitely many positive integers a for which

$$\frac{a^m + a - 1}{a^n + a^2 - 1}$$

is an integer.
4. Let n be an integer with $n \geq 2$. Let $1 = d_1 < d_2 < \cdots < d_k = n$ be all the divisors of n.
    (a) Prove that $D_n = d_1d_2 + d_2d_3 + \cdots + d_{k-1}d_k < n^2$;
    (b) Determine all n such that D is a divisor of n^2.
5. Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$(f(x) + f(z))(f(y) + f(t)) = f(xy - zt) + f(xt + yz)$$

for all real numbers x, y, z, t.

6. Let n be an integer with $n \geq 3$. Let $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ be unit circles in the plane, with centers $O_1, O_2, \ldots, O_n$, respectively. If no line meets more than two of the circles, prove that

$$\sum_{1 \leq i < j \leq n} \frac{1}{O_i O_j} \leq \frac{(n-1)\pi}{4}.$$

Solutions

1. For $0 \leq i, j \leq n-1$, let a_i denote the number of blue points with x-coordinate i, and let b_j denote the number of blue points with x-coordinate j. We need to show that two products are equal:

$$a_0 a_1 \cdots a_{n-1} = b_0 b_1 \cdots b_{n-1}. \qquad (\dagger)$$

If $a_i = 0$ for some i, then $(i, n-1-i)$ is red, and so are all the points to the left of $(i, n-1-i)$, implying that $b_{n-1-i} = 0$. In this case, $(\dagger)$ is true. Now we assume that all $a_i > 0$. Then all the points on line $x + y = n - 1$ are blue. We assign a weight of 1 to each of them. Then we assign a weight of $2/1$ to each blue point on line $x + y = n - 2$, a weight of $3/2$ to each blue point on line $x + y = n - 3$, and so on. In particular, each point on line $x + y = n - k$ $(1 < k \leq n)$ is assigned a weight of $k/(k-1)$. In column j, points $(j, n-1-j)$, $(j, n-2-j)$, $\ldots$, $(j, n - a_k - j)$ are the blue points. Those points have weights $1, 2/1, 3/2$, $\ldots$, a_k/a_{k-1}, so the product of the weights in column k is a_k. Because each blue point belongs to exactly one column, the product of all the weights of blue points is $a_0 a_1 \ldots a_{n-1}$. In the same way, we can show that the product of all the weights of blue points is $b_0 b_1 \ldots b_{n-1}$. Therefore $(\dagger)$ is true.

2. One of the hardest part of solving this problem is to use the condition $0^\circ < \angle AOB < 120^\circ$ effectively to deal with the configuration of this problem. Let P be the midpoint of segment OA, and let ℓ' be the perpendicular bisector of segment OA. Without loss of generality, let E be the intersection of ℓ' and ω on the same side of line OA as B, and let F be the intersection on the same side as C. Note that we can make this assumption because $180^\circ > \angle AOB > 0^\circ$, so A does not coincide with B or C. Observe that $AE = OE = OA$, so triangle AOE is equilateral. Similarly, triangle AOF is equilateral. Thus $\angle FOB = \angle AOB + 60^\circ < 180^\circ$. Thus, F lies on minor arc $\widehat{AC}$. Note also that $\angle AOD = \frac{1}{2}\angle AOB < 60^\circ = \angle AOE$. Hence C and D are on the opposite of sides line EF.

   Because arcs $\widehat{AE}$ and $\widehat{AF}$ both measure 60°, segment CJ is the internal angle bisector of $\angle ECF$. Since, $\angle DOB = \frac{1}{2}\angle AOB = \angle ACB$, $OD \parallel AC$. On the other hand, we are given $DA \parallel OJ$. Thus, quadrilateral $AJOD$ is a parallelogram with center P. Note also that $DEJF$ is a parallelogram as P is the midpoint of both of the segments EF and DJ. Therefore D and J are on the opposite side of line EF. Consequently, J lies inside of triangle CEF.

   We now show that J also lies on the internal angle bisector of $\angle FEC$. Since E and F are reflections of each other about P, it follows by reflection through P that $\angle FEJ = \angle DFE$. It remains to show that $\angle DFE = \frac{1}{2}\angle FEC$, which is equivalent to showing that $\angle DOE = \frac{1}{2}\angle FOC$. Indeed,

$$\begin{aligned}\angle DOE &= \angle AOE - \angle AOD = 60^\circ - \tfrac{1}{2}\angle AOB = \tfrac{1}{2}(180^\circ - (\angle AOB + 60^\circ)) \\ &= \tfrac{1}{2}(180^\circ - (\angle AOB + \angle FOA)) = \tfrac{1}{2}\angle FOC,\end{aligned}$$

   as desired.

3. We assume that reader can prove the following result: If $f(x)$ and $g(x)$ are rational polynomials with rational coefficients such that $g(x) \neq 0$ and there exist infinitely many positive integers a such that $f(a)/g(a)$ is an integer, then $f(x)$ is divisible by $g(x)$, that is, there is a polynomial $p(x)$ with rational coefficients such that $f(x) = g(x)p(x)$.

   The pair $(m, n) = (5, 3)$ is the unique solution. First verify that

$$\frac{a^5 + a - 1}{a^3 + a^2 - 1} = a^2 - a + 1,$$

   which is an integer whenever a is an integer. Now we will prove that $(m, n) = (5, 3)$ is the only possible solution. By the result stated above, we have

$$(x^n + x^2 - 1) \mid (x^m + x - 1). \tag{1}$$

   This implies that $n \leq m$, and that

$$\begin{aligned}(x^n + x^2 - 1) \mid &[(x + 1)(x^m + x - 1) - (x^n + x^2 - 1)] \\ &= x^{m+1} + x^m - x^n = x^n(x^{m-n+1} + x^{m-n} - 1).\end{aligned}$$

   Thus

$$(x^n + x^2 - 1) \mid (x^{m-n+1} + x^{m-n} - 1),$$

   and consequently $n \leq m - n + 1$ and $2 \leq n - 1 \leq m - n$. Because $0^n + 0^2 - 1 = -1$ and $1^n + 1^2 - 1 = 1$, there exists α in the interval $(0, 1)$ such that $\alpha^n + \alpha^2 - 1 = 0$. Because $n \leq m - n + 1$ and $2 \leq m - n$, we have $\alpha^n \geq \alpha^{m-n+1}$ and $\alpha^2 \geq \alpha^{m-n}$, so $0 \geq \alpha^{m-n+1} + \alpha^{m-n} - 1$. Equality must hold here because any root of $x^n + x^2 - 1$ must also be a root of $x^{m-n+1} + x^{m-n} - 1$, so we must have $n = m - n + 1$ and $2 = m - n$. Thus, $m = 5$ and $n = 3$, as claimed.

4. Note that if d is a divisor of n, so is n/d. Hence

$$\begin{aligned}\frac{D_n}{n^2} &= \frac{1}{n^2}\left(\frac{n}{d_1} \cdot \frac{n}{d_2} + \frac{n}{d_2} \cdot \frac{n}{d_3} + \cdots + \frac{n}{d_{k-1}} \cdot \frac{n}{d_k}\right) \\ &= \frac{1}{d_k} \cdot \frac{1}{d_{k-1}} + \frac{1}{d_{k-1}} \cdot \frac{1}{d_{k-2}} + \cdots + \frac{1}{d_2} \cdot \frac{1}{d_1} \\ &\leq \left(\frac{1}{d_{k-1}} - \frac{1}{d_k}\right) + \left(\frac{1}{d_{k-2}} - \frac{1}{d_{k-1}}\right) + \cdots + \left(\frac{1}{d_1} - \frac{1}{d_2}\right) \\ &= \frac{1}{d_1} - \frac{1}{d_k} < 1,\end{aligned}$$

   that is, $D_n < n^2$, as desired.

   Note also that $d_2 = p$ and $d_{k-1} = n/p$, where p is the least prime divisor of n. If n is a a prime, then $k = 2$ and $D_n = p$, which divides $n^2 = p^2$.

   If n is composite then $k > 2$, and $D_n > d_{k-1}d_k = n^2/p$, so $n^2/D_n < p$. If such an D_n were a divisor of n^2 then also n^2/D_n would be a divisor of n^2. But then $1 < n^2/D_n < p$, which is impossible as p is the least prime divisor of n^2.

   Therefore, D_n is a divisor of n if and only if n is prime.

5. The answers are $f(x) = 0$ for all $x \in \mathbb{R}$, $f(x) = 1/2$ for all $x \in \mathbb{R}$, and $f(x) = x^2$ for all $x \in \mathbb{R}$. These functions work because $(0+0)(0+0) = 0 = 0+0$, $(1/2+1/2)(1/2+1/2) = 1 = 1/2+1/2$, and

$$(x^2+z^2)(y^2+t^2) = |x+zi|^2|y+yi|^2 = |(x+zi)(y+ti)|^2$$
$$= |(xy-zt)+(xt+yz)i|^2 = (xy-zt)^2+(xt+yz)^2,$$

where $i^2 = -1$. There are at least two approaches to prove they are these only functions that satisfy the given conditions. The first approach makes a connection with complex numbers, in the light of the above identity. The second approach uses common skills of solving functional equation problems. We present only the first approach.

For a complex number $\omega = a + bi$, where $a, b \in \mathbb{R}$, let $\mathrm{Re}(\omega) = a$ and $\mathrm{Im}(\omega) = b$. In other words, $\mathrm{Re}(\omega)$ and $\mathrm{Im}(\omega)$ denote the real and imaginary part of ω. We define a function $g : \mathbb{C} \to \mathbb{R}$ such that

$$g(\omega) = f(\mathrm{Re}(\omega)) + f(\mathrm{Im}(\omega)).$$

For complex numbers $c = x + zi$ and $d = y + ti$, where $x, y, z, t \in \mathbb{R}$, we have $cd = (xy - zt) + (xt + yz)i$ and

$$g(c)g(d) = [f(x)+f(z)][f(y)+f(t)] = f(xy-zt)+f(xz+yt)$$
$$= g((xy-zt)+(xz+yt)i) = g(cd),$$

that is, g is multiplicative.

Hence $g(0)g(0) = g(0)$ and $g(1)g(1) = g(1)$, implying that $g(0), g(1) \in \{0, 1\}$.

If $g(0) = 1$, then $g(0) = f(0) + f(0) = 1$ and $f(0) = 1/2$. For $r \in \mathbb{R}$, $g(r)g(0) = g(0) = 1$, so $1 = g(r) = f(r) + f(0)$. Consequently, $f(r) = 1/2$ for all $r \in \mathbb{R}$.

If $g(0) = 0$, then $f(0) + f(0) = g(0) = 0$, so $f(0) = 0$. For $r \in \mathbb{R}$, $g(r) = f(r) + f(0) = f(r)$. Hence

$$g(r) = f(r) \quad \text{for all } r \in \mathbb{R}. \tag{1}$$

If $g(1) = 0$, then $f(r) = g(r) = g(1)g(r) = 0$, implying that $f(r) = 0$ for all $r \in \mathbb{R}$.

Now we are left to consider the case $g(0) = 0$ and $g(1) = 1$. For $\omega \in \mathbb{C}$, there exists a complex number ω' such that $\omega'^2 = \omega$. By (1),

$$g(\omega) = g(\omega')^2 \geq 0 \quad \text{for all } \omega \in \mathbb{C}. \tag{2}$$

Note that for $c, d \in \mathbb{R}$ and $\omega = c + di$,

$$g(\omega) = f(c) + f(d) = g(i\overline{\omega}),$$

where $\overline{\omega} = c - di$ is the conjugate of ω. Hence

$$g(\omega)^2 = g(\omega)g(\omega) = g(\omega)g(i\overline{\omega}) = g(i\omega\overline{\omega}) = g(i|\omega|^2).$$

Hence by (2), $g(\omega) = \sqrt{g(i|\omega|^2)}$. Therefore, for $\omega_1, \omega_2 \in \mathbb{C}$,

$$g(\omega_1) = g(\omega_2) \quad \text{if} \quad |\omega_1| = |\omega_2|. \tag{3}$$

We now show that $g(\sqrt{n}) = n$ for all positive integers n. We use induction on n. For $n = 1$, our claim is true because $g(\sqrt{1}) = g(1) = 1$. Assume that $g(\sqrt{n}) = n$ for some positive integer n. By (3) and then by (1), We have

$$g(\sqrt{n+1}) = g(\sqrt{n} + i) = f(\sqrt{n}) + f(1) = g(\sqrt{n}) + g(1) = n + 1,$$

which completes our induction.

We next show that $g(\sqrt{q}) = q^2$ for all positive rational numbers q. Indeed, we can write $|q| = e/f$ where e and f are positive integers. Then by (3),

$$g(q) = g(|q|) = g\left(\frac{e}{f}\right) = g(e)g\left(\frac{1}{f}\right) = g(e) \cdot \frac{g(1)}{g(f)} = \frac{e^2}{f^2} = q^2,$$

as claimed.

We now show that for $\omega_1, \omega_2 \in \mathbb{C}$, if $|\omega_1| \geq |\omega_2|$ then $g(\omega_1) \geq g(\omega_2)$. Indeed by (3), by (1), and then by (2), we have

$$\begin{aligned} g(\omega_1) &= g(|\omega_2| + i\sqrt{|\omega_1|^2 - |\omega_2|^2}) = f(|\omega_2|) + f(\sqrt{|\omega_1|^2 - |\omega_2|^2}) \\ &= g(|\omega_2|) + g(\sqrt{|\omega_1|^2 - |\omega_2|^2}) \geq g(|\omega_2|) = g(\omega_2). \end{aligned}$$

as claimed.

Finally, we claim that $g(\omega) = |\omega|^2$ for all $\omega \in \mathbb{C}$. For the sake of contradiction, we assume that there is $\omega \in \mathbb{C}$ such that $g(\omega) \neq |\omega|^2$. Because $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists $q \in \mathbb{Q}$ such that q is strictly between $|\omega|$ and $\sqrt{g(\omega)}$. If $|\omega| > q > \sqrt{g(z\omega)}$, then $g(\omega) \geq g(q) = q^2$ by previous two claims. Hence $\sqrt{g(\omega)} \geq q$, which is a contradiction. If $|\omega| < q < \sqrt{g(\omega)}$, then $g(\omega) \leq g(q) = q^2$ by previous two claims. Hence $\sqrt{g(\omega)} \leq q$, which is also a contradiction. Therefore our assumption was wrong and $g(\omega) = |\omega|^2$ for all $z \in \mathbb{C}$. By (1), $f(x) = g(x) = x^2$ for all $x \in \mathbb{R}$.

Therefore the only solutions are $f(x) = 0$, $f(x) = 1/2$, and $f(x) = x^2$.

6. For each circle Γ_i, consider the set of tangent lines to Γ_i. By the given condition, each tangent line can pass through at most one other circle. The points of Γ_i for which the tangents at those points do indeed intersect another circle form arcs on Γ_i. We relate the sum of the lengths of these arcs, summed over all the circles, to $\sum 1/O_iO_j$.

For any pair of circles Γ_i, Γ_j, there are four arcs, two on each circle, for which the tangents at points on the arcs intersect the other circle. These arcs are determined by the common internal and external tangents to Γ_i and Γ_j. For example, say T is the point of tangency of one of the internal tangents with Γ_i, and S is the point of tangency of Γ_i with the external tangent on the same side of O_iO_j as T. Then arc $\widehat{ST}$ is one of the four arcs, and the other three are congruent to it. Let P be the midpoint of segment O_iO_j, and let $\alpha = \angle O_iPT = \angle SO_iT$. Because Γ_i has unit radius, arc $\widehat{ST}$ has length α. On the other hand, considering right triangle O_iTP, we have $\sin\alpha = 2/O_iO_j$. Because $\sin\alpha < \alpha$, $2/O_iO_j$ is less than the length of arc $\widehat{ST}$. Multiplying this by four, the total contribution of Γ_i, Γ_j to the sum of all the arc lengths exceeds $8/O_iO_j$.

Next we find a bound on the sum of the arc lengths. As noted earlier, no two arcs may intersect. Thus, we immediately see that the sum of the arc lengths is at most the sum of the circumferences of all the circles, $2n\pi$. We can improve this bound by considering the convex hull of all the Γ_i. For any point on the boundary of the convex hull and on a circle Γ_i, the tangent to Γ_i at that point cannot intersect another

(*continued on p. 231*)

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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