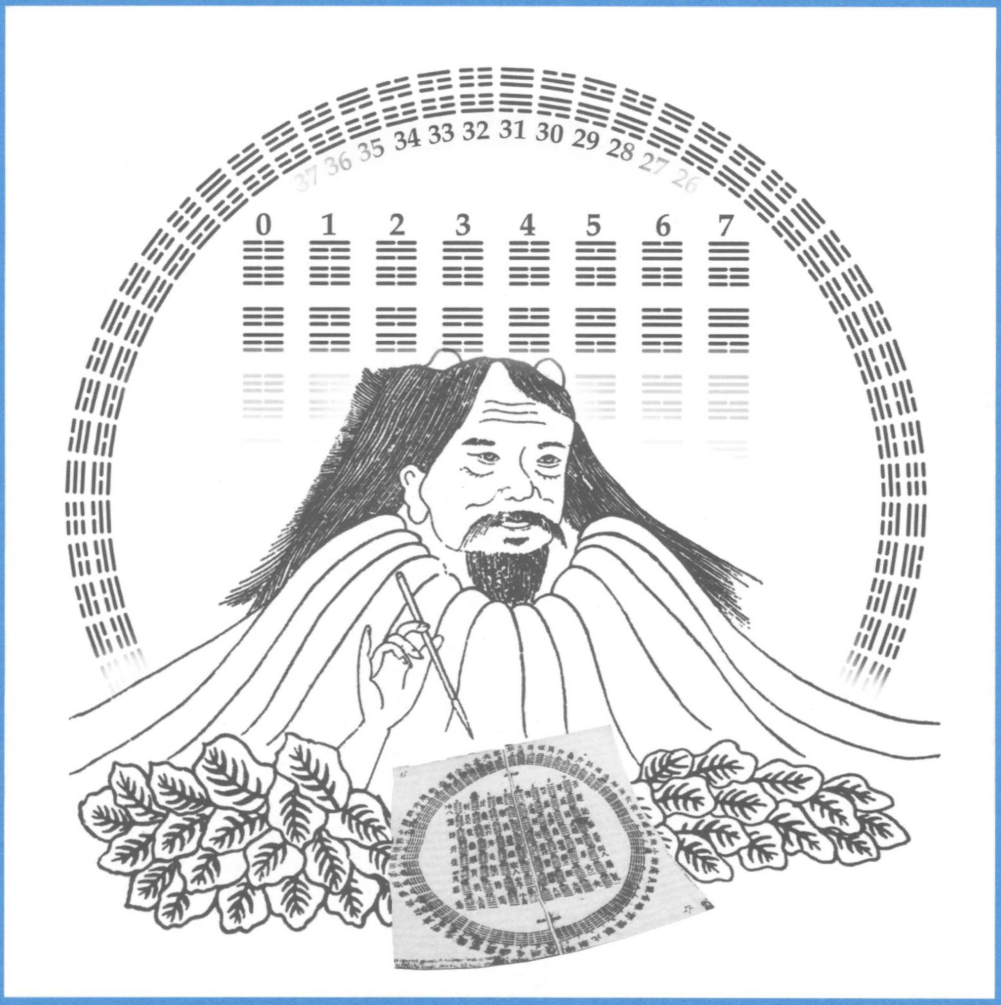
Vol. 76, No. 4, October 2003

MATHEMATICS MAGAZINE



Fuxi and the *Yijing*

- A Julia Set That Is Everything
- An Algorithm to Solve the Frobenius Problem
- Leibniz, the *Yijing*, and the Religious Conversion of the Chinese

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, *Fuxi and the Yijing*, by Anh Pham and Jason Challas. As Frank Swetz tells us in this issue, Fuxi, said to be the first emperor of China, is credited with first unveiling the *bagua*, the trigrams that were later stacked in pairs to form the hexagrams of the *Yijing*. Could Fuxi have known that the hexagrams depict a way to count from 0 to 63 in binary? Leibniz may have thought so. Fuxi is depicted here painting the "Natural Order" arrangement of the hexagrams, not the one commonly used. The hexagrams arching above Fuxi's head appear in numerical order, unlike the historical "Natural Order" diagram, where they run from 0 to 32 up the right-hand side, and 33 to 63 up the left.

Jason Challas lectures on zeroes and ones and computer art at Santa Clara University, where Ahn Pham is a student.

AUTHORS

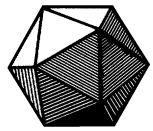
Julia A. Barnes is an Associate Professor of Mathematics at Western Carolina University. She received her Ph.D. from the University of North Carolina at Chapel Hill and her B.S. from the University of Central Florida. Her mathematical interests include complex dynamical systems and ergodic theory, and her cats are named Chaos and Fractal.

Lorelei M. Koss is an Assistant Professor of Mathematics at Dickinson College. She received her Ph.D. from the University of North Carolina at Chapel Hill and her B.A. and M.A. from Columbia University. Her mathematical interests include complex dynamical systems and ergodic theory. She and her coauthor, Julia Barnes, are "mathematical sisters" since they had the same graduate advisor.

Robert W. Owens received his B.S. from Santa Clara University and his Ph.D. from Michigan State University in 1975. His dissertation in approximation theory was completed under the supervision of V. P. Sreedharan. Except for brief leaves, he has been at Lewis and Clark College since 1975. In addition to approximation theory, his mathematical interests include numerical analysis, optimization, and the history and philosophy of mathematics. Work on his paper began with a question concerning Chicken McNuggets posed during a dinner conversation. His interest in travel has taken him many places, including Kenya and India where he has led semester-long overseas study programs for Lewis and Clark College.

Frank J. Swetz is Professor Emeritus of Mathematics and Education at Pennsylvania State University. He taught at the University's Harrisburg upper division campus, where he was instrumental in the founding and development of the Mathematical Sciences Program of study. His interest in "humanizing" the teaching of mathematics led him into research on the history of mathematics and ethnomathematics. He has been particularly active in documenting the history of Chinese mathematics and has had a prolonged involvement in mathematics education development in Southeast Asia. Dr. Swetz's most recent book is *Legacy of the Luoshu* (Open Court, 2001), a mathematical and cultural history of the magic square of order three. It is through researching this book that he became aware of Leibniz' entanglement in Chinese classical beliefs.

Vol. 76, No. 4, October 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

ARTICLES

A Julia Set That Is Everything

JULIA BARNES
Western Carolina University
Cullowhee, NC 28723
jbarnes@email.wcu.edu

LORELEI KOSS
Dickinson College
Carlisle, PA 17013
koss@dickinson.edu

We see unpredictable behavior around us every day: in the way weather patterns change, stock markets fluctuate, or wildfire spreads. We can also observe chaos in seemingly simple mathematical functions, and many recent undergraduate dynamical systems textbooks [**5, 10, 11, 15**] discuss chaotic systems. These texts all include chapters covering Julia sets, the part of the domain where a complex function behaves chaotically.

A Julia set is usually an intricate and beautiful object, and images of Julia sets can be found on posters, book covers, T-shirts, screen savers, and web pages. Many books [**10, 11, 15**] focus on Julia sets of complex polynomials of degree 2 or 3, and even the casual reader will observe from the pictures that all of these Julia sets appear to have area zero. For polynomials, it turns out that there must always be some large region where the function has very predictable behavior.

What if we examine other types of functions? Is there a complex function whose Julia set is everything? That is, does there exist a complex function that acts chaotically on its entire domain? The picture of such an example would be entirely black.

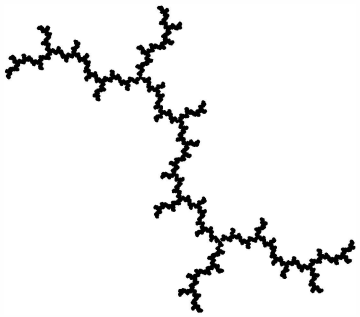


Figure 1 Julia set of $P_i(z) = z^2 + i$

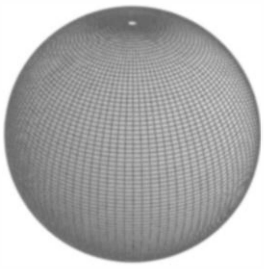


Figure 2 Julia set of $R(z) = \dfrac{(z^2+1)^2}{4z(z^2-1)}$

Làttes [**17**] and Böettcher [**7**] independently found an example in the early 1900s of a rational map whose Julia set happens to be the whole Riemann sphere. Although the construction is straightforward, the original papers are relatively inaccessible and the example is not well known. To explain it, we will generalize some ideas about iterat-

ing polynomials in the plane to iterating rational functions. Readers who are familiar with the background on Julia sets of polynomials and chaos can skip ahead to where we discuss the Julia sets of rational functions or further on to the construction of the Làttes/Böettcher example.

Julia set basics

In order to talk about chaos and the Julia set of a function, we need to analyze what happens to different points in its domain when we apply the function repeatedly. Suppose that $f(x)$ is the function we wish to study and x_0 is a point in the domain. We say that the first *iterate* of x_0 is $f(x_0)$, the second iterate is $f^2(x_0) = f(f(x_0))$, the third iterate is $f^3(x_0) = f(f(f(x_0)))$, etc. This requires that the function composition be defined, which is trivial for polynomials.

The *orbit* of x_0 under the function f is then the sequence of iterates, that is, $\{x_0, f(x_0), f^2(x_0), \ldots\}$. For example, if $f(x) = x^2$, then the orbit of 0 is $\{0, 0, 0, 0, \ldots\}$, and the orbit of $1/2$ is $\{1/2, 1/4, 1/16, \ldots\}$. We say that a point x_0 is *periodic* if there is a positive integer n such that $f^n(x_0) = x_0$.

The most common Julia sets discussed in undergraduate textbooks are generated from the family of complex polynomial functions, $P_c : \mathbb{C} \to \mathbb{C}$, where $P_c(z) = z^2 + c$ and c is a complex number. For *this* family of functions, the *filled Julia set* K_c is defined to be the collection of points whose orbits are bounded. That is, K_c is the set of points z for which there is a number B with $|P_c^n(z)| \leq B$ for all n (where $|.|$ represents the Euclidean distance from the origin in the complex plane). The *Julia set*, denoted $J(P_c)$, is then the boundary of K_c. We note that later in the paper we present an alternative definition of the Julia set that will apply to a broader class of functions.

Consider the function $P_0(z) = z^2$. Recall that any complex number z can be written in polar coordinates as $z = re^{i\theta}$. Then $P_0(re^{i\theta}) = (re^{i\theta})^2 = r^2e^{2i\theta}$; in other words, P_0 will square $|z|$ and double the angle θ. Let z_0 be a point inside the closed unit disk. Since $|z_0| \leq 1$, the orbit of z_0 is bounded by 1. Consequently, all points within the closed unit disk lie in K_0. What about points that are not in the closed unit disk? Let z_1 be a point that is outside the unit disk. Since $|z_1| > 1$, $|P_0^n(z_1)|$ will increase without bound as n goes to infinity. Therefore, no point outside of the unit disk can be in K_0. That means that K_0 is exactly the closed unit disk. The Julia set of P_0 is the boundary of K_0, which is simply the unit circle.

Most Julia sets are more elaborate than the Julia set for P_0 and cannot be determined by hand. Luckily, it is relatively easy to program a computer to generate Julia sets for polynomials. One method of programming uses the *Escape Criterion*: the computer takes sample points from the domain and checks whether a predetermined number of iterates remain within a given bound. If the program allows for enough iterations and large enough bounds, the computer can produce a fairly good approximation of the shape of both the filled Julia set and its boundary, the Julia set. Devaney [**10**] gives details on this procedure.

Another method of drawing a Julia set is the *Backward Iteration Algorithm*, which we used to draw the Julia sets in this paper. This method is based on the fact that the inverse images of any point (with at most two exceptions) accumulate at every point of the Julia set. Given a z_0, the Backward Iteration Algorithm plots randomly selected preimages $P^{-n}(z_0)$. After discarding roughly the first 50 points, plotting a large number of inverse iterates gives us a reasonable picture of the Julia set. Devaney [**11**] provides a more detailed explanation of the Backward Iteration Algorithm.

To understand why a Julia set is the region where chaotic behavior occurs, we need to provide a precise mathematical definition of chaos. The idea of the Julia set being

the region where chaotic behavior occurs will motivate our definition of Julia sets of rational maps.

Chaotic dynamical systems

What does it mean for a dynamical system to be chaotic? We assume that X is a metric space with metric d, though little will be lost by taking X to be the plane with its standard distance function. A subset Y of X is *dense* if it pervades X in a specific sense: for any $x \in X$ and any small number ϵ, there is a point $y \in Y$ that is within ϵ of x. We say that a function $F: X \to X$ has *dense periodic points* if the set of all periodic points is dense in X. A dynamical system is *transitive* if for any pair of points x and y and any $\epsilon > 0$ there is a third point z within ϵ of x whose orbit comes within ϵ of y. This means that, given any two points, we can find a point that is arbitrarily close to x whose orbit comes arbitrarily close to y.

How do these definitions apply to our previous example $P_0(z) = z^2$, whose Julia set is the unit circle? Are the periodic points of P_0 dense on the unit circle? Points on the unit circle have the form $e^{i\theta}$, so $P_0(e^{i\theta}) = e^{i2\theta}$ also lies on the unit circle. Any arc on the unit circle can be described as the collection of θ such that $\theta_1 < \theta < \theta_2$. In order to show that P_0 has dense periodic points on the unit circle, we would need to find a periodic point within that arc. The fixed points of P_0^n are the $2^{(n-1)}$-st roots of unity. These roots are evenly spaced around the unit circle, so for large enough n, one of them must lie in the given arc. To verify transitivity, we pick an n large enough so that P_0^n maps the arc $\theta_1 < \theta < \theta_2$ onto the entire circle. This is possible since P_0 doubles all angles under iteration.

At first glance, it is hard to see how transitivity or dense periodic points have anything to do with chaotic behavior. Intuitively, we view behavior as chaotic when two initial conditions that begin very close together can produce vastly different outcomes. We can formalize this idea mathematically by requiring that our dynamical system have sensitive dependence on initial conditions. A dynamical system has *sensitive dependence on initial conditions* if there is a $\beta > 0$ depending only on F such that for any x and any $\epsilon > 0$ there is a y within ϵ of x and a k such that $d(F^k(x), F^k(y)) > \beta$. In other words, F depends sensitively on initial conditions if for any x and any disk that we choose around x, we can always find a y in this disk whose orbit eventually separates from the orbit of x by a distance of at least β.

We are now ready to give a formal definition of chaos. Devaney **[10]** gives the following definition of a chaotic dynamical system.

DEFINITION. *We say that a dynamical system $F: X \to X$ is chaotic when all three conditions are satisfied:*

1. *F has dense periodic points,*
2. *F is transitive, and*
3. *F has sensitive dependence on initial conditions.*

Surprisingly, the property of sensitive dependence actually follows from the first two conditions in the definition, so a weaker definition is possible. This amazing result is proven independently in papers by Banks, et al. **[2]** and Glasner and Weiss **[14]**. We state it here for clarity.

THEOREM **[2, 14]**. *If $F: X \to X$ is transitive and has dense periodic points, then F has sensitive dependence on initial conditions.*

Since we have previously shown that P_0 is transitive and has dense periodic points, we can use Theorem 2 to conclude that P_0 is chaotic on its Julia set, the unit circle. Using Theorem 2 to prove that a general polynomial P_c acts chaotically on its Julia set is a little more complicated. Devaney [**10**] provides proofs of some more challenging examples.

Notice, however, that P_c is not chaotic on its entire domain (the complex plane). In fact, for any polynomial $P(z)$ we can find some number B with the property that for any z with $|z| > B$, we have $\lim_{n\to\infty} P^n(z) = \infty$. In addition, when $|z| > B$ we have $|P^{n+1}(z)| > |P^n(z)|$ for all n. This prevents transitivity from occurring on the set $\{z : |z| > B\}$. Therefore, we must look beyond the family of complex polynomials to find a function that is chaotic on its entire domain.

Rational maps of the sphere

To examine Julia sets of rational maps, we will view them as mapping a sphere to itself. We want to use the idea that a polynomial is chaotic on its Julia set to motivate the definition we use for the Julia set of a rational function.

A complex rational function is the quotient of two polynomials, $R(z) = P(z)/Q(z)$, where $P(z)$ and $Q(z)$ have no common factor. These functions give a natural way to extend the family of polynomials. However, they are not defined throughout the plane, since $Q(z) = 0$ for some complex numbers. In a typical undergraduate class, we might avoid this problem by simply eliminating from consideration all points z_0 where $Q(z_0) = 0$. However, even if we exclude z_0 from the domain we may still encounter problems when we iterate. For example, we would also have problems at any point whose orbit ever lands on z_0!

We approach the problem from a different perspective: Instead of eliminating points from the domain, we add infinity to the domain, and consider $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$. Then, for any point z with $Q(z) = 0$, we can define $R(z) = \lim_{z_n \to z} R(z_n)$ (which would equal ∞ if $P(z)$ and $Q(z)$ have no common factors). Similarly, $R(\infty) = \lim_{z\to\infty} R(z)$. This process of adding infinity to our domain is called a *compactification of the complex plane*. We can now easily iterate our rational function on the domain $\mathbb{C}_\infty$.

We also need to change the way we measure the distance between two points. The easiest way to do this is to view our new domain $\mathbb{C}_\infty$ as a sphere. Imagine we are looking at a sphere of radius 1 sitting in three-dimensional space centered at (0, 0, 0). The complex plane cuts through the sphere at the equator. For any point z in the complex plane, we can draw a line connecting z with the top of the sphere. This line will intersect the sphere at exactly one other point (see FIGURE 3). Notice that this provides an identification between all the points in the plane and all but one point on the sphere. (The north pole, if we view the sphere as a globe, doesn't get matched with a point in the plane.) By pairing the top of the sphere with ∞, we obtain a one-to-one corre-

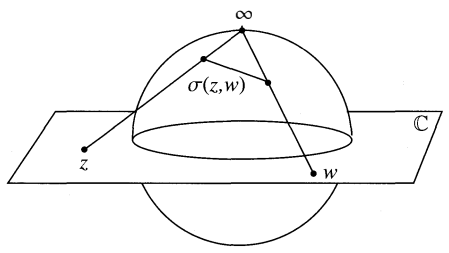


Figure 3 The identification between the sphere and $\mathbb{C}_\infty$

spondence between the sphere and our new domain $\mathbb{C}_\infty$, which is called the *complex Riemann sphere*.

Now our rational function $R(z)$ is well defined on $\mathbb{C}_\infty$ and can easily be iterated. We can measure the distance σ in $\mathbb{C}_\infty$ between two points as the length of the chord through the sphere connecting the two points. However, using this concept of distance, all orbits are bounded, so we cannot use the same definition of the Julia set that we used for P_c. Instead, we use a more general definition of the Julia set, one based on a property that we will observe in the Julia set of a polynomial.

We return to our previous example $P_0(z) = z^2$ (which is itself a rational map) and examine how small balls are mapped by P_0. If we pick a small ball inside the unit circle, iterates of that ball shrink and spiral into the origin. Similarly, if we choose a ball exterior to the unit circle the iterates spiral toward infinity as in FIGURE 4, and if we view the process on the Riemann sphere we see that the balls shrink as well. However, small balls intersecting the unit circle are stretched out and grow large both on the plane and the sphere, as in FIGURE 5. We see that only balls contained entirely in the *complement* of the Julia set of P_0 shrink under forward iteration.

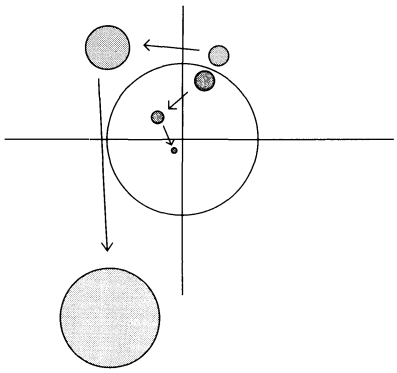


Figure 4 Iteration of balls inside or outside the unit disk for $P_0(z) = z^2$

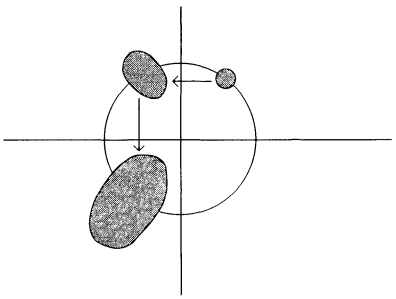


Figure 5 Iteration of balls intersecting the unit disk for $P_0(z) = z^2$

We need a formal definition to describe the property that balls shrink under forward iteration. The collection of iterates of the rational map R (that is $\{R, R^2, R^3, \ldots\}$) is called *equicontinuous at the point* z when for every $\epsilon > 0$ there is a $\delta > 0$ such that if $\sigma(z, y) < \delta$, then $\sigma(R^n(z), R^n(y)) < \epsilon$ for all $n \geq 0$.

Now instead of thinking about points that lie in the Julia set, we first consider the points that don't belong to the Julia set. The points that are *not* in the Julia set of a rational function R form the largest open set of points z for which the collection of iterates of R is equicontinuous. This roughly amounts to saying that the Julia set, $J(R)$, is the collection of points z where R depends sensitively on initial conditions. Using this definition for the Julia set of R, we can show that $J(R)$ has a dense collection of periodic points, and that R acts transitively on $J(R)$ (see Beardon's Theorems 4.3.2 and 6.9.2 **[6]** for details). Therefore, this definition of the Julia set of R gives us the set where $R(z)$ acts chaotically, just as the earlier definition of $J(P_c)$ did.

FIGURE 6 shows the Julia set of the rational map $R(z) = [(.7 + .7i)(z^2 + 1)]/z^2$. The Julia set of R resembles the Julia set of a polynomial in the sense that it appears to have zero area. It turns out that the Julia sets of many rational maps of the sphere have zero area. There are also many rational maps whose Julia set encompasses the entire

Figure 6 Julia set of $R(z) = (.7 + .7i)(z - \frac{1}{z})$

sphere. In the next two sections, we present one such example, but there are many other rational maps whose Julia set is $\mathbb{C}_\infty$ that are not constructed in this manner. The curious reader will wonder what else can happen with Julia sets of rational functions. Interestingly, it is not known if there exists a rational map whose Julia set has positive area without occupying the whole sphere.

Maps of the torus

We now examine a seemingly unrelated map that we will use later to construct a rational map of the sphere whose Julia set is $\mathbb{C}_\infty$. Start with the doubling map on the circle, which is our familiar function P_0 restricted to the unit circle. This function can be written in a variety of forms, and it will be easier for us to use a different notation.

Instead of the unit circle in the complex plane, we use the unit interval $[0, 1) = \{x \in \mathbb{R}: 0 \leq x < 1\}$. Define the function $S: [0, 1) \to [0, 1)$ by $S(x) = 2x \bmod 1$. Technically, this is a function on equivalence classes of real numbers modulo 1, but it is easily visualized as $[0, 1)$. If you take this interval and paste the point 0 to the point 1, you get a circle. The action of S on the interval is identical to the action of P_0 on the unit circle.

The map we need is defined on the *unit square*:

$$U = \{u = x + yi \in \mathbb{C}: 0 \leq x < 1, 0 \leq y < 1\} = [0, 1) \times [0, 1).$$

Let $T: U \to U$ be defined by $T(u) = T(x + yi) = (2x \bmod 1) + i(2y \bmod 1)$, which we will also denote by $2u \bmod [1, i]$. Notice that T is a generalization of the doubling map on the unit interval.

Just as the domain $[0, 1)$ could be identified with the circle, the region U can be identified with the *torus*. We can think of the torus as the surface (that is, the icing, but not the interior) of a donut. To see this identification, we first paste the top boundary a of the region U to the bottom boundary c, obtaining a cylinder. Then paste the right boundary b of the cylinder to the left boundary d, as in FIGURE 7. It is somewhat difficult to visualize how maps on the torus act, which is why we often use the region U in the plane when we are studying toral maps.

It turns out that the function $T: U \to U$ is topologically transitive and has dense periodic points in U. The proofs mimic the arguments that $S: [0, 1) \to [0, 1)$ has these properties.

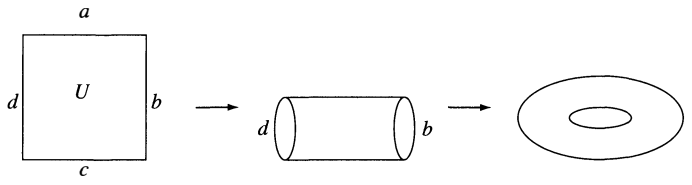


Figure 7 The identification of U and the torus

Làttes/Böettcher example with $J(R) = \mathbb{C}_\infty$

We are now prepared to discuss the Làttes/Böettcher example, which involves constructing a rational function on the sphere using the function T described above and a classical function called a Weierstrass elliptic function. The Weierstrass elliptic functions form a family, each of which maps the plane onto $\mathbb{C}_\infty$, with one function for each parallelogram with a corner at 0. We use the unique Weierstrass elliptic function $\wp$ whose parallelogram is the unit square U. Although the development of the Làttes/Böettcher example involves this famous elliptic function and the doubling map on the torus, we end up with a rational function for which an explicit formula is known.

Let $\Gamma = \{k + ij : k, j \in \mathbb{Z}\}$ represent the corners of all of the unit squares in the plane. For any $z \in \mathbb{C}$, the Weierstrass elliptic function with respect to Γ is defined by

$$\wp(z) = \frac{1}{z^2} + \sum_{v \in \Gamma - \{0\}} \left[\frac{1}{(z - v)^2} - \frac{1}{v^2} \right].$$

Although we have defined $\wp$ for any point in the plane, it follows from the definition that $\wp$ is periodic: $\wp(z) = \wp(z \bmod [1, i])$. Whatever $\wp$ does to the unit square U determines what it does to the entire plane, and we often think about $\wp$ only in terms of a function on U.

The formula for this elliptic function shows that it blows up at the origin; we call this a pole of order 2, and poles like this are known to be the only singularities of this function. We need three important properties. First, we know that $\wp(\mathbb{C}) = \wp(\bar{U})$, where $\bar{U}$ is the closure of U. Since $\wp(\mathbb{C})$ is open and $\wp(\bar{U})$ is closed, $\wp$ is onto $\mathbb{C}_\infty$. Second, if we think about $\wp$ as a map on the plane and $u \in U$, we see from the definition that $\wp(u) = \wp(-u)$. But $\wp(-u) = \wp(-u \bmod [1, i])$ and so the points u and $-u \bmod [1, i]$ in U have the same image under $\wp$. Third, there are exactly 4 points in the unit square U where $u \equiv -u \bmod [1, i]$: $1/2$, $i/2$, $(1 + i)/2$, and 0. Thus $\wp(1/2)$, $\wp(i/2)$, $\wp((1 + i)/2)$, and $\wp(0)$ are the only four points on the sphere that only have one preimage in U under $\wp$. (See Beardon [**6**, Section 4.3 and Appendix II], or Ahlfors [**1**, Section 7.3] for more details on $\wp$.)

Consider the composition $R(z) = \wp T \wp^{-1}(z)$. You might suspect that we may run into trouble with this composition since $\wp^{-1}(z)$ has 2 values in U for almost every $z \in \mathbb{C}_\infty$. Nevertheless, we can show that $R(z) = \wp T \wp^{-1}(z)$ is well defined. If z is any point on $\mathbb{C}_\infty$, let u and u' denote the two values of $\wp^{-1}(z)$ in U. From the properties of $\wp$ that we discussed earlier, the points u and u' must satisfy $u = -u' \bmod [1, i]$. But then $T(u) = -T(u') \bmod [1, i]$, and since $\wp$ is even we have $\wp(T(u)) = \wp(-T(u') \bmod [1, i])$. The composition $R(z) = \wp T \wp^{-1}(z)$ was first discussed in the early 1900s independently by Làttes and Böettcher and can be

illustrated by the following commutative diagram:

$$\begin{array}{ccc} U & \xrightarrow{T} & U \\ \downarrow{\scriptstyle \wp} & & \downarrow{\scriptstyle \wp} \\ \mathbb{C}_\infty & \xrightarrow{R} & \mathbb{C}_\infty. \end{array}$$

It can be shown using complex analysis that $R(z)$ is actually a rational function. The argument involves noticing that $\wp$, T, and $\wp^{-1}$ are all differentiable on small open balls, so R must be differentiable on all of $\mathbb{C}_\infty$ and thus must be a rational function. In fact, the equation for this rational function is the degree four map

$$R(z) = \frac{(z^2+1)^2}{4z(z^2-1)}$$

(see Beardon [**6**, Section 4.3]). As an exercise, the reader should verify that a typical $z \in \mathbb{C}_\infty$ has four preimages by following z through the composition $R^{-1} = \wp T^{-1} \wp^{-1}$.

Since T is chaotic on U, it is straightforward to show that R is chaotic on all of $\mathbb{C}_\infty$ because all of the necessary properties are preserved under composition by $\wp$ and $\wp^{-1}$. This proof technique is common in dynamical systems, and the details are not difficult to fill in.

As you might suspect, since $R(z)$ acts chaotically on its entire domain, the Julia set of R is all of $\mathbb{C}_\infty$. How can we use the definition of the Julia set to see that $J(R) = \mathbb{C}_\infty$? Let E be any disk on U. Since T stretches any set by a factor of two in each direction, we know that there must be an n large enough such that $T^n(E)$ covers all of U. Let $V = \wp(E)$. Then $R^n(V)$ covers the entire sphere. So there can't be any open set V on the sphere for which the collection of iterates of R is equicontinuous, and therefore the Julia set is the entire sphere.

Since $J(R) = \mathbb{C}_\infty$, any computer generated image of this Julia set would simply be the entire sphere (see FIGURE 2) and isn't very interesting to visualize. However, we know that R is chaotic on its entire domain, and thus the behavior of any point under iteration is extremely unpredictable. The picture of the Julia set might seem quite dull compared to other examples, but the function has very interesting dynamics because it is chaotic everywhere.

Other rational maps

The construction above can be generalized to other rational maps of the sphere either by changing the toral map T or the map from the torus to the sphere. In all of the following cases the Julia set of R is $\mathbb{C}_\infty$.

1. We can use the toral map $T_n(u) = nu \bmod [1, i]$ for any integer $n \geq 2$. Then $R(z) = \wp T_n \wp^{-1}(z)$ is a degree n^2 rational map of the sphere. This can be proved using arguments similar to the ones mentioned in the previous section. Barnes and Koss [**4**] and Milnor [**19**] give details.
2. Suppose $T_{a+bi}(u) = (a+bi)u \bmod [1, i]$ where $a, b \in \mathbb{Z}$ and $|a+bi| > 1$. Then $R(z) = \wp T_{a+bi} \wp^{-1}(z)$ is a degree a^2+b^2 rational map of the sphere. For example, $T_{1+i}(u) = (1+i)u \bmod [1, i]$ produces the degree two rational map of the sphere $R(z) = (z^2-1)/(2iz)$. See Barnes and Koss [**4**] or Milnor [**19**] for details.
3. Douady and Hubbard [**13**] and Koss [**16**] give other ways to generalize the acceptable toral maps on different complex tori.

4. We used the Weierstrass $\wp$-function to map the torus to the sphere. There are three other elliptic functions that, composed with a toral map, produce rational maps of the sphere. See Douady and Hubbard [**13**] or Koss [**16**] for details.

Note that the above list does not encompass all known rational maps whose Julia set is equal to all of $\mathbb{C}_\infty$. Lyubich [**18**] and Rees [**20**] have both constructed large families of functions with this property. However, their methods are very different from the ones demonstrated in this paper.

Acknowledgments. The second author was partially supported by NSF Grant 9970575. The programs to draw and print FIGURE 6 were written by Pete Vermeire, Jane Hawkins, Michael Taylor, and Lorelei Koss.

REFERENCES

1. L. Ahlfors, *Complex Analysis*, Third Edition, McGraw-Hill, Inc., New York, 1979.
2. J. Banks, et al., On Devaney's definition of chaos, *Amer. Math. Monthly*, **99** (1992), 332–334.
3. J. Barnes, Conservative exact rational maps of the sphere, *J. Math. Analysis and Applications*, **230** (1999), 350–374.
4. J. Barnes and L. Koss, One-sided Lebesgue Bernoulli maps of the sphere of degree n^2 and $2n^2$, *Internat. J. Math. & Math. Sci.*, **23** (2000), no. 6, 383–392.
5. M. Barnsley, *Fractals Everywhere*, 2nd Edition, Academic Press, Boston, MA, 1993.
6. A. Beardon, *Iteration of Rational Functions*, Springer-Verlag, New York 1991.
7. L. Böettcher, The principal convergence laws for iterates and their applications to analysis, *IZV. FIz.-Mat. Obshch. pri Imper. Kazanskom Univ.* **13** (1903), no. 1, 1–37; **14** (1904), nos. 3–4, 155–234.
8. *Chaos and Fractals: The mathematics behind the computer graphics*, Edited by R. Devaney and L. Keen, Proceedings of Symposia in Applied Mathematics, volume 39, American Mathematical Society, Providence, RI, 1994.
9. *Complex Dynamical Systems: The mathematics behind the Mandelbrot and Julia sets*, Edited by R. Devaney, Proceedings of Symposia in Applied Mathematics, volume 49, American Mathematical Society, Providence, RI, 1994.
10. R. Devaney, *An Introduction to Chaotic Dynamical Systems, 2nd edition*, Addison-Wesley Publishing Company, Inc., Redwood City, CA, 1989.
11. R. Devaney, *A First Course in Dynamical Systems*, Addison-Wesley Publishing Company, Inc., Reading, MA, 1992.
12. R. Devaney, The orbit diagram and the Mandelbrot set, *College Math J.*, **22** No. 1 (1991), 23–38.
13. A. Douady and J. Hubbard, A proof of Thurston's topological characterization of rational functions, *Acta Math.*, **171** (1993), 263–297.
14. E. Glasner and B. Weiss, Sensitive Dependence on Initial Conditions, *Nonlinearity*, **6** (1993), 1067–1075.
15. R. Holmgren. *A First Course in Discrete Dynamical Systems*,Springer Verlag Inc., New York, 1996.
16. L. Koss, Ergodic and Bernoulli properties of analytic maps of complex projective space, *Trans. Amer. Math. Soc.* **354** (2002), 2417–2459.
17. S. Làttes, Sur l'iteration des substitutions rationelles et les fonctions de Poincarè, *C. R. Acad. Sci. Paris*, **166** Ser. 1 Math. (1919), 26–28.
18. M. Lyubich, An analysis of the stability of the dynamics of rational functions, *Selecta Mat. Soviet.*, **9** no. 1 (1990), 69–90.
19. J. Milnor, *Dynamics in One Complex Variable: Introductory Lectures*, Vieweg Verlag, Braunscheig, 1999.
20. M. Rees, Positive measure sets of ergodic rational maps, *Éc. Norm. Sup.*, **19** (1986), 393–407.
21. H.-O. Peitgen, P.H. Richter, *The Beauty of Fractals. Images of Complex Dynamical Systems*, Springer Verlag, Berlin, 1986.

An Algorithm to Solve the Frobenius Problem

ROBERT W. OWENS
Department of Mathematical Sciences
Lewis & Clark College
Portland, OR 97219
owens@lclark.edu

McDonald's restaurants in the U.S.A. sell Chicken McNuggets in quantities of 6, 9, and 20. If you want exactly 26 or 27 McNuggets, you could buy them, but you cannot buy 25 or 28. You can buy exactly 41, but you cannot buy exactly 43 McNuggets. If you pick any number larger than 43 you can purchase exactly that many, so 43 is the largest number of Chicken McNuggets that you cannot obtain by combining orders of 6, 9 and 20. This is a special case of the *problem of Frobenius*: Given n positive integers with greatest common divisor one, find the largest integer not expressible as a nonnegative linear combination of the integers.

Not just a curiosity for the mathematically inclined who play with their food, the problem of Frobenius is also connected with probability [**4**], graph theory [**6, 7**], Gorenstein rings [**10**], the theory of concurrency in computer science [**14**], and the geometry of numbers [**9**]. Readers may know it as the "postage stamp problem."

The Frobenius number A few general theorems and many specific results are known concerning this problem. To help describe them, let n be a positive integer and let $a_1, \ldots, a_n$ be distinct positive integers. If an integer N can be expressed as a nonnegative linear combination of $a_1, \ldots, a_n$, we say that N is *dependent* on $a_1, \ldots, a_n$ or that N is *representable*; otherwise we say that N is *independent* of $a_1, \ldots, a_n$. If there is a largest positive integer independent of $a_1, \ldots, a_n$, we denote it by $g(a_1, \ldots, a_n)$ and call it the *Frobenius number* for $a_1, \ldots, a_n$.

If $\gcd(a_1, \ldots, a_n) = d > 1$, then any positive integer that is not a multiple of d is independent of $a_1, \ldots, a_n$, so $g(a_1, \ldots, a_n)$ is not defined in this case. However, if $\gcd(a_1, \ldots, a_n) = 1$, then $g(a_1, \ldots, a_n)$ is defined. To see this, notice that if we can find a_1 consecutive integers $N, N+1, \ldots, N+a_1-1$ that are representable, then all integers greater than or equal to N are representable, implying both the existence of the Frobenius number and an upper bound $g(a_1, \ldots, a_n) < N$. To establish the existence of the Frobenius number, it suffices to guarantee such a string of representable numbers. It is known that $\gcd(a_1, \ldots, a_n) = 1$ implies that we can express 1 as a linear combination of the a_i using integer coefficients, some of them negative. Group the positive and negative terms to write $1 = (K+1) - K$, where $K+1$ and K are both representable. The reader can check that $N = (a_1 - 1)K$ is the first in a string of a_1 consecutive representable numbers.

For example, in the Chicken McNuggets case, where $(a_1, a_2, a_3) = (6, 9, 20)$, we can express $1 = 7 * 9 - (7 * 6 + 1 * 20)$, so taking $K = 62$ and $a_1 = 6$, we find that $5 * 62 = 310, 311, \ldots, 315$ are 6 consecutive representable numbers. So $g(6, 9, 20) < 310$, a rather crude upper bound, but good enough to establish existence of the Frobenius number. We assume throughout this article that $\gcd(a_1, \ldots, a_n) = 1$.

We also assume that none of $a_1, \ldots, a_n$ is dependent on the others, since if a_k were dependent upon $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n$, then N would be independent of $a_1, \ldots, a_n$ if and only if N were independent of $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n$, so $g(a_1, \ldots, a_n) = g(a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n)$. Except for the uninteresting case in which $n = 1 = a_1$, these assumptions imply that $n > 1$ and $a_k > 1$ for all k. Finally, we assume throughout that the a_k are in strictly increasing order.

While not intending to survey all work on the Frobenius problem, we begin by mentioning highlights of the history that relate to this article. The Frobenius problem for two integers is completely solved by the formula $g(a, b) = ab - a - b$; moreover, exactly half of the integers between 0 and $g(a, b)$ are independent of a and b. Sylvester [**20**] established the first of these results (which will follow from a characterization theorem stated and proved in the next section) and an even stronger form of the second in a problems column of the *Educational Times* in 1884. The interested reader might pursue proofs of both results using elementary techniques.

Frobenius is reported to have posed the general problem often in lectures in the early 1900s, and since then his name has become attached to it. Interest in the problem continues to the present and has resulted in recent publications in both research [**3, 15**] and expository [**13, 19**] journals.

Curtis [**2**] established that for each $n > 2$ the Frobenius number cannot be expressed in terms of a finite set of polynomials, and Ramírez-Alfonsín [**15**] proved that the Frobenius problem for $n > 2$ is NP-hard. This is why the Frobenius problem still holds interest: If we involve three or more integers, there can be nothing like the simple polynomial formula that worked for two. Consequently, work on the Frobenius problem with $n > 2$ is directed towards other goals: determining formulae for special n-tuples of integers, including polynomial formulae for classes of integers; finding nonpolynomial algorithms to solve the $n = 3$ problem; finding bounds on the Frobenius number for arbitrary n, together with circumstances under which those bounds are achieved; and establishing the complexity of solving the Frobenius problem.

We start by introducing notation, terminology, and assumptions. Then we develop an algorithm to find the Frobenius number and prove that the algorithm is correct. We conclude with a brief summary of the assumptions and the algorithm, and work through three examples.

Preliminaries

There are two special cases in which the Frobenius problem can either be solved explicitly or reduced to a simpler problem. We mention these so that subsequently we may assume that such circumstances do not apply in the problems that we consider.

First, Roberts [**16**] showed that if the positive integers $a_1, \ldots, a_n$ form an arithmetic progression, then $g(a_1, \ldots, a_n) = \lfloor \frac{a_1 - 2}{n-1} \rfloor a_1 + (a_1 - 1)(a_2 - a_1)$, where $\lfloor \quad \rfloor$ denotes the floor function.

Second, if $\gcd(a_1, \ldots, a_{n-1}) = d$, then

$$g(a_1, \ldots, a_n) = d \cdot g\left(\frac{a_1}{d}, \ldots, \frac{a_{n-1}}{d}, a_n\right) + (d - 1)a_n.$$

This result was established by Johnson [**8**] for $n = 3$ and by Brauer and Shockley [**1**] in general.

Residue classes modulo a_1 Let us review the assumptions made up to now. The $n > 1$ integers whose Frobenius number we seek are denoted $a_1, \ldots, a_n$, with $1 < a_1 < \cdots < a_n$. None of $a_1, \ldots, a_n$ is dependent on the others, no $n - 1$ element subset of $a_1, \ldots, a_n$ has greatest common divisor greater than 1, and the numbers are not in arithmetic progression.

Brauer and Shockley [**1**] characterized the general solution of the Frobenius problem as follows. The method depends on showing that each residue class modulo a_1 contains nonnegative integers dependent upon $a_2, \ldots, a_n$. To see why, express 1 as a

linear combination of $a_2, \ldots, a_n$ using integer coefficients, and group them in positive and negative sums as before; this gives $1 = (N + 1) - N$, where $N + 1$ is dependent on a subset $\mathcal{A}$ of $a_2, \ldots, a_n$ and N is dependent on the complement of $\mathcal{A}$ in $a_2, \ldots, a_n$. Then the consecutive integers $a_1N, a_1N + 1, \ldots, a_1N + a_1 - 1$ are dependent on $a_2, \ldots, a_n$ and occupy different congruence classes mod a_1. For example, if $(a_1, a_2, a_3) = (5, 17, 23)$, then we can express $1 = 3 * 23 - 4 * 17$. With $N = 68$, we find that $5 * 68 = 340, 341, \ldots, 344$ are each dependent on 17 and 23 and congruent modulo 5, respectively, to $0, 1, \ldots, 4$.

In general, this means that for each k, $0 \leq k < a_1$, the set

$$S(k) = \{N \geq 0 \mid N \text{ is dependent upon } a_2, \ldots, a_n \text{ and } N \equiv k \pmod{a_1}\}$$

is nonempty. Brauer and Shockley's Lemma 3 [**1**, p. 217], which we will call the *characterization theorem*, tells how to use these sets to determine the Frobenius number. Not only is the characterization theorem the basis upon which several algorithms to solve the Frobenius problem have been shown to be correct, it is essential to proving the validity of the algorithm developed in this article.

CHARACTERIZATION THEOREM. *Let t_k be the smallest element of $S(k)$ and let $r = \max t_k$. Then $g = r - a_1$.*

This result is so central to subsequent work in this article that we pause to consider two simple examples and to present Brauer and Shockley's elegant proof. First notice that if $n = 2$, then all a_1 residue classes modulo a_1 are represented in the set $\{ka_2, 0 \leq k < a_1\}$, since a_1 and a_2 are relatively prime. Thus $r = (a_1 - 1)a_2$, yielding Sylvester's result, $g(a_1, a_2) = a_1a_2 - a_1 - a_2$.

Next consider the Frobenius problem with the integers 5, 17, and 23. It is easy to see that $g(5, 17, 23) = 41$, since 41 is independent of 5, 17, and 23, but the five consecutive integers 42 to 46 are dependent on 5, 17, and 23.

To find the Frobenius number using Brauer and Shockley's approach, group the integers dependent on 17 and 23 (17, 23, 34, 40, 46, 51, 57, 63, 68, 69, 74, 80, 86, ...) into residue classes modulo 5, obtaining

$$\begin{aligned}
S(0) &= \{40, 80, \ldots\}, && \text{so} \quad t_0 = \min\{x | x \in S(0)\} = 40,\\
S(1) &= \{46, 51, 86, \ldots\}, && \text{so} \quad t_1 = \min\{x | x \in S(1)\} = 46,\\
S(2) &= \{17, 57, \ldots\}, && \text{so} \quad t_2 = \min\{x | x \in S(2)\} = 17,\\
S(3) &= \{23, 63, 68, \ldots\}, && \text{so} \quad t_3 = \min\{x | x \in S(3)\} = 23, \quad \text{and}\\
S(4) &= \{34, 69, 74, \ldots\}, && \text{so} \quad t_4 = \min\{x | x \in S(4)\} = 34.
\end{aligned}$$

Then $r = \max t_k = 46$, so the Frobenius number is $46 - 5 = 41$.

Let us see why this algorithm produces the largest number independent of 5, 17, and 23. On the one hand, if 41 were dependent upon 5, 17, and 23, then $41 = 46 - 5 = 5x + 17y + 23z$, for some $x, y, z \geq 0$, so $46 - 5(x + 1) = 17y + 23z < 46$, contradicting that 46 is the smallest element in its residue class modulo 5. So 41 is independent of 5, 17, and 23. On the other hand, any number greater than 41, say 44 (which is congruent to 4(mod 5), is greater than $t_4 = 34$, which is the smallest number in its residue class modulo 5. Since 44 can be expressed as a number dependent on 17 and 23 plus a multiple of 5, it is dependent on 5, 17, and 23.

From the example it is a short step to the proof of Brauer and Shockley's result, which is reproduced verbatim [**1**, p. 217].

Proof. If $r - a_1$ could be represented in the form

$$r - a_1 = x_1a_1 + x_2a_2 + \cdots + x_ka_k \quad \text{with} \quad x_i \geq 0$$

then

$$r - (x_1 + 1)a_1 = x_2a_2 + \cdots + x_ka_k,$$

which contradicts the fact that r is the smallest such number in its residue class. Any number greater than $r - a_1$ is greater than or equal to the smallest number in its residue class that can be so represented and is therefore representable by

$$a_1x_1 + a_2x_2 + \cdots + a_kx_k. \qquad \blacksquare$$

Neither the example nor the proof requires that the integers a_k be in increasing order, but for ease of keeping track of residue classes, it is convenient that a_1 be smallest.

Toward an algorithm As illustrated in the example, the characterization theorem suggests a "brute force" algorithm. First catalog numbers dependent upon $a_2, \ldots, a_n$, from the smallest representable numbers onward, until all residue classes modulo a_1 are nonempty. Then pick the smallest element from each residue class. The largest of these numbers minus a_1 is the Frobenius number.

Most published algorithms solving the Frobenius problem implement a variation of this brute force algorithm, with increased efficiency sought in two ways. First, to avoid generating representable numbers explicitly, a set of numbers is found that *a priori* is guaranteed both to contain only representable numbers and also the smallest representable number in each residue class modulo a_1. Second, to avoid exhaustively searching this set for the element that yields the solution of the Frobenius problem, most of these representable numbers are discarded, with the solution of the Frobenius problem guaranteed to be among the remaining few elements.

For the $n = 3$ case, Brauer and Shockley [**1**] provided the first such algorithm by constructing an "L-shaped" region of lattice points in the plane, where each point is uniquely associated with one of the residue classes modulo a_1. The solution of the Frobenius problem then corresponds to one of two so-called *corner points*.

For the general case, Lewin [**11**] replaced Brauer and Shockley's L-shaped region with a triangular table of representable numbers modulo a_1, and he established an *a priori* bound on the number of rows needed to guarantee that all residue classes were represented. However, Lewin's method for determining which element of the table affords the solution of the Frobenius problem does not always provide correct values, as we will later see. A more subtle selection procedure is required, which suggests that the algorithm we present is not likely to be straightforward.

Algorithms based on other strategies have also been published. Rödseth [**17**] and Selmer and Beyer [**18**] developed algorithms for the $n = 3$ case based on a continued fractions expansions of a quotient associated with a_1, a_2, and a_3. At some point in the algorithm, a certain quotient—not a convergent—changes from positive to negative, and parameter values determined at this transition yield the solution of the Frobenius problem as the larger of two associated quantities. The characterization theorem is then used to establish that this algorithm solves the Frobenius problem.

These algorithms for the $n = 3$ case have attracted much attention because modified versions of them have been shown in one case to require fewer than $O(\log a_1)$ steps [**5**, p. 344] and in another case to be "a polynomial time algorithm" [**3**, p. 353]. Even if these results only hold for $n = 3$, they are encouraging in light of the more sobering theorems, noted earlier, of Curtis, saying that for each $n > 2$ the Frobenius number cannot be determined using a finite set of polynomials, and of Ramírez-Alfonsín, that the Frobenius problem for $n > 2$ is NP-hard.

The L-shaped region We describe Brauer and Shockley's algorithm in detail since it is this algorithm that we extend in this article. For notational convenience, let the n in-

tegers whose Frobenius number we seek be denoted $a, a_1, \ldots, a_{n-1}$. Noting Roberts's, Johnson's, and Brauer and Shockley's previously mentioned results, assume that no $n-1$ element subset of $\{a, a_1, \ldots, a_{n-1}\}$ has greatest common divisor greater than 1 and that the numbers are not in arithmetic progression. Even though the characterization theorem does not require $a, a_1, \ldots, a_{n-1}$ to be in increasing order, Roberts's result does, so we assume that $a, a_1, \ldots, a_{n-1}$ are in increasing order.

Since the algorithm presented in this article extends Brauer and Shockley's algorithm [**1**], our development follows that of Brauer and Shockley as closely as possible. Guided by the characterization theorem, both algorithms construct a set of "smallest" numbers dependent on $a_1, \ldots, a_{n-1}$, one for each of the residue classes modulo a. Each of these numbers can be associated with an $(n-1)$-dimensional lattice point with nonnegative integer coordinates, so the algorithms can be viewed as constructing a geometric object, a distinguished point of which yields the Frobenius number for the integers $a, a_1, \ldots, a_{n-1}$.

If $n = 3$, then $n - 1 = 2$, so in that case Brauer and Shockley obtain an easily visualized planar region, the L-shaped region previously mentioned. For $n > 4$, however, $n - 1 > 3$, which makes sketching images difficult. To motivate and explain the algorithm presented in this article, we consider a specific example with $n = 4$, so the resulting lattice points are in 3-space. The example that will accompany us throughout this article is the problem of finding the Frobenius number for the numbers $\{a, a_1, \ldots, a_{n-1}\} = \{31, 41, 47, 61\}$. Lewin [**11**] considered exactly this example with his algorithm producing the number 239, but the Frobenius number is actually 240.

Let $\mathbb{Z}$ denote the set of integers. For $\mathbf{x} = (x_1, \ldots, x_{n-1}) \in \mathbb{Z}^{n-1}$, by $\mathbf{x} > 0$ we mean that $x_k > 0$ for each k, $1 \leq k \leq n-1$; similarly, define $\mathbf{x} \geq 0$, $\mathbf{x} > \mathbf{y}$, and $\mathbf{x} \geq \mathbf{y}$ componentwise.

Let $H : \mathbb{Z}^{n-1} \to \mathbb{Z}$ be defined by $H(x_1, \ldots, x_{n-1}) = a_1x_1 + \cdots + a_{n-1}x_{n-1}$. In our example, $H : \mathbb{Z}^3 \to \mathbb{Z}$ is defined by $H(x_1, x_2, x_3) = 41x_1 + 47x_2 + 61x_3$.

If $\mathbf{x} = (x_1, \ldots, x_{n-1}) \geq 0$, then we can associate with $\mathbf{x}$ the representable number $H(\mathbf{x})$. Strictly speaking, $\mathbf{x}$ is a point in $\mathbb{Z}^{n-1}$ and $H(\mathbf{x})$ is an integer, but the association between $\mathbf{x}$ and $H(\mathbf{x})$ is direct enough so that, when no confusion ensues and convenience permits fewer words, we will refer to $\mathbf{x}$ as a representable number or that $\mathbf{x}$ is (or is not) a solution of the Frobenius problem.

The characterization theorem implies that the Frobenius number for $a, a_1, \ldots, a_{n-1}$ is $\max\{t_k | 0 \leq k < a\} - a$, where

$$t_k = \min\{H(\mathbf{x}) | H(\mathbf{x}) \equiv k \pmod{a} \quad \text{for some} \quad \mathbf{x} \geq 0\}.$$

Following Brauer and Shockley, we construct a distinguished set guaranteed to contain $\{t_k | 0 \leq k \leq a\}$, from which the solution can be obtained.

Two lattice points $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^{n-1}$ are *congruent* if $H(\mathbf{x}) \equiv H(\mathbf{y}) \pmod{a}$. In our example, $(7, 8, 9)$ and $(2, 5, 4)$ are congruent since $H(7, 8, 9) = 1212 \equiv 3 \pmod{31}$ and $H(2, 5, 4) = 561 \equiv 3 \pmod{31}$.

As we set our algorithm to move about this lattice seeking candidates for r, we will sometimes need the following operation: For each k, $1 \leq k \leq n-1$, let $\delta_k : \mathbb{Z}^{n-1} \to \mathbb{Z}^{n-1}$ be defined by

$$\delta_k(x_1, \ldots, x_{n-1}) = (y_1, \ldots, y_{n-1}),$$

where $y_k = x_k$ and $y_j = -x_j$ if $j \neq k$. Thus δ_k changes the sign of all coordinates of $(x_1, \ldots, x_{n-1})$ except for the kth coordinate, which remains unchanged. For example, $\delta_2(7, 8, 9) = (-7, 8, -9)$.

The algorithm

Reducing to a box Our first goal is to construct an $(n-1)$-dimensional box of lattice points within which the solution of the Frobenius problem must lie. Finding the lengths of the sides of the box requires a somewhat technical definition, but it will save time when we can limit our search to a box of the form $R = [0, b_1) \times [0, b_2) \times \cdots \times [0, b_{n-1})$.

For motivation, return for a moment to $g(5, 17, 23)$, where r turned out to be $46 = 2 * 23$. By inspection, we see that $3 * 17 \equiv 2 * 23 \pmod 5$, and that $3 * 17 > 2 * 23$. If we happened to think that $x * 17 + y * 23$ were a good candidate for r, we would be able to reduce it to $(x-3) * 17 + (y+2) * 23$, which is smaller, but still in the same equivalence class.

For each k, $1 \le k \le n-1$, let b_k be the smallest positive integer m such that there are nonnegative x_j, $1 \le j \le n-1$ and $j \ne k$, that solve

$$ma_k \equiv (a_1x_1 + \cdots + a_{k-1}x_{k-1} + a_{k+1}x_{k+1} + \cdots + a_{n-1}x_{n-1}) \pmod a,$$

$$ma_k - (a_1x_1 + \cdots + a_{k-1}x_{k-1} + a_{k+1}x_{k+1} + \cdots + a_{n-1}x_{n-1}) > 0.$$

That is, b_k is the smallest positive integer value of the kth coordinate of $\mathbf{x} = (x_1, \ldots, x_{n-1}) \ge 0$ for which there is a solution of $H \circ \delta_k(\mathbf{x}) \equiv 0 \pmod a$ and $H \circ \delta_k(\mathbf{x}) > 0$.

To see that the b_k are well defined, notice that if $\mathbf{x} = (x_1, \ldots, x_{n-1})$ has $x_k = a$ and $x_j = 0$ for $j \ne k$, then $H \circ \delta_k(\mathbf{x}) = aa_k$, so $H \circ \delta_k(\mathbf{x}) \equiv 0 \pmod a$ and $H \circ \delta_k(\mathbf{x}) > 0$. It follows that not only is b_k defined, but $0 < b_k \le a$.

By the definition of b_k, there exists a point $\mathbf{v}_k = (z_{k1}, \ldots, z_{k,n-1}) \ge 0$, with $z_{kk} = b_k$, that solves $H \circ \delta_k(\mathbf{v}_k) \equiv 0 \pmod a$ and $H \circ \delta_k(\mathbf{v}_k) > 0$. If there is more than one such $\mathbf{v}_k$, pick any of them.

In our example, $\{31, 41, 47, 61\}$, we see that $b_1 = 3$ and $\mathbf{v}_1 = (3, 0, 1)$, since $3 * 41 \equiv (0 * 47 + 1 * 61) \pmod{31}$, $3 * 41 - (0 * 47 + 1 * 61) > 0$, and no positive integer smaller than 3 satisfies such relationships. Similarly, $b_2 = 5$ and $\mathbf{v}_2 = (2, 5, 2)$, since $5 * 47 \equiv (2 * 41 + 2 * 61) \pmod{31}$ and $5 * 47 - (2 * 41 + 2 * 61) > 0$. Finally, $b_3 = 4$ and $\mathbf{v}_3 = (1, 3, 4)$, since $4 * 61 \equiv (1 * 41 + 3 * 47) \pmod{31}$ and $4 * 61 - (1 * 41 + 3 * 47) > 0$.

The solution of the Frobenius problem must occur at a point of the box $R = [0, b_1) \times [0, b_2) \times [0, b_3) = [0, 3) \times [0, 5) \times [0, 4)$. To see why, consider any point $\mathbf{x} \ge 0, \mathbf{x} \notin R$. For example, let $\mathbf{x} = (x_1, x_2, x_3) = (7, 8, 9)$. Then at least one coordinate of $\mathbf{x}$ is greater than or equal to the corresponding b_k value. With $\mathbf{x} = (7, 8, 9)$, all three coordinates exceed the corresponding b_k values, so pick any one of the excessive coordinates, say $x_1 = 7 \ge 3 = b_1$. By the construction of b_1 and $\mathbf{v}_1 = (3, 0, 1)$, $\mathbf{y} = \mathbf{x} - \delta_1(\mathbf{v}_1)$ satisfies (i) $\mathbf{y} \ge 0$, (ii) $H(\mathbf{y}) < H(\mathbf{x})$, and (iii) $\mathbf{x}$ and $\mathbf{y}$ are congruent points.

(i) $\mathbf{y} = (7, 8, 9) - \delta_1(3, 0, 1) = (7, 8, 9) - (3, 0, -1) = (4, 8, 10) \ge 0$.

(ii) $H(7, 8, 9) = 1212$, and $H(4, 8, 10) = 1150$, so $H(\mathbf{y}) < H(\mathbf{x})$.

(iii) $1212 \equiv 3 \pmod{31}$, and $1150 \equiv 3 \pmod{31}$.

So by moving from $\mathbf{x}$ to $\mathbf{y}$, we obtain a representable number, $H(\mathbf{y}) = 1150$, that is strictly less than $H(\mathbf{x}) = 1212$ but belongs to the same residue class modulo 31. Moreover, $\mathbf{y} \ge 0$ since $\mathbf{x} \ge 0$ and the only coordinate of $\mathbf{x}$ that decreased as we moved to $\mathbf{y}$ was the first coordinate. However, $x_1 \ge b_1$, so that $y_1 = x_1 - b_1 \ge 0$.

Now $\mathbf{y} \ge 0$ and $\mathbf{y} \notin R$, so we can repeat this process, and we can continue to repeat it as long as the resulting point is not in R. This process terminates, however, since the value of H at points in the nonnegative orthant are nonnegative, and with each translation, the represented number strictly decreases. One chain of translations that we could

obtain is $(7, 8, 9) \to (4, 8, 10) \to (1, 8, 11) \to (3, 3, 13) \to (4, 6, 9) \to (5, 9, 5) \to (6, 12, 1) \to (8, 7, 3) \to (10, 2, 5) \to (7, 2, 6) \to (4, 2, 7) \to (1, 2, 8) \to (2, 5, 4) \to (3, 8, 0) \to (5, 3, 2) \to (2, 3, 3) \in R$.

Before returning to the general case, observe that $(2, 1, 2) \in R$, $H(2, 1, 2) = 251 \equiv 3 \pmod{31}$ and $H(2, 3, 3) = 406 \equiv 3 \pmod{31}$. Thus, even though we followed $(7, 8, 9)$ back to R via a chain of residue preserving translations, the resulting point in R does not necessarily represent the smallest representable number in its residue class. In particular, $H(2, 3, 3) - 31 = 406 - 31 = 375$ is not the Frobenius number for 31, 41, 47, and 61. Moreover, we cannot get from $(2, 3, 3)$ to $(2, 1, 2)$ by translations of the type that led us from $(7, 8, 9)$ to $(2, 3, 3)$. There is more work ahead!

Let us return to the general case and establish that the solution of the Frobenius problem corresponds (via H) to a point of the $(n-1)$-dimensional box $R = [0, b_1) \times \cdots \times [0, b_{n-1})$. Reflecting on the example, there was nothing special about the specific points used, so the proof of the general result simply extends what was illustrated in the example.

Let $\mathbf{x} = (x_1, \ldots, x_{n-1}) \geq 0, \mathbf{x} \notin R$. Then there exists $k \in \{1, \ldots, n-1\}$ with $x_k \geq b_k$. With $\mathbf{y} = \mathbf{x} - \delta_k(\mathbf{v}_k)$, we have $H(\mathbf{y}) \equiv H(\mathbf{x}) \pmod{a}$ and $H(\mathbf{y}) = H(\mathbf{x} - \delta_k(\mathbf{v}_k)) = H(\mathbf{x}) - H(\delta_k(\mathbf{v}_k)) < H(\mathbf{x})$. Hence $\mathbf{x}$ and $\mathbf{y}$ are congruent points. Moreover, $\mathbf{y} \geq 0$ since $y_k = x_k - b_k \geq 0$ and $y_j \geq x_j \geq 0$, for $1 \leq j \leq n-1$ and $j \neq k$. So by moving from $\mathbf{x}$ to $\mathbf{y}$, the kth coordinate decreases by b_k, all other coordinates are nondecreasing, the number represented decreases, and $\mathbf{x}$ and $\mathbf{y}$ are congruent points. If $\mathbf{y} \notin R$, repeat this process, possibly using a coordinate other than x_k. This sequence of translations terminates at a point $\mathbf{y} \in R$ since the function H decreases with each translation and $H(\mathbf{z}) \geq 0$ if $\mathbf{z} \geq 0$. Let $H(\mathbf{x}) \equiv k \pmod{a}$, with $0 \leq k < a$. Then

$$H(\mathbf{x}) > H(\mathbf{y}) \geq t_k = \min\{H(\mathbf{z}) | H(\mathbf{z}) \equiv k \pmod{a} \quad \text{for some} \quad \mathbf{z} \geq 0\},$$

so by the characterization theorem, $H(\mathbf{x}) - a$ is not the Frobenius number for $a, a_1, \ldots, a_{n-1}$. Consequently, the solution occurs at a point of R.

Redundant points While the solution of the Frobenius problem occurs at a point in R, the solution is not simply $\max\{H(\mathbf{x}) | \mathbf{x} \in R\} - a$, since, as we will soon see, there are congruent points $\mathbf{x}, \mathbf{y} \in R$ with $H(\mathbf{x}) \neq H(\mathbf{y})$. A point $\mathbf{x} \in R$ is defined to be *redundant* if there exists a congruent point $\mathbf{y} \in R$ with $H(\mathbf{x}) > H(\mathbf{y})$. In the example, we not only showed that $(2, 3, 3)$ is redundant, but also that, as a result, $H(2, 3, 3) - 31 = 375$ is not the Frobenius number. Since the solution of the Frobenius problem never occurs at a redundant point, we look for a way to remove redundant points from consideration.

To see that R contains redundant points, let $\mathbf{a} = (a, a, \ldots, a) \in \mathbb{Z}^{n-1}$. Then $\mathbf{a} > 0$ and $\mathbf{a}$ is congruent to $(0, \ldots, 0) \in \mathbb{Z}^{n-1}$. Beginning at $\mathbf{a}$, we can obtain a chain of lattice points in $\mathbb{Z}^{n-1}$, with nonnegative coordinates, whose last element is not $(0, \ldots, 0)$, since at each step of the construction of the chain one coordinate decreases and at least one other coordinate increases. So there exists in R at least one point $\mathbf{x} \neq 0$ that is congruent to $(0, \ldots, 0)$. Not only is $\mathbf{x}$ redundant, but any point $\mathbf{y} \in R$ with $\mathbf{y} \geq \mathbf{x}$ is also redundant, since it immediately follows that $\mathbf{y} - \mathbf{x} \geq 0$, $\mathbf{y} - \mathbf{x} \in R$, $\mathbf{y}$ and $\mathbf{y} - \mathbf{x}$ are congruent, and $H(\mathbf{y} - \mathbf{x}) < H(\mathbf{y})$.

In the $n = 3$ case, Brauer and Shockley [1] proved that of all nonzero points $\mathbf{x} \in R$ that are congruent to $(0, 0)$, one has both the smallest first and the smallest second coordinate. Moreover, they explicitly found this special point $\mathbf{z}$ without searching. For completeness, we outline their development and the resulting algorithm using notation introduced in this article.

Suppose that we seek $g(a, a_1, a_2)$. As we have previously seen, we can find $\mathbf{v}_1 = (b_1, z_{12})$ and $\mathbf{v}_2 = (z_{21}, b_2)$ such that

(i) $0 < b_1, b_2 \le a$

(ii) $0 < z_{12}, z_{21}$

(iii) $a_1b_1 - a_2z_{12} \equiv 0 \pmod{a}$ and $a_1b_1 - a_2z_{12} > 0$, and

(iv) $-a_1z_{21} + a_2b_2 \equiv 0 \pmod{a}$ and $-a_1z_{21} + a_2b_2 > 0$.

Brauer and Shockley showed that $(x_3, y_3) = (b_1 - z_{21}, b_2 - z_{12})$ is the distinguished point in the nonnegative quadrant closest to the origin that is congruent to the origin. Then any point $(x, y) \in R = [0, b_1) \times [0, b_2)$ with $x \ge x_3$ and $y \ge y_3$ is redundant, and there are no other redundant points. So the set within which the solution of the Frobenius problem lies is an "L" shaped region of lattice points obtained by removing $[x_3, b_1) \times [y_3, b_2)$ from $[0, b_1) \times [0, b_2)$. Since $H(x, y)$ is an increasing function of both x and y, the maximum value of H on this L-shaped region occurs at one of the two "corner" points $(b_1 - 1, y_3 - 1)$ and $(x_3 - 1, b_2 - 1)$, which immediately yields the solution of the Frobenius problem when $n = 3$.

When $n > 3$ there need not be a single, distinguished point in R such as the one that made Brauer and Shockley's algorithm effective, but with a little more work their basic idea can be extended. The key is to find points $\mathbf{z}$ congruent to the origin with $H(\mathbf{z}) > 0$, and then any point $\mathbf{x} \in R$ for which $\mathbf{x} - \mathbf{z} \in R$ is redundant. (Unlike Brauer and Shockley's $n = 3$ case, there may be more than one such $\mathbf{z}$. In fact, there could be more than $n - 2$ such points.) These points may not be elements of R. Finally, for those who might enjoy pursuing this topic, we have yet to discern a pattern in the coordinates of the $\mathbf{z}$s in terms of the b_k and $\mathbf{v}_k$.

If $\mathbf{x}$ is redundant, if $\mathbf{y} \in R$ is congruent to $\mathbf{x}$ with $H(\mathbf{x}) > H(\mathbf{y})$, and if $\mathbf{z} = \mathbf{y} - \mathbf{x}$, then $H(\mathbf{z}) \equiv 0 \pmod{a}$ and $H(\mathbf{z}) < 0$. Conversely, and more importantly for this algorithm, if $\mathbf{z} \in \mathbb{Z}^{n-1}$ satisfies $H(\mathbf{z}) \equiv 0 \pmod{a}$ and $H(\mathbf{z}) < 0$, and if both $\mathbf{x}$ and $\mathbf{x} + \mathbf{z} \in R$, then $\mathbf{x}$ is redundant. The condition that $\mathbf{x}, \mathbf{x} + \mathbf{z} \in R = [0, b_1) \times \cdots \times [0, b_{n-1})$ implies that $\mathbf{z} \in (-b_1, b_1) \times \cdots \times (-b_{n-1}, b_{n-1})$. For a particular Frobenius problem, a *zero* is defined to be a point $\mathbf{z} \in (-b_1, b_1) \times \cdots \times (-b_{n-1}, b_{n-1})$ satisfying $H(\mathbf{z}) \equiv 0 \pmod{a}$ and $H(\mathbf{z}) < 0$.

In our example, there are three zeroes: $(-2, 3, -3)$, $(0, -2, -1)$, and $(0, -4, -2)$. Let us see how these zeroes help us identify and eliminate redundant points of R.

First consider $\mathbf{z} = (-2, 3, -3)$. Then $H(\mathbf{z}) = -124 \equiv 0 \pmod{31}$. Any point $\mathbf{x} \in R$ for which $\mathbf{x} + \mathbf{z} \in R$ is redundant since $H(\mathbf{x} + \mathbf{z}) = H(\mathbf{x}) + H(\mathbf{z}) < H(\mathbf{x})$ and $H(\mathbf{x} + \mathbf{z}) \equiv H(\mathbf{x}) \pmod{31}$. So our task is to determine which points $\mathbf{x} = (x_1, x_2, x_3) \in R$ can be eliminated since $\mathbf{x} + \mathbf{z} \in R$ also. Now

$$\mathbf{x} = (x_1, x_2, x_3) \in [0, 3) \times [0, 5) \times [0, 4),$$

so $0 \le x_1 < 3$, and

$$\mathbf{x} + \mathbf{z} = (x_1, x_2, x_3) + (-2, 3, -3) \in [0, 3) \times [0, 5) \times [0, 4),$$

so $0 \le x_1 - 2 < 3$, or $2 \le x_1 < 5$. Both conditions on x_1 hold, so $2 \le x_1 < 3$, or $x_1 = 2$. Another way to express this is $\max\{0, -(-2)\} \le x_1 < \min\{3, 2 - (-2)\}$, and a better way is $\max\{0, -z_1\} \le x_1 < \min\{b_1, b_1 - z_1\}$. Similarly, the conditions on x_2 and x_3 imply that $0 \le x_2 < 2$ and $3 \le x_3 < 4$. So if $\mathbf{x} \in [2, 3) \times [0, 2) \times [3, 4)$, then $\mathbf{x}$ is redundant and can be discarded. This eliminates only two points, $(2, 0, 3)$ and $(2, 1, 3)$, but a similar analysis with $\mathbf{z} = (0, -2, -1)$ allows us to eliminate the 27 redundant points in $[0, 3) \times [2, 5) \times [1, 4)$.

Each of the two sets of redundant points, $[2, 3) \times [0, 2) \times [3, 4)$ and $[0, 3) \times [2, 5) \times [1, 4)$, is a box within $R = [0, 3) \times [0, 5) \times [0, 4)$, with sides parallel to the sides of R, and extending "upwards and away from" the origin. The vertex of $[2, 3) \times [0, 2) \times [3, 4)$ closest to the origin is $(2, 0, 3)$, so the redundant points to be

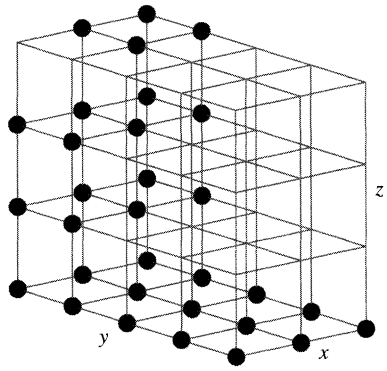


Figure 1 3-dimensional analog of Brauer and Shockley's L-shaped region

eliminated from R are $\mathbf{x} \geq (2, 0, 3)$. Similarly, we eliminate from R all $\mathbf{x} \geq (0, 2, 1)$. Geometrically, we chop subboxes out of box R, and the result is a 3-dimensional analog of Brauer and Shockley's L-shaped region in the plane, as in FIGURE 1.

Before leaving the example, we make four observations. First, the final zero $(0, -4, -2)$ generates no new redundant points since it is simply $2(0, -2, -1)$. Having one of the zeroes be a multiple of another was luck, not a general feature of the method. Second, the two sets of redundant points found are disjoint, but again this does not hold in general. Third, from the original set R containing 60 elements we have eliminated 29 elements, leaving 31, which is exactly the number of residue classes modulo 31 that are to be checked according to the characterization theorem. The number of points remaining will be at least a, and if there are more than a points, the congruent points that remain represent the same number. Finally, while $a = 31$ points remain at which the solution of the Frobenius problem might occur, $H(x_1, x_2, x_3) = 41x_1 + 47x_2 + 61x_3$ is an increasing function of each variable, so most of those remaining points can be ignored. We will exploit this observation later.

We now return to the general problem. Suppose that we seek $g(a, a_1, \ldots, a_{n-1})$, with $n \geq 3$. As described at the beginning of this section, determine $b_1, \ldots, b_{n-1}$ and the set of zeroes for the problem:

$$Z = \{\mathbf{z} \in (-b_1, b_1) \times \cdots \times (-b_{n-1}, b_{n-1}) | H(\mathbf{z}) \equiv 0 \pmod{a} \quad \text{and} \quad H(\mathbf{z}) < 0\}.$$

For each $\mathbf{z} \in Z$, eliminate from $R = [0, b_1) \times \cdots \times [0, b_{n-1})$ all points $\mathbf{x}$ for which $\mathbf{x} + \mathbf{z} \in R$ as follows. Let $\mathbf{x} = (x_1, \ldots, x_{n-1}) \in R$, let $\mathbf{z} = (z_1, \ldots, z_{n-1}) \in Z$, and suppose that $\mathbf{x} + \mathbf{z} \in R$. Then for $1 \leq k \leq n-1$, we have that $0 \leq x_k + z_k < b_k$, so $-z_k \leq x_k < b_k - z_k$, since $\mathbf{x} + \mathbf{z} \in R$. Also $0 \leq x_k < b_k$ since $\mathbf{x} \in R$. Thus $\max\{0, -z_k\} \leq x_1 < \min\{b_k, b_k - z_k\}$. This means that all points in

$$T(\mathbf{z}) = \bigtimes_{k=1}^{n-1} [\max\{0, -z_k\}, \min\{b_k, b_k - z_k\}]$$

are redundant, and consequently that $\bigcup_{\mathbf{z} \in Z} T(\mathbf{z})$ is the set of redundant points to be eliminated. Let T denote the complement of this union in R. It is T that generalizes the L-shaped region of Brauer and Shockley for the $n = 3$ case.

In the $n = 3$ case, T contains a points, one for each of the residue classes of representable numbers modulo a. For $n > 3$, T need not be so simple to describe geometrically, and it may include more than a points. By construction, the set T does not contain congruent points $\mathbf{x}, \mathbf{y}$ with $H(\mathbf{x}) \neq H(\mathbf{y})$, but this does not preclude the possibility of there being congruent points $\mathbf{x}, \mathbf{y} \in T$ with $\mathbf{x} \neq \mathbf{y}$ and $H(\mathbf{x}) = H(\mathbf{y})$ as occurs

in the last two examples of the next section. This causes no problem since both $\mathbf{x}$ and $\mathbf{y}$ yield the same representable number $H(\mathbf{x}) = H(\mathbf{y})$.

Generalized corner points Brauer and Shockley used the fact that $H(x, y)$ is an increasing function of both x and y to conclude that the maximum value of H on T occurs at one of the two corner points, which immediately yields the solution when $n = 3$. However, when $n > 3$, more care needs to be taken to find the ones to call "corner points." As previously noted, if $\mathbf{x} \leq \mathbf{y}$, then $H(\mathbf{x}) \leq H(\mathbf{y})$, so $\mathbf{x}$ can be ignored. Motivated by this observation, we say that $\mathbf{x}$ is *dominated* by $\mathbf{y}$ if $\mathbf{x} \leq \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$. Eliminate from T all points that are dominated by a point of T, and call the resulting set U; points in U will be called *corner points* of T for our problem. By the characterization theorem, the Frobenius number g for a particular choice of $a, a_1, \ldots, a_{n-1}$ is $g = \max\{H(\mathbf{x}) | x \in U\} - a$.

Returning a final time to the example that we have been following, the set of corner points is $U = \{(1, 1, 3), (2, 1, 2), (2, 4, 0)\}$, at which H achieves the values 271, 251, 270 (respectively.) So $g(31, 41, 47, 61) = 271 - 31 = 240$.

It was noted earlier that Lewin's algorithm [**11**] incorrectly works this example. That algorithm constructs a triangular table of representable numbers modulo 31, with an *a priori* bound on the number of rows needed to guarantee that all residue classes were represented. Lewin proved that as one moves from left to right or vertically downward, the represented numbers increase. He concluded that "the consecutive nature of the algorithm supplies a means ... for determining" the Frobenius number [**11**, p. 71]. That is, as one constructs the table from left to right, one row at a time, the last residue class modulo 31 to be represented will correspond to the maximum of the 31 minimum representable numbers modulo 31. This need not be so: $271 \equiv 23 \pmod{31}$ appears in row 5, column 14 of the table, and $270 \equiv 22 \pmod{31}$ appears in row 6, column 11 of the table, leading the algorithm to give the incorrect value $g(31, 41, 47, 61) = 270 - 31 = 239$.

Summary and examples

Assumptions Let $n > 1$. The n integers whose Frobenius number we seek are denoted $a, a_1, \ldots, a_{n-1}$, with $1 < a < a_1 < \cdots < a_{n-1}$. None of $a, a_1, \ldots, a_{n-1}$ is dependent on the others, no $n - 1$ element subset of $\{a, a_1, \ldots, a_{n-1}\}$ has greatest common divisor greater than 1, and the numbers are not in arithmetic progression.

The algorithm

1. For each k, $1 \leq k \leq n - 1$, find b_k, the smallest positive integer for the kth coordinate of $\mathbf{x} = (x_1, \ldots, x_{n-1}) \geq 0$ such that there is a solution of $H \circ \delta_k(\mathbf{x}) \equiv 0 \pmod{a}$ and $H \circ \delta_k(\mathbf{x}) > 0$. Let $R = [0, b_1) \times \cdots \times [0, b_{n-1})$.
2. Determine the set Z of *zeroes* for the problem, $Z = \{\mathbf{z} \in (-b_1, b_1) \times \cdots \times (-b_{n-1}, b_{n-1}) | H(\mathbf{z}) \equiv 0 \pmod{a}, H(\mathbf{z}) < 0\}$.
3. Determine T, the complement in R of $\bigcup_{\mathbf{z} \in Z} T(\mathbf{z})$, where
$$T(\mathbf{z}) = \bigtimes_{k=1}^{n-1} [\max\{0, -z_k\}, \min\{b_k, b_k - z_k\}).$$
4. Determine $U = \{\mathbf{u} \in T |$ there is no $\mathbf{t} \in T$ with $\mathbf{u} \leq \mathbf{t}$ and $\mathbf{u} \neq \mathbf{t}\}$, the set of *corner points*.
5. The Frobenius number for $a, a_1, \ldots, a_{n-1}$ is $g = \max\{H(\mathbf{x}) | \mathbf{x} \in U\} - a$.

The algorithm developed in this paper and the "brute force" algorithm were coded using *Mathematica* and run on a Macintosh. Both algorithms were used as an added check for accuracy since our results did not always agree with previously published results.

Despite the pejorative term brute force, this algorithm often solves the Frobenius problem quite quickly. A comparison of times to solve various problems using the two algorithms is problematic without knowing how efficiently each algorithm was coded, but qualitative information might be of interest. The brute force algorithm consistently found the Frobenius number much more quickly than the new algorithm when a (the smallest of the integers) was small, as in the example that we have been following through the article and in example 3 below. If the numbers were larger and comparable in size, as in example 2, the new algorithm outperformed the brute force algorithm by a wide margin.

EXAMPLE 1. $n = 3$ *and* $\{a, a_1, \ldots, a_{n-1}\} = \{137, 251, 256\}$.

This is a classic example solved by many authors in papers studying the problem of Frobenius; see [**1, 8, 17, 18**].

Solution. $b_1 = 13$, and $b_2 = 14$, so $R = [0, 13) \times [0, 14)$; the set of zeroes is $Z = \{(-8, -5)\}$; the set of corner points is $U = \{(7, 13), (12, 4)\}$, at which H achieves the values 5085 and 4036 (respectively). So $g(137, 251, 256) = 5085 - 137 = 4948$.

EXAMPLE 2. $n = 4$ and $\{a, \ldots, a_{n-1}\} = \{271, 277, 281, 283\}$. This example appears in [**12**], where an even earlier reference is given.

Solution. $b_1 = 2$, $b_2 = 6$, and $b_3 = 46$, so $R = [0, 2) \times [0, 6) \times [0, 46)$; the set of zeroes is $Z = \{(-1, -3, 3), (-1, -2, -43), (0, -5, -41), (1, -2, -44)\}$; the set of corner points is $U = \{(0, 4, 43), (0, 5, 40), (1, 1, 45), (1, 2, 42)\}$, at which H achieves the values 13293, 12725, 13293, 12725 (respectively). So $g(271, 277, 281, 283) = 13293 - 271 = 13022$. Notice that there are two corner points at which the maximum H-value is achieved.

EXAMPLE 3. $n = 5$ and $\{a, a_1, \ldots, a_{n-1}\} = \{34, 37, 38, 40, 43\}$.

Solution. $b_1 = 2$, $b_2 = 3$, $b_3 = 3$, $b_4 = 5$, so $R = [0, 2) \times [0, 3) \times [0, 3) \times [0, 5)$; the set of zeroes is $Z = \{(-1, -1, 0, -3), (-1, 0, -1, 1), (-1, 2, -2, -3), (0, -1, -2, -2), (0, -1, 1, -4), (0, 2, -1, -4), (1, -1, -1, -3)\}$; the set of corner points is $U = \{(0, 0, 2, 3), (0, 2, 0, 3), (0, 2, 1, 2), (0, 2, 2, 1), (1, 0, 0, 4), (1, 2, 0, 2)\}$, at which H achieves the values 209, 205, 202, 199, 209, 199 (respectively). So $g(34, 37, 38, 40, 43) = 209 - 34 = 175$.

This example also appears in [**11**], where the Frobenius number was found to be 163. Notice that $163 = 3 * 40 + 43$, clearly showing that result to be incorrect. For this example, Lewin reorders the numbers as $\{38, 34, 37, 40, 43\}$, observes that with $a = 38$ the remaining numbers form an arithmetic progression, for which his algorithm assumes a different form, producing the incorrect value 163. Curiously, Lewin's original algorithm applied to $\{a, a_1, a_2, a_3, a_4\} = \{38, 34, 37, 40, 43\}$ produces the correct value 175, but applied to $\{a, a_1, a_2, a_3, a_4\} = \{34, 37, 38, 40, 43\}$ produces 171.

REFERENCES

1. A. Brauer and J. E. Shockley, On a problem of Frobenius, *J. reine angew. Math.* **211** (1962), 215–220.
2. F. Curtis, On formulas for the Frobenius number of a numerical semigroup, *Math. Scand.* **57** (1990), 190–192.
3. J. L. Davison, On the linear Diophantine problem of Frobenius, *J. Number Theory* **48** (1994), 353–363.
4. W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume 1 (Third Edition), John Wiley & Sons, New York, 1968.

5. H. Greenberg, Solution to a linear Diophantine equation for nonnegative integers, *J. Algorithms* **9** (1988), 343–353.
6. B. R. Heap and M. S. Lynn, On a linear Diophantine problem of Frobenius: an improved algorithm, *Numer. Math.* **7** (1965), 226–231.
7. M. Hujter and B. Vizvari, The exact solution to the Frobenius problem with three variables, *J. Ramanujan Math. Soc.* **2** (1987), 117–143.
8. S. M. Johnson, A linear Diophantine problem, *Canad. J. Math.* **12** (1960), 390–398.
9. R. Kannan, Lattice translates of a polytope and the Frobenius problem, *Combinatorica* **12** (1992), 161–177.
10. E. Kunz, The value-semigroup of a one-dimensional Gorenstein ring, *Proc. Amer. Math. Soc.* **25** (1970), 748–751.
11. M. Lewin, An algorithm for a solution of a problem of Frobenius, *J. reine angew. Math.* **276** (1975), 68–82.
12. A. Nijenhuis, A minimal path algorithm for the "money changing problem," *Amer. Math. Monthly* **86** (1979), 10, 832–838.
13. H. O. Pollak, The postage stamp problem, *Consortium* **69** (1999), 3–5.
14. M. Raczunas and P. Chrzastowski-Wachtel, A diophantine problem of Frobenius in terms of the least common multiple, *Discrete Math.* **150** (1996), 347–357.
15. J. L. Ramírez-Alfonsín, Complexity of the Frobenius problem, *Combinatorica* **16** (1996), 143–147.
16. J. B. Roberts, Note on linear forms, *Proc. Amer. Math. Soc.* **7** (1956), 465–469.
17. O. J. Rödseth, On a linear Diophantine problem of Frobenius, *J. reine angew. Math.* **301** (1978), 171–178.
18. E. S. Selmer and O. Beyer, On the linear Diophantine problem of Frobenius in three variables, *J. reine angew.Math.* **301** (1978), 161–170.
19. H. S. Shulz, The postage-stamp problem, number theory, and the programmable calculator, *Math. Teacher* **92** (1999) 1, 20–26.
20. J. J. Sylvester, Mathematical questions, with their solutions, *Educational Times* **41** (1884), 21.

Transform Inverse

Once I sat in class so dreary,
while I pondered weak and weary.
Over maths both quaint and curious
of so soon forgotten lore—
While I nodded, nearly napping,
suddenly there came a tapping
As of someone gently rapping,
rapping on the classroom board.
"*The derivative*," he muttered,
rapping on the classroom board.
"*Only this and nothing more.*"

With all effort I could marshal,
there I stared—equations partial,
Etched in chalk across the endless,
black expanse of classroom board.
"*L denotes an operator,*
u is heat, we'll integrate or
Solve for u inside the bounds
b_1, b_2, b_3, b_4.
Now we need the temps initial:
You use u's sub-one through four."
He was raving, "*Want some more?*"

"*Using ξ might seem quite goofy,*
but will simplify the proof." He
Tried to justify this new ξ-
change. But why? But huh? Wherefore?
Warm, I felt, my heart slow-beating,
Still "$d\xi$" he kept repeating,
Still was bleating—oh! the agony! $dz d\xi$! Still more.

"*Here's the place.*" (perhaps he said
"*Laplace*"—I wasn't really sure.)
Still, still raving: "*Want some more?*"

"*Long-time heat is more involved. It*
's hard. Smooth operators solve it.
f with integrals—convolve it."
—What fresh L was f here for?
With a flick of chalk and loops he
changed the f to u, said: "*Oopsy!*"
u_0 is the state initial
that our plate attained before."
Too few letters left to use,
he wrote (I, woozy, wasn't sure):
Do u_0 ξ it's like b_4?"

Forum followed function's form; our
class, his plate approached the norm.
Our
Heat-death (b_0 proud!) came on,
the crowd became lukewarm.
Chalk and nablas, all were seen to
blur—regression to the mean. To
This he said "*This part gets tricky,*
so I'll wave my hands some more."
Time stood still. I heard the chalk, I
just stared blankly at the board—
Darkness there and nothing more.

—Glenn D. Appleby
Santa Clara University
Santa Clara, CA 95053

Leibniz, the *Yijing*, and the Religious Conversion of the Chinese

FRANK J. SWETZ
616 Sandra Avenue
Harrisburg, Pennsylvania 17109-5816
fjs2@psu.edu

Gottfried Wilhelm Leibniz was a synthesizer of ideas, a deep religious thinker and an avid Sinophile. Leibniz sought a *characteristica universalis*, or universal language, that would assist his quest to order all human knowledge. In his search he was drawn to the Chinese *Yijing*, or Book of Changes. The concepts and symbols of this ancient classic appeared to represent binary arithmetic, and through this suggested a mystical model that related God to creation. Through the *Yijing*, Leibniz also became involved with the Figurist and accommodationist strategies by which the Jesuits hoped to convert the Chinese people to Christianity. Here we examine this strange episode in the history of mathematics and discuss how Leibniz linked mathematics with philosophy, cosmology, theology, and metaphysics.

Perspective

Gottfried Wilhelm Leibniz (1646–1716) was a true polymath—a universal genius and intellectual leader of the European Enlightenment. Although usually recognized for his mathematical achievements in calculus, Leibniz is credited with diverse accomplishments beyond mathematics. In a time of rapidly changing social and political reforms, religious controversies, and fluctuating international alliances, Leibniz sought to consolidate, synthesize, and reconcile differing peoples and institutions. As a diplomat for the Hanoverian Court, he was involved in political intrigues. As a fervent religious thinker and ecumenicist, he sought religious reunification between Protestants and Catholics in the wake of the Thirty Years War. Although a Lutheran, Leibniz was widely respected in Catholic intellectual and political circles. In fact, while conducting research in Rome, he was even offered the Curatorship of the Vatican Library [**3**, p. 159], a position usually reserved for a Cardinal. He declined the position. Noted as a natural philosopher and scientist who shed the bonds of Scholasticism, he was also an active experimenter and inventor [**1**, pp. 87–90].

Despite his seemingly progressive outlook, Leibniz was a man of his times, involved in two contrasting worlds. Leibniz was drawn into an emerging era of scientific method, observation, and logical deduction, and yet simultaneously he was mired in the classical world of traditional authority. The rise in the sixteenth and seventeenth centuries of a scientific outlook, typically expressed mathematically, focused scholarly attention on the properties and relationships of number and geometric shape. Mystical beliefs from Neopythagoreanism as well as Neoplatonism were resurrected and incorporated into new theories. In his metaphysical and religious thinking, Leibniz was deeply influenced by two Catholic mystics, the Spanish Franciscan Raemon Lull (ca. 1232–1316) and the German Friar Nicholas of Cusa (1401–1464).

Lull was an intellectual rebel who explored the Hebrew Kabbala and delved into the works of Arabic authors. He attempted to develop an "art of finding the truth" by which he could arrive at a universal language, and through it a universal faith that would unite the major monotheistic religions: Judaism, Christianity, and Islam. Combining

logic with faith, the Catalan monk believed that everything could be related to God by examining how Creation was structured. Further, he theorized that God possessed nine specific truths [Dignites] which, when combined, six at a time, formed objects of Creation [**27**, p. 175]. Lull employed combinatorial tables and movable wheels to exhibit these objects. Such a combinatorial approach to understanding the Deity and His works had been pioneered in the Kabbala. Lull's theory was enunciated in *Ars combinatoria* [Combinatory Art] (ca. 1273) [**10**]. Henceforth, those who believed that all knowledge could be unified into a single, structured system were known as Lullists. Gottfried Wilhelm Leibniz was a Lullist.

Nicholas of Cusa was another metaphysical thinker and anti-traditionalist. His theories were frequently based on mathematical analogies applied in religious and philosophical contexts. Holding to a "concordance of opposites," he believed that apparent contradictions unite at infinity. Thus, the largest possible number must coincide with the smallest possible, one. Further, since numbers were discrete entities, all were contained in the ultimate unity and could simultaneously be produced from that unity. For Nicholas, God was infinity but also the unity that generated infinity. In an alternate geometric analogy, Friar Nicholas likened God to an infinite circle or sphere and at the same time, the center or generator of that circle or sphere. Such mathematical imagery appealed to Leibniz who believed God to be the primary unity, and who also utilized a sphere-circle analogy [**20**, p. 82].

Leibniz sought order, structure, and harmony in the world around him and labored that others might also perceive such patterns. His interest in combinatorics and probability theory lay in the fact that he believed they could render intelligible and rational what might otherwise appear random in nature. In his own *Ars Combinatoria* (1666), he discussed the development of an analytic language of reasoning, a *characteristica universalis*, which could express concepts as combinations comprising elements from a set of basic human thoughts. Thus, he delved into the structure of logic and language and sought a key for all knowledge [**17**]. If nature was mathematical, then the acts of creation had to be mathematical. Thus, the image of God could also be understood in the light of mathematics. In such a scenario, for Leibniz, reason and faith were combined. Of course, this conception was not new. In the fourth century, Augustine sanctioned Christian involvement with numerology by citing Wisdom 11:20 that God "has ordered all things in number, measure, and weight" [**4**, vol. 9, p. 73] and established the Deity in medieval iconography as the Great Mathematician. However, Leibniz was moving beyond a mere image of God. He sought to understand the relation of his God to all Creation.

Described by Frederick the Great as an academy within himself [**17**, p. 7], Leibniz was encyclopedic in his accumulation of knowledge and the dispensing of advice. The philosopher was an avid letter writer who sought the news on current events scientific, political, and philosophical. It would be through his correspondence with a French missionary in China that the philosopher-diplomat Leibniz would be drawn into a theological/cosmological controversy whose outcome, Leibniz believed, might conceivably convert the Chinese empire to Christianity.

Jesuits in China, Figurists, and the *Yijing*

By the seventeenth century, Europeans had known of the existence of China for centuries, but their knowledge of this strange land remained vague and fragmentary. Marco Polo, in the thirteenth century, supplied eyewitness reports of Cathay [**34**]. He described the greatest empire that had ever existed, ruled by the most powerful emperor in all of history. Polo's wondrous and, perhaps, fictional depiction fired the

imagination, but also raised further questions about this land, its people and customs. Answers to such questions could not be satisfied until the 1497–98 Portuguese voyages of Vasco de Gama opened direct trade routes to the Far East. Merchants, and then missionaries began to visit, observe and report on the Celestial Kingdom of the Chinese. Members of the Society of Jesus (the Jesuits), penetrated China in 1581 and by 1601, led by Matteo Ricci (1552–1610), had made their way to Beijing [**38**]. The Jesuits were an educational order of priests and an intellectual elite of the Catholic Church. They were schooled in theology and philosophy, but were also conversant in secular subjects that allowed them to be teachers and professors at institutions of higher learning. Their Chinese hosts were little attracted to the religious tenets and mystical doctrines of these foreign visitors, but they did appreciate the missionaries' scientific expertise and its possible use for the Empire [**32**]. In turn, the Jesuits found the Chinese to be a very disciplined and moral people, already possessing many of the spiritual attributes the Jesuits wished to advocate. As a result, the missionaries' strategy of proselytizing became two-fold: first, to ingrain themselves into the Chinese culture, hoping to find specific links between Chinese civilization and Christianity, and, second, to disseminate information back home in Europe, so as to maintain regal and popular support for their Chinese Mission [**27**].

Ricci and his colleagues became proficient in the Chinese language, calligraphy, customs, and its classical literature. They were accepted by the Emperor into the Court bureaucracy as mandarins and advisors. Among Matteo Ricci's contributions to this effort was a translation of the first six books of Christopher Clavius' *Euclidis Elementorum libri* XV (1574) into Chinese [**25**, p. 112]. Although a few Chinese scholars admired the organizational format of the *Elements*, the work had little impact on the Chinese use of geometry or their mathematical thinking. The remaining parts of Euclid did not appear in the language of the Middle Kingdom until 1851 [**41**].

Ricci believed that the Chinese society, so ancient and ethical in character, must have been exposed in antiquity to a Divine revelation, similar to that of the Judeo-Christian tradition. The Chinese were really lost "spiritual brothers" of European Christians. Thus, he felt that both Christians and Chinese shared a common "ancient theology." If this fact could be made obvious to the Chinese, he felt, perhaps they could be more easily converted. This belief has become known as *accommodationism* [**28**]. Other Catholic missionaries in China, for instance the Dominicans and Franciscans, as well as Church authorities, opposed such an approach to evangelization. Eventually, this theological issue would be decided by authorities in Rome.

Upon request from the Chinese Mission in 1685 for more scientific expertise, responding in the same year, Louis XIV of France sent a party of six Jesuit scientists to China. They embarked, accompanying the first French ambassador to the court of Siam, where one of the Jesuits was retained by the King of Siam. The remaining five mathematicians *de Roi* arrived in China in 1688. Among this delegation was Joachim Bouvet (1656–1730), a talented and versatile scholar who became the personal mathematics tutor to the ruling Kangxi Emperor and his children [**30**]. Bouvet soon became attracted to the metaphysical theories and cosmological beliefs contained in the Chinese classics. In particular, the alleged most ancient of the Chinese texts, the *Yijing* (known as the *Book of Changes*), (recognized as the I Ching under the old Wade-Giles system of transliteration of Chinese terms) held a special fascination for him. Bouvet believed this handbook of divination to be an instrument of Divine revelation, and was perhaps the missing link—the ancient theology—that would spiritually draw Christians and Chinese together.

Bouvet was led to believe that the origin of the *Yijing* rested with Fuxi (ca. 3000 BCE), said to be the first Emperor of China who, while on the bank of the Yellow River, experienced a strange visitation. The Emperor was confronted by a "dragon-

horse." In its track, the horse left eight distinct line configurations. Each configuration was a trigram composed of three line segments. These segments were either whole or broken into two halves. The complete set of configurations became known as the *bagua* or the Eight Trigrams. From the *bagua*, the Chinese believed Fuxi obtained language and a knowledge of all things [**40**, pp. 53–56]. The Chinese wood block print in FIGURE 1 depicts Fuxi residing over the *bagua*.

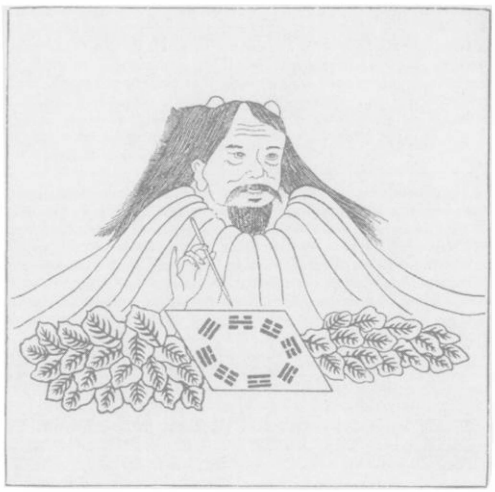


Figure 1 Fuxi and the *bagua*

During the historical period of the Zhou Dynasty (ca. 11 century BCE), the *bagua* were formalized (actually combined) into a more extensive system of six-line cryptic characters called *hexagrams*. Hexagrams and their interpretation became the substance of the *Yijing* [**48**]. While the *bagua* were communicated to Fuxi, it was believed that they ultimately came from a supreme universal power, the *Taiji*, which controlled all things through a system of dual, complementary forces, the *yin* and the *yang*. *Yin-yang* forces were graphically symbolized by broken and solid line segments respectively. When these forces were combined three at a time, they constituted the eight basic *bagua*, which in turn, combined with each other to form 64 hexagrams. FIGURE 2(top) illustrates the *yin-yang* evolution of the *bagua*; FIGURE 2(bottom) illustrates the standard *Yijing* tabulation of the hexagrams. (Each hexagram is read from the bottom upwards. Numbers indicate the sequential ordering of the hexagrams—note that the chart is read right to left, from the bottom to the top).

Since the source of all knowledge in the wise and ancient Celestial kingdom was traced to Fuxi and the *Yijing*, Bouvet considered this sage king to be the spiritual and intellectual father of China. Could he perhaps also be the Adam of Judeo-Christian belief? If so, then the *Yijing*, if properly understood, could yield the ancient theology sought by religious thinkers. In the seventeenth century, scholars who believed that parallels to Judeo-Christian beliefs, and even actual biblical images, could be found in the traditional literature of other cultures were called *Figurists* [**33**]. Joachim Bouvet was a Figurist who saw the tenets of Christian revelation prefigured within the Confucian classics. As he confided to a European correspondent, the Abbé Bignon in Paris:

> There is no mystery in the Chinese religion, no dogma in our theology, no maxim in the holiness of our morality, that is not expressed in these [Chinese] books with surprising clarity [**6**].

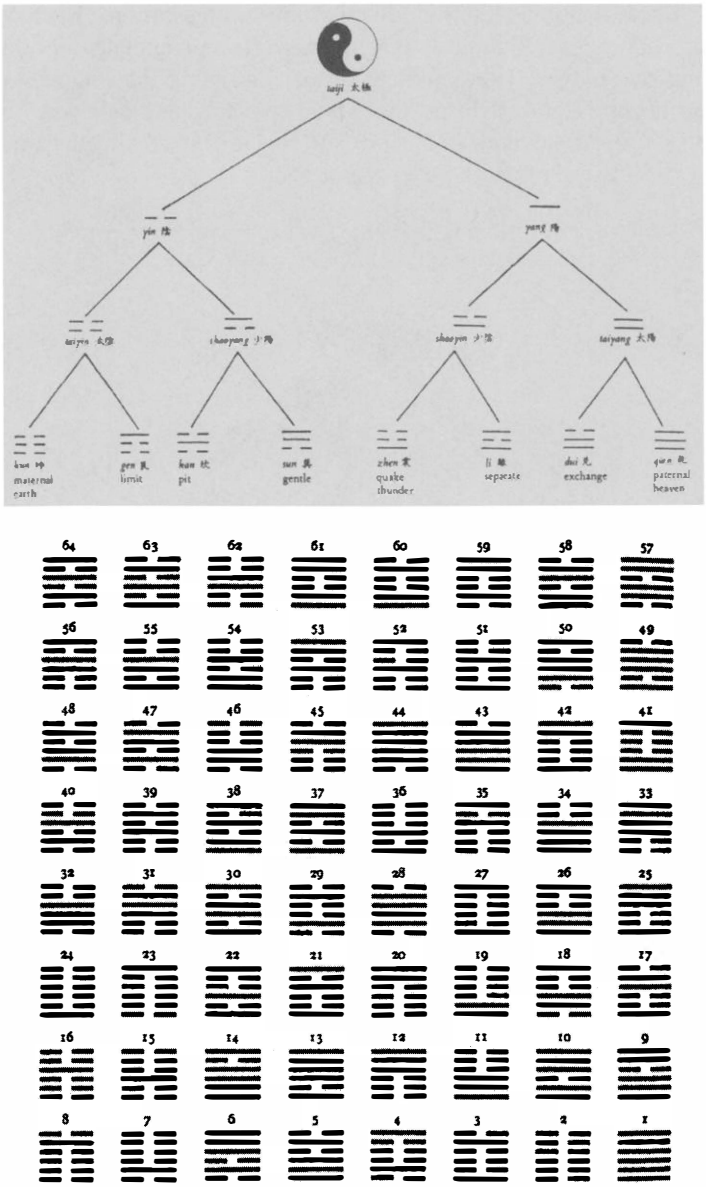


Figure 2 Evolution of the *bagua* and the tabulation of the hexagrams

In particular, he felt that the *Yijing* could supply the necessary link for conversion of the Chinese. Eventually, the missionary would find powerful support for his theories.

Leibniz's binary arithmetic

If God was the Great Mathematician, as many theologians/philosophers of the time believed, what numbers did He use to design the universe? The general concept of a number base, that is, the expression of numbers as the sums of powers of a particular number (the base), was known in the sixteenth century, including the use of bases other than ten [**13, 16**]. The English mathematician, Thomas Harriot (1560–1621) ex-

perimented with various nondecimal number systems including the binary system, but found no practical use for it at the time [**35**]. The Spanish bishop Juan Caramuel y Lobkowitz (1606–1682) discussed bases two through twelve in his *Mathesis bisceps* of 1670. Yet, what base would God use? What number base best exemplified Divine efficiency? Erhard Weigel (1625–1699), with whom Leibniz had studied in 1663, believed the answer was base four. Possibly this decision rested with the popular cosmological belief that all creation arose from the combination of four primal elements: Fire, Water, Air, Earth [**40**, p. 159]. Weigel published his base four theories in *Tetractyn* (1672) and later expanded them in *Philosophia mathematica* (1693).

Leibniz preferred binary (or *dyadic*) arithmetic, which to him seemed the simplest and most efficient system. In a paper found in his notes entitled *De Progressione Dyadia* and dated March 15, 1679, he discussed binary mathematics and constructing a binary digital calculating machine. Within a year, Leibniz outlined plans for such a machine to Duke Ernst August of Hanover, but also noted the technological difficulties in its construction. Writing in 1682 to his correspondent and collaborator, the mathematician Ehrenfried Walther Tschirnhaus, Leibniz discussed the "binary progression" and its possible applications in number theory. In particular, he was fascinated by the periodic behavior exhibited within the column entries of the vertical array of binary strings representing a known sequence of numbers such as the natural numbers, their squares, cubes, figurate numbers, and others. He believed such patterns would be useful and carried on a lengthy correspondence with Jacques Bernoulli on this matter [**13**, p. 45]. These ideas were discussed in a 1710 article in *Miscellanea Berolensia* [**9**]. However, Leibniz's interest in binary arithmetic was more than just mathematical. It was metaphysical and theological as well. In a 1679 paper entitled "On the General Characteristic" he affirms his faith in the transcendent power of numbers:

> There is an old saying that God created everything according to weight, measure and number. But there are things which cannot be weighed, those namely which have no force or power. There are also things which have no parts and hence admit of no measure. But there is nothing which is not subordinate to number. Number is thus a basic metaphysical figure as it were, and arithmetic is a kind of static of the universe by which the powers are discovered [**20**, p. 22].

For Leibniz, creation was creation *ex nihilo*, creation out of nothing: the primary void was represented by 0 and the creator God by 1, and the forming of the cosmos resulted from the coming together of 1 and 0. A glimpse of the theological implications of this duality is found in his *Of the True Theologica Mystica* (ca. 1690) where he noted:

> All creatures are from God and from nothing; their self-being is from God, their unbeing is from nothing. This is demonstrated in a wonderful manner by numbers, and the essence of things resembles number. True self-knowledge consists in distinguishing precisely our self-being from our unbeing [**26**, p. 148].

The first formal announcement of this mathematical and theological conception was given in 1697 by Leibniz in a congratulatory birthday message to his patron, Rudolph August, Duke of Brunswick. In this letter he suggests casting a commemorative medallion bearing a likeness of the Duke and promoting a binary theme, *Imago Creatonis* [in the image of creation]. See FIGURE 3. Clarifying his design, Leibniz writes:

> Because one of the main points of the Christian Faith, and among those points that have penetrated least into the minds of the worldly-wise and that are diffi-

Figure 3 Leibniz's design for a commemorative medallion

cult to make with the heathen is the creation of all things out of nothing through God's omnipotence, it might be said that nothing is a better analogy to, or even demonstration of such creation than the origin of numbers as here represented, using only unity and zero or nothing. And it would be difficult to find a better illustration of this secret in nature or philosophy; hence I have set on the medallion design IMAGO CREATONIS [in the image of creation].

It is no less remarkable that there appears there from, not only that God made everything from nothing, but also that everything that He made was good; as we can see here, with our own eyes, in this image of creation. Because instead of there appearing no particular order or pattern, as in the common representation of numbers, there appears here in contrast a wonderful order and harmony which cannot be improved upon. Inasmuch as the rule of alternation provides for continuation, so that one can write out without computation or the aid of memory as far as one wishes, if one alternates the last place 0,1,0,1,0,1, etc., putting these under each other; and then putting under each other in the second place (from the right) 0, 0, 1, 1, 0, 0, 1, 1, etc.; in the third 0, 0, 0, 0; 1, 1, 1, 1; 0, 0, 0, 0; 1, 1, 1, 1; etc; in the fourth 0, 0, 0, 0, 0, 0, 0, 0; 1, 1, 1, 1, 1, 1, 1, 1; 0, 0, 0, 0, 0, 0, 0, 0; 1, 1, 1, 1, 1, 1, 1, 1; and so forth, the period or cycle of change becomes again as large for each new place. Such harmonious order and beauty can be seen in the small table on the medallion up to 16 or 17; since for a larger table, say to 32, there is not enough room. One can further see that the disorder, which one imagines in the work of God, is but apparent; that if one looks at the matter with the proper perspective, there appears symmetry, which encourages one more and more to love and praise the wisdom, goodness, and beauty of the highest good, from which all goodness and beauty has flown. I am corresponding with Jesuit Father Grimaldi, who is currently in China and also president of the Mathematics Tribunal there, with whom I became acquainted in Rome, and who wrote me during his return trip to China from Goa. I have found it appropriate to communicate to him these number representations in the hope, since he had told me himself that the monarch of this mighty empire was a lover of the art of arithmetic and that he had learned to figure the European way from Father Verbiest, Grimaldi's predecessor, that it might be this image of the secret of creation which might serve to show him more and more the excellence of the Christian faith [**13**, pp. 31–33].

From this passage it is evident that, by this time, Leibniz had investigated the properties of binary arithmetic, attached a religious-metaphysical significance to the numbers 0 and 1, was fascinated by the mathematical possibilities of the patterns of peri-

odicity evident in the sequential ordering of binary numbers, and had communicated his thoughts on binary numbers to the Jesuits in China in order that they might use his theories to impress the emperor.

Leibniz's binary arithmetic and Fuxi

Fascinated by the strange and new, Gottfried Leibniz became a ready and avid Sinophile and enthusiastically gathered information on this curious land. For example, in 1668, he published a comment on Chinese medical practices [**20**, p. 11]. China's influence was soon felt in the philosophical and political issues of the day. Could the Chinese language, whose written words are composed of simple strokes, be the much sought after universal language? An Englishman, John Webb, thought so and published his thesis on this subject in 1669 [**47**]. Leibniz seemed to share this opinion and sought out information on the Chinese language. In 1674, Andreas Muller, a Berlin scholar, composed *Clavis Sinica* [Key to the Chinese Language]. Leibniz submitted a series of fourteen questions to the author about the nature and composition of written Chinese. Eventually, he would be able to seek information on the Chinese language directly from the Chinese Mission. In 1689, while visiting Rome, Leibniz befriended Claudio Filippo Grimaldi, a Jesuit who had served seventeen years as a missionary in China. Grimaldi continued to correspond with Leibniz after he returned to the Chinese Mission as its director. By April of 1697, Leibniz had accumulated enough information on China from Grimaldi and others to publish *Novissima Sinica* [Latest News from China] [**20**, p. 45–59]. Joachim Bouvet, S.J., read the *Novissima* while he was in Paris on furlough from the Chinese Mission. He wrote to Leibniz on October 18 and expressed his admiration for the publication and shared some "recent" China news. In this correspondence, Bouvet also included a biography he had written of the reigning Chinese Kangxi Emperor [**5**]. Leibniz was delighted to obtain a new China correspondent and answered this letter on December 2 of the same year [**42**]. He wrote to Bouvet of his quest for the *characteristica universalis*, noting its possible use for conveying the concepts of Christianity to the Chinese:

> According to my opinion the only real demonstration is by numbers and algebraic expressions. If we could explain by them abstract ideas, that would be the most scientific method. . . . With this method we could explain our ideas even to people of other tongues living in far-away places, and it would be of unimaginable usefulness. If we, for instance, should like to universally propagate the ideas of natural religion which evolved out of revealed religion, no method would be better to adopt [**49**, p. 215].

In a letter written February 28, 1698, while Bouvet was still in France, he expressed his Figurist beliefs to Leibniz, a receptive correspondent. In particular, Bouvet stressed the primacy of Fuxi and the *Yijing* with its hexagrams as a source of divinely revealed knowledge. Further, he referred Leibniz to a table of the hexagrams found in Philippe Couplet's *Confucius Sinarum Philosophus*, which had appeared in Paris in 1687 [**5**]. Leibniz became intrigued with this theory. Bouvet had initially learned of Leibniz's work with binary arithmetic from Grimaldi, but in a letter of February 15, 1701, Leibniz detailed his theory stressing the analogy between God (the primary unit), who created the universe out of nothing, and binary arithmetic. He included a table of binary numbers up to 32. Bouvet immediately noted the correspondence between the binary 0 and 1 and the broken (*yin*) and solid (*yang*) lines of Fuxi's trigrams. Bouvet's Fig-

urist theories were further confirmed! Responding to Leibniz on November 4, 1701, he announced this discovery:

> I am not at all surprised at the Characteristic plan that you are proposing for representing thoughts so that the same characters serve all together for calculating and for demonstrating in reasoning, etc. For this genre of writing seems to contain the true idea of the ancient hieroglyphs and of the Kabbala of the Hebrews as well as the characters of Fu Hsi who is regarded by the Chinese as the first inventor of letters or hieroglyphs of this nation, in the formation of which one commonly says that he used the 64 combinations of his system of whole and broken lines [**27**, p. 205].

Accompanying this letter, Bouvet sent a block print illustration of the "Natural Order" configuration of the *Yijing* hexagrams [**1**, p. 246]. This configuration was directly attributed to Fuxi and known to the Chinese as *Xiantiantu*, the "prior to heaven system." However, it was not the system commonly employed by the Chinese for *Yijing* interpretations. They preferred the *Houtiantu*, or later heaven system honorifically attributed to Wenwang (King Wen), founder of the Zhou dynasty (11th century BCE). The "Natural Order" (Bouvet's term) supplied the most obvious correspondence to Leibniz's system. For Leibniz's reference, the missionary denoted the top and bottom of the configuration in Greek: $\alpha\nu\omega$ [ano] and $\kappa\alpha\tau\omega$ [kato] respectively. See FIGURE 4.

By this time, two separate copies of the "Wen arrangement" of the *Yijing* hexagrams had been published in Europe. One was Martino Martini's *Sinicae Historiae decas prima* of 1658 [**20**, p. 66], and the other was Couplet's *Confucius Sinarum philosophus* of 1687 [**51**, p. 79]. In his previous correspondence, Bouvet had called Leibniz's attention to Couplet's illustration of the hexagrams as shown in FIGURE 5. Leibniz must have observed this diagram previously but did not notice its possible connection to binary numbers. Bouvet was the first to make this association.

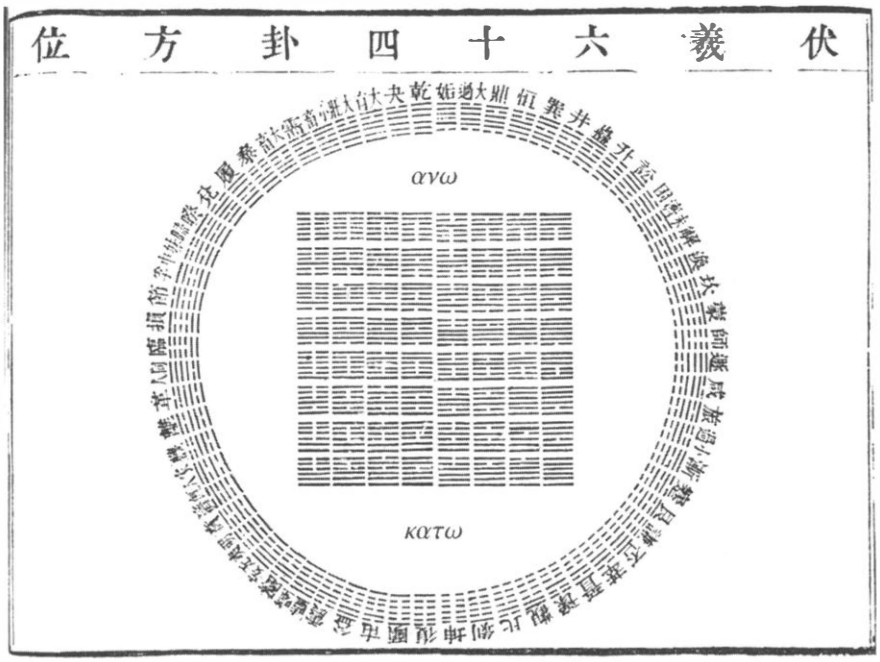


Figure 4 The "Natural Order" of the *Yijing* hexagrams

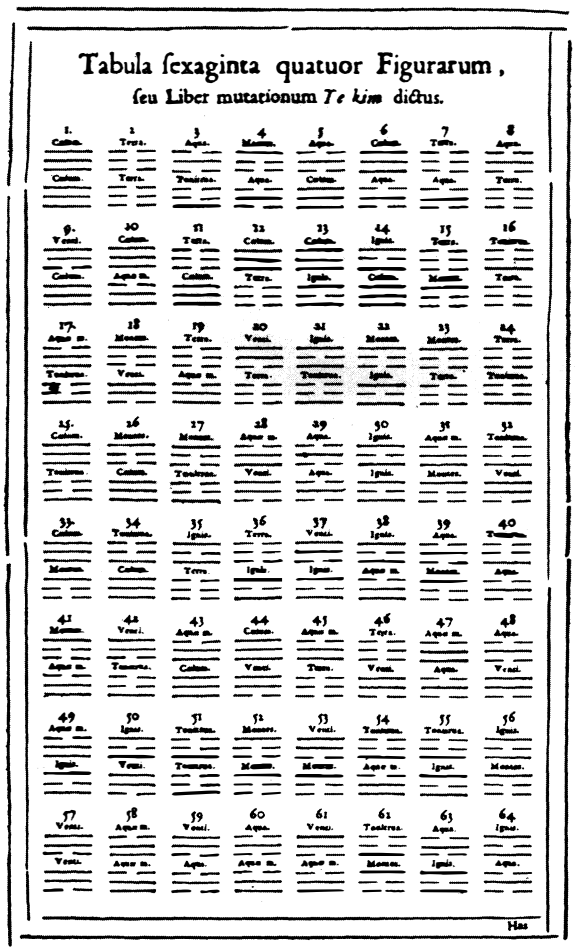

Tabula ſexaginta quatuor Figurarum,
ſeu Liber mutationum *Te kim* dictus.

Figure 5 Couplet's Table of Hexagrams

Leibniz studied the diagram of Fuxi's characters, and annotated his copy in red ink. Above each hexagram he inscribed its decimal value, and noted that within the square array the numbers followed a European ordering: from left to right, top to bottom. The circular configuration, however, confused him at first. Dividing the circular illustration in half vertically, Leibniz denoted the numbers 0 to 31 on the right half, beginning at the bottom and proceeding counterclockwise to the top; the ordering for numbers of the left half began with 32 at the bottom and advanced clockwise to 63 at the top [**1**, p. 246]. He believed that the circular arrangement was meant to represent the earth. Actually, in traditional Chinese cosmology, the earth is represented by a square, the heavens by a circle. Thus, the illustration in a Chinese context was intended to show that the hexagrams control both heaven and earth.

Leibniz responded to Bouvet and expressed his delight that the ancient figures of Fuxi corresponded so well with his own binary arithmetic. He even speculated that perhaps Fuxi, in his formulation of the foundational trigrams, had the biblical Genesis creation saga in mind:

> For 0 can symbolize the void which preceded the creation of heaven and earth .... At the beginning of the first day 1 existed, that is to say God. At the beginning of the second day, heaven and earth, being created on the first [that is, the end

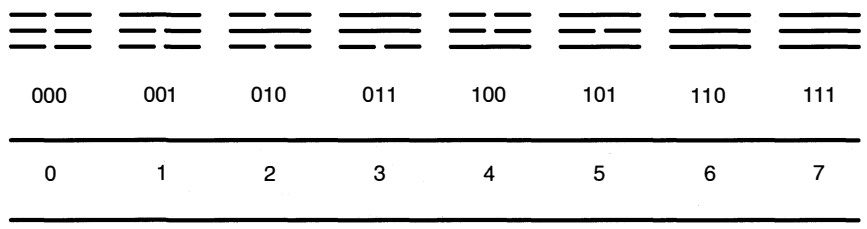
of the first day]. Finally at the beginning of the seventh day everything already existed. This is why the last [day] is the most perfect and the Sabbath, for all is created and complete. Thus 7 is written as 111 without 0. And it is only in this way of writing by 0's and 1's that we see the perfection of the seventh [day] which is considered holy. And it is even more remarkable that its character has some relation to the Trinity [**3**, p. 89; **5**].

Leibniz was especially elated that he had rediscovered this knowledge and further suggested that the fact of this European achievement would increase the respect of the Chinese for European science and the Christian religion, noting:

I believe that the scholars of China, when they become well versed in the consideration and above all see the artifice of Fu Hsi conforming to ours, they will be sufficiently inclined to believe that this great man had also wished to represent God as the author of all things, and the creation by which He drew those things out of nothingness. Thus this could be one of the most important articles of your catechism, taken from the classical authors of China and worthy of being explained to the emperor himself [**5**].

In 1702, Leibniz had acted to make his discoveries on binary arithmetic more widely known in the scientific community. He sent an announcement of his work and theories to Bernard de Bovier de Fontenelle, Secretary of the Académie Royale des Sciences, and included a proviso that no official mention of it be made until he could supply an application. By April of 1703, he had that application—the "Figures of Fohy"[Fuxi]. He announced his discovery to two correspondents: his friend, Carlo Mauritio Vota, Confessor of the King of Poland, and Hans Sloane, Secretary of the British Royal Society [**2**]. Leibniz also sent to Abbé Jean-Paul Bignon, the President of the Académie, an expanded version of his announcement for publication. That year's volume of the Mémoires [**22**] contained the communication by Gottfried Wilhelm Leibniz, "Explication de l'arithmetique binaire" in which he explained his binary notation, advised on its advantages for scientific investigations and expounded on its "Chinese connection."

What is astounding in this reckoning is that this arithmetic by 0 and 1 happens to contain the secret of the lines of an ancient king and philosopher named Fohy [Fuxi], who is believed to have lived more that 4000 years ago, and whom the Chinese regard as the founder of their empire and their sciences. There are several linear figures, which are attributed to him. They all come back to this arithmetic, but it suffices to show here the Figures of the Eight Cova [bagua], they are called, which pass as fundamental, and to adjoin to them the explanation, which is manifest provided that one notices firstly that a whole line —— means unity or 1 and secondly, that a broken line — — means zero or 0.



The Chinese lost the significance of these Cova or Lineations of Fohy, perhaps more than a 1000 years ago. They have made Commentaries on that, in which

they have gathered I know not what far out meanings. They are of a sort that it is necessary that their true explanation now come to Europeans. It is hardly more than two years ago that I sent to Reverend Father Bouvet, the celebrated French Jesuit, who resides in Peking, my method of counting by 0 and 1. He needed nothing further to make the observation that this was the key to the Figures of Fohy. When thus he wrote me on November 4, 1701, he sent me this princely philosopher's Grand Figure, which goes to 64 lineations and leaves no room for doubt about the truth of our interpretation, which is such that one could say that this Father has deciphered the Enigma of Fohy with the aid of that which I had communicated to him. Since these Figures are perhaps the most ancient monument of science which exists on this earth, this restitution of their meaning, after so long an interval of time, would seem most curious.

The agreement between the Figures of Fohy and my Table of Numbers is seen better if the initial zeros are supplied, which may seem superfluous, but which serves better to mark the periods of the columns, as I have supplied them in effect with little circles, to distinguish them from the necessary zeros. This accord gives me a high opinion of the profundity of the meditations of Fohy, because that which seems easy to us now was not so in those far-removed times. The binary or dyadic arithmetic is, in effect, very easy today with little thought going into it, because our manner of counting is conducive to it; it seems that one cuts off only the excess of it. But this ordinary arithmetic by tens does not seem very ancient, at least the Greeks and the Romans had ignored it, and have been deprived of its advantages. It seems that Europe owes its introduction at the time of Pope Sylvester II, to Gerbert, who had seen it with the Moors of Spain.

Now, as one believes in China, that Fohy is also the author of ordinary Chinese characters, which were severely altered in subsequent times, his essay on arithmetic calls for this judgment: it might well be possible to uncover again some considerable things by way of the rapport between the numbers and the ideas, if one could unearth the foundations of this Chinese writing, which, much more than is believed in China, has consideration of numbers established in itself. R.P. Bouvet has strongly urged to push this point and to expect good this kind. However, I do not know if there was ever an advantage in this Chinese writing approaching that which ought to exist necessarily in the feature that I project. It is that all reasoning, which one can pull from ideas, might be pulled from their characters by a manner of reckoning, which would seem one of the most important aids to the human intellect [**12**, pp. 41–42]

In his future writing, Leibniz continued to support an accomodationist approach for the conversion of the Chinese, and, more particularly, a Figurist position based on a mathematical/theological interpretation of the "Figures of Fohy" [**20**, p. 73].

Results of the cooperative alliance

Encouraged by Leibniz and hoping to pursue his Figurist theories further, Bouvet continued studying the *Yijing* and other ancient Chinese literary works. However, in 1708 he undertook a geographical and cartographical survey of China, which lasted seven years and consumed much of his time. In 1711, aided by intervention of the Chinese Emperor, Bouvet was joined in his *Yijing* research efforts by another French Jesuit, Jean-Francois Foucquet. A year later, they officially presented their theories of a *Yijing*-Christian synthesis to court scholars who, while sympathetic, remained unmoved by the suggestion of any Christian connection with the ancient work. Leibniz died in

1716 and did not see the playing out of this ecumenical intrigue, nor did he and Bouvet ever meet in person. Bouvet remained unsuccessful in winning converts by his Figurist approach, and died in 1730.

Associating the Figures of Fohy [Fuxi] with binary numbers ultimately proved to be an intellectual and mathematical cul-de-sac. Nevertheless, for a brief time, two creative thinkers believed they had discovered a mystical link binding the orient with the occident, the past with the present, and the Confucian society of China with the Christian civilization of Europe. Leibniz, the philosopher, thought he had found an ultimate source of human knowledge and the key to the much sought after *lingua universalis*. Bouvet, the Figurist, believed Fuxi to be the first prophet, the recipient of divine revelation.

Underlying their eagerness to become involved with Fuxi and the *Yijing* was an embedded and intriguing notion that they were rediscovering "lost knowledge." Many European scientists and philosophers of this period believed in the concept of a *prisca sapientia*, an ancient, advanced wisdom that existed among early priest-scientists such as the Chaldeans in Babylonia and the Pythagoreans in Greece. Over the centuries, this knowledge had become lost, but starting with the European Renaissance, was gradually being rediscovered. Descartes and Newton made references to their uncovering of the theories and methods of the ancients. (Even today the concept of *prisca sapientia* occasionally appears in popular writing. See, for example, Hancock [**14**]). Leibniz was certainly inclined to this way of thinking, as we see in his 1703 reply to Bouvet:

> Everyone believes that Fu Hsi [Fuxi] was one of the old Emperors of China, one of the best known philosophers of the world, and at the same time the founder of scholarship in the Chinese Empire and in the Far East. His I-ching table, handed down to the world, is the oldest monument of scholarship. But as far as I can see, for several thousand years there was nobody in a position to elucidate this four thousand year old object. Now it turns out that it is in absolute conformity with my new arithmetic. That I should have been able, after having received your letter, to solve it immediately and fittingly, whilst you have exerted all your knowledge in explaining its signs—that is really unimaginable [**49**, p. 215].

Bouvet, for his part, was driven by the possibility of revealing an "Ancient Theology." In their collaboration, they complemented and reinforced each other's beliefs. Further, unfortunately, faulty Sinology also deceived them.

Fuxi was a purely mythological being with no known historical foundation. Ancient Chinese chronologies abound with spurious ancient origins conceived to increase the prestige of sacred concepts and institutions [**7**]. *Yin-yang* concepts and symbols evolved over a period of centuries from a system of divination based on the casting of a set of rods. Resulting rod configurations were interpreted in two ways: *yin* or *yang* [**11**]. The actual concept of *yin-yang* is attributed to Boyang Fu (ca. 8th century BCE). It was incorporated into the *Yijing* but more fully developed in the writings of the scholar Zou Yan (ca. 350–270 BCE) and he is popularly credited with being the founder of this theory [**40**, p. 31].

The arrangement of the hexagrams sent to Leibniz by Bouvet, considered to be Fuxi's creation, was known to the Chinese as the prior to heaven system, and was not the hexagram ordering commonly used by the Chinese at this time. It is now realized that this ordering was due to the Sung philosopher Shao Yong (1011–1077). It first appeared in his *Huangji jingshi shu*, [Book of Sublime Principle Which Governs All Things Within the World], (1060) [**36**]. While based on a systematic and combinatorial arrangement, the ordering was arrived at without knowledge of binary numbers [**31**]. In their eagerness to see a binary number association in the Fuxi arrangement

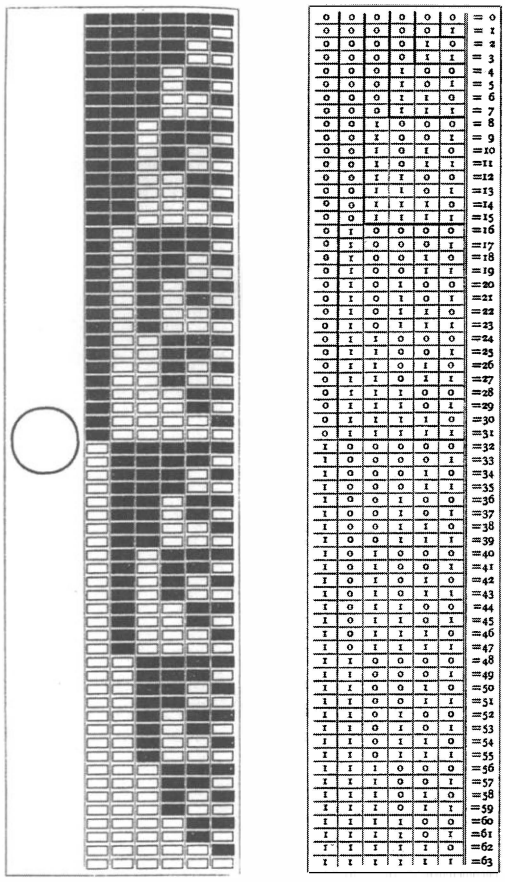


Figure 6 Shao's arrangement of the *Yijing* hexagrams

of hexagrams, both Bouvet and Leibniz read the array backwards! (They read them in the European manner—from left to right, top to bottom.) Leibniz also compared his Fuxi array of hexagrams with the Wenwang version published by Couplet and found the latter set "confused" and with no apparent order [**5**]. FIGURE 6 illustrates the visual segregation scheme employed by Shao to arrive at the "prior to heaven" ordering. In explaining its workings, he notes: "Thus, one divides to make two. Two divides to make four. Four divides to make eight... and thirty-two divides to make sixty-four" [**31**, p. 67]. In FIGURE 6, the circle represents the *Taiji*. In Leibniz's interpretation, black rectangles are yin and are represented in binary form by 0s; white rectangles are yang, represented by 1s.

Conclusion

Interpreting the figures of Fuxi as ancient rediscovered mathematical knowledge caused some stir in early eighteenth century intellectual circles. Wilhelm Tentzel, editor of the short-lived scholarly journal *Curieuse Bibliothec* took great pleasure in announcing the fact that the Chinese had lost the true meaning of the hexagrams and that a European genius had to rediscover the knowledge for them. Tentzel published

his opinions in 1705 [**43**]. However, within a few years, the novelty of Leibniz's theories concerning the *Yijing* and its hexagrams were all but forgotten.

Western interest in the hexagrams and their significance resurfaced in the early twentieth century with the appearance of James Legge's English translation of the Confucian classics [**19**] and Richard Wilhelm's German rendering of the *Yijing* [**50**]. C.G. Jung, one of the founders of analytic psychology, subscribed to the occult power of the hexagrams and attracted followers to this notion in his writings [**50**, preface]. Despite their fanciful interpretations, the hexagrams of the *Yijing* still possess a mathematical attraction [**44**] and have given rise to the yet unsolved problem of finding a mathematical logic for their ordering. Several interesting theories have been offered as solutions [**24, 37**].

The significance of the Leibniz-Bouvet correspondence and its relationship to the *Yijing*, its hexagrams, and binary arithmetic was first examined and studied by Gorai Kinzo. He published his findings in 1929 [**18**]. Kinzo's work was then translated into Chinese by Liu Pai-min [**23**] and became the basis of Hellmut Wilhelm's 1943 article on the subject [**49**]. Joseph Needham, when he compiled material on the *Yijing* for his monumental work *Science and Civilization in China*, relied heavily on Wilhelm's account [**29**, vol. 2: 304–345]. In 1973, Hans Zacher published his study of Leibniz's work on binary arithmetic, and included the correspondence between Leibniz and Bouvet [**51**]. Currently, Alan Berkowitz of Swarthmore College and Daniel Cook of Brooklyn College, CUNY are preparing for publication an annotated English language translation of the Leibniz-Bouvet correspondence. When their book is released it will provide a large audience with insight into this fascinating historical episode involving Gottfried Leibniz, the Chinese missionary Joachim Bouvet, binary arithmetic, religious conversion, and the *Yijing*.

REFERENCES

1. E. J. Aiton, *Leibniz: A Biography*, Adam Hilger, Boston, 1985.
2. ———, An unpublished letter of Leibniz to Sloane, *Annals of Science*, **38** (1981), 103–107.
3. ———, and Eikoh Shimo, Gorai Kinzo's study of Leibniz and the I-ching Hexagrams, *Annals of Science*, **38** (1981), 71–92.
4. Aurelius Augustine, *The Works of Aurelius Augustine*, Marcus Dods, trans., 15 vols., T & T Clark, Edinburgh, 1871–76.
5. Alan Berkowitz, trans., *The Leibniz-Bouvet Correspondence (1697–1707)*, unpublished draft translation obtained through private correspondence.
6. Joachim Bouvet, Correspondence with Abbé Bignon, MS 17240, China, Bibliothèque Nationale de France.
7. Paul Carus, Chinese Philosophy, *The Monist*, **6** (1896), 188–249.
8. Walter Davis, China, The Confucian Ideal and the European Age of Enlightenment, *Discovering China: European Interpretations in the Enlightenment*, Julia Ching and Willard Oxtoby, ed., University of Rochester Press, Rochester, 1992, 1–27.
9. Pierre Dangicourt, De periodis columnarum in serie nemerorum progressionis arithmeticae dyadice expressorum, *Miscellanea Berolensia* **1** (1710), 336–376.
10. J. Fauvel and R.J. Wilson, The lull before the storm: combinatorics and religion in the Renaissance, *Bull. Inst. Comb. Appl.* **11** (1994), 49–58.
11. Martin Gardner, The combinatorial basis of the "I Ching," the Chinese book of divination and wisdom, *Scientific American*, **230** (January, 1974), 108–113.
12. C.I Gerhardt, *Leibnizens mathematische Schriften*, 7 vols, A. Asher and Company, Berlin, 1849.
13. Anton Glaser, *History of Binary and Other Nondecimal Numeration*, Anton Glaser, Southampton, PA 1971.
14. Graham Hancock, *Fingerprints of the Gods*, Crown Publishers, New York, 1995.
15. Ho Peng Yoke, Li, Qi and Shu, *An Introduction to Science and Civilization in China*, Hong Kong University Press, Hong Kong, 1985.
16. Georges Ifrah, *The Universal History of Computing*, John Wiley and Sons, Inc., New York, 2001.
17. Hidé Ishiguro, *Leibniz: Philosophy of Logic and Language*, Cornell University Press, Ithaca, 1972.
18. Gorai Kinzo, *Confucianism and Its Influence on German Political Thought*, [in Japanese], Waseda University Press, Tokyo, 1929.

19. James Legge, *The I-Ching*, Dover Publications, New York, 1963. Reprint of 1899 edition.
20. Gottfried Wilhelm Leibniz, *Writings on China*, Daniel J. Cook and Henry Rosemont, trans. and ed., Open Court, Chicago, 1994.
21. ———, *Philosophical Papers and Letters*, Leroy Loemker, trans. and ed., D. Reidel Publishing Co., Dordrecht, 1969.
22. ———, Explication de l'arithmetique binaire, avec des remarques sur son utilité, et sur ce qu'elle donne le sens des annciennes figures Chinoises de Fohy, Memoires de l'Académie Royale des Science, vol. **3** (1703), 85–89.
23. P.M. Liu, Lai-pu-ni-tzu de Chou-I hsueh, *I-hseh t'ao-lan chi*, Li Cheng-kang (ed.), Shanghai, 1941, 99–113.
24. Peter Loy, A logical way of ordering the trigrams and hexagrams of the Yijing, *The Oracle—The Journal of Yijing Studies*, **2** (January, 2002), 2–13.
25. Jean-Claude Martzloff, *A History of Chinese Mathematics*, Springer, New York, 1997.
26. R. W. Meyers, *Leibniz and the Seventeenth-Century Revolution*, Bowes & Bowes, Cambridge, 1952.
27. David Mungello, *Curious Land: Jesuit Accomodation and the Origins of Sinology*, Franz Steiner, Stuttgart, 1985.
28. ———, *Leibniz and Confucianism: The Search for Accord*, University of Hawaii Press, Honolulu, 1977.
29. Joseph Needham, *Science and Civilization in China*, 7 vols., University Press, Cambridge, 1954–1983.
30. Rita Peng, The K'ang-hsi Emperor's absorption in Western Mathematics and its Extensive Applications of Scientific Knowledge, *Lishi Xuebao* (Taipei) **3**, (1975), 1–74.
31. James Rayan, Leibniz' Binary System and Shao Yong's Yijing, *Philosophy East & West*, **46**, (1996), 59–90.
32. Arnold H. Rowbotham, *Missionary and Mandarin: The Jesuits at the Court of China*, Russell & Russell, New York, 1966.
33. ———, The Jesuit Figurists and Eighteenth-Century religious thought, *Discovering China*, **7**, 39–54.
34. Milton Rugoff (ed.), *The Travels of Marco Polo*, Signet Classics, New York, 1961
35. John W. Shirley, Binary numeration before Leibniz, *American Journal of Physics*, **19**, (Nov, 1951), 452–454.
36. Yong Shao, Huangji jingshi shu (Sibu beiyao ed.) (1060).
37. Mondo Specter, The 'Zhou Yi' Hexagram Sequence: An Authentic, Intended Binary System Discloses a Rational, Bilateral Mathematical Symmetry, *8th International Conference on the History of Science in China*, Berlin, August, 1998, 30.
38. Jonathan Spence, *The Memory Palace of Matteo Ricci*, Viking Penguin, New York, 1984.
39. Z. D. Sung, *The Symbols of Yi King; or, The Symbols of the Chinese Logic of Change*, Paragon Book Reprint, New York, 1969. Reprint of 1934 Shangai edition.
40. Frank Swetz, *Legacy of the Luoshu: The 4,000 Year Search for the Meaning of The Magic Square of Order Three*, Open Court, Chicago, 2002.
41. ———, The introduction of mathematics in higher education in China, *Historia Mathematica*, **1**, (1974), 167–179.
42. R.M. Swiderski, Bouvet and Leibniz: A scholarly correspondence, *Eighteenth Century Studies*, **14** (1980).
43. W.E. Tentzel, *Curieuse Bibliothec*, Frankfort, 1705.
44. F. van der Blij, Combinatorial aspects of the hexagrams in the Chinese Book of Changes, *Scripta Mathematica*, **28**, (May, 1967), 37–49.
45. Ira Wade, *Voltaire and Candide: A Study in the Fusion of History, Art and Philosophy*, Kennikat Press, London, 1972.
46. Arthur Waley, Leibniz and Fu-Hsi, *Bull. of the London School of Oriental Studies*, **2**, (1921–23) 165–167.
47. John Webb, *An historical essay endeavoring a probability that the language of The Empire of China is the Primitive Language*, London, 1669.
48. Hellmut Wilhelm, *Change: Eight Lectures on the I Ching*, translated by Cary F. Baynes, Princeton University Press, Princeton, 1960.
49. ———, Leibniz and the I Ching, *Collectanea Commissiones Synodalis*, **16**, (1943), 205–219.
50. Richard Wilhelm, *I Ging, Das Buch der Wandlungen*, Jena, 1924.
51. Hans J. Zacher, *Die Hauptschriften zur Dyadik von G.W. Leibniz*, Vittorio Klostermann, Frankfort, 1973.

NOTES

Means Appearing in Geometric Figures

HOWARD EVES
University of Maine, Orono/University of Central Florida*

Practically every student of mathematics is acquainted with the arithmetic and geometric means of two given positive numbers, a and b. Not so many students realize that there are many other means of two given positive numbers. Which mean is required depends upon the problem at hand.

Following is a list of several means of a pair of positive numbers, a and b:

1. Arithmetic mean: $A(a, b) = (a + b)/2$
2. Geometric mean: $G(a, b) = \sqrt{ab}$
3. Harmonic mean: $H(a, b) = 2ab/(a + b)$
4. Heronian mean: $N(a, b) = (a + \sqrt{ab} + b)/3$
5. Contraharmonic mean: $C(a, b) = (a^2 + b^2)/(a + b)$
6. Root-mean-square: $R(a, b) = \sqrt{(a^2 + b^2)/2}$
7. Centroidal mean: $T(a, b) = 2(a^2 + ab + b^2)/3(a + b)$

Note that these are all *homogeneous*, in the sense that

$$M(ka, kb) = kM(a, b) \quad \text{for} \quad k > 0,$$

and *symmetric*, in that $M(a, b) = M(b, a)$.

The purpose of this paper is to show how some of these means occur in certain geometrical figures. One is surprised at the ubiquity of the harmonic mean.

The reader may be entertained by filling in the details.

Means in circles On the left in FIGURE 1, OP is the arithmetic mean of a and b, AD the geometric mean, and $AB+BC$ the heronian mean. On the right, AB' is the harmonic mean of a and b.

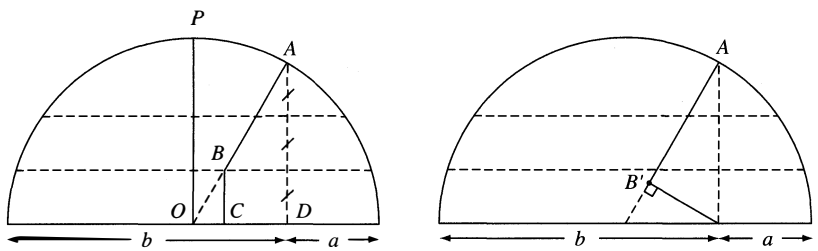


Figure 1 Left: The arithmetic, geometric, and heronian means. Right: The harmonic mean.

*Editor's Note: A new collection of Eves' popular books, the *Mathematical Circles* series, is available as a 3-volume set from the MAA. The editor also recommends Eves' *Mathematical Reminiscences.*

By superimposing the pictures, one can observe that $H(a, b) < G(a, b) < N(a, b) < A(a, b)$.

A wealth of harmonic means In each diagram of FIGURE 2, c is half the harmonic mean of a and b.

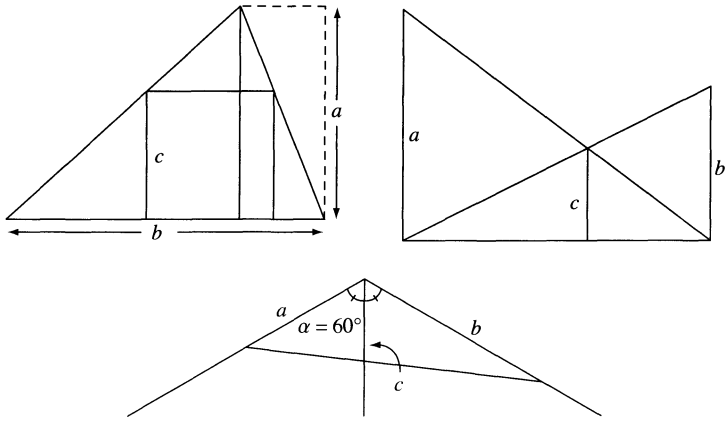


Figure 2 Three ways to construct half the harmonic mean

LEON BANKOFF'S THEOREM. *If A, B, C, D, E, F, and G are the vertices of a regular heptagon, then* $CD = H(AC, AD)/2$. (This is depicted in FIGURE 3 [**1**].)

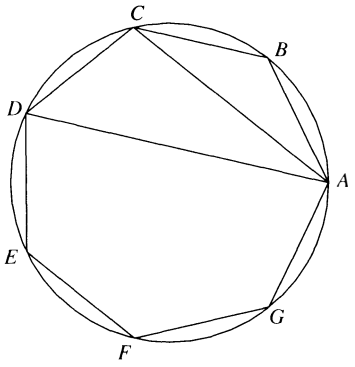


Figure 3 Leon Bankoff's harmonic mean in a heptagon

The harmonic mean also appears in a classic word problem: If one travels from A to B at rate r_1, and then returns from B to A at rate r_2, what is the average rate r for the round trip? The answer is the harmonic mean of r_1 and r_2.

In each diagram of FIGURE 4, AC is the harmonic mean of AB and AD.

The means in a trapezoid Suppose that a trapezoid has parallel sides of lengths a and b, pictured as vertical segments in FIGURE 5. The various means can be ranked as the lengths of vertical segments. The segment whose height is

1. the harmonic mean $H(a, b)$ passes through the intersection of the diagonals;
2. the geometric mean $G(a, b)$ divides the trapezoid into two similar trapezoids;

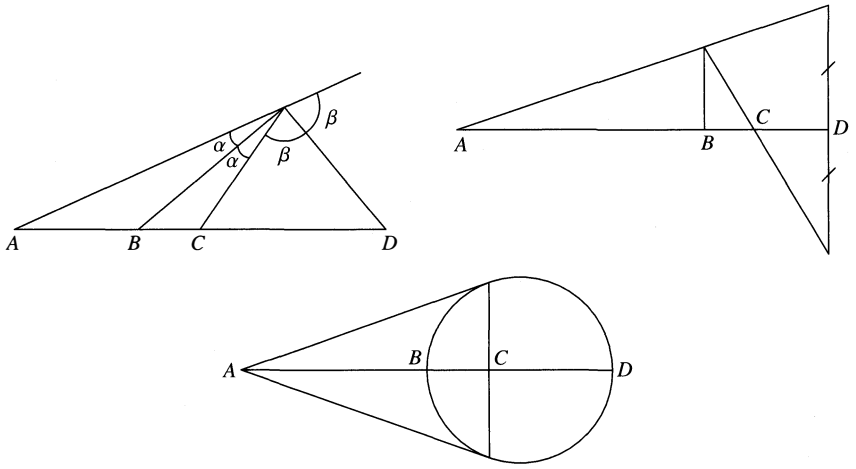


Figure 4 Three appearances of the harmonic mean

3. the heronian mean $N(a, b)$ is one-third the way from the arithmetic mean to the geometric mean;
4. the arithmetic mean $A(a, b)$ bisects the sides of the trapezoid.
5. the centroidal mean $T(a, b)$ passes through the centroid of the trapezoid;
6. the root-mean-square $R(a, b)$ bisects the area of the trapezoid;
7. the contraharmonic mean $C(a, b)$ is as far to the right of the arithmetic mean as the harmonic mean is to the left of it.

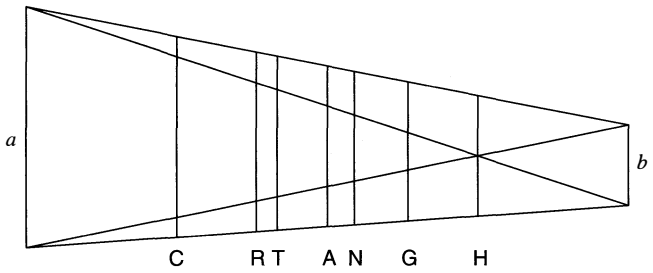


Figure 5 The means in a trapezoid

[*Editor's Note*: Shannon Umberger Patton created an animated version of this figure using *Geometer's Sketchpad* as part of her studies in the master's program at the University of Georgia. Inspired by an exercise in Eves' book [**2**], this is available at this MAGAZINE's website.]

REFERENCES

1. Leon Bankoff and Jack Garfunkel, The heptagonal triangle, this MAGAZINE, **46**:11 (1973), 7–19.
2. Howard Eves, *An Introduction to the History of Mathematics*, 6th ed., Saunders College Publishing, New York, 1990, p. 201.

Mathematical Modeling of Metal Leaves

YUKIO KOBAYASHI (corresponding author)
TAKASHI NIITSU
KOICHI TAKAHASHI
Faculty of Engineering, Soka University
1-236 Tangi-cho, Hachioji-shi
Tokyo 192-8577 Japan
koba@t.soka.ac.jp

SETSUKO SHIMOIDA
Shonan Institute of Technology
1-1-25 Nishikaigan Tujido Fujisawa-shi
Kanagawa-ken 251-8511 Japan

Fractal geometry, a mathematical subject, has nonetheless become useful in many branches of science. This note introduces a graduation project on fractal geometry that combines the areas of mathematics, computer science, and physical chemistry [**1**].

With a microscope, we can observe that particles suspended on the surface of water have a highly irregular motion. We describe this as *Brownian motion*. Molecular diffusion, which is extensively simulated in physical chemistry, is based on Brownian motion. If we assume *irreversible aggregation* (that is, once something gets stuck, it can not get un-stuck), then we can study particle motion with an experimental laboratory method called electrodeposition. Our paper compares this experimental method with a simulation method called diffusion-limited aggregation (DLA). DLA is a simulation method that models various physicochemical phenomena: it is a simple random-growth process generated by Brownian motion [**2**]. We use the mathematics of fractal geometry to compare the two methods.

DLA cluster formation by computer simulation In this article, we adopt a two-dimensional lattice model [**3**] (FIGURE 1) because, when a nonlattice simulation of a

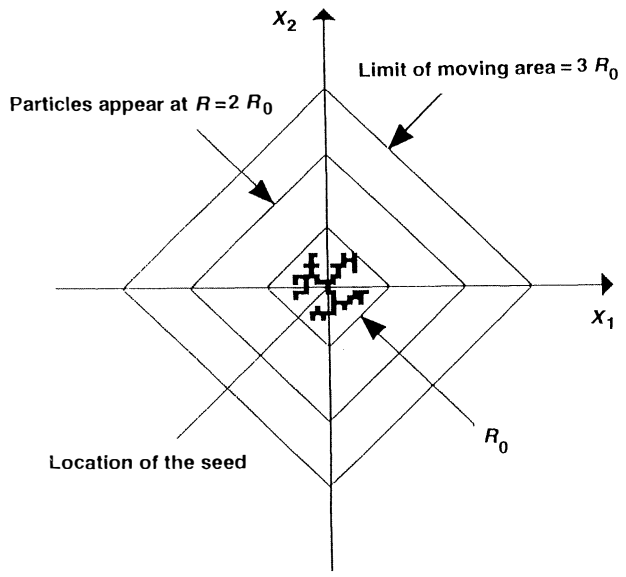


Figure 1 Two-dimensional lattice model

DLA cluster is used, it is difficult to test for an overlap with a growing cluster. The procedures of cluster formation by computer simulation are as follows: (1) Locate a seed particle on the origin of the lattice. (2) Release a particle from random location far from the seed particle and allow it to follow a Brownian trajectory selected using a random number generator for each step. (3) Form a two-particle cluster when the Brownian particle reaches a site adjacent to the seed particle. (4) Repeat the process until a cluster of sufficiently large size is formed; each time the moving Brownian particle reaches a position adjacent to the cluster, add the particle to the cluster. The flowchart given in FIGURE 2 outlines a BASIC program written by Takayasu [**4**], in which a starting point and a step are chosen using uniformly distributed random numbers.

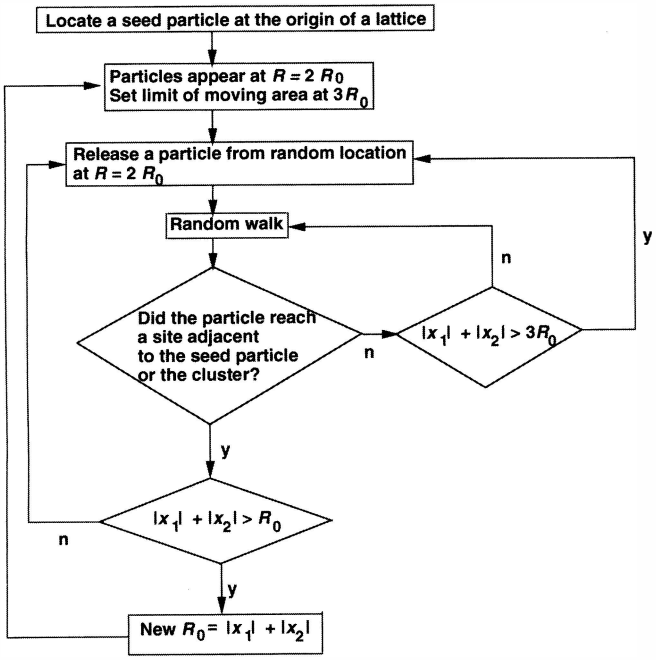


Figure 2 Flowchart for DLA cluster formation

We started with a particle at a random point on a circle that was centered on the seed particle and had a radius that was two times R_0. For the first time, we set R_0 at 5 lattice units. When a particle reached a point more than three times R_0 from the origin, it was killed and a new particle was started. When the distance from the seed to the most distant particle in the cluster was larger than R_0, we increased R_0 by 1 lattice unit. The distance from the seed to the lattice site (x_1, x_2) is $|x_1| + |x_2|$ lattice units.

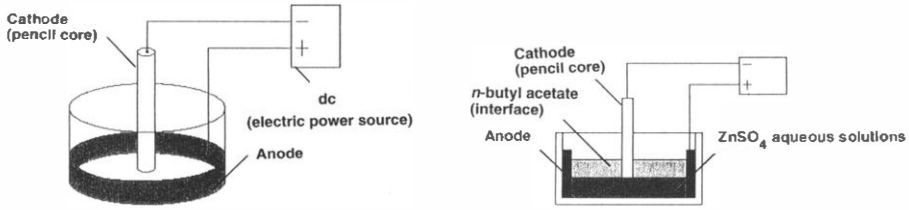


Figure 3 Schematics of experimental system

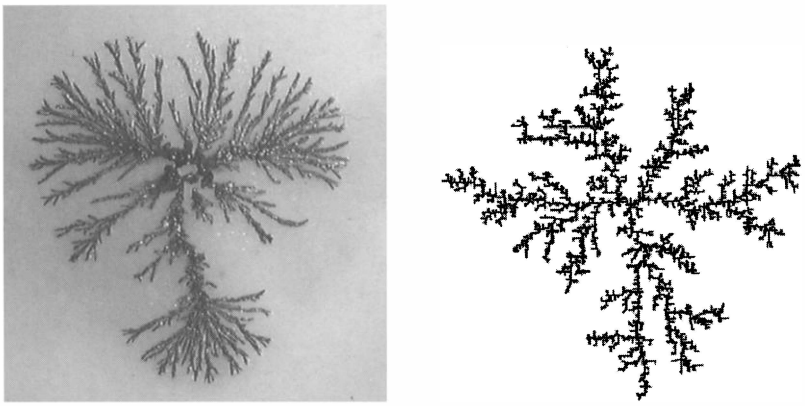


Figure 4 *(Left)* Zinc metal leaves grown by a voltage of 4.0V and a concentration of 0.5M. The fractal dimension is 1.61. *(Right)* DLA cluster generated by the algorithm. The fractal dimension is 1.55.

As seen in FIGURE 4, DLA cluster and metal leaves resemble each other closely. In our model, a particle zigzags randomly. On the other hand, zinc ions undergo irregular motion in random directions. The resemblance of metal leaves observed in an experimental system and fractal patterns generated by mathematical modeling is very satisfying.

Metal leaves grown by electrodeposition We provide a brief outline of the experiment: Following the procedures established by Matsushita et al. **[5]**, we generated the zinc metal leaves by electrodeposition. The simplicity of the system is shown in FIGURE 3: A petri dish of diameter 210mm and depth 58mm was filled with zinc sulfate aqueous solutions (concentration of 0.25M or 0.5M), on which *n*-butyl acetate was poured to make an interface. We contrived a carbon cathode using the core of a pencil and suspended it with fishing line. The pencil core was placed near the center of the petri dish so that the flat tip was placed precisely at the interface. Electrodeposition was initiated by applying a voltage of 4.0V or 6.0V between the carbon cathode and a zinc cylindrical anode. Photographs were taken both during and after the growth of the metal leaves.

Comparing DLA clusters and metal leaves The clusters and the metal leaves were analyzed using the box-counting method **[6]**. Following this method, each object was covered with boxes of side length δ. The number of boxes, N, required to cover the object is related to δ because of its box-counting dimension D. Let us consider a simple example. To cover a square, N smaller squares of the side length δ are required. Then,

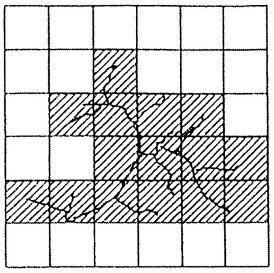


Figure 5 Determining the fractal dimension of a cluster using the box-counting method

$N\delta^2 = V$, where V is an area. In general, $N = V/\delta^D$ boxes are required to cover an object when the exponent D is the box-counting dimension of the object (FIGURE 5). Thus, the general formulation of D for objects of area V is

$$D = \frac{\log N - \log V}{\log(1/\delta)}.$$

A practical estimate can be obtained by drawing a best-fit line through the points at small values of δ and calculating the slope of this line (FIGURE 6).

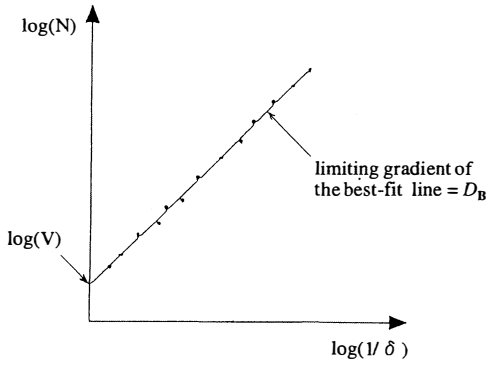


Figure 6 Estimating the box counting dimension of experimental data

The fractal dimensions (~1.6) agreed approximately between the clusters and the metal leaves independently of the experimental conditions (concentration, voltage). These fractal dimensions are almost in agreement with the fractal dimension measured by other methods [**5**]. This indicates that our model is a fairly good approximation to the physical process.

Acknowledgments. The experimental and simulated results constitute a part of the graduation thesis presented to Soka University by author Koichi Takahashi, who is employed by a fishing tackle store in Osaka.

REFERENCES

1. K. Takahashi, T. Niitsu, S. Shimoida, and Y. Kobayashi, Fractal structures of diffusion-controlled clusters and metal leaves: educational links between mathematics, computer architecture and physicochemical subjects, *9th International Congress on Mathematical Education*, Makuhari, 2000.
2. T. A. Witten Jr. and L. M. Sander, Diffusion-limited aggregation, a kinematic critical phenomenon, *Phys. Rev. Lett.* **47** (1981), 1400–1403.
3. P. Meakin, Diffusion-controlled cluster formation in 2–6-dimensional space, *Phys. Rev.*, A **27** (1983), 1495–1507.
4. H. Takayasu, *Fractal*, Asakura Book Co., Tokyo, 1986 [in Japanese].
5. M. Matsushita, M. Sano, Y. Hasekawa, H. Honjo, and Y. Sawada, Fractal structures of zinc metal leaves grown by electrodeposition, *Phys. Rev. Lett.* **53** (1984), 286–289.
6. P. S. Addison: *Fractals and Chaos*, IOP Publishing Ltd, UK, 1997.
7. D. Avnir, O. Biham, D. Lidar, and O. Malcai, Is the geometry of nature fractal?, *Science*, **279** (1998), 39–40.

A Simpler Dense Proof Regarding the Abundancy Index

RICHARD F. RYAN
Marymount College
Rancho Palos Verdes, CA 90275-6299
rryan@marymountpv.edu

Let a represent a (positive) integer, and $\sigma(a)$ denote the sum of the factors of a. The *abundancy index* of a, denoted by $I(a)$, is defined by $I(a) = \sigma(a)/a$. For example, $I(18) = 39/18 = 13/6$. R. Laatsch [**3**] showed that the set of abundancy indices (that is, the range of $I(x)$) is dense in the interval $[1, \infty)$. Let

$$T = \{\text{rational numbers, greater than 1, that are not abundancy indices}\}.$$

A proof that T is also dense in $[1, \infty)$, by P. A. Weiner [**5**], was published in October, 2000. In this note, we present a simpler proof of the density of T that bears a resemblance to the aforementioned proof by R. Laatsch. First, we review some properties of the abundancy index:

(A) If a_1 divides a then $I(a) \geq I(a_1)$. This is easily observed, due to the fact that

$$I(a) = \sum_{d|a} \frac{1}{d}.$$

(B) If $I(a) = r/s$ is in lowest terms, then s divides a. This follows since $s\sigma(a) = ra$ and $\gcd(r, s) = 1$.

(C) If $I(a) = r/s$ is in lowest terms then $r \geq \sigma(s)$. This follows from properties (B) and (A) since $r/s = I(a) \geq I(s) = \sigma(s)/s$. (The condition that r and s be relatively prime is an important one! Note that $I(2) = 6/4$ even though $6 < \sigma(4)$). Contrapositively, if r/s is in lowest terms with $\sigma(s) > r$ then r/s is not an abundancy index.

For simplicity, we let p_j denote the jth prime number (that is, $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc.).

THEOREM 1. *T is dense in $[1, \infty)$.*

Proof. Due to property (C), $(b + 1)/b \in T$ whenever b is composite. Additionally, $\lim_{b\to\infty}[(b + 1)/b] = 1$; thus 1 is a cluster point of T. If c is a real number greater than 1, then we will generate a sequence $\{c_k\}_{k=1}^{\infty}$ in the following way, with the stipulation that we will stop at any step in which $c_k = c$: Let $c_1 = 1 + 1/q_1$; in this "first step," q_1 is the least prime such that $1 + 1/q_1 < c$. Let $c_2 = c_1 + 1/q_2$, where q_2 is the least prime, greater than q_1, with the property that $c_1 + 1/q_2 \leq c$. Let $c_3 = c_2 + 1/q_3$, where q_3 is the least prime greater than q_2 for which $c_2 + 1/q_3 \leq c$, etc. Since the sequence $\{1/p_j\}_{j=1}^{\infty}$ converges to 0, the q_k that is needed to construct the corresponding c_k is guaranteed to exist. For each $k \geq 2$,

$$c_k = \frac{q_1q_2q_3\cdots q_k + q_2q_3\cdots q_k + q_1q_3\cdots q_k + \cdots + q_1q_2q_3\cdots q_{k-1}}{q_1q_2q_3\cdots q_k}.$$

Note that the numerator and denominator are relatively prime: q_1 is a factor of each term in the numerator except for the second, $q_2q_3\cdots q_k$; q_2 divides all terms on top

except for the third; etc. Each term in the numerator is a (different) factor of the denominator; however (when $k \geq 2$) the top does not contain the term "+1." Thus the numerator is less than $\sigma(q_1q_2q_3 \cdots q_k)$, and so by property (C), each term in the sequence $\{c_k\}_{k=2}^{\infty}$ is an element of T. The gap from c_k to c will be bridged due to the fact that

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{p_j} + \cdots$$

diverges to ∞ [**2**, p. 17]. Since c was allowed to be any number greater than 1, T is dense in $[1, \infty)$. ■

The abundancy index is a multiplicative function in the sense that $I(bc) = I(b)I(c)$ whenever b and c are relatively prime; this fact is useful when exploring the properties of perfect numbers. We say that a is *perfect* if $I(a) = 2$. It is well known that an even number is perfect if and only if it can be written in the form $2^{p-1}(2^p - 1)$, where p and $2^p - 1$ are primes. If an odd perfect number exists, it must be greater than 10^{300} [**1**]. P. A. Weiner showed that, if $I(b) = 5/3$, then $5b$ is odd and perfect. His result can be generalized in the following way:

THEOREM 2. *Suppose we have a fraction of the form* $(2n - 1)/n$, *where* $2n - 1$ *is prime.*

(i) *If* n *is even, but not a power of* 2, *then* $(2n - 1)/n \in T$.

(ii) *If* n *is odd and* $I(b) = (2n - 1)/n$, *then* b *is odd; moreover, if* $2n - 1$ *does not divide* b *then* $b(2n - 1)$ *is a perfect number.*

Proof. **(i)** Assume that n is even, divisible by an odd number greater than 1, and that $I(b) = (2n - 1)/n$ for some integer b. Due to property (B), b is even. Let m be the greatest integer such that 2^m divides b. There is a prime factor, q, of $\sigma(2^m)$ that also divides b since $\sigma(2^m) = 2^{m+1} - 1 \neq 2n - 1$. Thus

$$I(b) > I(2^m q) \geq \frac{2^{m+1} - 1}{2^m} \cdot \frac{2^{m+1}}{2^{m+1} - 1} = 2,$$

which contradicts $I(b) = (2n - 1)/n$.

(ii) Suppose that n is odd and b is even. Let m be the greatest integer such that 2^m divides b. Once again, there is a prime factor, q, of $\sigma(2^m)$ that divides b. Thus $I(b) > I(2^m q) \geq 2$ and we have a contradiction. Finally, if the prime number $2n - 1$ does not divide b then, since I is multiplicative, $I(b(2n - 1)) = 2$. ■

We will briefly consider cases in which $2n - 1$ is not necessarily prime, where n continues to represent a (positive) integer. If b is an even number that has an abundancy index of the form $(2n - 1)/n$, then $\sigma(2^m)$ divides $2n - 1$, where m retains the meaning that it had in the proof of theorem 2. Note that $I(10) = 9/5$, $I(44) = 21/11$, $I(2^3 \cdot 17^2 \cdot 307) = 1155/578$, and as previously implied, $I(2^l) = (2^{l+1} - 1)/2^l$ for each integer l. Other examples of even numbers with indices in the form $(2n - 1)/n$ are readily found. It is easy to verify, using an unsophisticated computer program, that 1 and $3^2 \cdot 7^2 \cdot 11^2 \cdot 13^2$ are the only odd integers less than 10^{16} having indices in the specified form ($I(3^2 \cdot 7^2 \cdot 11^2 \cdot 13^2) = 22021/11011$).

Resolving the existence issue for odd perfect numbers is an intriguing, if daunting, task. Our ever-increasing knowledge concerning properties of the abundancy index appears to be useful in this endeavor.

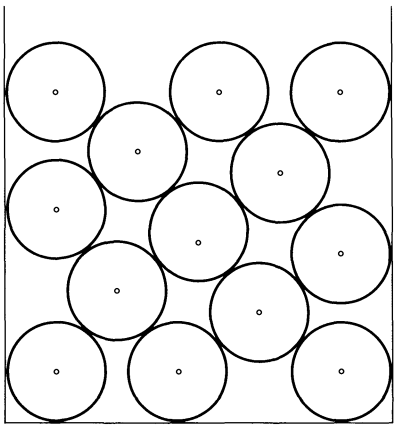
REFERENCES

1. R. P. Brent, G. L. Cohen, and H. J. J. te Riele, Improved techniques for lower bounds for odd perfect numbers, *Math. Comp.* **57** (1991), 857–868.
2. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., Oxford Univ. Pr., New York, 1979.
3. R. Laatsch, Measuring the abundancy of integers, this MAGAZINE **59** (1986), 84–92.
4. R. F. Ryan, Results concerning uniqueness for $\sigma(x)/x = \sigma(p^n q^m)/(p^n q^m)$ and related topics, *Int. Math. J.* **2** (2002), 497–514.
5. P. A. Weiner, The abundancy ratio, a measure of perfection, this MAGAZINE **73** (2000), 307–310.

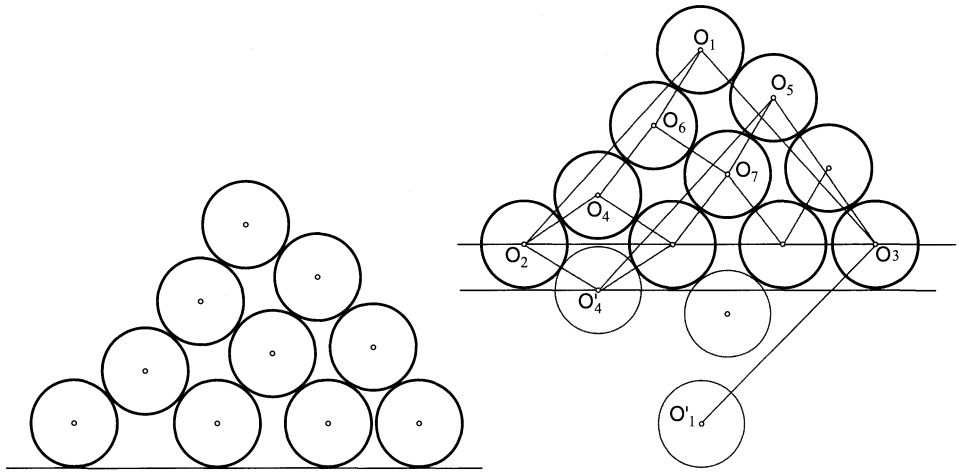
A Circle-Stacking Theorem

ADAM BROWN
Havergal College
Toronto, Ontario M5N 2H9
Canada
adam_brown@havergal.on.ca

By way of introduction, we first consider a preliminary problem—the wine rack problem. We have a wine rack and, fortunately, 13 bottles of wine. Three bottles are placed on the bottom row so that the outermost two are tangent to the vertical sides as depicted in the figure below. Then we place rows of two, three, two, and finally three more bottles on top of these. Surprisingly, no matter where the middle bottle is in the bottom row, the top row is always perfectly horizontal! Can you see why? This problem can be found in *Which Way Did the Bicycle Go?*, a nifty problems book published by the MAA. The problem was discovered by Charles Payan of the *Laboratoire de Structures Descretes et de Didactique* in France, one of the creators of CABRI geometry. He discovered this result while using CABRI, an incredibly powerful tool for experimenting with Euclidean geometry.



We will not present a proof here of this engaging result, but will offer it a companion instead. After playing around with this figure for a little while, one suspects that the middle circle in the middle row is half-way between the bottom two extreme circles. In fact, more turns out to be true: start with a row of equal circles on a line, and place a row of one fewer equal circles tangent to circles of the bottom row two at a time. Carry on in this way until one circle at the summit is reached. Then, provided the circles on



the bottom are positioned so that none of the succeeding circles can slip through any gaps, the center of the top circle lies half-way between the centers of the outermost bottom two.

The figure on the left shows the naked truth of the theorem, and the one on the right displays constructions and labeling. Notice that it suffices to show that O_1O_2 is equal in length to O_1O_3 for then triangle $O_1O_2O_3$ will be isosceles and the vertex O_1 will lie directly above the midpoint of the base. By joining the centers of tangent circles we obtain chains of rhombi, that is, quadrilaterals each of whose edge lengths equal twice the common radius of all the circles. Thus O_1O_5 is equal and parallel to O_6O_7, and so on. To transfer this segment all the way down to O_2, we reflect O_4 in O_2O_3 yielding O_4'. Thus O_2O_4' is parallel and equal to O_1O_5 which implies that O_1O_2 is parallel and equal to O_5O_4'. So, this sequence of translations has enabled us to transfer O_1O_2 onto a new equal and parallel segment. By carrying on this reasoning, we will transfer O_1O_2 onto O_3O_1', where O_1' is the reflection of O_1 in O_2O_3. Then, by reflection in O_2O_3, O_3O_1' is carried onto O_3O_1. Hence we have carried O_1O_2 onto O_1O_3 by a sequence of translations, followed by a reflection, and $O_1O_2 = O_1O_3$!

REFERENCE

1. Joseph D.E. Konhauser, Dan Velleman, and Stan Wagon, *Which Way Did the Bicycle Go?*, Mathematical Association of America, 1996.

Editor's Note: Alert reader Dwayne Jennings of Union University in Jackson, Tennessee, while enjoying our June cover, noticed that one of the lines indicating bunny ancestry is drawn in the wrong place. Please find the error and correct your copy accordingly by removing one line and adding another. (Hint: In the current version, the original pair is missing a pair of offspring.)

An Elementary Proof of Error Estimates for the Trapezoidal Rule

D. CRUZ-URIBE, SFO
Trinity College
Hartford, CT 06106-3100
david.cruzuribe@mail.trincoll.edu

C. J. NEUGEBAUER
Purdue University
West Lafayette, IN 47907-1395
neug@math.purdue.edu

Essentially every calculus textbook contains the trapezoidal rule for estimating definite integrals; this rule can be stated precisely as follows:

If f is continuous, then for each integer $n > 0$ the integral of f on $[a, b]$ is approximated by

$$T_n(f) = \frac{b-a}{2n}\big(f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)\big),$$

where $x_i = a + i(b-a)/n$, $0 \le i \le n$. Further, if f is twice differentiable, f'' is continuous, and $f''(t) \le M$ for $t \in [a, b]$, then

$$E_n^T(f) = \left| T_n(f) - \int_a^b f(t)\,dt \right| \le \frac{(b-a)^3}{12n^2} M. \tag{1}$$

(See, for instance, Stewart [**7**].) There are two problems, however, with this result. First, calculus books generally omit the proof, and instead refer the reader to an advanced text on numerical analysis. In such books the trapezoidal rule is usually derived as a corollary to a more general result for Newton-Cotes quadrature methods, and the proof, depending on polynomial approximation, is generally not accessible to calculus students. (See, for example, Ralston [**5**].)

Second, the error estimate given by (1) is not applicable to such well-behaved functions as $x^{3/2}$ and $x^{1/2}$ on [0, 1], since neither has a bounded second derivative. In other words, you can use the trapezoidal rule to approximate their integrals, but for a given n you have no idea, *a priori*, how good the approximation is.

In this note we give an elementary proof of inequality (1). The key idea is to use integration by parts "backwards." The argument is straightforward and should be readily understood by students in second-semester calculus. This approach is not new and goes back to Von Mises [**8**] and Peano [**4**]. (See also Ghizzetti and Ossicini [**3**].) However, our exact proof is either new or long forgotten—see Cruz-Uribe and Neugebauer [**2**] for a survey of the literature.

Our proof has two other advantages. First, we can omit the assumption that f'' is continuous, and replace it with the weaker assumption that it is (Riemann) integrable. Second, we can adapt our proof to give estimates for functions that do not have bounded second derivatives, such as $x^{3/2}$ and $x^{1/2}$.

Proving the inequality. The first step is to find an expression for the error on each interval $[x_{i-1}, x_i]$. If we divide the integral of f on $[a, b]$ into the sum of integrals on

each of these subintervals, we have

$$E_n^T(f) = \left| \sum_{i=1}^{n} \left(\frac{b-a}{2n}(f(x_{i-1}) + f(x_i)) - \int_{x_{i-1}}^{x_i} f(t)\,dt \right) \right|.$$

For each i, $0 \le i \le n$, define

$$L_i = \frac{b-a}{2n}(f(x_{i-1}) + f(x_i)) - \int_{x_{i-1}}^{x_i} f(t)\,dt. \tag{2}$$

Let c_i denote the center of the interval $[x_{i-1}, x_i]$: $c_i = (x_{i-1} + x_i)/2$. Then

$$x_i - c_i = c_i - x_{i-1} = \frac{b-a}{2n},$$

and if we apply integration by parts "backwards," we see that we can express the error in terms of the first derivative:

$$L_i = \int_{x_{i-1}}^{x_i} (t - c_i) f'(t)\,dt. \tag{3}$$

(Given (3), it is easy to apply integration by parts to show that (2) holds; it is more subtle to argue in the other direction.) If we integrate this result by parts, we can express the error in terms of the second derivative:

$$L_i = \frac{1}{2} \int_{x_{i-1}}^{x_i} \left(\left(\frac{b-a}{2n} \right)^2 - (t - c_i)^2 \right) f''(t)\,dt. \tag{4}$$

Therefore, if $| f''(t) | \le M$ for $t \in [a, b]$, we have that

$$\begin{aligned} E_n^T(f) &\le \sum_{i=1}^{n} | L_i | \\ &\le \frac{1}{2} \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \left| \left(\frac{b-a}{2n} \right)^2 - (t - c_i)^2 \right| | f''(t) |\,dt \\ &\le \frac{M}{2} \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \left(\left(\frac{b-a}{2n} \right)^2 - (t - c_i)^2 \right) dt. \end{aligned}$$

For each i, a straightforward calculation shows that

$$\int_{x_{i-1}}^{x_i} \left(\left(\frac{b-a}{2n} \right)^2 - (t - c_i)^2 \right) dt = \frac{1}{6} \left(\frac{b-a}{n} \right)^3.$$

Hence,

$$E_n^T(f) \le \frac{M}{2} \sum_{i=1}^{n} \frac{1}{6} \left(\frac{b-a}{n} \right)^3 = \frac{(b-a)^3}{12n^2} M,$$

which is precisely what we wanted to prove.

What if the second derivative is not bounded? The heart of the proof of (1) is using (4) to estimate the error. However, a very similar argument works if we start

with (3). Suppose $|f'(t)| \leq N$ for $t \in [a, b]$. Then

$$|L_i| \leq N \int_{x_{i-1}}^{x_i} |t - c_i| \, dt = \frac{N}{4} \left(\frac{b-a}{n} \right)^2,$$

which yields

$$E_n^T(f) \leq \frac{(b-a)^2}{4n} N.$$

For example, if $f(x) = x^{3/2}$ on [0, 1] then $|f'(x)| \leq 3/2$, so

$$E_n^T(f) \leq \frac{3}{8n}.$$

We can get a better estimate if the integrals of f'' and $|f''|$ exist as improper integrals, as is the case for $f(x) = x^{3/2}$. (Alternatively, we could use a more general definition of the integral, such as the Lebesgue integral or the Henstock-Kurzweiler integral. See [**1, 6**].) For in this case we can still apply integration by parts to derive (4). Since for each i,

$$0 \leq \left(\frac{b-a}{2n} \right)^2 - (t - c_i)^2 \leq \left(\frac{b-a}{2n} \right)^2,$$

it follows that

$$|L_i| \leq \frac{1}{2} \left(\frac{b-a}{2n} \right)^2 \int_{x_{i-1}}^{x_i} |f''(t)| \, dt.$$

Hence,

$$E_n^T(f) \leq \frac{(b-a)^2}{8n^2} \int_a^b |f''(t)| \, dt.$$

Again, if we let $f(x) = x^{3/2}$ on [0, 1], this yields the estimate

$$E_n^T(f) \leq \frac{3}{16n^2}.$$

Given a function such as $x^{1/2}$ on [0, 1], whose first derivative is integrable as an improper integral, but whose second derivative is not, we can use the same argument to derive similar estimates from (3):

$$E_n^T(f) \leq \frac{b-a}{2n} \int_a^b |f'(t)| \, dt.$$

Thus, if $f(x) = x^{1/2}$ on [0, 1],

$$E_n^T(f) \leq \frac{1}{2n}.$$

Further extensions. We prove these and related results for the trapezoidal rule in [**2**]. We also show that all of these error estimates are sharp: for each n we construct a function f such that equality holds.

We also use the same techniques to prove error estimates for Simpson's rule in terms of the first and second derivative. The classical error estimate for Simpson's rule involve the fourth derivative [**5, 7**], and it is an open problem to extend our ideas to give an elementary proof of this result.

REFERENCES

1. G. Bartle, *A modern theory of integration*, Graduate Studies in Mathematics, 32, American Mathematical Society, Providence, 2001.
2. D. Cruz-Uribe, SFO and C. J. Neugebauer, Sharp error bounds for the trapezoidal rule and Simpson's rule, *J. Inequal. Pure Appl. Math.* **3** (2002), Issue 4, Article 49. Available at `jipam.vu.edu.au/v3n4/index.html`.
3. A. Ghizzetti and A. Ossicini, *Quadrature formulae*, Academic Press, New York, 1970.
4. G. Peano, Resto nelle formule di quadratura espresso con un integrale definito, *Atti Accad. naz. Lincei, Rend., Cl. sci. fis. mat. nat.* (5) **22-I** (1913), 562–9.
5. A. Ralston, *A first course in numerical analysis*, McGraw-Hill, New York, 1965.
6. W. Rudin, *Real and Complex Analysis,* Third Ed., McGraw-Hill, New York, 1987.
7. J. Stewart, *Calculus*, 4th Ed., Brooks/Cole, Pacific Grove, 1999.
8. R. von Mises, Über allgemeine quadraturformeln, *J. Reine Angew. Math.* **174** (1935), 56–67; reprinted in *Selected Papers of Richard von Mises*, Vol. 1, 559–574, American Mathematical Society, Providence, 1963.

Triangles with Integer Sides

MICHAEL D. HIRSCHHORN
University of New South Wales
Sydney NSW 2052
Australia
m.hirschhorn@unsw.edu.au

It is well known [**1, 2, 3, 4, 5, 6**] that the number T_n of triangles with integer sides and perimeter n is given by

$$T_n = \begin{cases} \langle (n+3)^2/48 \rangle & \text{if } n \text{ is odd} \\ \langle n^2/48 \rangle & \text{if } n \text{ is even,} \end{cases}$$

where $\langle x \rangle$ is the integer closest to x. The object of this note is to give as quick a proof of this as I can, starting with:

LEMMA 1. *The number S_n of scalene triangles with integer sides and perimeter n is given for $n \geq 6$ by*

$$S_n = T_{n-6}.$$

Proof. If $n = 6, 7, 8$, or 10, both are 0. Otherwise, given a scalene triangle with integer sides $a < b < c$ and perimeter n, let $a' = a - 1$, $b' = b - 2$, $c' = c - 3$. Then a', b', c' are the sides of a triangle of perimeter $n - 6$. Moreover, the process is reversible, so the result follows. ∎

COROLLARY. *$T_n - T_{n-6} = I_n$, where I_n denotes the number of isosceles (including equilateral) triangles with integer sides and perimeter n.*

LEMMA 2. *If $n \geq 1$, then*

$$I_n = \begin{cases} \langle (n-3)/4 \rangle & \text{if } n \text{ is even} \\ \langle n/4 \rangle & \text{if } n \text{ is odd.} \end{cases}$$

Proof. If $n = 1, 2, \text{or } 4$, $I_n = 0$. Otherwise, if $n \equiv 0 \pmod 4$, write $n = 4m$. The isosceles triangles with integer sides and perimeter n have sides

$$\{2, 2m-1, 2m-1\}, \{4, 2m-2, 2m-2\}, \ldots, \{2m-2, m+1, m+1\}.$$

Thus, there are $m-1$ such triangles, and $I_n = m - 1 = \langle (n-3)/4 \rangle$. The other three cases are similar. ∎

COROLLARY.

$$T_n - T_{n-6} = \begin{cases} \langle (n-3)/4 \rangle & n \text{ even} \\ \langle n/4 \rangle & n \text{ odd.} \end{cases}$$

LEMMA 3. *For $n \geq 12$*

$$T_n - T_{n-12} = \begin{cases} (n-6)/2 & n \text{ even} \\ (n-3)/2 & n \text{ odd.} \end{cases}$$

Proof.

$$\begin{aligned} T_n - T_{n-12} &= T_n - T_{n-6} + T_{n-6} - T_{n-12} \\ &= \begin{cases} \langle (n-3)/4 \rangle + \langle (n-9)/4 \rangle = (n-6)/2 & n \text{ even} \\ \langle n/4 \rangle + \langle (n-6)/4 \rangle = (n-3)/2 & n \text{ odd.} \end{cases} \end{aligned}$$

∎

Proofs of the next two lemmas are very easy.

LEMMA 4. *Let $f(n)$ be defined by*

$$f(n) = \begin{cases} n^2/48 & n \text{ even} \\ (n+3)^2/48 & n \text{ odd.} \end{cases}$$

Then

$$f(n) - f(n-12) = \begin{cases} (n-6)/2 & n \text{ even} \\ (n-3)/2 & n \text{ odd.} \end{cases}$$

LEMMA 5. *Let $\delta_n = T_n - f(n)$. Then for $n \geq 12$, $\delta_n = \delta_{n-12}$.*

THEOREM.

$$T_n = \langle f(n) \rangle.$$

Proof. It is easy to check that $|\delta_n| \leq 1/3$ for $0 \leq n \leq 11$, so by Lemma 5, we have $|\delta_n| \leq 1/3$ for all n. The result follows. ∎

REFERENCES

1. George E. Andrews, A note on partitions and triangles with integer sides, *Amer. Math. Monthly* **86** (1979), 477.
2. Michael D. Hirschhorn, Triangles with integer sides, revisited, this MAGAZINE **73** (2000), 59–64.
3. R. Honsberger, *Mathematical Gems III*, vol. 9, Dolciana Mathematical Expositions, Mathematical Association of America, Washington, DC, 1985.

4. Tom Jenkyns and Eric Muller, Triangular triples from ceilings to floors, *Amer. Math. Monthly* **107** (2000), 634–639.
5. J. H. Jordan, R. Welch and R. J. Wisner, Triangles with integer sides, *Amer. Math. Monthly* **86** (1979), 686–689.
6. Nicholas Krier and Bennet Manvel, Counting integer triangles, this MAGAZINE **71** (1998), 291–295.

A Generalization of Odd and Even Functions

LAWRENCE J. CRONE
American University
Washington, DC 20016-8050
lcrone@american.edu

The familiar concept of odd and even functions can be formulated in terms of the composition of functions. For each real number λ, let M_λ be the function that multiplies by λ. That is, $M_\lambda(x) = \lambda x$. Suppose f_o is an odd function and f_e is an even function. Then $f_o(-x) = -f_o(x)$, and $f_e(-x) = f(x)$, which can be written

$$f_o \circ M_{-1} = M_{-1} \circ f_o, \tag{1}$$

and

$$f_e \circ M_{-1} = f_e. \tag{2}$$

In other words, the odd functions are those that commute with M_{-1}, and the even functions are those that are invariant with respect to M_{-1}. We generalize these concepts by allowing the function M_{-1} in equations (1) and (2) to be replaced by any of an appropriate class of functions.

Algebra of functional composition In this note we will only consider functions that map zero onto zero and which are continuous on an interval of the form $(-\epsilon, \epsilon)$, where ϵ may depend on the function. We are assured that the composition of any two such functions is defined, and that the composite function also has these properties. For convenience we call such functions *continuously fixed at zero.* The functions M_1 and M_0 act as identity and zero functions, respectively, since for each f, $M_1 \circ f = f = f \circ M_1$, and $M_0 \circ f = M_0 = f \circ M_0$. Although functional composition is not commutative, it is associative, and we usually just write $f \circ g \circ h$ rather than $(f \circ g) \circ h$, or $f \circ (g \circ h)$. There is a right distributive law: $(f + g) \circ h = f \circ h + g \circ h$. Functional composition is not in general left distributive, but we do have a special case: $M_\lambda \circ (g + h) = M_\lambda \circ g + M_\lambda \circ h$. We will also need to use the fact that if f is strictly monotonic on an open interval containing zero, then, restricted to that interval, f has a unique functional inverse, which we write f^{-1}. The functional inverse satisfies the equation $f \circ f^{-1} = M_1 = f^{-1} \circ f$. If f and g are invertible, then so is $f \circ g$ and $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$. (The basic properties of composition of functions and of functional inverses is covered in most calculus texts. For example see Larson [**1**] and Stewart [**2**].)

Symmetry functions The function M_{-1} which appears in the definition of odd and even functions has two essential properties: $M_{-1} \circ M_{-1} = M_1$, and M_{-1} is monotonically decreasing. We will say that a function continuously fixed at zero is a symmetry function, or simply, a symmetry, if it is its own inverse, and is strictly decreasing in

an open interval containing zero. There are many symmetry functions. For example, for each real number α the function $-x/(1+\alpha x)$ is a symmetry. When $\alpha = 0$, this reduces to M_{-1}. Also, for each $\alpha > 0$ the function

$$S(x) = \begin{cases} -\alpha x & x \geq 0 \\ -x/\alpha & x < 0 \end{cases}$$

is a symmetry.

The graphs of symmetry functions are symmetric with respect to the line $y = x$. Symmetry functions are defined by equations that are symmetric in x and y. For example, the function $y = -x/(1+\alpha x)$ is the solution to the symmetric equation $x + y + \alpha xy = 0$. However, not every implicit solution of a symmetric equation is a symmetry. The equation $x^2 + y^2 = 1$ does not define any function that maps zero onto zero. The equation $xy = 0$ has the solution $y = 0$, but that is not a symmetry.

***S*-odd and *S*-even functions** For each symmetry S, we say that a function is S-odd if it satisfies the condition

$$f \circ S = S \circ f \tag{3}$$

and that a function is S-even if it satisfies

$$f \circ S = f. \tag{4}$$

The S-odd and S-even functions corresponding to the symmetry $S = M_{-1}$ are the ordinary odd and even functions. More interesting is the symmetry

$$S(x) = \begin{cases} -2x & x \geq 0 \\ -x/2 & x < 0 \end{cases}.$$

For every real β, the function

$$f_e(x) = \begin{cases} \beta x & x \geq 0 \\ -\beta x/2 & x < 0 \end{cases}$$

is S-even. The situation is a little more complicated when we look for corresponding S-odd functions. For positive β,

$$f_o(x) = \beta x$$

and

$$g_o(x) = \begin{cases} -\beta x & x \geq 0 \\ -\beta x/4 & x < 0 \end{cases}$$

are S-odd.

The function M_0 is both S-odd and S-even for every symmetry. The function M_1 is S-odd for every symmetry, but is never S-even. For any symmetry, S itself is always S-odd, and $M_1 + S$ is always S-even. The function $M_1 - S$ turns out to be extremely interesting, and is central to the discussion following equation (15). There are many S-odd and S-even functions for every symmetry. In fact we will show that for each symmetry, there is a one-to-one correspondence between the S-odd functions and the odd functions, and between the S-even and the even functions. We will derive formulas that give all the S-odd and S-even functions for arbitrary symmetries. The example

section below shows the computation of an S-odd and an S-even function in a more complicated case.

The sets of S-odd and S-even functions share many of the properties of the odd and even functions. For instance, if S is any symmetry then:

1. If f and g are S-odd, then $f \circ g$ is S-odd.
2. If f is S-even and g is S-odd, then $f \circ g$ is S-even.
3. If f is any function and g is S-even, then $f \circ g$ is S-even.
4. If f and g are S-even, then fg is S-even.
5. If f and g are S-even, then $f + g$ is S-even.
6. If f is S-even and α is any real number, then αf is S-even

The proofs are all elementary, and none of them depend on the symmetry properties of S. These are properties of functions that are commuting and invariant with respect to any function. There is another fact about odd and even functions that does not generalize so easily. Every function can be written in a unique way as a sum of an odd function and an even function: $f = f_o + f_e$. We say that f_o is the odd part of f and f_e is the even part of f. It is not hard to see that $f_o(x) = (f(x) - f(-x))/2$, and $f_e(x) = (f(x) + f(-x))/2$. Perhaps the most exciting aspect of this study is the fact that this result generalizes to S-odd and S-even functions for an arbitrary symmetry S

THEOREM 1. *Let S be a symmetry, and let f be continuously fixed at zero. Then f can be written in a unique way as a sum of an S-odd function and an S-even function.*

Proof. Suppose f has the decomposition

$$f = f_o + f_e, \tag{5}$$

where f_o is S-odd and f_e is S-even. Then

$$f \circ S = S \circ f_o + f_e. \tag{6}$$

Subtracting equation (6) from equation (5) gives

$$f - f \circ S = (M_1 - S) \circ f_o. \tag{7}$$

The function $M_1 - S$ is monotonic increasing at zero. Therefore we may compose both sides of equation (7) from the left with $(M_1 - S)^{-1}$ to solve for f_o:

$$f_o = (M_1 - S)^{-1} \circ (f - f \circ S). \tag{8}$$

From equation (5) we have

$$f_e = f - (M_1 - S)^{-1} \circ (f - f \circ S). \tag{9}$$

This shows that if the decomposition exists, it is unique, and the S-odd and S-even parts must be given by equations (8) and (9). It remains to demonstrate that equations (8) and (9) give S-odd and S-even functions, respectively. We first establish several identities, which will prove useful. Since $(f - f \circ S) \circ S = (f \circ S - f)$, we have

$$(f - f \circ S) \circ S = M_{-1} \circ (f - f \circ S). \tag{10}$$

This equation remains true if f is replaced by M_1.

$$(M_1 - S) \circ S = M_{-1} \circ (M_1 - S). \tag{11}$$

The inverse of this equation will also be useful.

$$S \circ (M_1 - S)^{-1} = (M_1 - S)^{-1} \circ M_{-1}. \tag{12}$$

■

Proof that f_o given by equation (8) is S-odd:

$$\begin{aligned} f_o \circ S &= (M_1 - S)^{-1} \circ (f - f \circ S) \circ S \\ &= (M_1 - S)^{-1} \circ M_{-1} \circ (f - f \circ S) && \text{by equation (10)} \\ &= S \circ (M_1 - S)^{-1} \circ (f - f \circ S) && \text{by equation (12)} \\ &= S \circ f_o. \end{aligned}$$

■

Proof that f_e given by equation (9) is S-even: First note that by equation (8) $(M_1 - S) \circ f_o = f - f \circ S$, and therefore $f \circ S = f - (M_1 - S) \circ f_o$. this gives

$$\begin{aligned} f_e \circ S &= (f - f_o) \circ S \\ &= f \circ S - f_o \circ S \\ &= f - (M_1 - S) \circ f_o - f_o \circ S \\ &= f - f_o \\ &= f_e \end{aligned}$$

■

EXAMPLE. We leave it as an exercise to show that if $S = M_{-1}$, then f_o and f_e are the ordinary odd and even parts of f.

For a less familiar example, we now determine the S-odd and S-even parts of the function $M_2(x) = 2x$ with respect to the symmetry $S(x) = -x/(1 + x)$. We will need to know the inverse function $(M_1 - S)^{-1}(x)$, so we first compute

$$x = (M_1 - S)(y) = y + y/(1 + y) = (y^2 + 2y)/(1 + y),$$

and solve for y:

$$\begin{aligned} x + xy &= y^2 + 2y \\ y^2 - 2(x/2 - 1)y - x &= 0. \end{aligned}$$

The quadratic formula gives

$$\begin{aligned} y = (M_1 - S)^{-1}(x) &= (x/2 - 1) + \sqrt{(x/2 - 1)^2 + x} \\ &= (x/2 - 1) + \sqrt{x^2/4 + 1}. \end{aligned} \tag{13}$$

We need to take the "+" sign in the quadratic formula to make zero map to zero. The next step is to compute

$$(M_2 - M_2 \circ S)(x) = 2x - 2(-x/(1 + x)) = 2(x + x/(1 + x)). \tag{14}$$

Composing equations (13) and (14) gives

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

Finally, since $f_e = M_2 - f_o$, we have

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

The reader is invited to verify algebraically that these functions are indeed S-odd and S-even. This can also be verified graphically. The closure of the L-shaped region in FIGURE 1, and the rectangle in FIGURE 2 demonstrate that equations (3) and (4) hold at x_0.

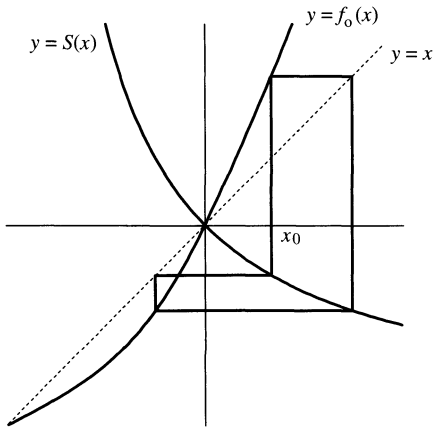


Figure 1 The L-shaped region closes because f_o is S-odd

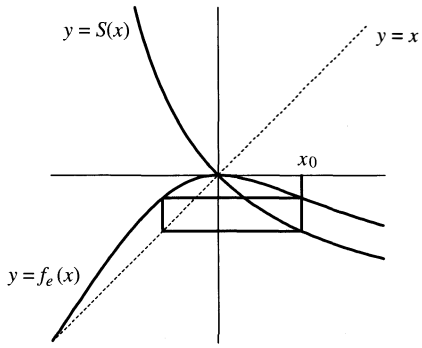


Figure 2 The rectangle closes because f_e is S-even

It is only slightly messier to compute the S-odd and S-even parts of M_λ with respect to $S_\alpha(x) = -x/(1+\alpha x)$. In this case the S-odd and S-even parts are given by

$$f_o(x) = \frac{\alpha\lambda(x + \frac{x}{1+\alpha x}) - 2 + \sqrt{\alpha^2\lambda^2(x + \frac{x}{1+\alpha x})^2 + 4}}{2\alpha},$$

and

$$f_e(x) = \frac{\alpha\lambda(x - \frac{x}{1+\alpha x}) + 2 - \sqrt{\alpha^2\lambda^2(x + \frac{x}{1+\alpha x})^2 + 4}}{2\alpha}.$$

Notice that the computations involved are relatively straightforward, with the possible exception of the computation of the inverse of $(M_1 - S)$. In particular, it would not be hard to determine the S-odd and S-even parts of $f(x) = \sin x$ or $\ln(x + 1)$ with respect to $S(x) = -x/(1 + x)$.

Conjugation There is a wealth of information hidden in equation (11). Multiplying equation (11) on the left by $(M_1 - S)^{-1}$ gives

$$S = (M_1 - S)^{-1} \circ M_{-1} \circ (M_1 - S). \tag{15}$$

We see that every symmetry is of the form $h^{-1} \circ M_{-1} \circ h$, where h is an invertible function. We say that an invertible function h determines a conjugation, which maps the set of functions continuously fixed at zero into itself. That is, if f is any such function, then the congugate of f is $\hat{f}$, where

$$\hat{f} = h^{-1} \circ f \circ h. \tag{16}$$

This map is one-to-one and onto, and the inverse map is the conjugation determined by h^{-1}. Conjugations have the nice property that the conjugate of the composition of two functions is the composition of their conjugates. If h is invertible, and f and g are arbitrary, then

$$h^{-1} \circ (f \circ g) \circ h = (h^{-1} \circ f \circ h) \circ (h^{-1} \circ g \circ h). \tag{17}$$

Conjugation is an essential tool in the study of the commutativity of functions. In particular, we state two simple corollaries of equation 17:

COROLLARY 1. *f commutes with g if and only if $h^{-1} \circ f \circ h$ commutes with $h^{-1} \circ g \circ h$.*

COROLLARY 2. *f is invariant with respect to g if and only if $h^{-1} \circ f \circ h$ is invariant with respect to $h^{-1} \circ g \circ h$.*

Our first application of these ideas is to characterize symmetry functions.

THEOREM 2. *A function S is a symmetry function if and only if there is a function h that is strictly monotonic on an open interval containing zero such that*

$$S = h^{-1} \circ M_{-1} \circ h.$$

Proof. Equation (15) shows that S has the form $h^{-1} \circ M_{-1} \circ h$. If we have $S = h^{-1} \circ M_{-1} \circ h$, then

$$S \circ S = h^{-1} \circ M_{-1} \circ h \circ h^{-1} \circ M_{-1} \circ h = h^{-1} \circ M_{-1} \circ M_{-1} \circ h = h^{-1} \circ h = M_1.$$

Because h is strictly monotonic and M_{-1} is strictly decreasing, S is strictly decreasing.

The conjugating function h in Theorem 2 is not unique. If k is any odd function and $g = k \circ h$, then $g^{-1} \circ M_{-1} \circ g = h^{-1} \circ M_{-1} \circ h$.

Theorem 2 gives us a recipe for constructing symmetries. Theorem 3 gives us recipes for the construction of all the S-odd and S-even functions corresponding to a given symmetry. ■

THEOREM 3. *Let $S = h^{-1} \circ M_{-1} \circ h$ be a symmetry, and let f be arbitrary. Then f is S-odd if and only if $h \circ f \circ h^{-1}$ is odd, and f is S-even if and only if $h \circ f \circ h^{-1}$ is even.*

Proof. If $S = h^{-1} \circ M_{-1} \circ h$, then $M_{-1} = h \circ S \circ h^{-1}$. The result follows from Corollaries 1 and 2. ■

These are practical recipes in the sense that, given h, we can compute S and the S-odd and S-even functions. If we start with a symmetry S, on the other hand, we know that $M_1 - S$ is an appropriate choice for h. Note that Theorem 5 establishes a one-to-one correspondence between the S-odd functions and the odd functions, and between the S-even functions and the even functions.

As an example of the power of Theorem 3, consider the following observations. We remarked above that M_0 is both S-odd and S-even for every symmetry S. It is easy to check that M_0 is the only function that is both odd and even. Since every conjugate of M_0 is M_0, the only function that is S-odd and S-even for any symmetry is M_0. It is also easy to see that no even function can be strictly monotonic in every open interval containing zero. By Theorem 3 this is true for S-even functions with respect to any symmetry. The odd functions x, $-x$, and $x^2 \sin(1/x)$ demonstrate that S-odd functions can be monotonically increasing, decreasing, or neither near zero.

We might suspect that conjugation gives an easier way to determine the S-odd and S-even parts of a function f. Let $h = (M_1 - S)^{-1}$. Then $\hat{S} = h^{-1} \circ S \circ h = M_{-1}$. The conjugate of f, $\hat{f} = h^{-1} \circ f \circ h$, has a unique decomposition into odd and even parts: $\hat{f} = \hat{f}_o + \hat{f}_e$. The inverse conjugates f_o and f_e of $\hat{f}_o$ and $\hat{f}_e$ are S-odd and S-even, respectively. But unless $S = M_{-1}$ it is not true that $f = f_o + f_e$.

We conclude with a rather surprising theorem.

THEOREM 4. *Let S be a symmetry. Then the S-odd parts of the functions M_λ for all real λ commute with each other.*

Proof. By equation (8) the S-odd part of M_λ is $(M_1 - S)^{-1} \circ M_\lambda \circ (M_1 - S)$. The result follows from Corollary 1 since $M_\lambda \circ M_\alpha = M_{\lambda\alpha} = M_\alpha \circ M_\lambda$. ■

REFERENCES

1. Roland E. Larson, Robert P. Hostetler, and Bruce H. Edwards, *Calculus of a Single Variable, 6th Edition*, Houghton Mifflin Company, 1998.
2. James Stewart, *Calculus, 4th Edition*, Brooks/Cole Publishing Company, 1999.

The Number of 2 by 2 Matrices over $\mathbb{Z}/p\mathbb{Z}$ with Eigenvalues in the Same Field

GREGOR OLŠAVSKÝ
Penn State Erie, The Behrend College
Erie, PA 16563
gxo1@psu.edu

In this note, I count the number of 2 by 2 matrices over the field of integers mod p (where p is an odd prime), with the added restriction that all eigenvalues must also belong to $\mathbb{Z}/p\mathbb{Z}$. One consequence of the count is that, as p gets larger, the number of such matrices approaches half the total number of 2 by 2 matrices over $\mathbb{Z}/p\mathbb{Z}$. We use some easy matrix theory, as well as a simple result from group theory about counting conjugacy classes [**1**].

Given any 2 by 2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

the eigenvalues of A are the roots of the quadratic

$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = 0,$$

which is called the *characteristic equation* of A. The left-hand side is a polynomial of degree two, the *characteristic polynomial*:

$$p(\lambda) = \lambda^2 - (a+d)\lambda + ad - bc.$$

If the entries of the matrix A are real numbers, the roots of the characteristic polynomial are not necessarily real. That is, the eigenvalues of a matrix might not belong to same field that provides its entries. I suppose the classic example is the rotation matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

whose characteristic polynomial is $\lambda^2 + 1$ and whose eigenvalues are i and $-i$.

Of course, for 2 by 2 matrices with real entries, the number of matrices having real eigenvalues is infinite (as is the number not having real eigenvalues). But for matrices over a given finite field like $\mathbb{Z}/p\mathbb{Z}$ (which we abbreviate $\mathcal{F}_p$), it is actually possible to count the number of matrices whose eigenvalues lie in the same field.

The easy case: singular matrices If a matrix over $\mathcal{F}_p$ is singular, then all its eigenvalues must also belong to $\mathcal{F}_p$, since a singular matrix always has zero as an eigenvalue. This can be seen from the characteristic equation. Since $ad - bc = 0$,

$$p(\lambda) = \lambda^2 - (a+d)\lambda = \lambda(\lambda - (a+d)) = 0, \qquad a, d \in \mathcal{F}_p.$$

There are p^4 2 by 2 matrices total; by eliminating the nonsingular ones, we will be able to complete this easy case. A relatively easy count [**2**] shows that there are $p(p-1)^2(p+1)$ nonsingular 2 by 2 matrices with entries in $\mathcal{F}_p$, so this leaves $p^3 + p^2 - p$ singular matrices, all having their eigenvalues in $\mathcal{F}_p$.

The not-so-easy case: nonsingular matrices Let's consider the nonsingular matrices with entries from $\mathcal{F}_p$, known as the general linear group of 2 by 2 matrices over $\mathcal{F}_p$:

$$GL(2, \mathcal{F}_p) = \left\{ A : A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, a, b, c, d, \in \mathcal{F}_p \quad \text{and} \quad \det(A) \neq 0 \right\}.$$

The order of this group was previously given [**2**] as

$$\left|GL(2, \mathcal{F}_p)\right| = p(p-1)^2(p+1).$$

Since $\det(A)$ (the product of the eigenvalues) is nonzero for any matrix A in $GL(2, \mathcal{F}_p)$, zero cannot be an eigenvalue of such a matrix. Thus, if the eigenvalues of A are in $\mathcal{F}_p$, the only possibilities are $1, 2, \ldots, p-1$. If a quadratic polynomial has one root in $\mathcal{F}_p$, then it must have both roots there. So the matrices we exclude from our count are those with no eigenvalue in $\mathcal{F}_p$, that is, the ones whose characteristic polynomials cannot be factored (that is, are irreducible) over $\mathcal{F}_p$.

When I first tackled this problem, I considered counting a simpler case: the subgroup $TL(2, \mathcal{F}_p)$ of $GL(2, \mathcal{F}_p)$ consisting of the upper triangular matrices. This was natural, since a quick look at the characteristic equation shows that the eigenvalues of

an upper triangular matrix sit on the main diagonal. Hence, every matrix in $TL(2, \mathcal{F}_p)$ has its eigenvalues in $\mathcal{F}_p$. The order of $TL(2, \mathcal{F}_p)$ is easily seen to be $p(p-1)^2$.

For $p = 3$, I formed the left cosets in $GL(2, \mathcal{F}_3)$ by multiplying $TL(2, \mathcal{F}_3)$ by various matrices from $GL(2, \mathcal{F}_3)$. Any such coset necessarily has the same cardinality as $TL(2, \mathcal{F}_3)$ and $GL(2, \mathcal{F}_3)$ is partitioned into $p + 1$ disjoint subsets, because

$$\frac{\text{order of the group}}{\text{order of the subgroup}} = \frac{p(p-1)^2(p+1)}{p(p-1)^2} = p + 1.$$

Interestingly enough, in every coset except $TL(2, \mathcal{F}_3)$, exactly half of the matrices have their eigenvalues in $\mathcal{F}_3$. To see whether this might hold for larger values of p, I used a computer program in C++ written by Charles Burchard, which showed that the pattern continued for every prime we tested. I conjectured that this would be true in general and used it to determine a preliminary count.

How do we use the conjectured result? Since every matrix in $TL(2, \mathcal{F}_p)$ has eigenvalues in $\mathcal{F}_p$, there are $p(p-1)^2$ such matrices. For the other p distinct cosets, suppose that exactly half have their eigenvalues in $\mathcal{F}_p$. Then we have a total of

$$\frac{p}{2}(p(p-1)^2) = \frac{p^2}{2}(p-1)^2$$

matrices with eigenvalues in $\mathcal{F}_p$ in all these cosets. Thus the total number of matrices having eigenvalues in $\mathcal{F}_p$ is

$$p(p-1)^2 + \frac{p^2}{2}(p-1)^2 = \frac{p}{2}(p-1)^2(p+2).$$

Of course, a computer test does not prove the general result for all primes p. A proof is in order.

THEOREM. The number of matrices in $GL(2, \mathcal{F}_p)$ with eigenvalues in $\mathcal{F}_p$ is

$$\frac{p}{2}(p^3 - 3p + 2).$$

Proof. Let A be an element of $GL(2, \mathcal{F}_p)$ whose eigenvalues λ_1 and λ_2 lie in $\mathcal{F}_p$, actually in $\mathcal{F}_p^*$, the set of nonzero elements in $\mathcal{F}_p$. Since the eigenvalues of A lie in $\mathcal{F}_p$, a standard result from linear algebra says that A is conjugate in $GL(2, \mathcal{F}_p)$ to its Jordan canonical form. We plan to count the number of distinct conjugacy classes and then count the number of elements in each class.

If $\lambda_1 \neq \lambda_2$, then the Jordan canonical form for A is

$$A(\lambda_1, \lambda_2) = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Of course, it could also be written $A(\lambda_2, \lambda_1)$ because $A(\lambda_1, \lambda_2)$ and $A(\lambda_2, \lambda_1)$ are conjugate in $GL(2, \mathcal{F}_p)$. Thus, the number of distinct conjugacy classes of elements of $GL(2, \mathcal{F}_p)$ with distinct eigenvalues both in $\mathcal{F}_p$ is the number of two-element subsets of $\mathcal{F}_p^*$, that is, $(p-1)(p-2)/2$.

An elegant way to count the elements in a group conjugate to a given element is to use the *centralizer* of that element, the set of elements that commute with it. The centralizer of $A(\lambda_1, \lambda_2)$ is easily found to be

$$C_{GL(2,\mathcal{F}_p^*)}(A(\lambda_1, \lambda_2)) = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} : a, b \in \mathcal{F}_p^* \right\}.$$

Since the order of this centralizer is $(p-1)^2$, the conjugacy class of $A(\lambda_1, \lambda_2)$ contains $|G|/(p-1)^2 = p(p+1)$ elements [**1**]. Thus, the number of elements of $GL(2, \mathcal{F}_p)$ with distinct eigenvalues both in $\mathcal{F}_p$ is

$$\frac{p(p+1)(p-1)(p-2)}{2}.$$

If $\lambda_1 = \lambda_2 = \lambda$, then the Jordan canonical form is either

$$B(\lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \quad \text{or} \quad C(\lambda) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

Because $B(\lambda)$ is in the center of $GL(2, \mathcal{F}_p)$, the conjugacy class of $B(\lambda)$ contains exactly one element, namely itself. There are exactly $p-1$ conjugacy classes of this type, one for each $\lambda \in \mathcal{F}_p^*$.

It is not hard to compute that

$$C_{GL(2,\mathcal{F}_p)}(C(\lambda)) = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a, b \in \mathcal{F}_p, a \neq 0 \right\}$$

Since the order of this centralizer is $p(p-1)$, the conjugacy class of $C(\lambda)$ contains exactly

$$\frac{|G|}{p(p-1)} = (p-1)(p+1)$$

elements. There are exactly $p-1$ conjugacy classes of this type, one for each $\lambda \in \mathcal{F}_p^*$. Adding up all the cases, we find that the number of matrices in $GL(2, \mathcal{F}_p)$ with eigenvalues in $\mathcal{F}_p$ is

$$\frac{(p-1)(p-2)}{2} p(p+1) + (p-1) + (p-1)(p-1)(p+1),$$

which simplifies to $\frac{p}{2}\left(p^3 - 3p + 2\right)$. ■

Grand finale If we add the number of singular and nonsingular matrices whose eigenvalues lie in $\mathcal{F}_p$, we have:

$$p^3 + p^2 - p + \frac{p}{2}\left(p^3 - 3p + 2\right) = \frac{p^2}{2}\left(p^2 + 2p - 1\right).$$

One implication of this result is that the ratio of matrices in $\mathcal{F}_p$ having eigenvalues in $\mathcal{F}_p$ to the total number of such matrices is

$$\frac{\frac{p^2}{2}\left(p^2 + 2p - 1\right)}{p^4}.$$

As p approaches infinity, this fraction decreases to a limiting value of $1/2$.

Acknowledgment. Special thanks to Prof. Michael Barry from Allegheny College for help with the proof.

REFERENCES

1. I. N. Herstein, *Topics in Algebra*, 2nd ed., Xerox College Publishing, Lexingon, MA, 1975, p. 83.
2. G. Olšavský, Groups formed from 2 × 2 matrices over $\mathbb{Z}_p$, this MAGAZINE, **63**:4 (1990), 269–272.

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

Assistant Editors: RĂZVAN GELCA, Texas Tech University; ROBERT GREGORAC, Iowa State University; GERALD HEUER, Concordia College; VANIA MASCIONI, Ball State University; PAUL ZEITZ, The University of San Francisco

Proposals

To be considered for publication, solutions should be received by March 1, 2004.

1677. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let triangle ABC be inscribed in circle $\mathcal{C}$. Let A^* be the point on $\mathcal{C}$ that bisects the arc BC that contains A. Let B^* and C^* be defined in a similar way.

(a) Prove that $\overline{A^*B^*}$ is parallel to $\overline{C^*C}$, $\overline{B^*C^*}$ is parallel to $\overline{A^*A}$, and $\overline{C^*A^*}$ is parallel to $\overline{B^*B}$.

(b) The points A, B, C, A^*, B^*, C^* partition $\mathcal{C}$ into six arcs. The length of the shortest of these can be taken as a measure of the "crowdedness" of the six points. How should A, B, C be chosen so that the points A, B, C, A^*, B^*, C^* are least crowded?

1678. *Proposed by Kent Holing, Statoil Research Center, Trondheim, Norway.*

Let $P[h, w, l]$ denote the following problem: *A box of height h and width w is placed on the floor so it is flush with a wall and the sides of length w are perpendicular to the wall. A ladder of length l leans against the wall and just touches the box along the upper edge. Let x be the length of the portion of the ladder that is between the wall and the point of contact with the box.*

(a) Characterize all ordered triples (h, w, l) of positive integers such that the problem $P[h, w, l]$ has at least one integral solution for x.

(b) Characterize all ordered triples (h, w, l) of positive integers such that the problem $P[h, w, l]$ has exactly one solution for x, and that solution is integral.

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

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

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

1679. *Proposed by José Luis Díaz-Barrero and Juan José Egozcue, Universitat Politècnica de Catalunya, Barcelona, Spain.*

Let f_k denote the k^{th} Fibonacci number, that is, $f_0 = 0$, $f_1 = 1$, and $f_{k+2} = f_{k+1} + f_k$, $k \geq 0$. Prove that for any positive integer n,

$$\sum_{k=1}^{n} \binom{n}{k} \log\left(f_k^{f_{2n}}\right) \leq (2^n - 1) \sum_{k=1}^{n} \binom{n}{k} \log\left(f_k^{f_k}\right).$$

1680. *Proposed by H. A. Shah Ali, Tehran, Iran.*

For positive integers k and n, with $k \leq n$, define the k^{th} elementary symmetric function $S_{k,n}$ of the n numbers $x_1, x_2, \ldots, x_n$ by

$$S_{k,n} = S_{k,n}(x_1, \ldots, x_n) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k}.$$

Let $c \in (0, 1]$ be given and assume that $x_i \geq 0$ for $1 \leq i \leq n$, and that $\sum_{k=1}^{n} S_{k,n} = c$. Find the minimum and maximum values of $S_{k,n}$ for each $1 \leq k \leq n$, and determine necessary and sufficient conditions for the extrema to occur.

1666. *Proposed by Razvan A. Satnoianu, City University, London, England.*

Let $f, g : [0, \infty) \longrightarrow [0, \infty)$ be functions with f increasing, g entire, and $g^{(n)}(0) \geq 0$ for all nonnegative integers n. Prove that if $x \geq y \geq z \geq 0$ (or $y \geq z \geq x \geq 0$ or $z \geq x \geq y \geq 0$), then

$$f(x)\,(g(x) - g(y))\,(x - z) + f(y)\,(g(y) - g(z))\,(y - x) + f(z)\,(g(z) - g(x))\,(z - y) \geq 0.$$

(This problem was originally published with the conditions on the variables incorrectly stated.)

Quickies

Answers to the Quickies are on page 325.

Q933. *Proposed by Jody M. Lockhart and William P. Wardlaw, U. S. Naval Academy, Annapolis, MD.*

Let f_k denote the kth Fibonacci number (with $f_0 = 0$ and $f_1 = 1$) and let p be a positive prime. Prove that if p divides f_k, then for each positive integer r, p^{r+1} divides $f_{p^r k}$.

Q934. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

A tetrahedron has base with sides of length a, b, c and an altitude of length h to this base. Determine the minimum possible surface area for the tetrahedron.

Solutions

A Trapezoid in a Triangle **October 2002**

1653. *Proposed by Larry Hoehn, Austin Peay State University, Clarksville, TN.*

In $\triangle ABC$, let F and G be points on $\overline{AB}$ with F between G and A, and let D and E be on $\overline{AC}$ with E between D and A. Prove that if $\triangle BDE \sim \triangle CGF$, then quadrilateral $DEFG$ is a trapezoid.

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

Because $\angle BDE = \angle CGF$, we have $\angle BDC = \angle CGB$. Thus, quadrilateral $BCDG$ is cyclic, and $\angle GDC = \pi - \angle ABC$. By a similar argument, quadrilateral $BCEF$ is cyclic and $\angle FEC = \pi - \angle ABC$. It follows that $\overline{FE}$ is parallel to $\overline{GD}$, and hence that $DEFG$ is a trapezoid.

II. *Solution by Michel Bataille, Rouen, France.*

Identify the points with complex numbers, with A as the origin. For the moment let $G = \beta B$, $D = \gamma C$, $F = \lambda G$, and $E = \mu D$, for some real numbers $\beta, \gamma \neq 0$ and $\lambda, \mu \neq 1$. (Note that for the problem statement, $0 < \beta, \gamma, \lambda, \mu < 1$.) We prove that

(a) If $\triangle BDE$ and $\triangle CGF$ are oppositely similar, then $DEFG$ is a trapezoid.

(b) If D, E, F, G are arranged as described in the problem statement, then $\triangle BDE$ and $\triangle CGF$ cannot be directly similar.

The result will then follow.

(a) If $\triangle BDE$ and $\triangle CGF$ are oppositely similar, then there are complex numbers a, b, with $a \neq 0$, such that $C = a\overline{B} + b$, $G = a\overline{D} + b$, and $F = a\overline{E} + b$. It follows that

$$a = \frac{G - C}{\overline{D} - \overline{B}} = \frac{\beta B - C}{\gamma\overline{C} - \overline{B}} \quad \text{and} \quad a = \frac{F - G}{\overline{E} - \overline{D}} = \frac{(\lambda - 1)\beta B}{(\mu - 1)\gamma\overline{C}}, \quad \text{from which,}$$

$$(\beta B - C)(\mu - 1)\gamma\overline{C} = (\gamma\overline{C} - \overline{B})(\lambda - 1)\beta B.$$

Rearranging gives

$$\frac{B}{C}(\mu - \lambda) = \frac{\mu - 1}{\beta} - \frac{\lambda - 1}{\gamma}\frac{|B|^2}{|C|^2}.$$

Now B/C cannot be real, for otherwise A, B and C are collinear. Thus it must be the case that $\mu = \lambda$. It follows that $\overrightarrow{FE} = \lambda\overrightarrow{GD}$, so $DEFG$ is a trapezoid (convex if $\mu = \lambda > 0$, and self-crossing if $\mu = \lambda < 0$.)

(b) If $\triangle BDE$ and $\triangle CGF$ are directly similar, then there are complex numbers a, b, with $a \neq 0$, such that $C = aB + b$, $G = aD + b$, and $F = aE + b$. As above, it follows that $(\mu - \lambda)\beta\gamma BC = \gamma(\mu - 1)C^2 - \beta(\lambda - 1)B^2$. Thus B/C satisfies the quadratic equation

$$(1 - \lambda)\beta X^2 - (\mu - \lambda)\beta\gamma X - (1 - \mu)\gamma = 0.$$

Because B/C is not real, we must have $(\mu - \lambda)^2\beta^2\gamma^2 + 4(1 - \lambda)(1 - \mu)\beta\gamma < 0$. However, this certainly cannot occur if $\beta\gamma > 0$ and $0 < \lambda, \mu < 1$, which is the situation described in the problem statement.

Also solved by Mangho Ahuja, Ayoub B. Ayoub, Herb Bailey, Roy Barbara (Lebanon), Jean Bogaert (Belgium), Knut Dale (Norway), Daniele Donini (Italy), Petar Drianov (Canada), Ovidiu Furdui, Habib Y. Far, Julien Grivaux (France), John G. Heuver, Geoffrey A. Kandall, Ken Korbin, Victor Y. Kutsenok, Neela Lakshmanan

(India), Hien Nguyen, Ünsal Okuducu (Turkey), Raul A. Simon (Chile), Achilleas Sinefakopoulos, H. T. Tang, R. S. Tiberio, Eric Wepsic, Michael Woltermann, Peter Y. Woo, Li Zhou, and the proposer.

A Traveling Salesman! **October 2002**

1654. *David Singmaster, London, England.*

A salesman's office is located on a straight road. His N customers are all located along this road to the east of the office, with the office of customer k at distance k from the salesman's office. The salesman must make a driving trip whereby he leaves the office, visits each customer exactly once, then returns to the office. Because he makes a profit on his mileage allowance, the salesman wants to drive as far as possible during his trip. What is the maximum possible distance he can travel on such a trip, and how many different such trips are there? Assume that if the travel plans call for the salesman to visit customer j immediately after he visits customer i, then he drives directly from i to j.

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

Write $N = 2n$ if N is even and $N = 2n + 1$ if N is odd. For each k, $1 \le k \le n$, the mile between $k - 1$ and k can be traveled at most $2k$ times because there are only k possible places to the west of customer k from which the salesman can drive east through this mile. Similarly, for each k, $n + 1 \le k \le N$, the mile from $k - 1$ to k can be traveled at most $2(N - k + 1)$ times. Thus an upper bound for the mileage for any trip is

$$\begin{aligned} &2(1 + 2 + \cdots + n) + 2((N - n) + (N - n - 1) + \cdots + 1) \\ &\quad = n(n + 1) + (N - n)(N - n + 1) \\ &\quad = \begin{cases} 2(n + 1)^2 & N = 2n + 1 \\ 2n(n + 1) & N = 2n. \end{cases} \end{aligned}$$

If $N = 2n + 1$ is odd, then a mileage total of $2(n + 1)^2$ is attained by starting from the office at 0, visiting a customer at a point in $B = \{n + 1, n + 2, \ldots, N\}$, then visiting a customer in $A = \{1, 2, \ldots, n\}$, then alternating visits to points in A and B, and ending with a leg from a point in B back to 0. Because the points in B can be visited in $(n + 1)!$ orders and the points in A can be visited in $n!$ orders, there are $n!(n + 1)!$ trips that generate maximal mileage. Now let $N = 2n$ be even, let $A = \{1, 2, \ldots, n - 1\}$ and $B = \{n + 1, n + 2, \ldots, N\}$. Note that for a trip of maximal length, each of the miles $n - 1$ to n and n to $n + 1$ must be crossed $2n$ times. This can be done only if the two miles are crossed consecutively each time the salesman travels east or west. Thus a trip of maximal mileage can be described as follows: leave 0, travel to a location in B, then alternate between B and A, and at the end of the trip, travel from a point in B to 0. During one of the $2n$ legs the salesman makes the stop at n. Because the points of A can be visited in $(n - 1)!$ ways and the points of B in $n!$ ways, there are $2n(n - 1)!n! = 2(n!)^2$ different trips of maximal mileage.

II. *Solution by Herb Bailey and David Mutchler, Rose–Hulman Institute of Technology, Terre Haute, IN.*

Let x_i, $1 \le i \le n$ be the location of the customer that the salesman visits at stop i, so x_i is the distance from the office (at the origin) to customer i. The total distance L that the salesman travels is

$$L = x_1 + |x_2 - x_1| + |x_3 - x_2| + \cdots + |x_n - x_{n-1}| + x_n = \sum_{i=1}^{n} c_i x_i,$$

where each c_i is $2, 0$, or -2. We make some observations:

(a) x_i is to the east of both x_{i-1} and x_{i+1} if and only if $c_i = 2$.

(b) x_i is to the west of both x_{i-1} and x_{i+1} if and only if $c_i = -2$.

(c) x_i is between x_{i-1} and x_{i+1} if and only if $c_i = 0$.

Next we prove that if $c_i = 2$ and $c_j = 2$, then there must be a k between i and j with $c_k = -2$. Without loss of generality, we may assume that $i < j$. Suppose that $c_k = 0$ or 2 for each k between i and j. First assume that $x_i < x_j$. Because $c_i = 2$, it follows from a) that $x_{i+1} < x_i$. Also, because $c_{i+1} = 0$ or 2, we can use c) or a) to conclude that $x_{i+2} < x_{i+1}$, and hence that $x_{i+2} < x_i$. Continuing in this manner we find that $x_j < x_i$, which contradicts our assumption that $x_i < x_j$. If $x_j < x_i$, similar reasoning shows that $x_{j-1} < x_j$, and $x_{j-2} < x_j$, and eventually $x_i < x_j$, contradicting the assumption $x_j < x_i$.

To maximize L, we maximize the number of c_ks that have value 2 and assign maximal values to the corresponding x_ks. By the result proved above, for each pair (i, j) with $c_i = c_j = 2$, there must be a k between i and j such that $c_k = -2$. We assign minimal values to the corresponding x_ks. There are two cases to consider.

CASE 1, n ODD. Set $c_i = 2$ for i odd and $c_i = -2$ for i even. A maximal value of L is obtained by assigning values $\frac{n-1}{2} + j$, $1 \le j \le (n+1)/2$ to the x_i for i odd, and assigning values j, $1 \le j \le (n-1)/2$ to the remaining x_is. This assignment can be done in $[(n+1)/2]![(n-1)/2]!$ ways. The corresponding maximal value for L is

$$L = 2\sum_{k=1}^{(n+1)/2}\left(\frac{n-1}{2} + k\right) - 2\sum_{k=1}^{(n-1)/2} k = \frac{(n+1)^2}{2}.$$

CASE 2, n EVEN. To maximize L, set $c_i = 2$ for $n/2$ of the is, set $c_i = -2$ for some i between each pair of c_ks that are equal to 2, and set the remaining c_i to 0. Again assigning maximal values to the x_i for which $c_i = 2$, minimal values to those for which $c_i = -2$ and value $n/2$ to the x_i for which c_i=0, we find the maximum value of L is

$$L = 2\sum_{k=1}^{n/2}\left(\frac{n}{2} + k\right) - 2\sum_{k=1}^{(n-2)/2} k + 0 \cdot \frac{n}{2} = \frac{n(n+2)}{2}.$$

Because the $c_i = 0$ can be placed in the list at any of the n positions, the number of trips of maximal length is $n[n/2]![(n-1)/2]! = 2[(n/2)!]^2$.

Also solved by Jean Bogaert (Belgium), Con Amore Problem Group (Denmark), Chip Curtis, Daniele Donini (Italy), Marty Getz and Dixon Jones, Jerrold W. Grossman and R. Suzanne Zeitman, The Ithaca College Solvers, Tom Jager, Victor Y. Kutsenok, Kevin McDougal, Rob Pratt, Southwest Missouri State University Problem Solving Group, Philip D. Straffin, Christopher N. Swanson, Allen Terny, Paul J. Zwier, and the proposer. There were two incorrect submissions.

A Symmetric Inequality **October 2002**

1655. *Proposed by Costas Efthimiou, Department of Physics, University of Central Florida, Orlando, FL.*

Let $0 < a, b, c < 1$ with $ab + bc + ca = 1$. Prove that

$$\frac{a}{1-a^2} + \frac{b}{1-b^2} + \frac{c}{1-c^2} \ge \frac{3\sqrt{3}}{2},$$

and give the conditions under which equality holds.

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

Let $a = \tan\alpha$, $b = \tan\beta$, and $c = \tan\gamma$, with $0 < \alpha, \beta, \gamma < \frac{\pi}{4}$. Because $ab + bc + ca = 1$, $\tan(\alpha + \beta + \gamma) = \dfrac{a + b + c - abc}{1 - ab - bc - ca}$ is undefined. Thus $\alpha + \beta + \gamma = \frac{\pi}{2}$. Because $\tan x$ is strictly convex on $(0, \frac{\pi}{2})$,

$$\frac{2a}{1-a^2} + \frac{2b}{1-b^2} + \frac{2c}{1-c^2} = \tan(2\alpha) + \tan(2\beta) + \tan(2\gamma)$$
$$\geq 3\tan\left(\frac{2(\alpha+\beta+\gamma)}{3}\right) = 3\sqrt{3}.$$

Equality holds if and only if $\alpha = \beta = \gamma$, that is, if and only if $a = b = c = \frac{1}{\sqrt{3}}$.

Note. Roy Barbara proved the following generalization: Let $n \geq 2$ be a positive integer and let k be a constant with $0 < k < n(n-1)/2$. If $0 < a_1, a_2, \ldots, a_n < 1$ and $\sum_{1 \leq i < j \leq n} a_i a_j = k$, then $\sum_{i=1}^{n} \dfrac{a_i}{1 - a_i^2} \geq \dfrac{n\sqrt{2kn(n-1)}}{n(n-1) - 2k}$, with equality if and only if $a_1 = a_2 = \cdots = a_n = \sqrt{\dfrac{2k}{n(n-1)}}$.

Also solved by Raza Akhlaghi, Mangho Ahuja, Tsehaye Andebrhan, Michael Andreoli, Dionne T. Bailey and Elsie Campbell and Charles Diminnie and Karl Havlak, Herb Bailey, Roy Barbara (Lebanon), Michel Bataille (France), Michael S. Becker and Charles K. Cook, Mihály Bencze (Romania), Rafael D. Benguria (Chile), Anthony C. D. Blackman (West Indies), Jean Bogaert (Belgium), Pierre Bornsztein (France), Paul Bracken (Canada), Erhard Braüne (Austria), Minh Can, Mario Catalani (Italy), J. D. Chow, Con Amore Problem Group (Denmark), Chip Curtis, Daniele Donini (Italy), Petar Drianov (Canada), Zuming Feng, Arthur H. Foss, Ovidiu Furdui, Marty Getz and Dixon Jones, Julien Grivaux (France), John G. Heuver, Enkel Hysnelaj (Australia), D. Kipp Johnson, Stephen Kaczkowski, Parviz Khalili, Murray S. Klamkin (Canada), Victor Y. Kutsenok, Neela Lakshmanan (India), Elias Lampakis (Greece), Kee-Wai Lau (China), Sai Luk Lau (China), Charles McCracken, McDaniel College Problems Group, Northwestern University Math Problem Solving Group, Peter E. Nüesch (Switzerland), Robert Poodiak, B. Ravikumar, Phillip P. Ray, Rolf Richberg (Germany), Ralph Rush, Ossama A. Saleh and Stan Byrd, Achilleas Sinefakopoulos, University of West Alabama Problems Group, Eric Wepsic, Roman Wong, Michael Vowe (Switzerland), David Zhu, Paul Zwier, and the proposer. There was one solution with no name, one incorrect submission, and one illegible submission.

A Symmetric Inequality **October 2002**

1655. *Proposed by Costas Efthimiou, Department of Physics, University of Central Florida, Orlando, FL.*

Linear Combinations **October 2002**

1656. *David M. Bloom, Hartsdale, NY.*

Let $a_1, a_2, \ldots, a_k$, $k \geq 2$ be pairwise relatively prime positive integers, and let

$$b_j = \left(\prod_{i=1}^{k} a_i\right) \Big/ a_j, \quad 1 \leq j \leq k, \quad \text{and}$$
$$M = (k-1)a_1 a_2 \cdots a_k - (b_1 + b_2 + \cdots + b_k).$$

Prove that for each positive integer n, one and only one of the integers n and $M - n$ can be expressed in the form $x_1 b_1 + x_2 b_2 + \cdots + x_k b_k$, where the x_i are nonnegative integers. (*Note: This problem generalizes Problem 1561, this* MAGAZINE, *Dec. 1999, 411–412.*)

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

First observe that because the numbers $b_1, b_2, \ldots, b_k$ are relatively prime, every integer can be written as a linear combination of the b_is. Next, because

$$\lambda_1 b_1 + \lambda_2 b_2 + \cdots + \lambda_k b_k = \sum_{i \neq j,k} \lambda_i b_i + (\lambda_j - a_j \lambda) b_j + (\lambda_k + a_k \lambda) b_k$$

for $1 \leq j \leq k-1$ and all $\lambda_1, \lambda_2, \ldots, \lambda_k, \lambda$, we may assume that, for

$$n = x_1 b_1 + x_2 b_2 + \cdots + x_k b_k \quad \textit{and} \quad M - n = y_1 b_1 + y_2 b_2 + \cdots + y_k b_k,$$

the integers x_j and y_j satisfy $0 \leq x_j, y_j \leq a_j - 1$ for $1 \leq j \leq k-1$. From $M - n - (M - n) = 0$ we have

$$(k-1)a_1 a_2 \cdots a_k - b_1(x_1 + y_1 + 1) - b_2(x_2 + y_2 + 1) - \cdots - b_k(x_k + y_k + 1) = 0. \qquad (*)$$

From this it follows that a_j divides $x_j + y_j + 1$ for all j. However, $1 \leq x_j + y_j + 1 \leq 2a_j - 1$ for $1 \leq j \leq k-1$, so it must be the case that $x_j + y_j + 1 = a_j$ for such j. Hence $b_j(x_j + y_j + 1) = b_j a_j = a_1 a_2 \cdots a_k$ for $1 \leq j \leq k-1$, so $(*)$ yields

$$b_k(x_k + y_k + 1) = 0.$$

Thus $x_k + y_k + 1 = 0$, and it follows that exactly one of x_k and y_k is negative and the other is nonnegative. Thus one and only one of the integers n and $M - n$ can be expressed as a linear combination of the b_is with all coefficients nonnegative.

Also solved by Jean Bogaert (Belgium), Daniele Donini (Italy), Li Zhou, and the proposer.

An Identity of Sums **October 2002**

1647. *Leon Quet, Denver, CO.*

Prove that for all $x > 0$,

$$\sum_{j=1}^{\infty} \sum_{k=1}^{j} \frac{x^{(j+1)^{-2}} - x^{k^{-2}}}{(j+1)^2 - k^2} = \frac{\pi^2}{8} - \frac{3}{4} \sum_{j=1}^{\infty} \frac{x^{j^{-2}}}{j^2}.$$

Solution by Rolf Richberg, Stolberg, Germany.

Given $x > 0$, let $a_n = x^{n^{-2}} - 1$ for positive integer n, so that

$$a_n = O(n^{-2}) \quad \text{as} \quad n \to \infty. \qquad (1)$$

Now,

$$\sum_{j=1}^{\infty} \sum_{k=1}^{j} \frac{x^{(j+1)^{-2}} - x^{k^{-2}}}{(j+1)^2 - k^2} = \sum_{j=1}^{\infty} \sum_{k=1}^{j} \frac{a_{j+1} - a_k}{(j+1)^2 - k^2}, \qquad (2)$$

which, in view of (1), equals

$$\sum_{j=1}^{\infty} a_{j+1} \sum_{k=1}^{j} \frac{1}{(j+1)^2 - k^2} - \sum_{k=1}^{\infty} a_k \sum_{j=k}^{\infty} \frac{1}{(j+1)^2 - k^2}, \qquad (3)$$

where, in the second expression, we have changed the order of summation. Using partial fractions, that is,

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

the inner sums in (3) are readily expressed using the harmonic numbers, $H_n = \sum_{\nu=1}^{n} 1/\nu$. Using these, (2) and (3) yield

$$\sum_{j=1}^{\infty}\sum_{k=1}^{j}\frac{x^{(j+1)^{-2}}-x^{k^{-2}}}{(j+1)^2-k^2} = \sum_{j=1}^{\infty}\frac{a_{j+1}}{2j+2}\left(H_{2j+1}-\frac{1}{j+1}\right)-\sum_{k=1}^{\infty}\frac{a_k}{2k}H_{2k}$$
$$= \sum_{j=2}^{\infty}\frac{a_j}{2j}\left(H_{2j}-\frac{3}{2j}\right)-\sum_{k=1}^{\infty}\frac{a_k}{2k}H_{2k}$$
$$= -\frac{3}{4}\sum_{j=1}^{\infty}\frac{a_j}{j^2} = \frac{\pi^2}{8}-\frac{3}{4}\sum_{j=1}^{\infty}\frac{x^{j^{-2}}}{j^2}.$$

Also solved by Michel Bataille (France), Jean Bogaert (Belgium), Chip Curtis, Daniele Donini (Italy), Robert L. Doucette, Julien Grivaux (France), Li Zhou, Paul Zwier, and the proposer.

Answers

Solutions to the Quickies from page 319.

A933. First observe that if $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, then $A^k = \begin{pmatrix} f_{k-1} & f_k \\ f_k & f_{k+1} \end{pmatrix}$ for every positive integer k. If p divides f_k for some $k \geq 1$, then $f_k \equiv 0 \pmod{p}$, and $f_{k+1} \equiv f_{k-1} + f_k \equiv f_{k-1} \pmod{p}$, so $A^k \equiv \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \pmod{p}$, for some integer c. It follows that $A^k = cI + pB$ for some integer matrix B. By the binomial theorem, there is an integer matrix M such that

$$A^{p^r k} = (cI + pB)^{p^r} = c^{p^r} I + p^{r+1} M \equiv c^{p^r} I \pmod{p^{r+1}},$$

where we have used the fact that $\binom{p^r}{j} p^j$ is divisible by p^{r+1} for $j = 1, 2, \ldots, p^r$. Thus, $A^{p^r k} \pmod{p^{r+1}}$ is a diagonal matrix, and it follows that $f_{p^r k} \equiv 0 \pmod{p^{r+1}}$.

A934. Let the distances from the foot of the altitude to the sides of length a, b, c be x, y, z, respectively. Then the sum of the surface areas of the three lateral faces is

$$\frac{1}{2}a\sqrt{x^2+h^2}+\frac{1}{2}b\sqrt{y^2+h^2}+\frac{1}{2}c\sqrt{z^2+h^2}.$$

By Minkowski's inequality, this quantity is greater than or equal to

$$\frac{1}{2}\sqrt{h^2(a+b+c)^2+(ax+by+cx)^2} = \frac{1}{2}\sqrt{h^2(a+b+c)^2+4F^2},$$

where F is the area of the base. Equality occurs if and only if $x = y = z = r$, where r is the inradius of the base, that is, if and only if the incenter of the base is the foot of the given altitude. Thus the minimal surface area is $\frac{1}{2}(a+b+c)(r+\sqrt{h^2+r^2})$.

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.

Wade, Nicholas, A mathematician crunches the Supreme Court, *New York Times* (24 June 2003) D3; http://www.nytimes.com/2003/06/24/science/24MATH.html . Banzhaf, John F., III, Analyzing Supreme math (letter), *New York Times* (1 July 2003) F4; http://www.nytimes.com/2003/07/01/science/01LETT.html . Sirovich, Lawrence, A pattern analysis of the second Rehnquist Supreme Court, *Proceedings of the National Academy of Sciences of the USA* 100 (13) (24 June 2003) 7432–7437; http://www.pnas.org/cgi/doi/10.1073/pnas.1132164100 .

"The court acts as if composed of 4.68 ideal justices... [and] the decision space of the Rehnquist court [1995–2002] requires only two dimensions for its description," says author Sirovich (Mt. Sinai School of Medicine). The first claim is based on only about half of the 2^9 potential voting patterns occurring. The second comes from singular value decomposition of the matrix of decisions, a technique used in face recognition systems; the two dimensions correspond to unanimity and to the most common of the court's 5–4 voting patterns. Analysis of the 1960s Warren court produces similar results. Letter author Banzhaf notes, however, "they are judging, rather than casting independent or random ballots... trying to determine how well... a given factual situation is consistent with generally agreed-upon legal principles and methods of analysis. Therefore agreement, rather than disagreement, should be the norm." Perhaps the two dimensions reflect selection: The court may accept mainly cases where a priori there is either broad agreement or broad disagreement.

Pesic, Peter, *Abel's Proof: An Essay on the Sources and Meaning of Mathematical Unsolvability*, MIT Press, 2003; viii + 211 pp, $24.95. ISBN 0–262–16216–4.

A humanities colleague teaching an interdisciplinary studies course "500 Years Ago, 1503 c.e." sought to incorporate some perspective on mathematics (his father is a mathematician). How about Cardano (b. 1501)? The impetus for his formula for the cubic came from conversations of del Ferro and Pacioli in 1502, and other course units are on the Mona Lisa (Cardano's father was a friend of da Vinci) and on medical "cures" (Cardano was a doctor). My advice: Show the solution formulas for a linear equation, quadratic, cubic, and quartic. Then tell them, "That's it—no formulas for general equations of higher degree. Not just more and more complicated—no more formulas." [Technical note: No *algebraic* formulas. In 1984, Hiroshi Umemura proved Kronecker's conjecture that "Every algebraic equation has a root that can be expressed in terms of a modular function and hyperelliptic integrals"—Krzystof Maurin, as quoted by Abe Shenitzer, *American Mathematical Monthly* 103 (1996) 258.] I told my colleague, "It's too hard to show how and why, but you can romance them with the stories of Galois and Abel (b. 1803), who were the students' age when they did the work." Just on cue, Pesic's book appears, ideal for such a colleague. It largely consigns the mathematics to sidebars but reproduces Abel's 1824 paper with a commentary. This easy-to-read book builds on the common denominator of the public's mathematical experience to give a sense of the achievements of Abel and Galois and a glimpse of some of the larger consequences of unsolvability.

Davis, Benjamin Lent, and Diane Maclagan, The card game SET, *Mathematical Intelligencer* 25 (3) (Summer 2003) 33–40. Peterson, Ivars, Ivars Peterson's Math Trek: SET Math, `http://www.maa.org/mathland/mathtrek_08_25_03.html` .

You may already know the popular card game SET and appreciate it as a way to teach children to learn about and enjoy sets. Each card displays a design with four attributes, each of which may take on one of three values; the deck has one each of the 3^4 different cards. Play consists of competing to be first to recognize (in a tableau) groups of three cards that have each attribute all the same or all different; such a group is called a SET. Authors Davis and Maclagan start from the question, What is the largest collection of cards that does not contain a SET? They translate the question into affine collinearity in a vector space, generalize to other dimensions, and classify games of this kind. Along the way appear affine transformations, the pigeon-hole principle, Fourier transforms, projective planes, Steiner triples, and even the classification of finite simple groups! (The game SET can be ordered from `www.setgame.com/set/` .)

Peterson, Ivars, There are no magic knight's tours on the chessboard, `http://mathworld.wolfram.com/news/2003-08-06/magictours/` .

A *knight's tour* of an $n \times n$ board is a sequence of knight moves that visits each square once; each square is numbered in order of the tour. A vast literature discusses algorithms and classifications of such tours. A further problem is whether any tour produces an array of numbers that is a *magic square*, in which the numbers in each row, column, and main diagonal add to the same constant; if the rows and columns do but not the diagonals, the square is *semimagic*. Magic tours are impossible for odd n but are known for all $n = 4k$ with $k > 2$. Whether such a tour exists on an 8×8 board is a longstanding question. The answer is no, determined by complete enumeration of all semimagic tours (140 of them) via distributed computing organized by J.C. Meyrignac and Guenter Stertenbrink.

Devlin, Keith, Devlin's Angle: The shame of it (May 2003); `www.maa.org/devlin/devlin_05_03.html` . When is a proof? (June 2003); `www.maa.org/devlin/devlin_06_03.html` .

"The rot set in [in] 1993, when Andrew Wiles was forced to withdraw his dramatic claim to have proved Fermat's Last Theorem." Devlin begins by feeling ashamed of the "sloppy behavior" of fellow mathematicians who have recently had to retract claims of results (Poincaré conjecture, relative of twin primes conjecture). He distinguishes "right wing" proofs (logically correct arguments) from "left wing" proofs (which convince a typical mathematician). [I find his terminology a lamentable extension into mathematics of the political polarization of this country.] He tells the result of four years of efforts by 12 referees on Thomas Hales's 1998 claimed proof of the Kepler sphere-packing conjecture: The team was 99% certain of its correctness; however, they could not certify the argument completely and *will not check it further* (having run out of energy for the project). Hales hopes to rescue the result from this limbo with computer-aided proof-checking; but wouldn't that lead only to a "left wing" proof?

Kiltinen, John O., *Oval Track and Other Permutation Puzzles, and Just Enough Group Theory to Solve Them*, MAA, 2003; xviii + 305 pp + CD-ROM with software (Windows and Macintosh), $42.40 (P) (for MAA members: $33.95). ISBN 0–88385–725–1.

Rubik's Cube (b. 1974) is back; the first world championships since 1982 just took place in Toronto (winner's average time: 20.2 sec). This book discusses a different but related puzzle, formerly marketed as To p Spin. Disks are rotated in turn around an oval track track and its adjacent turntable. The puzzle involves putting the disks back in order from a scrambled position. The puzzle is purely permutational in the pieces, in contrast to Rubik's Cube, which involves the orientations of the cubelets. This book shows how to use the track puzzle (and its computerized realization) to help teach abstract algebra, with exercises, extensions to other puzzles, abundant references, and accompanying software.

NEWS AND LETTERS

Carl B. Allendoerfer Awards — 2003

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in Mathematics Magazine. The Awards are named for Carl B. Allendoerfer, a distinguished mathematician at the University of Washington and President of the Mathematical Association of America, 1959–60.

Ezra Brown The Many Names of (7, 3, 1), Mathematics Magazine, 2002, pp. 83–94.

The article by Ezra Brown is a story about a single object that is "all at once a difference set, a block design, a Steiner triple system, a finite projective plane, a complete set of orthogonal Latin squares, a double regular round-robin tournament, a skew-Hadamard matrix, and a graph consisting of seven mutually adjacent hexagons drawn on the torus." The intellectual picnic that Brown creates is written in a clear and engaging style and traces the connections prompted by (7, 3, 1).

From the first incidence of (7, 3, 1) in the set $Q_7 = \{1, 2, 4\}$ (the difference set S of 3 nonzero integers mod 7 such that every nonzero integer n mod 7 can be represented as a difference of elements of S in exactly 1 way) through connections in Galois Theory, Brown masterfully follows the combinatorial threads of (7, 3, 1). He leads the reader in a very natural way from one topic to the next and gives the needed definitions and hints for exploring further connections with references from Euler to R. A. Fisher.

Like many good stories, Brown ends his article with additional questions: "Does (7, 3, 1) have any other names?" "Where can I find out more about difference sets?" and clues about how to find the answers. Indeed, there are also suggestions for a sequel or two. The question "Did the Reverend Thomas Kirkman ever get credit for anything?" contains a description of what is known as Kirkman's Schoolgirls Problem. "The solution to this problem is a particularly interesting Steiner triple system on 15 varieties, with parameters (35, 15, 7, 3, 1)—but that's another story." With the gifted telling of the story of (7, 3, 1) Brown makes us want another story.

Biographical note Ezra (Bud) Brown has degrees from Rice and Louisiana State, and has been at Virginia Tech since the first Nixon Administration, with time out for sabbatical visits to Washington, DC (where he has spent his summers since 1993) and Munich. His research interests include number theory, graph theory, and combinatorics, and he particularly enjoys discovering connections between apparently unrelated areas of mathematics. He received the MAA MD-DC-VA Section Award for Outstanding Teaching in 1999 and MAA Pólya Awards in 2000 and 2001. He enjoys working with students who are engaged in research. Over the years, he has put together seventeen nomination dossiers which led to college and university recognition of his colleagues' superior teaching, advising and outreach. He enjoys singing (everything from grand opera to rock'n'roll), playing jazz piano, and talking about his granddaughter Phoebe Rose. With his wife Jo's help, he has become a fairly tolerable gardener. He occasionally bakes biscuits for his students.

Response from Ezra Brown In my graph theory course in graduate school, Brooks Reid first introduced me to (7, 3, 1) as a difference set. "And by the way," he said, "it is also a finite projective plane, a doubly regular tournament, and a symmetric balanced incomplete block design." I marveled at the notion that the same mathematical object could have so many different names—or perhaps the rhythm of Brooks Reid's words got to me. At any rate, as time passed, my interest in (7, 3, 1) deepened with the discovery of each new name. The article was written with the hopes that its readers might share my fascination with (7, 3, 1) and combinatorial designs in general, and it is a privilege and a pleasure to be honored with the Allendoerfer Prize.

Dan Kalman Doubly Recursive Multivariate Automatic Differentiation, Mathematics Magazine, 2002, pp. 187–202.

"Automatic differentiation is a way to find the derivative of an expression without finding an expression for the derivative." This opening sentence and the rest of Kalman's model initial paragraph in which he explains what automatic differentiation is and is not, set the stage for the remainder of the article in which a rather intricate situation is masterfully explained. The automatic differentiation "computation is equivalent to evaluating a symbolic expression for $f'(x)$, but no one has to find that expression—not even the computer system."

The author begins with a brief review of the one-variable, one-derivative case presented by L. B. Rall in Mathematics Magazine in 1986. He then discusses a beautiful and mathematically intriguing extension that can handle any number of variables and derivatives. By means of derivative structures and substructures (vectors for the one variable case, triangles for two variables, pyramids for three variables, and extending to derivative structures in general) and operations on them, the author shows how we can obtain an object containing the function value and the values of all partial derivatives through a particular order. The examples are well-chosen and the author's style is friendly and clear throughout.

The paper ends with a discussion on implementation and efficiency in which Kalman points out some of the inefficiencies of the recursive approach and suggestions about how one might improve upon these inefficiencies. The final paragraph claims: "More generally, as computational speed continues to increase, the importance of execution efficiency will continue to decline, particularly for problems with small numbers of variables. In these cases, the directness and simplicity of the current development offers an attractive paradigm for implementing an automatic differentiation system." Indeed, Kalman's presentation of double recursive automatic differentiation offers us an outstanding paradigm for masterful expository writing.

Biographical note Dan Kalman has been a member of the mathematics faculty at American University, Washington, DC, since 1993. Prior to that he worked for 8 years in the aerospace industry and taught at the University of Wisconsin, Green Bay. During the 1996–1997 academic year he served as an Associate Executive Director of the MAA. Kalman has a B.S. from Harvey Mudd College, and a Ph.D. from University of Wisconsin, Madison. Kalman has been a frequent contributor to all of the MAA journals, and served as an Associate Editor for *Mathematics Magazine*. He has written one book, *Elementary Mathematical Models*, published in the MAA's Classroom Resource Materials series. His current interests include creating interactive computer activities for mathematics instruction, using Mathwright software.

Response from Dan Kalman I am very grateful to receive this Carl B. Allendoerfer Award. It is particularly gratifying because this paper reports my best mathematical work. The original inspiration occurred in one of those rare flashes of understanding

which suddenly illuminates an entire development. At the moment of discovery I was actually thinking about something quite different. All at once I was struck by a new insight, in a completely unexpected direction. It was totally compelling: I knew with utter certainty that this idea would prove correct, before examining a single example, though it concerned a question I had never before considered. It is hard to convey the astonishment and euphoria of that experience. My thanks go to Robert Lindell, a colleague at the Aerospace Corporation, whose collaboration paved the way for my discovery. Working with him was a privilege. I would also like to thank Frank Farris, editor of Mathematics Magazine, whose guidance greatly improved the structure of the paper. Finally, thanks to the MAA for publications and programs which so profoundly enhance our profession.

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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