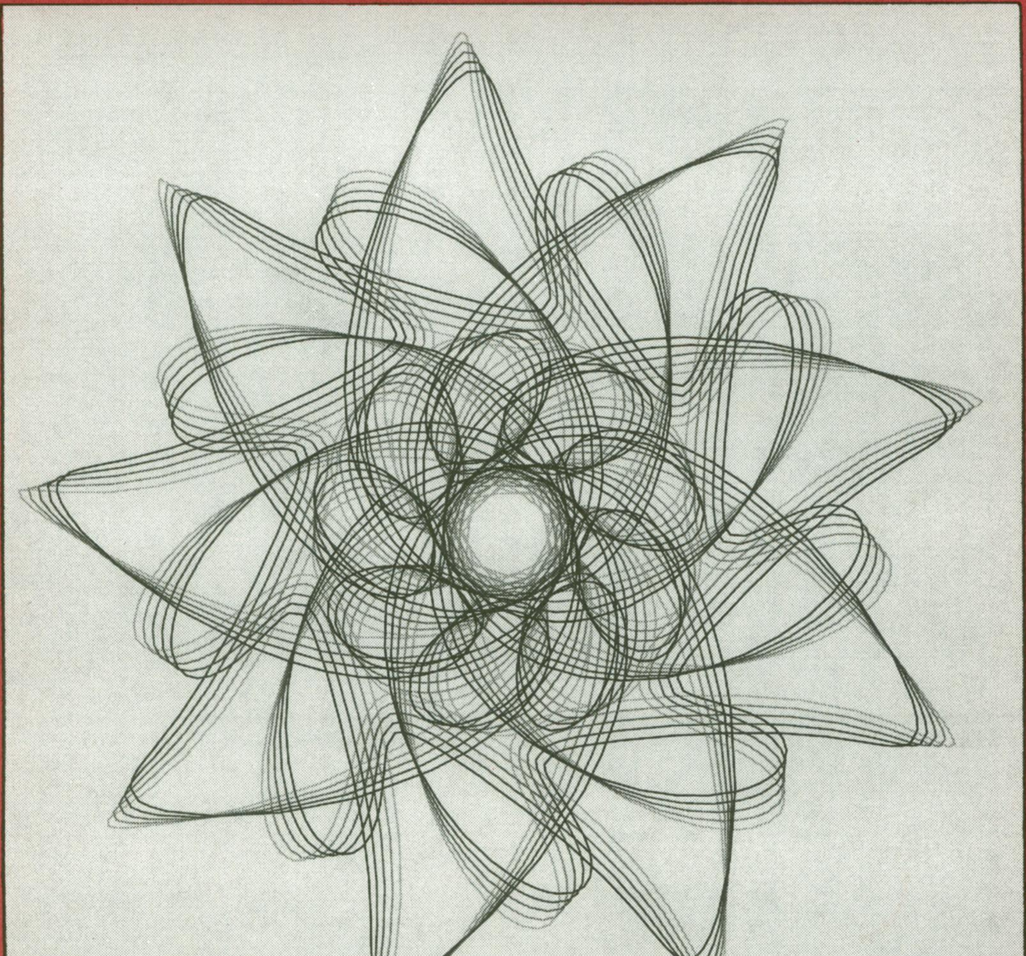
Vol. 69, No. 3 June 1996

MATHEMATICS MAGAZINE



A Curve with 9-Fold Symmetry (see pp. 185–189)

- The Group of Rational Points on the Unit Circle
- A Simple Introduction to Integral Equations

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

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

Cover illustration: Spirograph, by Robert Miner. Copyright by The Geometry Center (`http://www.geom/umn.edu`) for the Regents of the University of Minnesota.

AUTHORS

Lin Tan is a native of Hangzhou, China. After high school, he did farm work for two years before entering Zhejiang University in 1978, where he was granted an M.S. degree in 1981. He received his Ph.D. from UCLA in 1986, under the supervision of Robert Steinberg. He taught at Indiana University before joining the faculty of West Chester University in 1989. His primary interest is algebraic groups and invariant theory. His hobbies include continued fractions, history of mathematics and automorphic forms.

Glenn Ledder first discovered integral equations through his research in mathematical modeling. He received his B.S. in ceramic engineering from Iowa State University in 1977. A sequence of brief careers led him eventually to Union College, in Schenectady, New York, where his interest in mathematics was reawakened. He received his Ph.D. from Rensselaer Polytechnic Institute in 1990 and has been on the faculty at the University of Nebraska-Lincoln since 1989. His other interests include singing, canoeing, and preparing to run his first (and only) marathon.

Vol. 69, No. 3 June 1996

MATHEMATICS MAGAZINE

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

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

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

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

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

The Group of Rational Points on the Unit Circle

LIN TAN
West Chester University
West Chester, PA 19383

Introduction

Recently, Fermat's Last Theorem was proved. A long chain of arguments, based on many mathematicians' deep work, culminated in Wiles' last and decisive step (**[28, 25]**).

FERMAT'S LAST THEOREM. *Let* $n, a, b, c \in \mathbb{Z}$ *with* $n > 2$. *If* $a^n + b^n = c^n$, *then* $abc = 0$.

There have been several attempts to present the basic idea of this marvelous proof to a wider mathematical audience (**[2, 7, 8]**). Another recent paper **[3]** provides more details.

Wiles' proof is based on the theory of elliptic curves, i.e., curves defined by cubic equations. A big part of this theory is devoted to understanding the "rational points" (points whose coordinates are rational numbers) on these curves. The set of rational points on an elliptic curve has a natural group structure, which will be described very briefly later. It is often very difficult, however, to find all the rational points on an elliptic curve.

In this paper we take a much easier and more familiar example—the unit circle—and show how to compute the group structure of its rational points. Next, some applications are given. Finally, for comparison, we give a brief summary of known results on the group structure for rational points on elliptic curves.

Rational Points on the Unit Circle

Let C be the unit circle in the real plane, defined by $x^2 + y^2 = 1$. The *rational points* on C are those for which both coordinates are in $\mathbb{Q}$. For example, $(\frac{3}{5}, \frac{4}{5})$, $(-\frac{5}{13}, \frac{12}{13})$ and $(0, 1)$ are rational points, while $\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$ is not. We denote the set of rational points on C by $C(\mathbb{Q})$. A rational point $(\frac{a}{c}, \frac{b}{c})$ on C corresponds to an *integer* solution to $X^2 + Y^2 = Z^2$, with $X = a$, $Y = b$, and $Z = c$. (More generally, a rational point on the curve $x^n + y^n = 1$ corresponds to an integer solution to $X^n + Y^n = Z^n$.)

C is an abelian group under the "angle addition" $\oplus$, defined by

$$(x_1, y_1) \oplus (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1) \tag{1}$$

for $(x_1, y_1), (x_2, y_2) \in C$. The identity element is $(1, 0)$, and the inverse of (x, y) is $(x, -y)$. Note that (1) is merely the familiar "addition formula" in trigonometry using the correspondence $\theta \mapsto (x, y) = (\cos\theta, \sin\theta)$; (1) is also the usual formula for the multiplication in the field of complex numbers.

It is clear from (1) that $C(\mathbb{Q})$ is a subgroup of C. This raises a natural question:

What is the group structure of $C(\mathbb{Q})$?

We begin our search for an answer with some elementary facts from number theory and algebra (see, e.g., [**12**]).

Every *Pythagorean triple* (a, b, c) (i.e., a triple of integers a, b, c with $c \neq 0$, satisfying $a^2 + b^2 = c^2$) corresponds to the rational point $(\frac{a}{c}, \frac{b}{c})$ on $C(\mathbb{Q})$. Two Pythagorean triples (a, b, c) and (a', b', c') correspond to the same point on $C(\mathbb{Q})$ if and only if $(a, b, c) = r(a', b', c')$ for some $r \in \mathbb{Q} \setminus \{0\}$. Therefore, if (a, b, c) is *primitive* (i.e., if $c > 0$ and the greatest common divisor of a, b, and c is 1), then every Pythagorean triple corresponding to $(\frac{a}{c}, \frac{b}{c})$ must have the form (na, nb, nc) for some nonzero integer n.

From elementary number theory we know that the parametrization $(m^2 - n^2, 2mn, m^2 + n^2)$, with $m, n \in \mathbb{Z}$, not both 0, gives all Pythagorean triples (a, b, c) with $c > 0$. Those m and n that satisfy $(m, n) = 1$ and $m - n \equiv 1 \pmod 2$ produce all the primitive triples. (See, e.g., [**23**, p. 13].) It is perhaps illuminating to have a geometric interpretation of this parametrization. The expressions $m^2 - n^2$ and $2mn$ remind us of the double-angle formulas for cosine and sine. They come from the well-known "rational parametrization" $\rho \colon \mathbb{R} \to C$ of the unit circle, defined by

$$\rho(t) = \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right). \tag{2}$$

(See FIGURE 1; for more details, see [**23**, p. 11].) In particular,

$$\rho\left(\frac{n}{m}\right) = \left(\frac{m^2 - n^2}{m^2 + n^2}, \frac{2mn}{m^2 + n^2} \right).$$

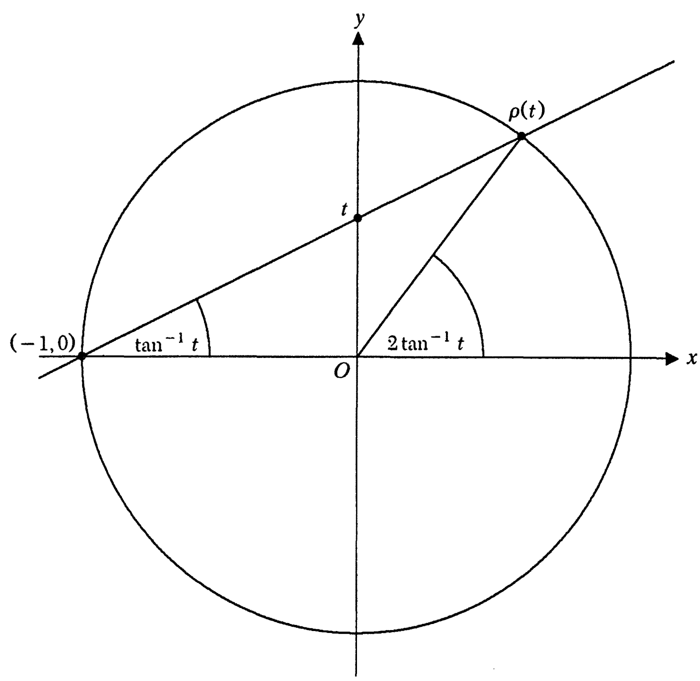


FIGURE 1

As the lines through $(-1,0)$ sweep through the rational slopes, their intersections with C, other than $(-1,0)$, sweep through $C(\mathbb{Q})$. Notice too that $\frac{1-t^2}{1+t^2} = \cos(2\tan^{-1}(t))$, and that $\rho|_{\mathbb{Q}}\colon \mathbb{Q} \to C(\mathbb{Q})\setminus\{(-1,0)\}$ is onto, since $\rho(\frac{v}{1+u}) = (u,v)$.

Next we consider the ring $\mathbb{Z}[i] = \{m + ni \mid m, n \in \mathbb{Z}\}$ of Gaussian integers. For $m + ni \in \mathbb{Z}[i]$ with $m \neq 0$, the line connecting $m + ni$ and O intersects the vertical line $x = 1$ at $(1, \frac{n}{m})$ (see FIGURE 2). Combining the above observations (by superimposing FIGURE 1 and FIGURE 2, with the line $x = 1$ in FIGURE 2 aligned with the y-axis in FIGURE 1), we get a map $f\colon \mathbb{Z}[i]\setminus\{0\} \to C(\mathbb{Q})$ defined by

$$f(m+ni) = \left(\frac{m^2-n^2}{m^2+n^2}, \frac{2mn}{m^2+n^2}\right); \tag{3}$$

$f(ni)$ is defined to be $(-1,0)$. It is clear from (3) that any Gaussian integer on the line of slope n/m through O yields the same value for $f(m+ni)$.

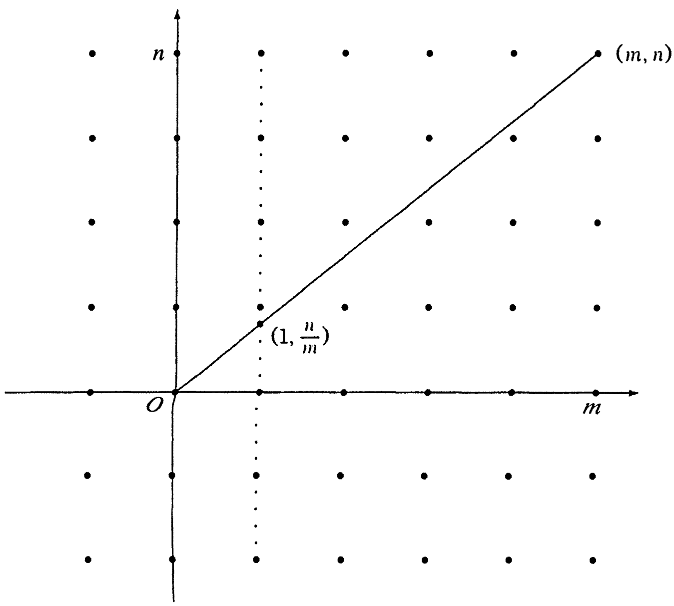


FIGURE 2

The keys to our approach are the simple facts that f is an onto map from $\mathbb{Z}[i]\setminus\{0\}$ (which is almost a group under multiplication but for the lack of inverses, i.e., a semigroup) to the group $C(\mathbb{Q})$, and that f preserves multiplication. This, again, is the double-angle formulas (see FIGURE 1): If $m + ni$ corresponds to angle $\theta = \tan^{-1}(n/m)$, then $f(m+ni) \in C(\mathbb{Q})$ corresponds to 2θ. The following properties are also clear:

$$f(m-ni) \oplus f(m+ni) = (1,0), \tag{4}$$

$$f(m+ni) = (1,0) \text{ (resp. } (-1,0)) \quad \text{if and only if } n = 0 \text{ (resp. } m = 0). \tag{5}$$

The homomorphism $f\colon \mathbb{Z}[i]\setminus\{0\} \to C(\mathbb{Q})$ will enable us to explore $C(\mathbb{Q})$, using the well-known properties of $\mathbb{Z}[i]$. The Lemma below shows the advantage of doing so.

The Group Structure of $C(\mathbb{Q})$

Now for $k = 1, 2, 3$, let $(c_k, s_k) \in C(\mathbb{Q}), (c_k, s_k) = f(m_k + n_k i)$. Then $(c_1, s_1) = (c_2, s_2) \oplus (c_3, s_3)$ if and only if $(m_1 + n_1 i)l' = (m_2 + n_2 i)(m_3 + n_3 i)l$ for some $l, l' \in \mathbb{Z} \setminus \{0\}$. Thus, the elements $f(m + ni) \in C(\mathbb{Q})$ with $m + ni$ irreducible in $\mathbb{Z}[i]$ suffice to generate $C(\mathbb{Q})$.

Let us recall some basic properties of $\mathbb{Z}[i]$. (For more details, see **[12]**.) First, $\mathbb{Z}[i]$ is a unique factorization domain (in fact, a Euclidean domain) with units ± 1 and $\pm i$. Second, the norm N in $\mathbb{Z}[i]$, defined by

$$N(m + ni) = (m + ni)(m - ni) = m^2 + n^2,$$

is a homomorphism from the multiplicative semigroup $\mathbb{Z}[i] \setminus \{0\}$ to the multiplicative semigroup $\mathbb{N}$ of natural numbers. Third, $\mathbb{Z}[i]$ has three types of irreducible elements:

(i) For a rational prime (i.e., a prime number in $\mathbb{Z}$) $p \equiv 1 (\mathrm{mod}\, 4)$, p can be written, uniquely up to sign and order, as the sum $m^2 + n^2$ of two squares. For such m and n, $m + ni$ is irreducible in $\mathbb{Z}[i]$ of norm p.
(ii) For a rational prime $p \equiv 3 (\mathrm{mod}\, 4)$, p remains irreducible in $\mathbb{Z}[i]$ of norm p^2.
(iii) $1 + i$ and $1 - i$ are irreducible, each with norm $(1 + i)(1 - i) = 2$.

For irreducible elements p of type (ii), $f(p) = (1, 0)$, the identity in $C(\mathbb{Q})$. For irreducible elements $1 \pm i$ of type (iii), $f(1 \pm i) = (0, \pm 1)$, which have order 4 in $C(\mathbb{Q})$. For each $p \equiv 1 (\mathrm{mod}\, 4)$, we pick, for definiteness, integers m_p and n_p so that $p = m_p^2 + n_p^2, m_p > n_p > 0$. Then

$$\left\{ \left(\frac{m_p^2 - n_p^2}{m_p^2 + n_p^2}, \frac{2 m_p n_p}{m_p^2 + n_p^2} \right) \right\}_{p \equiv (\mathrm{mod}\, 4)} \cup \quad \{(0, 1)\}$$

is a set of generators for $C(\mathbb{Q})$. Now we show that no relation exists among these generators, other than that $(0, 1)$ has order 4. First we handle generators of type (i):

LEMMA. *There is no non-trivial relation in*

$$\left\{ \left(\frac{m_p^2 - n_p^2}{m_p^2 + n_p^2}, \frac{2 m_p n_p}{m_p^2 + n_p^2} \right) \right\}_{p \equiv 1 (\mathrm{mod}\, 4)}$$

Proof. Let $p_1, \ldots, p_k$ be rational primes, all congruent to 1 modulo 4. Let $m_j = m_{p_j}, n_j = n_{p_j}$, and suppose that

$$a_1 f(m_1 + n_1 i) \oplus \cdots \oplus a_k f(m_k + n_k i) = (1, 0) \tag{6}$$

for integers $a_1, \ldots, a_k$. Then

$$(m_1 + n_1 i)^{a_1} \ldots (m_k + n_i i)^{a_k} l = l' \tag{7}$$

for some $l, l' \in \mathbb{Z} \setminus \{0\}$. Let $l = q_1 q_2 \ldots$ and $l' = q_1' q_2' \ldots$ with q_j and q_j' rational primes. We may assume, by unique factorization, that $l = 1$. So

$$(m_1 + n_1 i)^{a_1} \ldots (m_k + n_k i)^{a_k} = q_1' q_2' \ldots . \tag{8}$$

By unique factorization again, each q_j' is associated with a product of two factors, each the complex conjugate of the other, of the left side of (8). The images of such a pair of factors under f are inverses of each other by (4) and (5). Unless all $a_j = 0$ this leads to a contradiction, since all m_j and n_j are positive. Thus (6) is a trivial relation. □

If, on the other hand, a relation involves $(0,1)$ $(=f(1+i))$, then the "quadruple" of this relation has the form (6), and is therefore trivial by the Lemma. It follows that the original relation is a consequence of the relation $4\cdot(0,1) = (1,0)$.

To finish the analysis of $C(\mathbb{Q})$ we invoke a special case of Dirichlet's Theorem on primes in arithmetic progression. This case was proved by Euler **[5]** in 1775:

There are infinitely many primes congruent to 1 *modulo* 4.

Combining the results above, we get the structure theorem for $C(\mathbb{Q})$.

THEOREM 1. *The abelian group $C(\mathbb{Q})$ is the direct sum of infinitely many cyclic subgroups*:

$$C(\mathbb{Q}) \cong C_2 \oplus \left(\bigoplus_{p \equiv 1 \pmod 4} C_p \right),$$

where C_2 is generated by $(0,1)$ (an element of order 4), and C_p is the infinite cyclic group generated by

$$\left(\frac{m_p^2 - n_p^2}{m_p^2 + n_p^2}, \frac{2 m_p n_p}{m_p^2 + n_p^2} \right),$$

with m_p, n_p being the unique solution to $m_p^2 + n_p^2 = p$, $m_p > n_p > 0$.

Examples and Corollaries

We illustrate the significance of Theorem 1 with some examples and corollaries.

Example A. C_5, C_{13}, C_{17}, C_{29}, C_{37}, and C_{41} are generated, respectively, by $(\frac{3}{5}, \frac{4}{5}) = f(2+i)$, $(\frac{5}{13}, \frac{12}{13}) = f(3+2i)$, $(\frac{15}{17}, \frac{8}{18}) = f(4+i)$, $(\frac{21}{29}, \frac{20}{29}) = f(5+2i)$, $(\frac{35}{37}, \frac{12}{37}) = f(6+i)$, and $(\frac{9}{41}, \frac{40}{41}) = f(5+4i)$.

Example B. $(\frac{-76}{1445}, \frac{1443}{1445}) = (0,1) \oplus (-1)\cdot(\frac{3}{5}, \frac{4}{5}) \oplus 2\cdot(\frac{15}{17}, \frac{8}{17})$. This decomposition can be computed as follows. From the prime factorization $1445 = 5 \cdot 17^2$, we have $(\frac{-76}{1445}, \frac{1443}{1445}) = n\cdot(0,1) \oplus (\pm 1)\cdot(\frac{3}{5}, \frac{4}{5}) \oplus (\pm 2)\cdot(\frac{15}{17}, \frac{8}{17})$. Now we choose the coefficient of $(\frac{3}{5}, \frac{4}{5})$ to be 1 (resp. -1) if the denominators are 17^2 in $(\frac{-76}{1445}, \frac{1443}{1445}) \oplus (-1)\cdot(\frac{3}{5}, \frac{4}{5})$ (resp. in $(\frac{-76}{1445}, \frac{1443}{1445}) \oplus (1)\cdot(\frac{3}{5}, \frac{4}{5})$), when reduced. The coefficient ± 2 is chosen similarly. Finally, the coefficient of $(0,1)$ is chosen to get the right signs and the right order of the two coordinates.

The following corollaries are straightforward consequences of Theorem 1.

COROLLARY 1. *Let α and β be real numbers, and suppose $P_\alpha = (\cos\alpha, \sin\alpha)$ and $P_\beta = (\cos\beta, \sin\beta)$ are in $C(\mathbb{Q})$, and that $\frac{\alpha}{\beta} \in \mathbb{Q}$. Then there exist $r, s \in \mathbb{Z}$, $P_\gamma \in C(\mathbb{Q})$, and $c_\alpha, c_\beta \in C_2$, such that $P_\alpha = rP_\gamma \oplus c_\alpha$ and $P_\beta = sP_\gamma \oplus c_\beta$. In particular, if $P_\alpha \in C(\mathbb{Q})$ and α is a rational multiple of π, then $P_\alpha \in C_2$.* (**[26]**)

Proof. Let $\alpha/\beta = r/s \in \mathbb{Q}$, with r and s relatively prime. Then $sP_\alpha = rP_\beta$, because in $C(\mathbb{Q})$ we have $n\cdot(\cos\theta, \sin\theta) = (\cos(n\theta), \sin(n\theta))$. So the first assertion follows from Theorem 1 by comparing the C_p-components of P_α and P_β: Each C_p-component of P_α (resp. P_β) is "divisible" by r (resp. by s), and we can construct P_γ by defining its C_p-component as $\frac{1}{r}\cdot$\{the C_p-component of P_α\}, which is equal to $\frac{1}{s}\cdot$\{the

C_p-component of P_β}. (Because C_2 is finite, we cannot compare the C_2-components here, and so must include the "c_α" and "c_β" terms.)

The second assertion follows from the first by taking $\beta = \pi$. □

COROLLARY 2. *Let $P_\alpha = (\cos\alpha, \sin\alpha)$ and $P_\beta = (\cos\beta, \sin\beta)$ be in $C(\mathbb{Q})$ and suppose that $\alpha - \beta$ is a rational multiple of π. Then $\alpha = \beta + \frac{k\pi}{2}$ for some $k \in \mathbb{Z}$.*

Proof. This follows directly by applying Corollary 1 to $\alpha - \beta$. □

COROLLARY 3. *On the "square geoboard" (the lattice $\mathbb{Z} \times \mathbb{Z}$ in $\mathbb{R}^2$), the only angles of rational measure (in degrees) that can be formed by three lattice points are integer multiples of* 45°. *(This was conjectured in* 1921 **[26]** *and first proved in* 1945 **[20]**.*)*

Proof. The claim follows directly from Corollary 2 and the double-angle correspondence (3) between the geoboard $\mathbb{Z}[i]$ and $C(\mathbb{Q})$. □

On a geoboard, therefore, one cannot "construct" angles measuring 30°, 60°, 22.5°, 36° and 20°. Thus, Corollary 3 solves a "teaser" in the premier issue of the *Math Horizons* **[16]**.

We mention in passing that although for $r \in \mathbb{Q}$, $\cos(r\pi)$ and $\sin(r\pi)$ are both irrational (the only possible exception is $r = \frac{n}{6}$ for $n \in \mathbb{Z}$), these values are known to be algebraic **[24, 14, 9]**. By contrast, the values $\cos r$ and $\sin r$, for $r \in \mathbb{Q} \setminus \{0\}$, are all transcendental **[15, 27]**. (The first result follows easily from De Moivre's Theorem; the second is much harder.)

Remarks. (a) The unique factorization property of $\mathbb{Z}[i]$ gives an algebraic interpretation of the double-angle correspondence f in (3). In fact, from $a^2 + b^2 = c^2$, we get

$$(a+bi)(a-bi) = c^2. \qquad (9)$$

Writing $a + bi$, $a - bi$ and c as products of irreducible elements in $\mathbb{Z}[i]$ shows that $a + bi$ is a square in $\mathbb{Z}[i]$, i.e., $a + bi = (m+ni)^2 = (m^2 - n^2) + (2mn)i$ for some $m, n \in \mathbb{Z}$. Hence $\tan^{-1}(\frac{b}{a}) = 2\tan^{-1}(\frac{n}{m})$. For instance, $3 + 4i = (2+i)^2$. Thus the simple factorization (9) links the additive and multiplicative properties of numbers, and leads to the starting point of Kummer's work on Fermat's Last Theorem. (Cf. **[1**, Ch. 3] and **[4**, Ch. 5].)

(b) $C(\mathbb{Q})$ can be identified with the (multiplicative) subgroup

$$\mathbb{Q}(i)_1 = \left\{ \frac{a}{c} + \frac{b}{c} i \in \mathbb{Q}(i) \,\middle|\, \left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 = 1 \right\}$$

of $\mathbb{Q}(i)$ consisting of the norm 1 elements, where $\mathbb{Q}(i)$ is realized as the field of fractions of $\mathbb{Z}[i]$. Then for $\frac{a}{c} + \frac{b}{c}i \in \mathbb{Q}(i)$, $(a, b, c) = 1$, and the prime decomposition in $\mathbb{Z}$ of $c = p_1 p_2 \cdots q_1 q_2 \cdots$ where $p_i \equiv 1 (\mathrm{mod}\, 4)$ and $q_j \equiv 3 (\mathrm{mod}\, 4)$, we have, by unique factorization, that

$$\frac{a+bi}{c} = \frac{(r_1 + s_1 i)(r_2 + s_2 i) \cdots}{p_1 p_2 \cdots}.$$

As remarked above,

$$\frac{a+bi}{c} = \frac{(m_1 + n_1 i)^2 (m_2 + n_2 i)^2 \cdots}{p_1 p_2 \cdots} = \frac{m_1 + n_1 i}{m_1 - n_1 i} \frac{m_2 + n_2 i}{m_2 - n_2 i} \cdots.$$

By multiplying both $m_j + n_j i$ and $m_j - n_j i$ by an element in C_2, we can assume that each $m_j > |n_j| > 0$. Thus, we have the generation part of Theorem 1. The argument using $\mathbb{Q}(i)_1$ on the relation part is similar. The map $\mathbb{Q}(i)^* \to \mathbb{Q}(i)_1$ defined by (3) is the map in Hilbert's Satz 90 ([**10**]), applied to the current special case.

Rational Points on the Hyperbola

Our analysis of the rational points on the unit circle $x^2 + y^2 = 1$ can be modified to study the rational points on the hyperbola H defined by $x^2 - y^2 = 1$. In this case, we have a rational parametrization ρ: $\mathbb{R} \backslash \{\pm 1\} \to H$ defined by

$$\rho(t) = \left(\frac{1+t^2}{1-t^2}, \frac{2t}{1-t^2} \right), \tag{10}$$

obtained by intersecting the line of slope t, through $(-1, 0)$, with H. We define an addition $\oplus$ on H by

$$(x_1, y_1) \oplus (x_2, y_2) = (x_1 x_2 + y_1 y_2, x_1 y_2 + x_2 y_1) \tag{11}$$

for $(x_1, y_1), (x_2, y_2) \in H$; the identity element is $(1, 0)$; and the inverse of (x, y) is $(x, -y)$. It is clear from (11) that the rational points $H(\mathbb{Q})$ form a subgroup of H. In the current case, $\mathbb{Z}[i]$ is replaced by $\mathbb{Z}[\varepsilon] \cong \mathbb{Z}[X]/(X^2 - 1)$. Note that $\mathbb{Z}[\varepsilon]$ is not an integral domain, since $(\varepsilon - 1)(\varepsilon + 1) = 0$. To remedy this, we take the subset $R = \{m + n\varepsilon | m > n\}$ of $\mathbb{Z}[\varepsilon]$. Then R is closed under multiplication and has the unique factorization property. Since H consists of two branches $H_1 = \{(x, y) \in H | x > 0\}$ and $H_2 = \{(x, y) \in H | x < 0\}$, the subgroup $H(\mathbb{Q})$ has a corresponding decomposition

$$H(\mathbb{Q}) = H_1(\mathbb{Q}) \cup H_2(\mathbb{Q}). \tag{12}$$

It can be readily verified that H_1 is a subgroup of H and $H_1(\mathbb{Q})$ is a subgroup of $H(\mathbb{Q})$. Thus, (12) is the coset decomposition of $H(\mathbb{Q})$ with respect to $H_1(\mathbb{Q})$, with $H_2(\mathbb{Q}) = (-1, 0) \oplus H_1(\mathbb{Q})$ and $H(\mathbb{Q}) = H' \oplus H_1(\mathbb{Q})$, where $H' = \{(\pm 1, 0)\}$. The map f: $R \to H_1(\mathbb{Q})$ defined by

$$f(m + n\varepsilon) = \left(\frac{m^2 + n^2}{m^2 - n^2}, \frac{2mn}{m^2 - n^2} \right) \tag{13}$$

is an onto *homomorphism*. The following theorem, whose proof is left to the reader, gives a complete description of the group structure of $H(\mathbb{Q})$.

THEOREM 2. *The abelian group $H(\mathbb{Q})$ is the direct sum of infinitely many cyclic subgroups*

$$H(\mathbb{Q}) = H' \oplus \left(\bigoplus_{p \text{ prime}} H_p \right),$$

where, for each prime p, H_p is the infinite cyclic group generated by $\left(\frac{p^2+1}{2p}, \frac{p^2-1}{2p}\right)$.

Note that for $p > 2$, $\left(\frac{p^2+1}{2p}, \frac{p^2-1}{2p}\right) = f\left(\frac{p+1}{2} + \frac{p-1}{2}\varepsilon\right)$, and $p = 2$, $\left(\frac{5}{4}, \frac{3}{4}\right) = f(3 + \varepsilon)$.

Example C. $\left(-\frac{409}{120}, -\frac{391}{120}\right) = (-1, 0) \oplus 2 \cdot \left(\frac{5}{4}, \frac{3}{4}\right) \oplus (-1) \cdot \left(\frac{5}{3}, \frac{4}{3}\right) \oplus \left(\frac{13}{5}, \frac{12}{5}\right)$.

General Conic Sections

The curves C and H studied above are but two special cases of the conic sections (defined by integral polynomials of degree 2). The structure of the rational points on a general conic section may be more complicated. Indeed, some conic sections have no rational points at all. For example, the curve $ax^2 + by^2 = c$, where $a, b, c \in \mathbb{Z}$, not all of the same sign, and abc is square-free, has rational points if and only if $-bc$, $-ca$, $-ab$ are quadratic residues modulo $|a|, |b|, |c|$ respectively. This is Legendre's Criterion **[13]**. Analyzing the general case requires the arithmetic theory of quadratic forms. Once we know that a single rational point P exists, we can get all the rational points by sweeping the secant lines through P of rational slope; this gives a rational parametrization of the curve. (This is a special case of the Hilbert-Hurwitz Theorem **[11]** on the rational parametrization of curves of "genus 0".) In the cases of the unit circle and the hyperbola, we took P to be $(-1,0)$. These cases are exceptionally interesting because of their natural group structure that can be described in a simple way.

Conclusion

We conclude with a brief discussion of the situation for cubic curves.

Let E be a nonsingular cubic curve defined over $\mathbb{Q}$, and let $E(\mathbb{Q})$ be the set of rational points on E. We may assume that E is defined by an equation in Weierstrass form: $y^2 = x^3 + ax^2 + bx + c$. Now call the unique point at infinity $\mathcal{O}$; it will be the identity element of our group. The "negative" of a point (x, y) is $(x, -y)$. In general, a line intersects E in three points, since a general cubic equation has three roots. We define the addition on E and $E(\mathbb{Q})$ by decreeing that collinear points on E "add up" to $\mathcal{O}$. The reader is referred to **[23**, Ch. I**]** for detailed discussion. The picture on the cover of **[23]** illustrates the definition of the addition on E. On such a curve E, we have these group-theoretic properties:

(i) *E is an abelian group.* (This has been known since Euler.)
(ii) *$E(\mathbb{Q})$ is a finitely generated abelian group.* (This is the celebrated Mordell's Theorem **[10]**, first conjectured by Poincaré **[21]** in 1901.)
(iii) *The torsion subgroup of $E(\mathbb{Q})$ can have only the following forms, each of which is realizable: A cyclic group of order n with $1 \leq n \leq 10$ or $n = 12$, or the product of a cyclic group of order 2 and a cyclic group of order $2n$ with $1 \leq n \leq 4$.* (This is Mazur's Theorem **[17, 18]**.)
(iv) *E has only finitely many points with integer coordinates.* (This is Siegel's Theorem **[22]**.)

Note that in the unit circle case, the torsion subgroup coincides with the set of integer points on the curve; this is not the case for cubic curves. The difference as regards finite generation between degree 2 and degree 3 (or higher) should not be surprising: Fermat's equation $X^n + Y^n = Z^n$ has infinitely many solutions if $n = 2$ and no nontrivial solution if $n \geq 3$. The corresponding curves have different topology as well—they have "genus" 0 and 1 (or $n - 2$) respectively.

We hope that our elementary discussion of $C(\mathbb{Q})$ will stimulate the reader's interest in the deeper theory of quadratic forms and cubic curves, and in the discoveries on these subjects by Fermat, Euler, Lagrange, Legendre, Minkowski, Mordell, Hasse, Weil, Siegel, Mazur, Wiles, and others.

Acknowledgement. We thank Sebastian Koh and Joe Silverman for sharing insights with us, the referees and the editor for many valuable comments and suggestions, and Shiv Gupta and Cliff Johnston for technical help.

REFERENCES

1. Z. Borevich and I. Shafarevich, *Number Theory*, Academic Press, New York, 1966.
2. D. Cox, Introduction to Fermat's Last Theorem, *Amer. Math. Monthly* 101(1994), 3–14.
3. H. Darmon, F. Diamond and R. Taylor, Fermat's Last Theorem, in: *Current Developments in Mathematics, 1995*, International Press, Cambridge, MA, 1995, pp. 1–107.
4. H. Edwards, *Fermat's Last Theorem*, Springer-Verlag, New York, 1977.
5. L. Euler, De summa seriei ex numeris primis formitatae ubi numeri primi formae $4n - 1$ habent signum positivum formae autem, $4n + 1$ signum negativum, *Opera Omnia*, Teubner, Leipzig, I, 4, pp. 146–162.
6. G. Faltings, Endlichkeitssätze für abelsche Varietäten über Zahlköopern, *Invent. Math.* 73 (1983), 349–366. (English translation: *Arithmetic Geometry*, G. Cornell and J. Silverman (*ed.*), pp. 9–27. Springer-Verlag, New York, 1986.)
7. G. Faltings, The proof of Fermat's Last Theorem by R. Taylor and A. Wiles, *Notices of the Amer. Math. Soc.* 42(1995), 743–746.
8. F. Gouvêa, A marvelous proof, *Amer. Math. Monthly* 101(1994), 203–222.
9. R. Hamming, The transcendental character of cox x, *Amer. Math. Monthly* 52(1945), 336–337.
10. D. Hilbert, Die Theorie der algebraischer Zahlköper, *Jahesber. Deutsch. Math. Verein.* 4(1897), 175–546.
11. D. Hilbert and A. Hurwitz, Über die diophantischen Gleichungen vom Geschlecht Null, *Acta Math.* 14(1890), 217–224.
12. K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, 2nd edition, Springer-Verlag, New York, 1990.
13. A. Legendre, *Recherches d'Analyse Indéterminée.* Histoire de l'Académie Royale des Sciences, 1785. (Also in: A. Legendre, *Thèorie des Nombres*, 3^{me} ed., t.I, §§III, IV, Mazer, Leipzig, 1886.)
14. D. Lehmer, A note on trigonometric numbers, *Amer. Math. Monthly* 40(1933), 165–166.
15. F. Lindermann, Über die Zahl π, *Math. Ann.* 20(1882), 213–225.
16. *Math Horizons* 1(1993), 21.
17. B. Mazur, Modular curves and the Eisenstein ideals, *IHES Publ. Math.* 47(1977), 33–186.
18. B. Mazur, Rational isogenies of prime degree, *Invent. Math.* 44(1978), 129–162.
19. L. J. Mordell, On the rational solutions of the indeterminate equations of the third and fourth degrees, *Proc. Camb. Philos. Soc.* 21(1922), 179–192.
20. J. Olmsted, Rational values of trigonometric functions, *Amer. Math. Monthly* 52(1945), 507–508.
21. H. Poincaré, Sur la propriétés arithmétiques des courbes algébriques, *Journal de Mathematiques Pures et Appliquees* 7(1901), 161–233.
22. C. L. Siegel, Über einige Anwendungen diophantischer Approximationen, *Abh. Preuss. Akad. Wiss. Phys.-Mat. Kl.* (1929) (*Collected Works*, Springer-Verlag, New York, 1966, 209–266.)
23. J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, New York, 1992.
24. E. Swift, Note on trigonometric functions, *Amer. Math. Monthly* 29(1922), 404–405.
25. R. Taylor and A. Wiles, Ring theoretic properties of certain Hecke algebras, *Annals of Math.* 141(1995), 553–572.
26. R. Underwood, On the irrationality of certain trigonometric functions, *Amer. Math. Monthly* 28(1921), 374–376.
27. K. Weierstrass, Zu Lindermanns Abhandluig "Über die Ludolphsche Zahl," *Sitzgsber. Preuss. Akad. Wiss.* (1885), 1067–1085. (*Math. Werke*, II, 341–362.)
28. A. Wiles, Modular elliptic curves and Fermat's Last Theorem, *Annals of Math.* 141(1995), 443–551.

A Simple Introduction to Integral Equations

GLENN LEDDER
University of Nebraska-Lincoln
Lincoln, NE 68588-0323

Introduction

Integral equations are equations in which the unknown function appears inside a definite integral. They are closely related to differential equations. Initial value problems and boundary value problems for ordinary and partial differential equations can often be written as integral equations (see [7] for an introduction to the technique), and some integral equations can be written as initial or boundary value problems for differential equations. Problems that can be cast in both forms are generally more familiar as differential equations, owing to the larger collection of analytical procedures for solving differential equations.

Many applications are best modeled with integral equations, but most of these problems require a lengthy derivation. A relatively simple example is the model for population dynamics, with birth and death rates that depend on age. This model was first formulated in 1922 by Alfred J. Lotka [8], who is best known for the Lotka-Volterra predator-prey population model. A similar model has been used to model the spread of the HIV virus among IV drug users [4]. A simple age-dependent population model will be developed at the end of this article.

Integral equations are also important in the theory and numerical analysis of differential equations; this is where the mathematics student is most likely to encounter them. For example, Picard's existence and uniqueness theorem for first-order initial value problems is conveniently proved using integral equations [1]; the proof is constructive and can be used to formulate a method for numerical solution of initial value problems.

Systematic study of integral equations is usually undertaken as part of a course in functional analysis (see [6]) or applied mathematics (see [9]). This advanced setting is required for a full appreciation of integral equation theory, but it makes the subject accessible only to the advanced student. By contrast, several of the important results in the theory of integral equations can be demonstrated using nothing more than elementary analysis, and the elaboration of this statement is the goal of the present discussion. In fact, all but one of the results presented here will be derived using nothing more than the material presented in a standard advanced calculus course. This gives less advanced students of mathematics a chance to encounter some integral equation concepts that play an important role in the theory and application of continuous mathematics.

Definitions, Notation, and Assumptions

A general treatment of linear integral equations is given in Kanwal [5]. We consider only linear integral equations of the second kind; that is, equations of the form

$$y(x) = f(x) + \int_a^B k(x,s)\,y(s)\,ds, \tag{1}$$

where f and k are prescribed functions, y is unknown, a is a given constant, and R is either x or a constant $b > a$. The function k is called the **kernel** of the integral equation. When $f \equiv 0$, equations of the form (1) are called homogeneous; the theory of homogeneous equations is of particular importance, as we shall see in the discussion that follows. The distinction between the variable upper limit and the constant upper limit is also quite important; however, it will be convenient at times to keep the equations general by using the generic upper limit R, since many of the proofs are the same for both cases. Equations with $R = x$ are called **Volterra equations**, while equations with a fixed integration interval are called **Fredholm equations**. Volterra equations are often associated with initial value problems (as in the case of population models), while Fredholm equations are often associated with boundary value problems. This is one reason that Volterra equations are generally easier to work with than Fredholm equations; another reason will become clear in the discussion of the fundamental lemmas that follow.

We will assume throughout the discussion that the nonhomogeneous term f and the solution y are continuous on $[a, b]$ and the kernel k is continuous on the square $[a, b] \times [a, b]$. Some of the results given here are true under more general circumstances, but the proofs require functional analysis **[2, 5, 6, 9]**. Since we are dealing with continuous functions, the statement "$y = \phi$" will be taken to mean that $y(x) = \phi(x)$ for all $x \in [a, b]$.

The theoretical development that follows will make frequent use of two bounds related to the kernel; these are defined by

$$M_V = \max_{[a,b]\times[a,b]} |k(x,s)| \text{ and } M_F = \sup_{a \le x \le b} \int_a^b |k(x,s)|\, ds. \tag{2}$$

Note that these bounds always exist because k is continuous on the given closed domain. The function defined by $g(x) = \int_a^b |k(x, s)|\, ds$ is not necessarily continuous on $[a, b]$; however, it can be shown that any singularities it may possess are removable. Clearly, $0 \le g(x) \le (b - a)M_V$, so

$$M_F \le (b - a) M_V. \tag{3}$$

In the theorems that follow, the tighter bound M_F will be used for Fredholm equations, while the simpler bound M_V will be required for Volterra equations. These bounds are related to norms used in functional analysis: M_V is the sup norm of the function k; M_F is like the sup norm with regard to x, but is like the L_1 norm with regard to s **[6]**. The bound M_F is used here in place of the L_2-norm of k, which is defined by

$$\|k\| = \left[\int_a^b \int_a^b |k(x,s)|^2\, dx\, ds\right]^{1/2}. \tag{4}$$

The norm of k is a tighter bound than M_V, but may be smaller or larger than M_F, depending on the specific kernel. The main advantage of $\|k\|$ relative to M_F is that it applies to a broader class of functions, since L_2 functions need not be continuous. The main advantage of M_F is the natural way it is used in arguments, as will be seen in Lemma 2 below.

As an example of the different bounds, consider the kernel $k(x, s) = e^{xs}$ on the interval $[-1, 1]$. The Volterra bound is $M_V = e$. The function appearing as the integral in the Fredholm bound M_F is $g(x) = \int_{-1}^{1} e^{xs}\, ds = 2x^{-1} \sinh x$. This function has a removable singularity at $x = 0$, with $\lim_{x \to 0} g(x) = 2$. Thus, the Fredholm bound is

$M_F = 2\sinh 1$. The L_2-norm is $\|k\| = (2\int_0^2 t^{-1}\sinh t\,dt)^{1/2}$. For comparison, the numerical values of these bounds are, to 2 significant figures, $\|k\| \approx 2.24$, $M_F \approx 2.35$, $(b-a)M_V \approx 5.44$.

In either the Volterra case or the Fredholm case, we take y_0 to be some continuous function and define a sequence by

$$y_n(x) = f(x) + \int_a^R k(x,s)\,y_{n-1}(s)\,ds \quad \text{for } n = 1, 2, \ldots. \tag{5}$$

These sequences will be referred to as **Volterra** or **Fredholm sequences**. In particular, the next section will deal with homogeneous ($f \equiv 0$) Volterra and Fredholm sequences.

Two Fundamental Lemmas

We begin with a basic result about homogeneous Volterra sequences.

LEMMA 1. *Let y_0 be any continuous function on $[a,b]$, let $x \in [a,b]$, and let $c = \max_{[a,b]}|y_0(x)|$. Let y_n be the Volterra sequence $y_n(x) = \int_a^x k(x,s)y_{n-1}(s)\,ds$. Then $|y_n(x)| \le cM_V^n(x-a)^n/n!$ and so $\lim_{n\to\infty} y_n(x) = 0$ for all $x \in [a,b]$.*

Proof. Let $x \in [a,b]$. Then

$$|y_1(x)| \le \int_a^x |k(x,s)|\,|y_0(s)|\,ds \le cM_V \int_a^x ds = cM_V(x-a).$$

Substituting this result back into (5) gives $|y_2(x)| \le \int_a^x |k(x,s)|\,|y_1(s)|\,ds \le cM_V^2 \int_a^x (s-a)\,ds = cM_V^2(x-a)^2/2$. Similarly, by induction we have the general result $|y_n(x)| \le cM_V^n(x-a)^n/n!$, from which it follows that y_n converges to 0. □

A corresponding result can be proven for homogeneous Fredholm sequences.

LEMMA 2. *Let y_0 be any continuous function on $[a,b]$, let $x \in [a,b]$, and let $c = \max_{[a,b]}|y_0(x)|$. Let y_n be the Fredholm sequence $y_n(x) = \int_a^b k(x,s)y_{n-1}(s)\,ds$. Then $|y_n(x)| \le cM_F^n$ and so $\lim_{n\to\infty} y_n(x) = 0$ for all $x \in [a,b]$ whenever $M_F < 1$.*

Proof. Let $x \in [a,b]$. Then

$$|y_1(x)| \le \int_a^b |k(x,s)|\,|y_0(s)|\,ds \le c\int_a^b |k(x,s)|\,ds \le cM_F.$$

Substituting this result back into (5) gives

$$|y_2(x)| \le \int_a^b |k(x,s)|\,|y_1(s)|\,ds \le cM_F \int_a^x |k(x,s)|\,ds \le cM_F^2.$$

Similarly, by induction we have the general result $|y_n(x)| \le cM_F^n$, from which it follows that y_n converges to 0 whenever $M_F < 1$. □

There is a subtle difference between these two lemmas, with important implications. The convergence of the sequence y_n for Volterra equations is caused by the repeated integration of the polynomial $(x-a)^n$. The sequence converges to 0 for any finite value of M_V; consequently, iterative solution of Volterra equations will be successful for any continuous k and efficient when $(b-a)M_V$ is of moderate value. By contrast, the convergence of y_n for Fredholm equations is caused by accumulation of factors M_F; the sequence of iterates will behave like the sequence of terms in a geometric series. As a result, iterative solution of Fredholm equations will be successful when $M_F < 1$ and efficient *only when* M_F *is small.*

The bound M_V could have been used in Lemma 2, but the hypothesis of the lemma is strengthened by the use of M_F instead, according to (3). A more significant strengthening of the lemma would be to use the L_2 norm of k (defined in (4)) in place of M_F. The lemma would then hold for some kernels that are not continuous, but the proof requires more sophistication (see [5]).

A Simple Example

Let $a=0$, $b=1$, $f=0$, $k=\lambda xs$, and $y_0=1$, where λ is a parameter. Note that $M_v=|\lambda|$ and $M_F=|\lambda|/2$. Table 1 compares the error bounds given by Lemmas 1 and 2 with the actual sequence of iterates defined by (5) for the Volterra and Fredholm cases.

TABLE 1. Comparison of the convergence of $\{y_n\}$ for Volterra and Fredholm equations, with $a=0$, $b=1$, $f=0$, $k=\lambda xs$, $y_0=1$

	Error Bounds from Lemmas	Actual Results (from (5))
Volterra	$\lvert y_n(x)\rvert \le \frac{(\lvert\lambda\rvert x)^n}{n!}$	$y_n(x)=\frac{(\lambda x^3)^n}{2\cdot 5\cdot 8\cdot \cdots \cdot(3n-1)}$
Fredholm	$\lvert y_n(x)\rvert \le \left(\frac{\lvert\lambda\rvert}{2}\right)^n$	$y_n(x)=\frac{\lambda^n x}{2\cdot 3^{n-1}}$

1. The actual results may be considerably better than the error bounds given by the lemmas. For the Volterra case, the denominator actually grows much faster than $n!$. For the Fredholm case, the accumulating factor in the denominator is 3 rather than 2.
2. The rate of convergence may be quite different for different values of x. In the Volterra case of the example, the numerator has the factor x^{3n}, compared to the factor x^n of the error bound. Convergence in the example will be slower near $x=1$ than elsewhere.
3. The Fredholm sequence may still converge if $M_F \ge 1$ (in the example, we get convergence for $|\lambda|<3$, or $M_F<3/2$). This is not important in practice, because when $M_F \approx 1$ the convergence will be too slow to be of practical value. The stronger theorems obtained with functional analysis are theoretically better, but the difference is of little importance in practice.

FIGURES 1 and 2 illustrate the convergence of the sequences of Volterra and Fredholm iterates. Note that in the Volterra case with $\lambda=8$, the value $y_n(1)$ actually increases with n at first. For smaller values of x, the convergence is apparent immediately. The figures clearly confirm the observations given above.

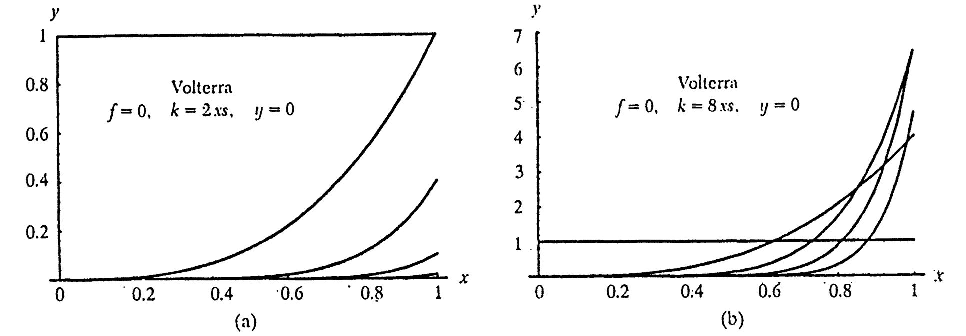


FIGURE 1
The iterates y_0 to y_4 for the homogeneous example Volterra equation, with $y_0 = 1$. A: $k = 2xs$, B: $k = 8xs$.

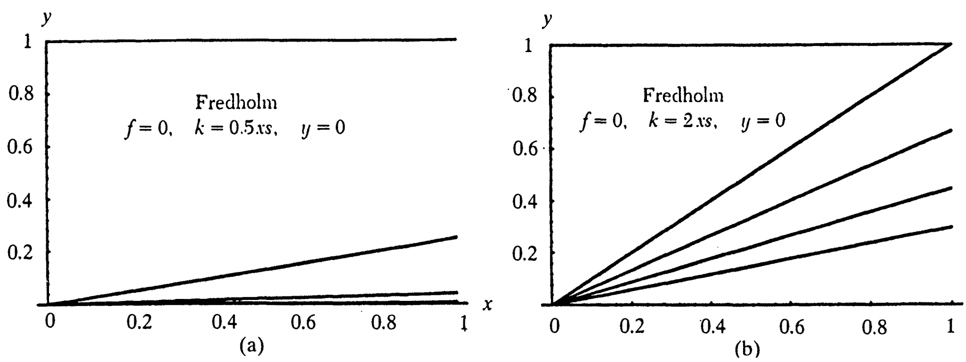


FIGURE 2
The iterates y_0 to y_4 for the homogeneous example Fredholm equation, with $y_0 = 1$. A: $k = 0.5xs$, B: $k = 2xs$.

Some Theoretical Results

Using the lemmas of the previous section, we can develop theoretical results about the existence and uniqueness of continuous solutions. The first theorem deals with existence and uniqueness of solutions to the homogeneous Volterra and Fredholm equations.

THEOREM 1. (*for homogeneous equations*)

(1) *$y = 0$ is the unique solution of $y(x) = \int_a^x k(x,s)y(s)\,ds$.*

(2) *If $M_F < 1$, then $y = 0$ is the unique solution of $y(x) = \int_a^b k(x,s)y(s)\,ds$.*

Proof. Let $R = x$ for the Volterra case and $R = b$ for the Fredholm case. We first note that $y = 0$ is a solution of $y(x) = \int_a^R k(x,s)y(s)\,ds$. Now let ϕ be a solution and let $y_0 = \phi$. Then $y_1(x) = \int_a^R k(x,s)\phi(s)\,ds = \phi(x)$, and so clearly $y_n = \phi$ for all n. Thus, $\lim_{n\to\infty} y_n(x) = \phi(x)$ for all $x \in [a,b]$. But, by Lemma 1 or Lemma 2, $\lim_{n\to\infty} y_n(x) = 0$ for all $x \in [a,b]$. Thus, $\phi = 0$. □

Homogeneous Volterra equations have no non-trivial solutions, but the possibility of non-trivial solutions to homogeneous Fredholm equations is only ruled out by Theorem 1 when $M_F < 1$. A simple example of a homogeneous Fredholm equation with non-trivial solutions is provided by $y(x)=\int_0^1 y(s)\,ds$, for which any constant function is a solution. Here $M_F=1$, so, the theorem is not contradicted.

Suppose the kernel $k(x,s)$ includes a parameter λ that appears as a multiplicative factor. Then the value of M_F will be proportional to λ. Values of λ for which the integral equation has non-trivial solutions are called **eigenvalues** of the associated integral operator **[5]**. (Note that some authors, such as Stakgold **[9]**, use the term eigenvalue for the reciprocal of λ rather than λ.) Theorem 1 says that Volterra operators have no eigenvalues and that any eigenvalues of Fredholm operators are bounded away from 0.

We now direct our attention to the question of existence and uniqueness for nonhomogeneous integral equations. Theorem 2 uses uniqueness for homogeneous equations to prove uniqueness for nonhomogeneous equations.

THEOREM 2. (*uniqueness of solutions*)

(1) $y(x)=f(x)+\int_a^x k(x,s)y(s)\,ds$ *has at most one solution.*

(2) *If* $M_F<1$, *then* $y(x)=f(x)+\int_a^b k(x,s)y(s)\,ds$ *has at most one solution.*

Proof. Let $R=x$ for the Volterra case and $R=b$ for the Fredholm case. Let ϕ_1 and ϕ_2 be solutions and let $z=\phi_1-\phi_2$. Then $z(x)=\phi_1(x)-\phi_2(x)=\int_a^R k(x,s)[\phi_1(s)-\phi_2(s)]\,ds=\int_a^R k(x,s)z(s)\,ds$. Applying Theorem 1 to this homogeneous integral equation gives $z\equiv 0$. Thus, $\phi_1\equiv\phi_2$. □

We now demonstrate the existence of solutions. Our proof uses Lemmas 1 and 2 and requires more background in analysis than the first two theorems.

THEOREM 3. (*existence of solutions and convergence of* $\{y_n\}$). *Let* $y_0=f$ *and define the sequence* $\{y_n\}$ *by* $y_n(x)=f(x)+\int_a^R k(x,s)y_{n-1}(s)\,ds$ *for* $n=1,2,\dots$, *with* $R=x$ *or* $R=b$.

(1) *If* $R=x$, *then* $\{y_n\}$ *converges to the unique solution of* $y(x)=f(x)+\int_a^x k(x,s)y(s)\,ds$.

(2) *If* $R=b$ *and* $M_F<1$, *then* $\{y_n\}$ *converges to the unique solution of* $y(x)=f(x)+\int_a^b k(x,s)y(s)\,ds$.

Proof. We begin by defining a sequence formed by differences of consecutive terms of $\{y_n\}$: Let $E_0=y_0$ and $E_n=y_n-y_{n-1}$ for $n=1,2,\dots$. Note that for any positive integers n and m,

$$|y_{n+m}-y_n|=|y_{n+m}-y_{n+m-1}+y_{n+m-1}-y_{n+m-2}+\cdots+y_{n+1}-y_n|$$
$$\le|E_{n+m}|+|E_{n+m-1}|+\cdots+|E_{n+1}|<\sum_{m=1}^{\infty}|E_{n+m}|.$$

To complete the proof, it will suffice to show that $g_n\equiv\sum_{m=1}^{\infty}|E_{n+m}|$ exists and converges to 0 as $n\to\infty$. This will prove that $\{y_n\}$ is a Cauchy sequence and therefore convergent to some function ϕ. Hence, $\phi(x)=\lim_{n\to\infty}y_n(x)=f(x)+\lim_{n\to\infty}\int_a^R k(x,s)y_{n-1}(s)\,ds=f(x)+\int_a^R k(x,s)\cdot\lim_{n\to\infty}y_{n-1}(s)\,ds=f(x)+\int_a^R k(x,s)\phi(s)\,ds$, so ϕ is a solution (unique by Theorem 2) of the integral equation.

As a first step to using Lemmas 1 and 2 to prove that $g_n\equiv\sum_{m=1}^{\infty}|E_{n+m}|$ converges to 0, we establish the result $E_n=\int_a^R k(x,s)E_{n-1}(s)\,ds$ by induction. Note first that $E_1=y_1-y_0=f(x)+\int_a^R k(x,s)y_0(s)\,ds-f(x)=\int_a^R k(x,s)E_0(s)\,ds$. Now suppose that $E_j=\int_a^R k(x,s)E_{j-1}(s)\,ds$ for $j=1,\cdots,n-1$. Then $E_n(x)=y_n(x)-y_{n-1}(x)=[f(x)+\int_a^R k(x,s)y_{n-1}(s)\,ds]-[f(x)+\int_a^R k(x,s)y_{n-2}(s)\,ds]=\int_a^R k(x,s)[y_{n-1}(s)-y_{n-2}(s)]\,ds=\int_a^R k(x,s)E_{n-1}(s)\,ds$, proving the claim.

(1) For the Volterra case, we first define the real number β by $\beta = \sum_{m=1}^{\infty}[M_V^m(b-a)^m/m!] = e^{M_V(b-a)} - 1$. We may obtain a bound for g_n as follows:

$$g_n = \sum_{m=1}^{\infty} |E_{n+m}| \le \sum_{m=1}^{\infty} cM_V^{n+m}(x-a)^{n+m}/(n+m)!$$

$$< c\left[M_V^n(x-a)^n/n!\right] \cdot \sum_{m=1}^{\infty}\left[M_V^m(x-a)^m/m!\right] \le \beta c M_V^n (b-a)^n/n!.$$

In this calculation, the first inequality follows from Lemma 1, the second follows from $n!m! < (n+m)!$, and the third from systematic replacement of x by b. Thus, $\{g_n\}$ converges to 0 and the theorem is proved for the Volterra case.

(2) For the Fredholm case, we have, with the aid of Lemma 2 (given $M_F < 1$), the bound

$$g_n = \sum_{m=1}^{\infty} |E_{n+m}| \le \sum_{m=1}^{\infty} cM_F^{n+m} = cM_F^{n+1}\sum_{m=1}^{\infty} M_F^{m-1} = c(1-M_F)^{-1}M_F^{n+1}.$$

Thus, $\{g_n\}$ converges to 0 and the theorem is proved for the Fredholm case. □

Iterative Solution of Integral Equations

Theorem 3 is constructive in nature; thus, it has as a practical consequence the description of an iteration scheme that converges to the solution. Theoretically, we can solve any Volterra equation and any Fredholm equation with $M_F < 1$ by constructing the sequence $\{y_n\}$ and computing enough terms to achieve any desired tolerance. In practice, one can easily write routines using a computer algebra system, such as *Mathematica* or *Maple*, to compute the sequence. The usefulness of the sequence will be dependent on how rapidly the sequence converges.

We now give a brief illustration of the convergence of $\{y_n\}$ to the solution of the example equations

$$y(x) = 1 - 4x^3 + \int_0^x 8xsy(s)\,ds, \qquad y(x) = 1 - x + \int_0^1 2xsy(s)\,ds. \qquad (6)$$

These equations were chosen for comparison with the homogeneous examples: a, b, and k are the same as in FIGURES 1B and 2B, and the nonhomogeneous term f was chosen for each example to yield the solution $y \equiv 1$. Note that since $M_F = 1$, Theorem 3 doesn't actually apply for the Fredholm case; nevertheless, the iterations in this example converge as long as $M_F < 3/2$.

FIGURE 3 shows the iterates y_0 to y_4 for these two cases. The graphs are, of course, quite similar to the corresponding graphs of FIGURES 1B and 2B. The proof of Theorem 3 suggests that convergence for nonhomogeneous equations might be noticeably slower than for homogeneous equations, since the error in y_n is bounded by an *infinite sum* of homogeneous error bounds. In practice, this distinction does not make very much difference, since the terms in the infinite sums for the error bounds decay fairly fast with n (particularly for the Volterra case). Examples with a relatively large value of M_V and a near-critical value of M_F were deliberately chosen so that convergence is slow enough for a good illustration in FIGURE 3. With smaller values of λ, the convergence is pleasingly rapid to the practitioner who wants an accurate numerical approximation of the solution.

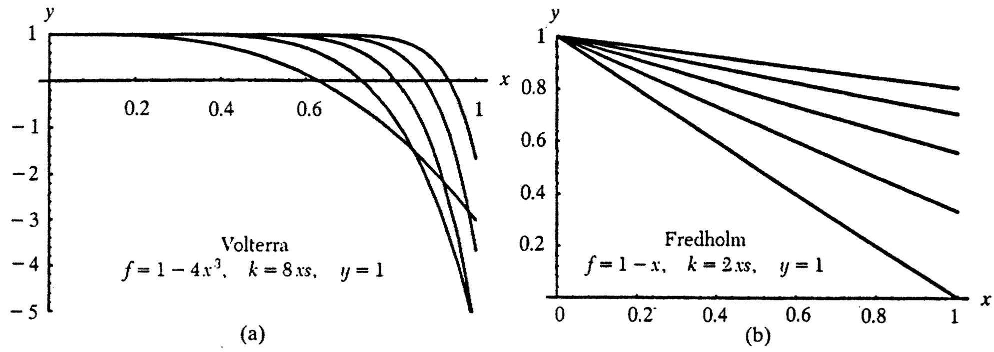


FIGURE 3

The iterates y_0 to y_4 for the nonhomogeneous example equations. A: Volterra, B: Fredholm.

A Simple Age-Dependent Population Model

As an illustration of the application of integral equations, consider a single-sex population with reproduction rate dependent on age. We assume familiarity with the standard model for growth of a single-sex population $P(t)$ with unlimited resources [1]. Population growth is taken to be proportional to the population, leading to the differential equation model

$$\frac{dP}{dt} = B(t) = rP(t), \qquad P(0) = P_0, \tag{7}$$

with $B(t)$ the net birth rate, r a constant representing the fertility of the species, and P_0 the initial population. As is typically done in modeling large populations, we take P to be a continuous function approximating the population rather than the actual population, which is an integer function.

One improvement that can be made in the foregoing model is to allow the fertility coefficient r to vary with the age of the individual. We will still have $dP/dt = B$, but with no obvious algebraic expression for B in terms of P. We assume that r is a positive function of age a that is increasing up to age A and decreasing thereafter. For the sake of simplicity, we also assume that all individuals in the population are immortal. The reader is reminded that the use of population models is not restricted to biological populations. The age-dependent model with immortality being considered here might be applied to the population of published papers on mathematics. We are also assuming that each new issue of a journal spawns subsequent papers at a rate proportional to the number of papers in the issue, with a proportionality factor that reaches a maximum at a time A after the original publication and decreases thereafter. In the remaining development, we will continue to use the language of population dynamics.

We will need to keep track of the age distribution of the population, so let $Q(a,t)$ be the population at time t of individuals younger than age a and let u be the unknown population density function, defined by

$$u(a,t) = \frac{\partial Q}{\partial a}(a,t) \geq 0. \tag{8}$$

In general, the population P_{a_1,a_2} between ages a_1 and a_2 is given by

$$P_{a_1,a_2}(t) = Q(a_2,t) - Q(a_1,t) = \int_{a_1}^{a_2} \frac{\partial Q}{\partial a}(a,t)\,da = \int_{a_1}^{a_2} u(a,t)\,da. \tag{9}$$

Specifically,

$$P(t) \equiv P_{0,\infty}(t) = Q(\infty,t) = \int_0^\infty u(a,t)\,da. \tag{10}$$

The improper integral in (10) converges whenever the population is finite. We assume that the initial distribution is given by $u(a,t) = u_0(a)$, with u_0 known and $P_0 = \int_0^\infty u_0(a)\,da$.

Suppose the rate of births to a cohort of individuals of age a is proportional to the population of the cohort, with proportionality factor $r(a) \geq 0$. The rate of births at time t to the cohort with ages from a to $a + da$ is approximately $r(a)u(a,t)\,da$, so the overall birth rate $B(t)$ is given by

$$B(t) = \int_0^\infty r(a)u(a,t)\,da. \tag{11}$$

The integral in (11) converges under the given assumptions, since

$$\begin{aligned} 0 \leq \int_0^\infty r(a)u(a,t)\,da &= \int_0^A r(a)u(a,t)\,da + \int_A^\infty r(a)u(a,t)\,da \\ &\leq \int_0^A r(a)u(a,t)\,da + r(A)P_{A,\infty}(t). \end{aligned}$$

(Note also that (11) reduces to $B(t) = rP(t)$ when r is constant, in agreement with (7).) Once the birth rate is known, the population $P(t)$ can be determined by integration of (7):

$$P(t) = P_0 + \int_0^t B(s)\,ds. \tag{12}$$

It remains to find another relationship between B and u to combine with (11).

In the absence of deaths, all members of the population present at time 0 persist for all subsequent time, aging equally as time passes. Thus,

$$P_{a+t,\infty}(t) = P_{a,\infty}(0) \tag{13}$$

or, equivalently,

$$Q(\infty,t) - Q(a+t,t) = Q(\infty,0) - Q(a,0).$$

Differentiating with respect to a gives

$$u(a+t,t) = u(a,0),$$

or

$$u(a,t) = u(a-t,0) = u_0(a-t), \quad \text{for } a \geq t \geq 0. \tag{14}$$

Similarly, given a and t with $a < t$, the group of individuals under age a at time t

consists of those born between times $t-a$ and t; hence

$$Q(a,t)=\int_{t-a}^{t} B(\tau)\,d\tau, \quad \text{for } a<t. \tag{15}$$

Differentiating yields

$$u(a,t)=B(t-a), \quad \text{for } a<t. \tag{16}$$

Substituting (14) and (16) into (11) gives

$$B(t)=\int_{t}^{\infty} r(a)u_0(a-t)\,da+\int_{0}^{t} r(a)B(t-a)\,da.$$

After changing variables in both integrals, we have a Volterra equation of the second kind for the birth rate:

$$B(t)=f(t)+\int_{0}^{s} r(t-s)B(s)\,ds, \tag{17}$$

with

$$f(t)=\int_{0}^{\infty} r(t+s)u_0(s)\,ds.$$

Equation (17) is of convolution type and so can be solved analytically using the Laplace transform [**1**, **7**]. However, a satisfactory approximation can be obtained for specific functions r and u_0 by the iteration scheme given above.

A more general model of age-dependent populations is presented by Hoppensteadt [**3**], who includes the effect of an age-dependent death rate. Equations (13) and (15) must be replaced by a first order partial differential equation for u, which can then be solved by the method of characteristics. The result is a linear Volterra equation of the second kind for B similar to (17), but with a kernel and nonhomogeneous term that include the effects of the death rate. Hoppensteadt derives several results about such models, using the Laplace transform method to obtain an explicit solution for B.

REFERENCES

1. W. E. Boyce and R. C. DiPrima, *Elementary Differential Equations*, 5th ed., Wiley, New York, 1992.
2. L. M. Delves and J. L. Mohamed, *Computational Methods for Integral Equations*, Cambridge University Press, Cambridge, UK, 1985.
3. F. Hoppensteadt, *Mathematical Theories of Populations: Demographics, Genetics and Epidemics*, SIAM, Philadelphia, 1975.
4. J. M. Hyman and E. A. Stanley, AIDS and a risk-based model: mathematical formalism, *Los Alamos Science* 18(1989), 24–27.
5. R. P. Kanwal, *Linear Integral Equations: Theory and Technique*, Academic Press, New York, 1971.
6. E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, New York, 1978.
7. J. D. Logan, *Applied Mathematics: A Contemporary Approach*, Wiley, New York, 1987.
8. A. J. Lotka, The stability of the normal age distribution, *Proc. Nat. Acad. Sci.* 8(1922), 339–345.
9. I. Stakgold, *Green's Functions and Boundary Value Problems*, Wiley, New York, 1979.

NOTES

Proof of a Conjecture of Lewis Carroll

NORBERT HUNGERBÜHLER
ETH-Zentrum, Departement Mathematik
CH-8092 Zürich, Switzerland

Introduction Charles Lutwidge Dodgson, who is better known as Lewis Carroll, the author of "Alice's Adventures in Wonderland" (1865), was also a many-sided mathematician. His work treats—among many other subjects—investigations on right triangles with sides of integer length. In *Life and Letters of Lewis Carroll* ([**1**, p. 343]) we find the diary entry

> Dec. 19 (Sun).—Sat up last night till 4 a.m., over a tempting problem, sent me from New York, "to find 3 equal rational-sided rt.-angled Δ's." I found *two*, whose sides are 20, 21, 29; 12, 35, 37; but could not find *three*.

Thus Carroll's search was for triples of integer-sided right triangles, all having the same area. Despite his failure, Carroll conjectured that infinitely many such triples exist. (Two triples are identified if they are equal up to scaling by a rational factor.) Apparently, Carroll was not aware of Fermat's observation [**4**] that if z is the hypotenuse and b and d are the legs of a rational right triangle, then we obtain a new rational right triangle of the same area with sides

$$z' = \frac{z^4 + 4b^2d^2}{2z(b^2 - d^2)}; \qquad b' = \frac{z^4 - 4b^2d^2}{2z(b^2 - d^2)}; \qquad d' = \frac{4z^2bd}{2z(b^2 - d^2)}.$$

It is not clear, however, whether iterating this procedure actually produces at least three non-congruent triangles. Nor is it obvious whether we can find infinitely many *different triples* by this procedure. Problems of this type have been treated by several authors (see, e.g., [**2**], [**3**] and [**7**]), but I am not aware of any explicit answer to Carroll's conjecture.

We prove in this Note that Carroll's conjecture is true.

Preliminaries It is well known (see, e.g., [**6**]) that the integer equation

$$x^2 + y^2 = z^2$$

has the general solution

$$x = 2\lambda mn; \qquad y = \lambda(m^2 - n^2); \qquad z = \lambda(m^2 + n^2) \tag{1}$$

with $\lambda, m, n \in \mathbb{N}$. If we restrict m and n to be relatively prime and not both odd, and set $\lambda = 1$, we obtain an infinite family of primitive triangles (i.e., $\gcd(x, y, z) = 1$).

Since x and y are the legs the triangle has area

$$A = \frac{1}{2}xy = mn(m^2 - n^2). \tag{2}$$

Thus Carroll's conjecture is an assertion on the solutions of the third-order equation (2).

Proof of the conjecture The following theorem answers Carroll's conjecture.

THEOREM. *There exists an infinite family of triples of integer-sided right triangles with the same area. In particular, if m is a prime number of the form $m = 6N + 1$, $N \in \mathbb{N}$, then there exist positive integer solutions n and l of the equation $m^2 = n^2 + nl + l^2$, and a primitive triple of integer-sided right triangles of area $A = mnl(n + l)$. Different values of m lead to different primitive triples.*

Proof. The theory of diophantic quadratic forms (see, e.g., **[6]**) asserts that the equation

$$m^2 = n^2 + nl + l^2 \tag{3}$$

has nonzero integer solutions if m has the form $m = \prod_{i=1}^{r} p_i$, where the p_i are primes of the form $p_i = 6N_i + 1$. In particular, solutions n and l exist if m itself is prime of the form $m = 6N + 1$. Now Dirichlet's theorem (see, e.g., **[5]**) says that every arithmetic sequence $\{a_N\}_{N \in \mathbb{N}}, a_N = \alpha N + \beta$, where α and β are relatively prime, includes infinitely many prime numbers. In particular, there are infinitely many prime numbers of the form $6N + 1$.

The solutions n and l of (3) trivially satisfy $n \neq \pm l$. Multiplying (3) by $n - l$ and rearranging terms gives

$$m^2 n - n^3 = m^2 l - l^3 = \frac{A}{m}, \tag{4}$$

where the last equality defines the number A. We may assume that $A > 0$ (otherwise, replace (n, l) with $(-n, -l)$). We may interpret n and l as two roots of the cubic polynomial $p(x) = x^3 - m^2 x + A/m$. If i is the third root, then Viëta's theorem (see, e.g., **[5]**) implies that

$$n + l + i = 0 \tag{5}$$

and

$$nli = -\frac{A}{m}.$$

Hence, all the roots n, l and i are integers, with $\pm n \neq i \neq \pm l$. Since $A/m > 0$, two roots (say n and l) are positive and one is negative. Thus

$$mn(m^2 - n^2) = A; \qquad ml(m^2 - l^2) = A; \qquad -im(i^2 - m^2) = A.$$

In other words the three integer-sided right triangles D_1, D_2, D_3, with sides

$$\begin{array}{lll} x_1 = 2mn & y_1 = m^2 - n^2 & z_1 = m^2 + n^2 \\ x_2 = 2ml & y_2 = m^2 - l^2 & z_2 = m^2 + l^2 \\ x_3 = -2mi & y_3 = i^2 - m^2 & z_3 = i^2 + m^2 \end{array} \tag{6}$$

have common area $A = -mnli$. Since the z_i are all different, so are the triangles.

Notice that (5) implies that at least one of n, l and i is even. It is easy to see that none of n, l and i is a multiple of m. Since m is prime, it follows that at least one of D_1, D_2 and D_3 is primitive. Hence the triple D_1, D_2, D_3 is primitive.

Now, it follows from $m < A$ that we obtain infinitely many primitive triples in this way. To see that different values of m lead to different primitive triples, choose the triangle D of a triple that has maximum hypotenuse z; denote its legs by x and y. Since m is prime, it follows from (6) that

$$\text{either}\quad m = \sqrt{\frac{z-x}{2}} \in \mathbb{N} \quad \text{or} \quad m = \sqrt{\frac{z-y}{2}} \in \mathbb{N}.$$

Hence m is determined by the triple and this completes the proof.

Remark The method above does not produce all primitive triples. The primitive triple with minimum area which is of a different type is

$$\begin{array}{lll} x_1 = 4080 & y_1 = 1001 & z_1 = 4201 \\ x_2 = 1430 & y_2 = 2856 & z_2 = 3194 \\ x_3 = 528 & y_3 = 7735 & z_3 = 7753 \end{array}$$

It is generated by (1), using the numbers

$$(m_1, n_1) = (51, 40); \quad (m_2, n_2) = (55, 13); \quad (m_3, n_3) = (88, 3).$$

The common area is $A = 2042040$. (This example was found by a *Mathematica* program.)

REFERENCES

1. Stuart Dodgson Collingwood, *The Life and Letters of Lewis Carroll*, Gale Research Co., Detroit, 1967.
2. J. Collins, *The Gentleman's Math. Companion*, 2, No. 11 (1808), 123.
3. J. Cunliffe, *New Series of the Math. Repository (ed. Th. Leybourn)*, 3, II (1814), 60.
4. P. Fermat, *Œuvres III, 254–255; Fermat's Diophanti Alex. Arith.*, 1670, 220.
5. M. Hazewinkel, *Encyclopaedia of Mathematics*, Kluwer Academic Publishers, Dordrecht, 1987.
6. L. J. Mordell, *Diophantine Equations*, Academic Press, London, 1970.
7. C. Tweedie, *Proc. Edinb. Math. Soc.*, 24 (1905–6), 7–19.

Wheels on Wheels on Wheels —Surprising Symmetry

FRANK A. FARRIS
Santa Clara University
Santa Clara, CA 95053

While designing a computer laboratory exercise for my calculus students, I happened to sketch the curve defined by this vector equation:

$$(x, y) = (\cos(t), \sin(t)) + \frac{1}{2}(\cos(7t), \sin(7t)) + \frac{1}{3}(\sin(17t), \cos(17t)).$$

I was thinking of the curve traced by a particle on a wheel mounted on a wheel mounted on a wheel, each turning at a different rate. The first term represents the largest wheel, of radius 1, turning counter-clockwise at one radian per second. The second term represents a smaller wheel centered at the edge of the first, turning 7 times as fast. The third term is for the smallest wheel centered at the edge of the second, turning 17 times as fast as the first, clockwise and out of phase. See FIGURE 1. As you can notice from FIGURE 2, this curve displays a 6-fold symmetry, a fact that one would probably not guess by looking at the formulas.

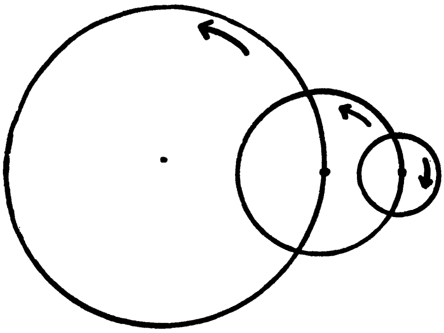


FIGURE 1

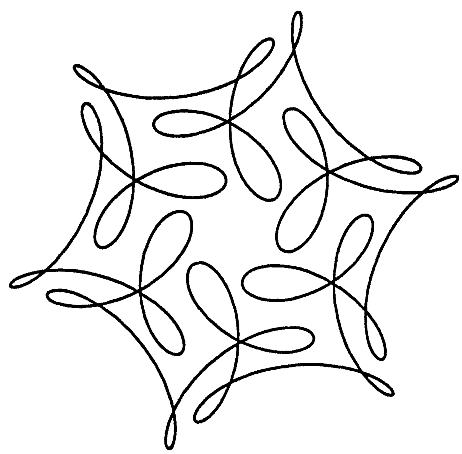


FIGURE 2

The symmetry of this curve, and a condition for the symmetry of any continuous curve, is illuminated by introducing complex notation, in terms of which the formulas above represent a terminating Fourier series. This connection enables the statement and proof of a general theorem relating the symmetry of a parametric curve to the frequencies present in its Fourier series.

1. Analysis of Examples In complex notation, the curve above is written

$$f(t) = x(t) + iy(t) = e^{it} + \frac{1}{2}e^{7it} + \frac{i}{3}e^{-17it}.$$

The source of the symmetry turns out to be this: 1, 7, and -17 are all congruent to 1 modulo 6. When t is advanced by one-sixth of 2π, each wheel has completed some number of complete turns, plus one-sixth of an additional turn, resulting in symmetry. Examine what happens when time is advanced by one-sixth of a period for a representative wheel:

$$e^{(6j+1)i(t+2\pi/6)} = e^{(6j+1)it}\,e^{2\pi i/6}.$$

This wheel is back where it started, but rotated one-sixth of the way around. When each wheel in the superposition has this same behavior, symmetry will result.

A similar result is obtained using any integer m instead of 6, and keeping track of all the wheels at once.

If

$$f(t) = \sum a_j e^{n_j it} \text{ with } n_j = b_j m + 1,$$

then

$$f\left(t + \frac{2\pi}{m}\right) = \sum a_j e^{n_j i(t+2\pi/m)} = \sum a_j e^{n_j it}\,e^{n_j i 2\pi/m} = e^{2\pi i/m} f(t),$$

has m-fold symmetry. Imagine the trigonometric identities we have avoided by using complex rather than real notation! Notice that an infinite sum would behave the same way as long as it converges for each value of t.

That this is not the end of the story can be seen from another example:

$$f(t) = e^{-2it} + \frac{e^{5it}}{2} + \frac{e^{19it}}{4}.$$

The 7-fold symmetry apparent in FIGURE 3 arises because -2, 5, and 19 are all congruent to 5 modulo 7. Here there is an additional, perhaps less interesting, symmetry; since the coefficients are real, $f(-t)$ is the complex conjugate of $f(t)$. The analog of the computation above for a single wheel is now:

$$e^{(7j+5)i(t+2\pi/7)} = e^{(7j+5)it}\,e^{(5\cdot 2\pi i/7)}.$$

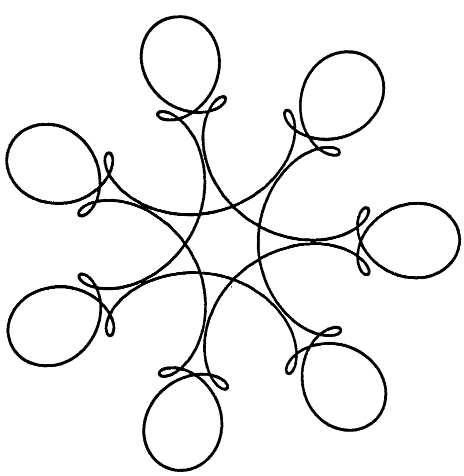


FIGURE 3

Here the wheel has rotated five-sevenths of a full rotation, after time has advanced one-seventh of the fundamental period 2π. If we superimpose several wheels with the same behavior, symmetry will result.

Imagine the plane divided into sectors of size $2\pi/7$, numbered $0, 1, \dots 6$. After completing its first loop in sector 0, the curve moves on to trace the same pattern in sector 5, then in sector 10, which is more simply called sector 3, and so on. Proceeding by multiples of 5, reduced modulo 7, the curve fills in every sector.

This motivates a definition. We say that a function $f(t)$ exhibits *m-fold symmetry* if, for some integer k we have

$$f\left(t + \frac{2\pi}{m}\right) = e^{k2\pi i/m} f(t). \tag{1}$$

It also makes sense to require that k be prime modulo m, for if k times j were congruent to zero mod m, applying (1) j times would give:

$$f\left(t + j\frac{2\pi}{m}\right) = f(t).$$

This would be considered a periodicity of the function rather than symmetry. For instance, suppose we seek a 6-fold symmetry from a wheel with exponent $(2it)$. Advancing time by a sixth of the fundamental period gives:

$$e^{2i(t+2\pi/6)} = e^{2it} e^{2\cdot 2\pi/6}.$$

Numbering six sectors from 0 to 5, we find that the wheel traces its pattern in sectors 0, 2, and 4, returning to sector 0 without ever tracing in sectors 1, 3, and 5. This would be a 3-fold, rather than a 6-fold, symmetry.

The equations above show that m-fold symmetry occurs in a sum of this type when all the frequencies are congruent modulo m to the same number k, which must be a prime modulo m.

Knowing this, it is amusing to experiment by assembling terms to produce a curve of given symmetry.* For FIGURE 4, I chose frequencies $2, -16, -7, 29$ (all congruent to $2 \bmod 9$) to produce a curve with 9-fold symmetry; I then adjusted the coefficients to make the pattern more pleasing. They are 1, $i/2$, $1/5$, and $i/5$.

2. Symmetry and Fourier Series The discussion of examples virtually proves one direction of the following theorem. We choose $C^0[0, 2\pi]$ as a simple setting for proving the following theorem.

THEOREM. *If a continuous function f is not identically zero then f has m-fold symmetry, in the sense of satisfying* (1), *if, and only if, the nonzero coefficients in the Fourier series for f,*

$$f(t) \sim \sum_{n=-\infty}^{\infty} a_n e^{nit},$$

correspond to frequencies, n, which are all congruent to the same prime modulo m.

*Dr. Erich Neuwirth, of the University of Vienna, has kindly prepared a spreadsheet (Microsoft Excel 5.0) for experimenting with curves of this type. It can be downloaded from http://www.smc.univie.ac.at/~neuwirth/wheels.

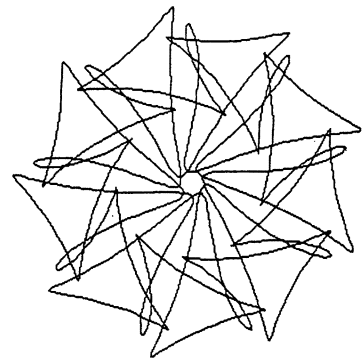


FIGURE 4

Proof. Since the Fourier series for f converges pointwise everywhere [**1**], the discussion above shows that the sum of a series whose frequencies are all congruent to the same prime modulo m does exhibit m-fold symmetry. It remains to prove that functions with the given symmetry do indeed have Fourier series of the type discussed above, with the frequencies all congruent modulo m to a prime modulo m.

Assume that a continuous function f has the symmetry defined in equation (1), with k a prime modulo m. In the integral formula for the Fourier coefficients for f, we will break up the integral into a sum of m integrals:

$$2\pi a_n = \int_0^{2\pi} f(t)\, e^{-int}\, dt = \sum_{j=0}^{m-1} \int_{j2\pi/m}^{(j+1)2\pi/m} f(t)\, e^{-int}\, dt.$$

We make the change of variables, $u = t - j2\pi/m$ to make all the limits of integration range from 0 to $2\pi/m$. In term j of the resulting sum of integrals, we use (1) j times, obtaining:

$$\int_0^{2\pi/m} f\left(u + j\frac{2\pi}{m}\right) e^{-inu}\, e^{-inj2\pi/m}\, du = \int_0^{2\pi/m} f(u)\, e^{-inu}\, e^{ji(k-n)2\pi/m}\, du.$$

The m integrals are now identical and may be factored out. We find:

$$2\pi a_n = \Pi \int_0^{2\pi/m} f(u)\, e^{-inu}\, du \sum_{j=0}^{m-1} e^{ji(k-n)2\pi/m}.$$

The sum can be rewritten as

$$\sum_{j=0}^{m-1} \left(\omega^{(k-n)}\right)^j,$$

where ω is a primitive m-th root of unity. Such a sum is zero unless all the terms are one, in which case m divides $k - n$ and n is congruent to k modulo m. Thus the only frequencies with nonzero coefficients are those congruent to k modulo m.

3. Pedagogical Directions My intent in constructing the original example was to interest students in some pleasing curves that would be too difficult to sketch by hand. Showing students a few examples and assigning them the problem of discovering criteria for symmetry would make an interesting calculus project, serving as an effective advertisement for the power of complex notation.

Other questions present themselves. For any prime m, call

$$F_{k,m} = \{f \varepsilon L^2(0, 2\pi) | a_n = 0 \text{ unless } n \equiv k \bmod m\}.$$

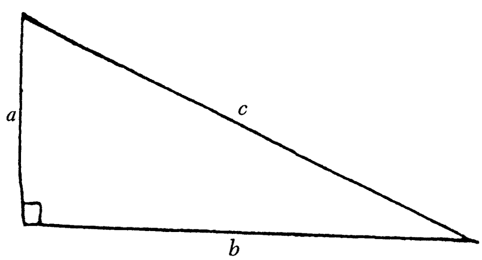
For each m, these spaces give an orthogonal decomposition of L^2 that generalizes the writing of a function as the sum of even and odd parts. Operators projecting onto such subspaces are similar to the Hardy projector. These would provide interesting examples for students first encountering Hilbert spaces.

REFERENCE

1. H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, 1972.

Proof Without Words: Parametric Representation of Primitive Pythagorean Triples

$$\frac{a}{2}, b, c \in Z^+, (a, b) = 1$$



$$\frac{c+b}{a} = \frac{n}{m}, (n, m) = 1 \Rightarrow \frac{c-b}{a} = \frac{m}{n}$$

$$\Rightarrow \frac{c}{a} = \frac{n^2+m^2}{2nm}, \frac{b}{a} = \frac{n^2-m^2}{2nm}$$

$$\Rightarrow n \not\equiv m (\bmod 2)$$

$$\therefore (a, b, c) = (2nm, n^2 - m^2, n^2 + m^2)$$

—RAYMOND A. BEAUREGARD
E. R. SURYANARAYAN
UNIVERSITY OF RHODE ISLAND
KINGSTON, RI 02881

Note: For details and related results, see the authors' article "Pythagorean Triples: The Hyperbolic View," *College Math. J.*, May 1996.

Other questions present themselves. For any prime m, call

$$F_{k,m} = \{f \varepsilon L^2(0,2\pi) | a_n = 0 \text{ unless } n \equiv k \bmod m\}.$$

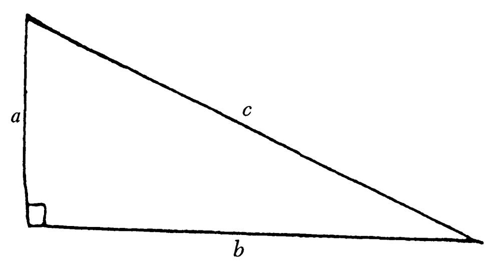
For each m, these spaces give an orthogonal decomposition of L^2 that generalizes the writing of a function as the sum of even and odd parts. Operators projecting onto such subspaces are similar to the Hardy projector. These would provide interesting examples for students first encountering Hilbert spaces.

REFERENCE

1. H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, 1972.

Proof Without Words: Parametric Representation of Primitive Pythagorean Triples

$$\frac{a}{2}, b, c \in Z^+, (a,b) = 1$$



$$\frac{c+b}{a} = \frac{n}{m}, (n,m) = 1 \Rightarrow \frac{c-b}{a} = \frac{m}{n}$$

$$\Rightarrow \frac{c}{a} = \frac{n^2+m^2}{2nm}, \frac{b}{a} = \frac{n^2-m^2}{2nm}$$

$$\Rightarrow n \not\equiv m (\bmod 2)$$

$$\therefore (a,b,c) = (2nm, n^2-m^2, n^2+m^2)$$

—Raymond A. Beauregard
E. R. Suryanarayan
University of Rhode Island
Kingston, RI 02881

Note: For details and related results, see the authors' article "Pythagorean Triples: The Hyperbolic View," *College Math. J.*, May 1996.

On the Asymptotic Solution of a Card-Matching Problem

FINN F. KNUDSEN
Norwegian Institute of Technology
University of Trondheim
N-7034 Trondheim, Norway

IVAR SKAU
Telemark College
N-3800 Bø Norway

Introduction Recently S. G. Penrice **[8]** has shown how an asymptotic result concerning the number of *derangements* in a classical card-matching problem can be obtained by using two "hard" celebrated inequalities, known as Minc's conjecture and Van der Waerden's conjecture. Both were verified years after they were proposed. It will be demonstrated here that this result in fact follows from *easy* inequalities and a straightforward application of the principle of inclusion and exclusion. We show this by proving a more general result. However, as will be explained, the inequalities for the derangement probability established by Penrice yield an asymptotic result that otherwise seems difficult to verify.

For a standard deck of cards, with $k=4$ suits and $n=13$ denominations in each suit, the problem is equivalent to a card game played in Las Vegas: The cards of a shuffled deck are dealt one at a time and face up. At the same time the player calls the denominations in the order ace, two, three, ..., queen, king; ace, two, ... A **match** occurs when the player calls the same denomination as the card dealt; the suits need not match. The player wins if no match occurs. If one continues the comparison through the whole deck, the number of matches X is a random variable on the sample space $S_{13\times 4}$, which consists of all possible shufflings of the cards.

What can we say, to begin, about the derangement probability $p_0 = P(X=0)$, which is the probability that the player wins? Using the Penrice inequalities (6) below,

$$0.0156 \le p_0 \le 0.0190.$$

Playing the game 10^7 times on a computer (using a program written for us by Trym Staal Eggen) gave the statistical estimate

$$\hat{p}_0 \pm SD(\hat{p}_0) = 0.01624 \pm 0.00004.$$

This sample of 10^7 is but a tiny drop compared to the immense ocean of all 52! permutations. But the program has obviously given us a fairly representative drop, since

$$p_0 = 0.0162327275\left(\lessapprox e^{-4} = \lim_n p_0\right).$$

According to **[12]**, this value has been achieved by R. St. E. Johns (in an unpublished solution) from an elaborate recurrence formula.

The extended problem Consider two decks of cards, each with a total of nk consecutively numbered cards $\{1,2,\dots,nk\}$. The first deck has n kinds of cards, k of each kind; with proper ordering of the cards, say $T_i = \{(i-1)k+1, \dots, ik\}$, $i = 1, 2, \dots, n$. The second deck contains the same n kinds of cards, but only $l(\le k)$ of each kind, say $U_i = \{(i-1)k+1, \dots, (i-1)k+l\}$, $i = 1, 2, \dots, n$. In addition, the second deck contains $(k-l)n$ supplementary blank (not match-giving) cards. The decks are shuffled separately, and the orderings of the decks are compared. A **match** occurs when two cards of the same kind occupy the same position in their respective decks. Because relative, rather than absolute, positions of the cards are important, the first deck may be put in standard order and matched against all possible orders of the second. Each of these permutations (events) may be uniquely represented on a *square board*—with the elements to be permuted as column heads and the positions as row heads (or vice versa)—as n dots or *non-taking rooks* placed in the n^2 cells (positions) on the board, *such that no two are in the same row or column*. Further, by putting a cross at each of the match-giving "restricted" positions, we get a helpful illustration of the problem and may minimize the formalism. The "restriction" board for our problem is illustrated in FIGURE 1, with $(n, k, l) = (4, 5, 3)$. The dots mark the permutation $\sigma = (20, 18, 5, 17, 8, 7, \dots, 19, 13)$ with matches in rows 5 and 7.

We shall be concerned with the probability distribution of the "hit" variable X, i.e., the random variable that counts the number of matches in such a card-matching experiment.

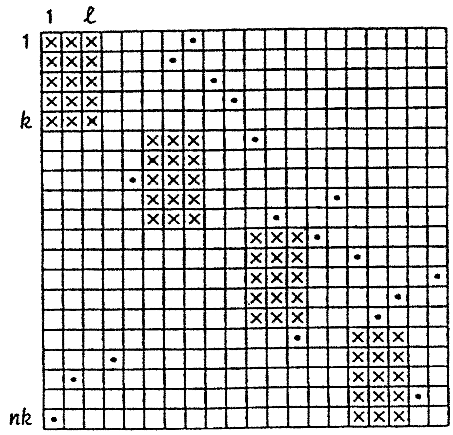


FIGURE 1

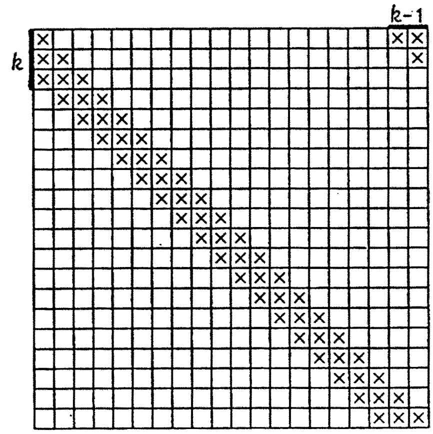


FIGURE 2

As an historical example, let us return to the previous description of the Las Vegas game, which corresponds to $(n, k, l) = (13, 4, 4)$, and stop the comparison after the first series of denomination; ace, two, ..., queen, king. What we have now is the "game of thirteen" (jeu du treize), proposed by Pierre Rémont De Montmort in 1708. By introducing blanks, this game can be fit into our framework as $(n, k, l) = (13, 4, 1)$; consider the denomination ace, blank, blank, blank, two, blank, ..., king, blank, blank, blank. The derangement $(X = 0)$ case of the game of thirteen and its generalization $(n, k, 1)$ were studied, solved, and communicated in letters between Montmort and the Bernoullis in 1710–11. For a detailed historical account, see **[11]**.

The principle of inclusion and exclusion To find the probability $P(X = x)$, we have to count the number of permutations with exactly x matches, E_x; then $P(X = x) = E_x/(nk)!$. The idea is to express E_x by the easily accessible *unfiltered counts*; B_i, $i = x, x+1, \dots$. To define these "overcounting" numbers; fix *i restricted* *positions*, no two in the same row or column. Obviously $(kn-i)!$ is the number of permutations having these i matching positions *irrespective* of the other $(kn-i)$ positions. B_i counts these $(kn-i)!$ permutations for *all* such samples of i restricted positions. In other words: letting R_i denote the number of ways of putting i non-taking rooks (i.e., no two in the same row or column) on restricted positions, then $B_i = R_i(nk-i)!$.

(To unravel the more general considerations below, the uninitiated reader should consider the ("hat check") $k=l=1$ case, and show that $B_i = \binom{n}{i}(n-i)!$. Next, for the generalized game of thirteen $(n,k,1)$, first choose i of the n columns with restrictions, then one cross in each of these i columns, and show that $B_i = \binom{n}{i}/k^i(nk-i)!$.)

Now, consider the permutation in FIGURE 1: This is counted *once* by $B_0(=(nk)!)$, *twice* by B_1, *once* by B_2, and of course does not contribute to B_i for $i>2$. More generally, each permutation with exactly m matches (restricted positions) is counted $\binom{m}{i}$ times by B_i since this is the number of ways to fix i of these m restricted positions. Hence we get the easy relationship between the exact and unfiltered counts:

$$B_i = \sum_{m=i}^{nl} \binom{m}{i} E_m, \qquad i = 0,1,2,\ldots,nl.$$

Note that the nonzero elements of the (upper triangular) matrix of the linear system above is a section of Pascal's triangle (rotated). The inverse matrix consists of the same elements equipped with alternating signs, i.e.,

$$\sum_{i=r}^{m} \binom{m}{i}(-1)^{i-r}\binom{i}{r} = \delta_{m,r}, \quad \text{where } \delta_{m,m}=1 \text{ and } \delta_{m,r}=0 \text{ if } m \neq r.$$

This is (in transposed form) the so-called Inverse Binomial Theorem. For a quick verification, note that

$$t^m = (1+(t-1))^m = \sum_{i=0}^{m}\binom{m}{i}(t-1)^i = \sum_{r=0}^{m}\left(\sum_{i=r}^{m}\binom{m}{i}(-1)^{i-r}\binom{i}{r}\right)t^r = \delta_{m,r}t^r,$$

where the last equality follows by identification of coefficients. Thus we get what the principle of inclusion-exclusion (sieve method) demands:

$$E_r = \sum_{i=r}^{nl}(-1)^{i-r}\binom{i}{r}B_i, \qquad r = 0,1,2,\ldots,nl.$$

(Alternative proofs can be found in any combinatorics textbook; see, e.g., **[6]**.)

Since $P(X=x) = E_x/(nk)!$, this gives

$$P(X=x) = \sum_{i=0}^{nl-x}(-1)^i\binom{i+x}{x}\frac{B_{i+x}}{(nk)!}. \tag{1}$$

Note that $B_j/(nk)! = R_j/(nk)_j$, where we use the notation $(s)_t = s(s-1)\ldots(s-t+1)$ for the falling factorial. Thus our main result depends on estimates of R_i.

To count R_i, note first that i_j elements (rows) in T_j can be chosen and simultaneously "occupied" with elements from U_j in exactly $\binom{k}{i_j}(l)_{i_j}$ ways, since only one element is permitted in each row or column. To fix i matches we therefore choose

$i_j \geq 0$ elements (rows) in T_j, $j = 1, 2, \ldots, n$, such that $i_1 + i_2 + \ldots + i_n = i$, and "occupy" them with U-elements from their respective families. This corresponds to placing i_j non-taking rooks on the j^{th} rectangular block of restrictions, $j = 1, 2, \ldots, n$. The number of ways to do this is then

$$R_i = \sum_{i_1 + \ldots + i_n = i} \prod_{j=1}^{n} \binom{k}{i_j} (l)_{i_j}. \tag{2}$$

Note: For the generalized game of thirteen $(n, k, 1)$, (1) gives

$$\begin{aligned} P(X=x) &= \sum_{i=0}^{n-x} (-1) \binom{i+x}{x} \binom{n}{i+x} \frac{k^{i+x}}{(nk)_{i+x}} \\ &= \frac{1}{x!} \sum_{i=0}^{n-x} \frac{(-1)^i}{i!} \frac{k^{i+x}(n)_{i+x}}{(nk)_{i+x}} \to \frac{e^{-1}}{x!} \quad \text{as } n \to \infty. \end{aligned} \tag{3}$$

For $k = 1$, (3) reduces to the well-known solution for the "hat check" problem:

$$P(X=x) = \frac{1}{x!} \sum_{i=0}^{n-x} \frac{(-1)^i}{i!} \to \frac{e^{-1}}{x!} \quad \text{as } n \to \infty.$$

These two problems have the same Poisson limiting distribution with parameter $\lambda = 1$. This is a special case of the theorem to be proved in the next section.

The limit distribution To find the limit distribution ($n \to \infty$) for the general (n, k, l) case, we will use the following inequalities:

LEMMA.

$$\binom{n}{i} l^i k^i \leq R_i \leq l^i \binom{nk}{i}. \tag{4}$$

Proof. There are $\binom{n}{i}$ terms in (2) where exactly i i_j's equal 1; their sum equals the left-hand side of (4). As for the right-hand side, we note that $(l)_{i_j} \leq l^{i_j}$, and hence

$$R_i \leq l^i \sum_{i_j + \ldots + i_n = i} \prod_{j=1}^{n} \binom{k}{i_j} = l^i \binom{nk}{i}.$$

The last equality is left to the reader to see combinatorially, to prove either by induction (Hint: Recall that $\Sigma_{i_1 + i_2 = i} \binom{M}{i_1} \binom{N}{i_2} = \binom{M+N}{i}$), or by expanding an appropriate polynomial equation.

THEOREM. *The limit distribution of the number of matches is Poisson with parameter $\lambda = l$, i.e.,*

$$\lim_{n \to \infty} P(X=x) = \frac{l^x}{x!} e^{-l}, \qquad x = 0, 1, 2, \ldots.$$

(*Note the independence of k.*)

Proof. Since $B_i/(nk)! = R_i/(nk)_i$, (4) gives

$$\frac{B_0}{(nk)!} = 1, \quad \frac{B_1}{(nk)!} = l, \quad \text{and} \quad \frac{l^i}{i!}\frac{(n-1)\dots(n-i+1)}{\left(n-\frac{1}{k}\right)\dots\left(n-\frac{i-1}{k}\right)} \leq \frac{B_i}{(nk)!} \leq \frac{l^i}{i!}, \quad i \geq 2. \tag{5}$$

This tells us that $B_i/(nk)! = l^i/i!a(n,i)$, where $a(n,i) \to 1$ as $n \to \infty$ and $0 \leq a(n,i) \leq 1$. Combined with (1), this gives

$$P(X=x) = \sum_{i=0}^{nl-x} (-1)^i \binom{i+x}{x} \frac{B_{i+x}}{(nk)!} = \frac{l^x}{x!} \sum_{i=0}^{nl-x} (-1)^i \frac{l^i}{i!} a(n,i+x)$$
$$\to \frac{l^x}{x!} e^{-l} \quad \text{as } n \to \infty,$$

where the limit is justified by the Dominated Convergence Theorem. (A more detailed examination shows that

$$P(X=x) = \frac{l^x}{x!} e^{-l}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \text{ when } k > 1.)$$

As the following table (calculations from [5]) for the "Las Vegas distribution" $(n,k,l) = (13,4,4)$ reveals, the Poisson distribution is already apparent for moderate values of n:

	x	0	1	2	3	4	5	6	7
Las Vegas	$P(X=x)$	.0162	.0689	.1442	.1982	.2013	.1613	.1052	.060
Poisson	$\frac{4^x}{x!}e^{-4}$	.0183	.0733	.1465	.1954	.1954	.1563	.1042	.0595

The benefits of Penrice's inequalities To describe Penrice's method, we set $l = k$ and consider FIGURE 1 as a $kn \times kn$ matrix, $A_{k,n}$, with 0's in each cell in the $k \times k$ blocks of restricted positions and 1's elsewhere. The product of the numbers in the positions of a permutation is thus 0 if it has a matching (restricted) position, and 1 if it has none. Thus the *sum* of all such products gives us the number of derangements of the problem. This is the sum of the terms in the determinant, *all taken with positive sign*, and is called the **permanent** of the actual matrix. This is more generally defined for an $m \times m$ matrix $A = (a_{ij})$ by

$$\text{per}(A) = \sum_{\pi} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{m\pi(m)},$$

where Σ runs over all permutations π of $\{1, 2, \dots, m\}$.

Since the permanent is not invariant under row and column operations, it is hard to evaluate. However, we see easily that it is multilinear in the rows and columns. This fact, together with the very special nature of our *derangement* permanents, makes them amenable to being attacked—from both sides—by two famous inequalities:

MINC'S CONJECTURE. *Let $A = (a_{ij})$ be an $m \times m$ (0,1)-matrix (i.e., $a_{ij} = 0$ or 1) with nonzero row sums $r_1, \dots, r_m$. Then*

$$\text{per}(A) \leq (r_1!)^{1/r_1} \dots (r_m!)^{1/r_m}.$$

This inequality was proved in 1973 [2], 10 years after it was proposed in [7].

VAN DER WAERDEN'S CONJECTURE. *Let $A = (a_{ij})$ be a doubly stochastic $m \times m$ matrix, i.e., all $a_{ij} \geq 0$ and all row sums and column sums are equal to 1. Then*

$$\operatorname{per}(A) \geq \frac{m!}{m^m},$$

with equality if and only if all $a_{ij} = 1/m$.

This inequality was proved in 1981, [3] and [4], 55 years after it was proposed in [13].

Now, note that $A_{k,n}/(kn-k)$ is a doubly stochastic matrix. Hence, by the multilinearity and Van der Waerden's inequality,

$$\operatorname{per}(A_{k,n}) = (kn-k)^{kn} \operatorname{per}\left(\frac{1}{kn-k} A_{k,n}\right) \geq (kn-k)^{kn} \frac{(kn)!}{(kn)^{kn}} = (kn)!\left(1-\frac{1}{n}\right)^{kn}.$$

Applying Minc's inequality on the other hand and dividing by $(kn)!$, Penrice thus squeezes the derangement probability by

$$\left(1-\frac{1}{n}\right)^{kn} \leq P(X=0) \leq \frac{(kn-k)!^{kn/(kn-k)}}{(kn)!}$$
$$\sim \left(1-\frac{1}{n}\right)^{kn+1/2} (kn-k)^{1/2(n-1)} \qquad (n \to \infty) \tag{6}$$

where on the right-hand side we have used Stirling's formula.

To see the benefits of these inequalities, let k be "movable," i.e. $k = k(n)$. From (6) we now get

$$P(X=0) \sim e^{-k-1/2(k-\mathcal{O}(\ln k)/n)-\mathcal{O}(k/n^2)} \qquad (n \to \infty)$$

and hence

$$P(X=0) \sim e^{-k} \qquad \text{if, and only if,} \qquad k = o(n). \tag{7}$$

This suggests that $k = n$ represents the actual break point for the distribution to be *asymptotically* Poisson, $X \overset{a}{\sim} P_o(\lambda)$, in the sense that $P(X=x)/(\lambda^x/x!)e^{-\lambda} \to 1$ as $n \to \infty$ for $x = 0, 1, 2, \ldots$, where $\lambda = \lambda(k,n)(= \lambda(n))$ is independent of x. At any rate, (7) tells us that the distribution at least must *have ceased* to be asymptotically Poisson with parameter $\lambda = k(= EX)$ at this point. This seems otherwise difficult to verify.

In [9] (Ex. 7, page 177–8) the *complete* distribution of the number X of matches in the general n, k-case has been studied. In this detailed and rather complicated analysis the author suggests that a Poisson distribution be adapted through an expansion (in powers of $(nk)^{-1}$). No instructions are given as to the control/estimate of the "tail" in this expansion. Therefore it does not seem to follow immediately from this treatment that the limit $(n \to \infty)$ is a Poisson distribution. However, the first (three) terms calculated in this proposed expansion might suggest that the threshold for $X \overset{a}{\sim} P_o(\lambda)$ is represented by $k = n$; i.e., if *only* these terms were taken into consideration, we would get $X \overset{a}{\sim} P_o(\lambda)$ if, and only if, $k = o(n)$. However, we have not been able to determine whether the general term in this rather complicated expansion displays the same pattern.

Note that Penrice's argument also applies to the generalized problem $(l \leq k)$: The *derangement* probability is now equal to the permanent of the matrix resulting from the restriction board (FIGURE 1), with zeros in the rectangular blocks of "restricted" positions and 1's in the other entries. This matrix may be converted to a doubly stochastic matrix by multiplying columns containing zeros by $1/(kn-k)$ and the other columns by $1/kn$. A modification of his proof yields in this case, by applying Minc's inequality (on the right side) to the *transposed* matrix,

$$\left(1-\frac{1}{n}\right)^{ln} \le P(X=0) \le \frac{((kn-k)!)^{ln/(kn-k)}((kn)!)^{n(k-l)/kn}}{(kn)!}$$

$$\sim \left(1-\frac{1}{n}\right)^{ln} (kn)^{l/2(l/k(n-1))}$$

from which it follows that

$$P(X=0) \sim e^{-l-1/2(l/n)(1-\mathcal{O}(\ln k/k))-\mathcal{O}(l/n^2)}.$$

Hence

$$P(X=0) \sim e^{-l} \quad \text{if and only if } l = o(n).$$

Exercise A k-ply generalization of the (reduced) "problème des ménages" ($k=2$ gives the famous classical "ménages" problem; see **[9]**, Chapter 8, where the case $k=3$ is treated in problems 24–28). Suppose n persons are seated at a round table. Then they randomly change places. What is the "derangement" probability, $P(X=0)$, that *no* person is reseated at any one of the k places *at* and *to the right* of his original place?

(a) Show that the restriction board for this problem—with natural numbering of persons and places—may be given by Figure 2 (where $k=3$ for $n=20$ persons).

(b) Use Penrice's method to prove that

$$P(X=0) \sim e^{-k} \quad \text{if and only if } k = o(\sqrt{n}).$$

Generalizations The "homogeneity" assumption (invariable l's and k's) is certainly not of fundamental—but chiefly of methodological—nature: When throwing off this "straitjacket," we will have to manage without such simple inequalities as those that gave us the limit theorem. For more on this, see, e.g., **[10]**, which uses a direct elementary method. A treatment based on the Chen-Stein method is given in **[1]**.

REFERENCES

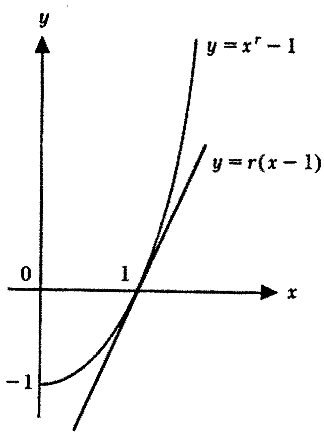
1. A. D. Barbour, Lars Holst and Svante Janson, *Poisson Approximation*, Clarendon Press, Oxford, 1992, pp. 77–80.
2. L. M. Bregman, Certain properties of non-negative matrices and their permanents, *Dokl. Akad. Nauk SSSR* 211(1973), 27–30. Russian. (English translation: *Soviet Math. Dokl.* 14(1973), 945–949.)
3. G. P. Egoritchev, The solution of Van der Waerden's problem for permanents, *Adv. in Math.* 42(1981), 299–305.
4. D. I. Falikman, A proof of the Van der Waerden conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki* 29(1981), 931–938. (Russian.) (English translation: *Math. Notes of the Academy of Science of the USSR* 29(1981), 475–479.)
5. R. E. Greenwood, Probabilities of certain solitaire card games, *Journal of the American Statistical Association* 48(1953), 88–93.
6. R. P. Grimaldi, *Discrete and Combinatorial Mathematics*, Addison-Wesley Publishing Co., Reading, MA, 1989, 307.
7. H. Minc, Upper bounds for permanents of (0, 1)-matrices, *Bull. Amer. Math. Soc.* 69(1963), 789–791.
8. S. G. Penrice, Derangements, permanents and Christmas presents, *The American Mathematical Monthly* 98(1991), 617–620.

9. John Riordan, *An Introduction to Combinatorial Analysis*, John Wiley and Sons, Inc., New York, NY, 1958, 177.
10. I. Skau, Permanents and "hatcheck" problems, some variations on a classical theme, in preparation (Norwegian).
11. Lajos Takács, The problem of coincidences, *Archiv. Hist. Exact Sci.* 21(1980), 229–244.
12. *The Mathematical Gazette*, Vol. 74, March (1990), 64.
13. B. L. Van der Waerden, Aufgabe 45, *Jber. Deutsch. Math. Verein.* 35(1926), 117.

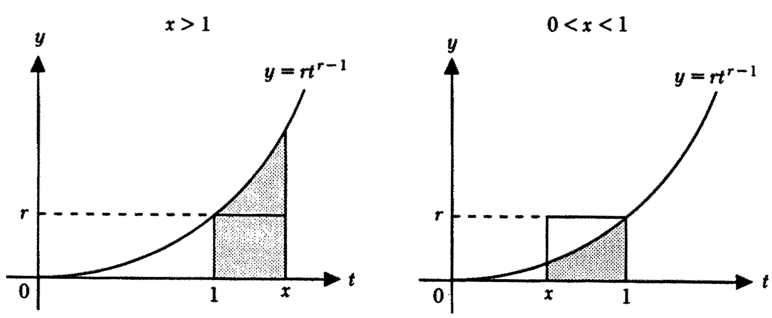
Proof Without Words: Bernoulli's Inequality (two proofs)

$$x > 0,\ x \neq 1,\ r > 1 \Rightarrow x^r - 1 > r(x-1)$$

I. (first-semester calculus)



II. (second-semester calculus)



$x^r - 1 = \int_1^x rt^{r-1}dt > r(x-1)$ $\qquad$ $1 - x^r = \int_x^1 rt^{r-1}dt < r(1-x)$

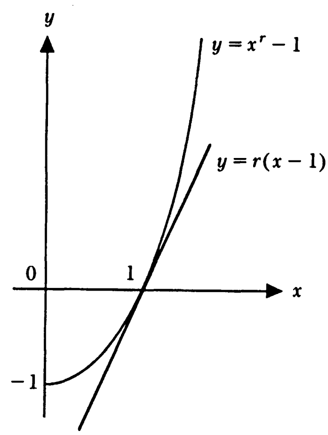
—Roger B. Nelsen
Lewis and Clark College
Portland, OR 92719

9. John Riordan, *An Introduction to Combinatorial Analysis*, John Wiley and Sons, Inc., New York, NY, 1958, 177.
10. I. Skau, Permanents and "hatcheck" problems, some variations on a classical theme, in preparation (Norwegian).
11. Lajos Takács, The problem of coincidences, *Archiv. Hist. Exact Sci.* 21(1980), 229–244.
12. *The Mathematical Gazette*, Vol. 74, March (1990), 64.
13. B. L. Van der Waerden, Aufgabe 45, *Jber. Deutsch. Math. Verein.* 35(1926), 117.

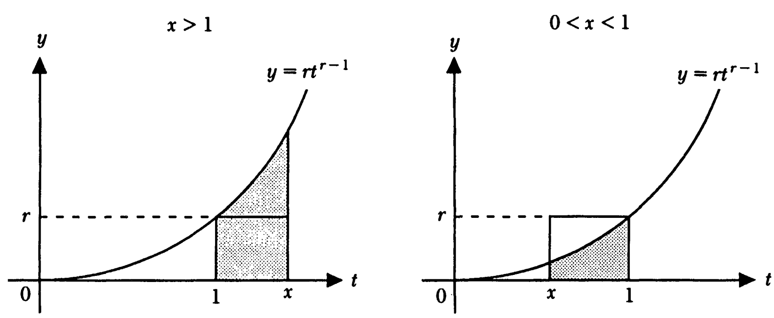
Proof Without Words: Bernoulli's Inequality (two proofs)

$$x > 0,\ x \neq 1,\ r > 1 \Rightarrow x^r - 1 > r(x-1)$$

I. (first-semester calculus)



II. (second-semester calculus)



$x^r - 1 = \int_1^x rt^{r-1}dt > r(x-1)$

$1 - x^r = \int_x^1 rt^{r-1}dt < r(1-x)$

—Roger B. Nelsen
Lewis and Clark College
Portland, OR 92719

The x^x Spindle

MARK D. MEYERSON
U.S. Naval Academy
Annapolis, MD 21402-5002

> *"What is that that twists round so briskly?" asked the maiden, and taking the* **spindle** *into her hand she began to spin.*
> —from **The Sleeping Beauty**
> by the Brothers Grimm

The simple expression x^x leads to a bizarre graph. (See FIGURE 1.) While the graph is smooth for $x > 0$, for $x < 0$ it is only defined if x can be written as $-p/q$ where p and q are positive integers and q is odd. The graph determines a collection of dots and holes for $x < 0$, which lie along and are dense on each of the two curves defined by $y = \pm|x|^x$.

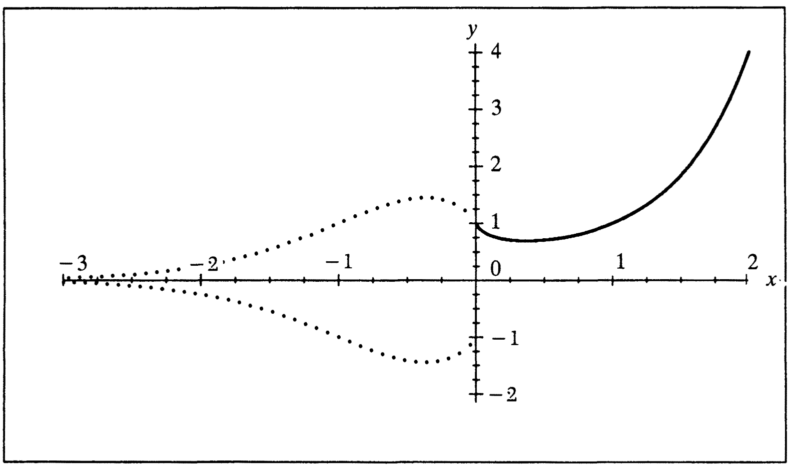


FIGURE 1
The usual graph of $y = x^x$ for x and y real. For $x < 0$, negative integer multiples of $1/25$ are plotted.

We can get a better understanding of x^x by looking at complex values. We'll see that the graph of $z = x^x$, where x is still real but z is allowed to be complex, appears as a *spindle* shape. Part of such a graph appears in FIGURE 2. The word spindle is doubly appropriate; not only is the general shape spindlelike, but the graph consists of a countable infinity of curves or *threads* wrapped around the shape. In the following we investigate this x^x spindle.

1. Why the dots in FIGURE 1 for $x < 0$? The dots in FIGURE 1 arise basically because (working with real numbers) we can take the odd roots of negative numbers

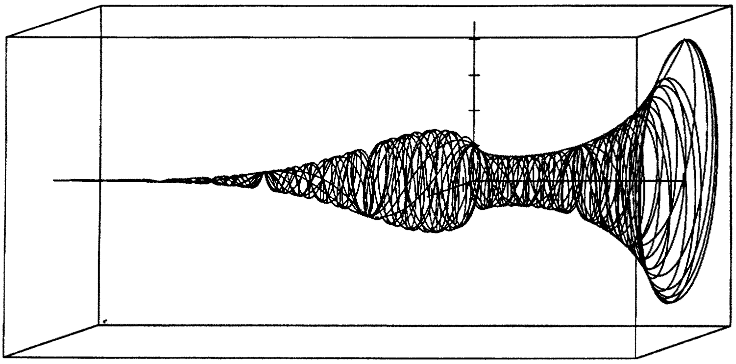


FIGURE 2
The graph of $z = x^x$ for x real and z complex. Twenty-one of the threads on the graph are plotted for x in $[-4\ldots2]$.

but not the even roots. For q odd, $\sqrt[q]{x} = -\sqrt[q]{-x}$, and so

$$\left(-\frac{p}{q}\right)^{-p/q} = \left(-\sqrt[q]{\frac{p}{q}}\right)^{-p} = \pm\left(\frac{p}{q}\right)^{-p/q}$$

where the $\pm$ sign is $-$ if p is odd and $+$ if p is even.

For example,

$$\left(-\frac{3}{5}\right)^{-3/5} = -\sqrt[5]{125/27} \approx -1.3587 \quad \text{and} \quad \left(-\frac{2}{5}\right)^{-2/5} = \sqrt[5]{25/4} \approx 1.4427.$$

Note that there seems to be no "correct" value for x^x when x is negative and irrational.

Further confusion arises from the usual definitions of exponentiation in that when writing

$$x^{p/q} = \left(\sqrt[q]{x}\right)^p = \sqrt[q]{x^p},$$

we need to ensure that p/q is in lowest terms. For otherwise we get paradoxes like

$$-1 = (-1)^{1/1} = (-1)^{2/2} = \sqrt{(-1)^2} = \sqrt{1} = 1?$$

Our definition of exponentiation in terms of logarithms (in the next section) will help eliminate this difficulty.

2. The basic definitions To analyze x^x we write it as is often done to define irrational exponents: namely $x^x = (e^{\log x})^x = e^{x \log x}$, where log is the natural logarithm. Since part of the problem in understanding x^x seems to be taking even roots of negative numbers, it's natural to look at complex numbers (which are often defined by starting with i as the *square* root of -1). The other fundamental fact we need is Euler's relation:

$$e^{i\theta} = \cos\theta + i \sin\theta.$$

This formula can be justified by power series, but we just give a small bit of evidence that it's right. According to Euler's relation we should have $e^{i\pi/2} = \cos(\pi/2) +$

$i \sin(\pi/2) = i$ and $e^{i\pi} = \cos \pi + i \sin \pi = -1$. So then

$$i \cdot i = e^{i\pi/2} \cdot e^{i\pi/2} = e^{i\pi} = -1,$$

which checks. Exponentiation with base e (the exp function) is extended to all complex numbers by the usual exponentiation rules; in particular, $e^{a+ib} = e^a e^{ib} = e^a \cos b + ie^a \sin b$. We want log to be the inverse of exp, but for $z = e^{a+ib}$ we have $z = e^{a+i(b+2n\pi)}$ for any integer n. So $\log z$ must be *multiple valued* and equal to any of $a + i(b + 2n\pi)$.

To distinguish between this complex multiple-valued log and the real single-valued one, we'll use capitalized Log to stand for the real-valued logarithm (also known as the *principal part* of log). Thus, $\text{Log}\, x$ is only defined for $x > 0$ and gives a unique real value. For $x > 0$, we have for n even, $x = e^{\text{Log}\, x} = e^{\text{Log}\, x + i\pi n}$, and so $\log x$ can be $\text{Log}\, x + i\pi n$ for any even integer n. For $x < 0$, $x = |x|(-1) = |x| e^{i\pi} = e^{\text{Log}|x| + i\pi n}$ for n odd, and $\log x$ can be $\text{Log}|x| + i\pi n$ for any odd integer n. It's convenient to include the absolute value in the positive case, too.

(There are no other possible values for $\log z$ since if $c + id$ is a value for $\log z$, then exponentiating gives $e^{a+ib} = e^{c+id}$. Dividing by e^{c+id} gives $e^{a-c} e^{i(b-d)} = 1$. Looking at magnitude gives $e^{a-c} = 1$, so $a = c$, and hence $e^{i(b-d)} = 1$. But then $b - d$ must be an integer multiple of 2π.)

Multiplying by x we find that $x \log x$ will be $x\, \text{Log}|x| + i\pi nx$ where n must be even if $x > 0$ and odd if $x < 0$. And exponentiating, we finally have for x^x the values

$$e^{x\, \text{Log}|x| + i\pi nx} = |x|^x e^{i\pi nx} = |x|^x(\cos(n\pi x) + i \sin(n\pi x))$$

with the same rules for the parity of n depending on the sign of x ($n \bmod 2 = (1 - \text{sgn}(x))/2$).

3. Method of graphing To graph our (multivalued) functions, we'll consider a complex plane perpendicular to each point of the x-axis. This gives us a 3-dimensional $x - u - v$ coordinate system. We'll find it convenient to use the "engineering" placement of axes instead of the usual "mathematical" placement: The positive x- and u-axes are to the right and up, respectively (as they are in 2 dimensions for the graph of $u = x^x$), while the positive v-axis slants forward (and will be foreshortened). The domain of $z = \log x$, $z = x \log x$, and $z = x^x$ will lie in the real x-axis. The range will lie in the complex $u + iv$ plane. We'll sometimes look at graphs as we've described them in 3 dimensions; other times we'll look right down an axis for a 2-dimensional projection (namely a top view down the positive u-axis, a front view down the positive v-axis, or a side view down the positive x-axis). All four views appear in FIGURE 3. Note that the complex plane appears rotated from its usual position: The real (u)-axis is up and the imaginary (v)-axis is out.

The *graph* of $z = \log x$ is precisely the set of points (x, u, v) such that $u + iv$ is one of the multiple values of $\log x$. Similarly, the graphs of $z = x \log x$ and $z = x^x = e^{x \log x}$ are the sets of points (x, xu, xv) and (x, e^{xu}, e^{xv}), respectively, where $u + iv$ is a value of $\log x$.

4. The graph of $z = \log x$ The graph of $z = \log x$ is in FIGURE 3. Note the *separated threads*: For each integer n there is a *thread* (or branch) of the graph, labeled t_n in the 3-dimensional graph, such that the imaginary part of t_n (the v value) is equal to $in\pi$. If we restrict the values of log to t_n we have

$$\log x = \text{Log}|x| + in\pi$$

where if $x > 0$ then n must be even, and if $x < 0$ then n must be odd.

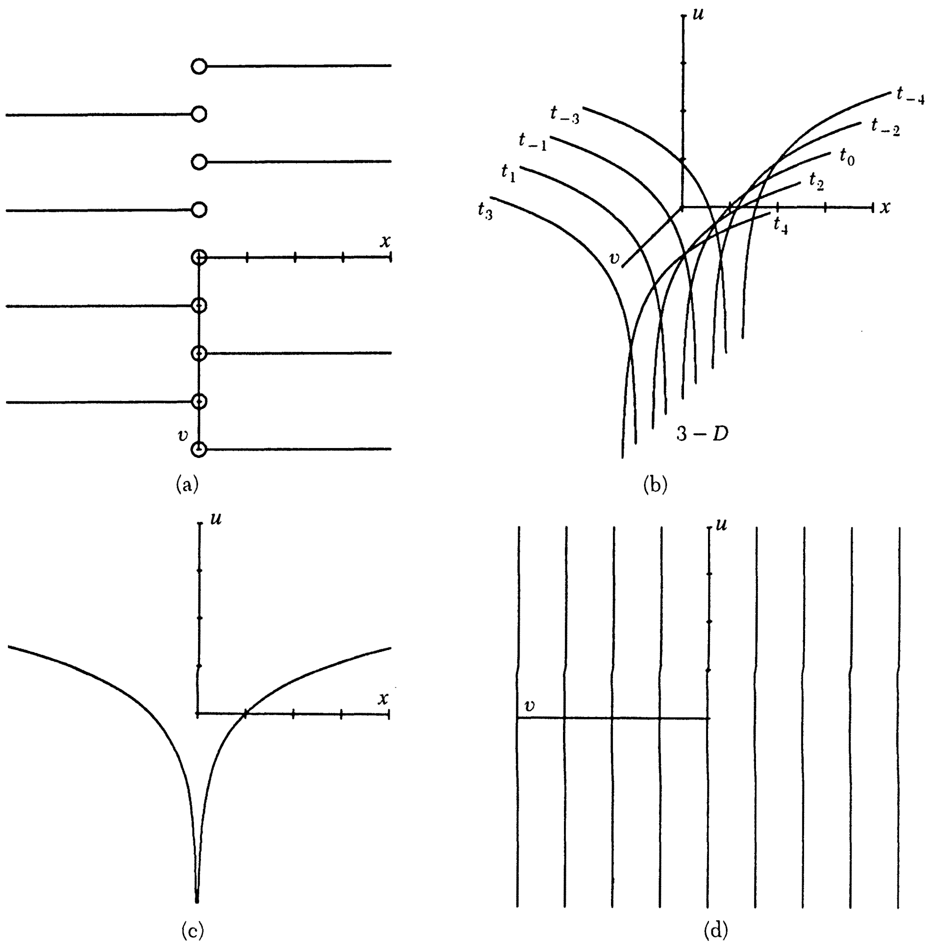


FIGURE 3

The graph of $z = \log x$ for x real and z complex. Parts of threads t_n are plotted, for $n = -4, -3, \ldots, 3, 4$.

5. The graph of $z = x \log x$ For each x value, $x \log x$ takes the values

$$x(\text{Log}\,|x| + in\pi) = x\,\text{Log}\,|x| + in\pi x,$$

where n is any odd integer for $x < 0$ and n is any even integer for $x > 0$. The front view in FIGURE 4 shows the graph of $y = x \log |x|$ for x and y real. Since $u = x\,\text{Log}\,|x|$ and $v = n\pi x$, the top view (which graphs $v = n\pi x$) is a family of half-lines emanating from the origin with slopes equal to integer multiples of π. The 3-D view shows that the graph is spiderlike, with the legs lined up just right for the top and front views. The 'body' of the spider, the origin, is missing from the graphs. As we move from this body in the direction of increasing x, the legs drop for a bit and then go up. We'll let t'_n represent the thread for $z = x \log x$ that corresponds to the thread t_n for $z = \log x$.

6. The graph of $z = x^x$ Now we're ready to look at the graph of x^x. Recall that $x^x = e^{x \log x}$ takes the values $e^{x\,\text{Log}\,|x| + i\pi n x}$. Let's take a few threads, t''_n (corresponding to t_n in $z = \log x$ and t'_n in $z = x \log x$, respectively), at a time.

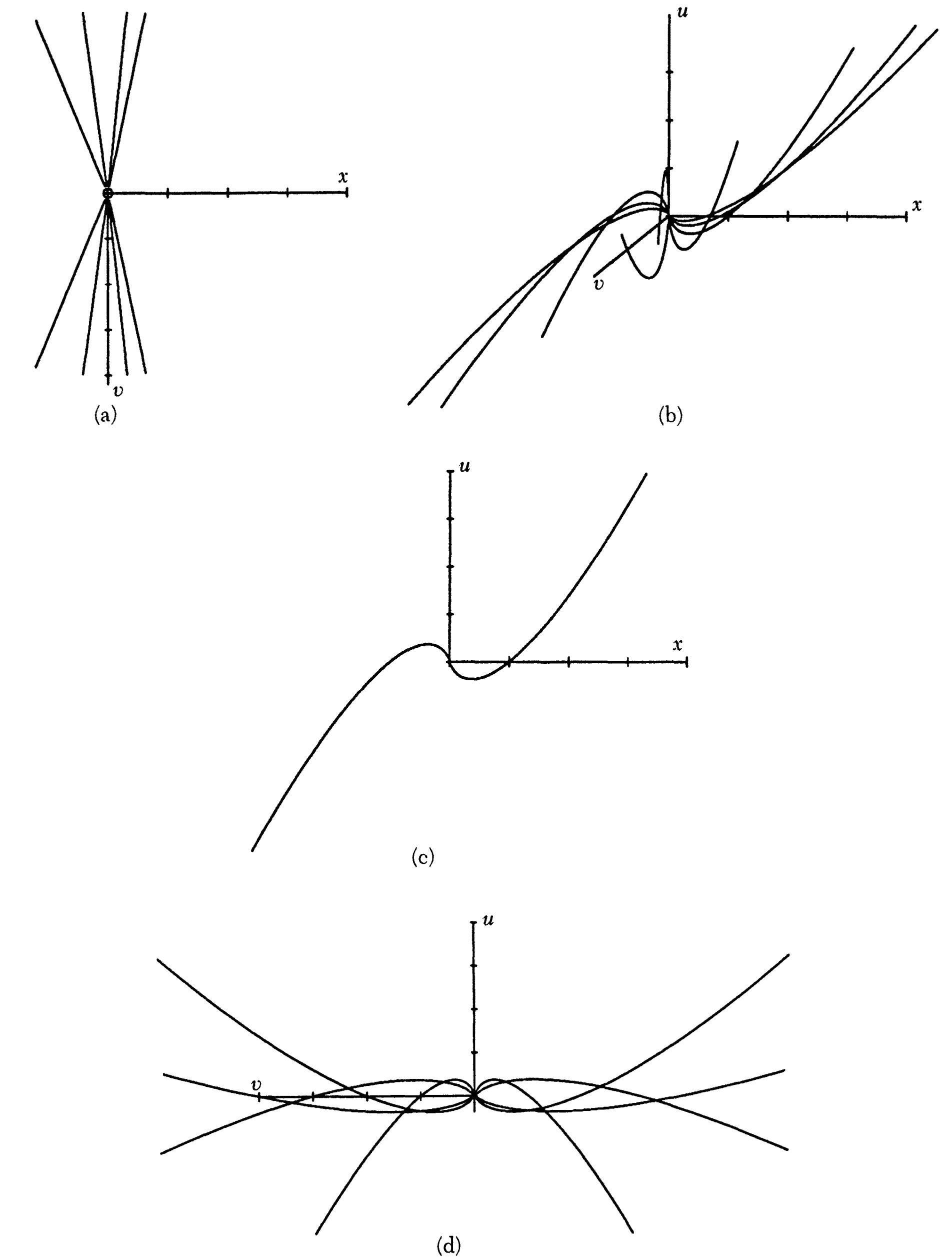


FIGURE 4
The graph of $z = x \log x$ for x real and z complex. Parts of threads t'_n are plotted, for $n = -4, -3, \ldots, 3, 4$.

For $n = 0$, we get a special thread. In the $\log x$ graph, t_0 is the usual graph of the real logarithm function, lying in the vertical plane through the x-axis. So, in the $x \log x$ graph, t'_0 has only real values, and is only defined for $x > 0$—it's the right half of the 'front' view in FIGURE 4. And thus in the $x^x = e^{x \log x}$ graph, t''_0 will take only positive real (u) values and is defined for positive x. It's the usual planar curve for real x^x and $x > 0$.

Next, let's consider $n=2$. In the $\log x$ graph, t_2 is just like t_0 except shifted 2π units in the v (or i) direction—out from the plane of the paper. So in the $x \log x$ graph, t_2' will have the same front view as t_0', but the top view will be the half-line with slope 2π. This ray lies on the line $v = 2\pi u$ and will point down since positive v goes down in the top view. So for t_2', $x \log x = x \operatorname{Log} x + i2\pi x$ and $x > 0$. Finally, in the graph of x^x, t_2'' takes the values $e^{x \operatorname{Log} x} e^{i2\pi x}$ for $x > 0$. As x increases from near 0, the magnitude of x^x along t_2'' is just the usual real value of x^x (as on t_0''). But the angle (or 'argument') of the complex value of x^x increases steadily by 2π radians for each unit increase in x. In other words, t_2'' lies on the surface of revolution formed by rotating t_0'' about the x-axis. Furthermore, t_2'' makes a complete rotation about the axis in every unit of x. Also note that at integer values of x, t_2'' will be real and will intersect t_0''. When viewed from $x = +\infty$, as x increases, t_2'' rotates in a counterclockwise direction about the x-axis.

For all other positive even values of n, t_n'' is similar to t_2''—the only difference is that it completes its revolutions about the x-axis $n/2$ times as fast. So for these n, t_n'' will be positive and real (and meet t_0'') for $x = 2p/n$ where p is any positive integer.

For negative even values of n, the only change is that t_n'' rotates in a clockwise direction when viewed from $+\infty$. Thus t_n'' is a 'mirror-image' of t_{-n}'', reflected through the mirror of the $x-u$ plane (the plane of the paper).

Note that for n even but not 0, t_n'' takes negative real values precisely for $x = (2p-1)/n$ where p is a positive integer. These threads 'pierce' the real $x-u$ plane at negative u values just halfway (measured in the x-direction) between the positive u values. In this way, we get negative even roots of positive reals. For example, $(\frac{1}{2})^{1/2}$ on t_0'' (and t_4'', t_{-8}'', etc.) gives the positive square root of $1/2$, while on t_2'' (and t_6'', t_{-2}'' etc.) it gives $-1/\sqrt{2}$.

For odd n, the main difference is that $x < 0$ and that the magnitude (still given by $|x|^x$—absolute value is now needed) goes to zero instead of $+\infty$ as x moves away from the origin. The threads still wrap around the x-axis in the same way. Consider t_1''. On this thread, $x < 0$ and $x^x = e^{x \operatorname{Log}|x| + i\pi x} = |x|^x e^{\pi i x}$, so it lies on the surface of revolution formed by rotating the real graph of $y = |x|^x$ for $x < 0$ about the x-axis. Also, t_1'' does a complete rotation about the x-axis as x moves through any interval of length 2. Viewed from $x = +\infty$, the movement is counterclockwise as x increases.

The remaining threads are similar to t_1''. For n odd and positive, t_n'' lies on the same surface described above, it just does a complete rotation as x moves through a distance of $2/n$. For n odd and negative, this 'period' is unchanged, but the rotation becomes clockwise.

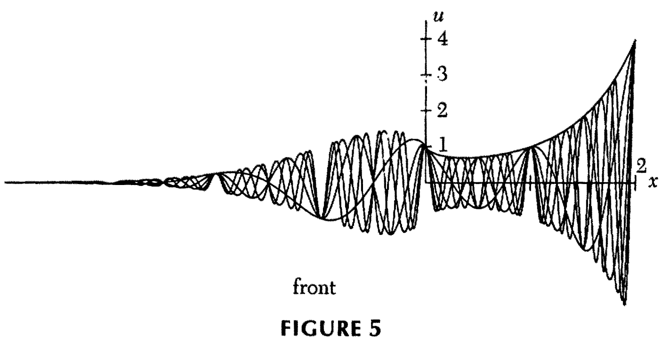


FIGURE 5

Front view of the graph of $z = x^x$ for x real and z complex. Threads t_n'' are plotted, for $n = -10, -9, \ldots, 9, 10$ for x in $[-4 \ldots 2]$.

We've seen a few of these threads in a 3-dimensional view; FIGURE 5 shows the threads from the front. Since t''_n and t''_{-n} have the same real u values for each x value they are indistinguishable in this front view.

FIGURE 6 shows these threads again, but from the top.

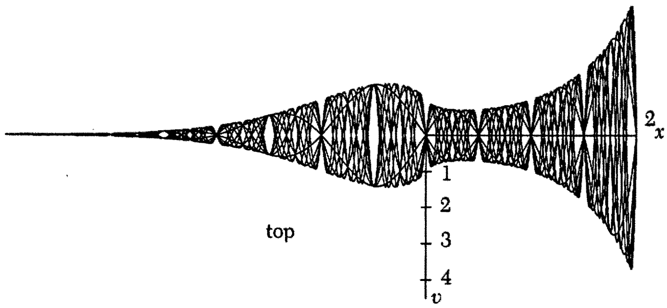


FIGURE 6

Top view of the graph of $z = x^x$ for x real and z complex. Threads t''_n are plotted, for $n = -10, -9, \ldots, 9, 10$ for x in $[-4 \ldots 2]$.

7. Gaps and particular x-values Notice the apparent gaps in FIGURES 5 and 6 at certain x-values. Let's start at $x = 0$. Here, x^x is undefined. But on any thread, as x gets close to 0, the values of $x \log x$ get close to 0 (the body of the spider), and so the values of $x^x = e^{x \log x}$ get close to 1. So all the threads seem to emanate from the point $(0, 1 + 0i)$.

Next, consider $x = 1$. For positive x, we only consider threads t''_n with n even and each of these complete a rotation in $2/n$ units. Thus all these threads pass through $(1, 1 + 0i)$ and there is a single value of x^x for $x = 1$. Similarly for x equal to any positive integer this is true.

More generally, consider a positive fraction, $x = p/q$ in lowest terms. The nth thread (n even) will complete $n/2$ rotations for every unit, so when it gets to x, it will have completed $(n/2) \cdot (p/q)$ rotations. Thus all the "even" threads will pass through the q points one gets by starting at the real position (for $n = 0$) and rotating in the complex plane about the circle of radius $|x|^x$ in steps of p/q of a complete rotation.

For x negative, the threads still emanate from $(0, 1 + 0i)$ and complete rotations in $2/n$ units, but now n is odd. For example, at $x = -1, -3, \ldots, t''_n$ will have completed $(n/2)x$ rotations and so will take the value -1. And for $x = -2, -4, \ldots, t''_n$ will take the value $+1$. For $x = -p/q$ in lowest terms, t''_n will have completed $-(n/2) \cdot (p/q)$ rotations, and so will take one of q equally spaced values.

We will look at this question again, from a slightly different perspective, in the next section.

8. Circular slices and more on gaps For fixed $x \neq 0$, the values of x^x all lie on a circle in the $u - v$ plane of radius $|x|^x$. How are they spaced?

For x rational, we can write $x = \pm p/q$ in lowest terms. Then the nth thread, t''_n, meets this circle at x in the point $e^{in\pi x} = e^{\pm in\pi p/q}$. There exist $2q$ distinct values of $e^{in\pi x}$ as n takes on the values $1, 2, \ldots, 2q$: namely $e^{i\pi p/q}, e^{i\pi 2p/q}, \ldots, e^{i\pi 2q \cdot p/q} = 1$. All these are distinct because of our lowest terms assumption. However, depending upon the sign of x, n is always even or always odd; and so only every other value above is taken. So we get exactly q distinct values for x^x, all equally spaced on the circle of radius $|x|^x$.

Where on the circle do these points lie? In all cases the set of points will be symmetric with respect to the real (u)-axis. This can be seen by analyzing the above situation in the rational case or more easily (even in the irrational case) by remembering that each t''_n has such a symmetry with t''_{-n}. For x rational and positive, one of the points will be on t''_0; i.e., will be positive and real (see FIGURE 7). There will thus be a negative real value if, and only if, the number of points, q, is even.

For $x<0$ there are three cases to consider. If p is even and q is odd, then t''_q meets the circle at $|x|^x e^{i\pi q(-p/q)} = |x|^x e^{-i\pi p} = |x|^x$. So the spindle has a positive real value, but since there are an odd number (q) of equally spaced points on the circle, there is no negative real value. For p and q both odd, t''_q meets the circle at $|x|^x e^{i\pi q(-p/q)} = |x|^x e^{-i\pi p} = -|x|^x$. Thus x^x has a negative real value and no positive real one. If p is odd and q is even, we never get real values, but get angles from the real axis of $\pm(1/q)\pi, \pm(3/q)\pi, \dots, \pm((q-1)/q)\pi$. These three cases appear in order in FIGURE 8. Note p and q can't both be even because we assume x is written in lowest terms.

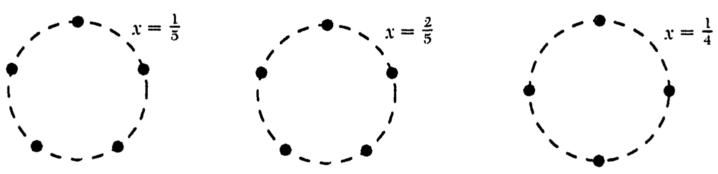


FIGURE 7

The q points of intersection with the complete spindle with the $u-v$ plane for a few positive rational values of x. Note that real values (u) are vertical.

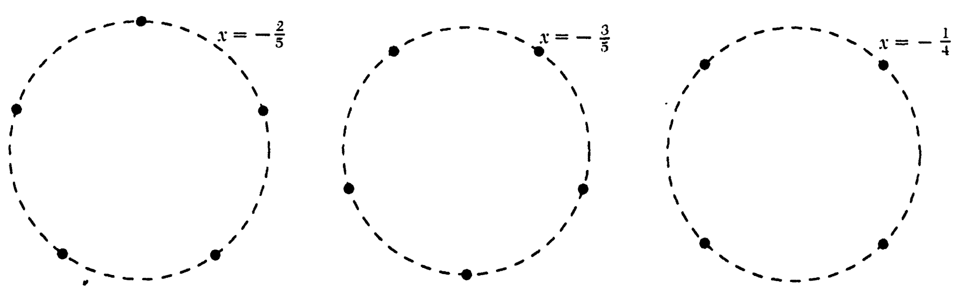


FIGURE 8

The q points of intersection with the complete spindle with the $u-v$ plane for a few negative rational values of x. Note that real values (u) are vertical.

9. The irrational case If x is irrational, then every thread takes a different value at x and the set of values is dense in the circle of radius $|x|^x$. For t''_n takes the value with angle $n\pi x$, so if t''_m (with $m \neq n$) takes the same value, then $n\pi x$ and $m\pi x$ will differ by an integer multiple of 2π, say $n\pi x - m\pi x = k2\pi$. So $x = 2k/(n-m)$, contradicting the irrationality of x. To see the density of values, note that for any $\varepsilon > 0$, since all the threads take distinct values at irrational x, there must be two different threads, say t''_n and t''_m, whose angles at x are within ε of each other (this is the Pigeon Hole Principle). Letting $k = m - n$, any consecutive threads in

$t''_n, t''_{n+k}, t''_{n+2k}, t''_{n+3k}, \dots$ will be within ε but will eventually "march" all around the circle. Thus any angle of ε will be pierced by one of these threads.

10. A closer look In FIGURE 9 we show about twice as many threads as we've seen before. Note that the gaps, while still present and as deep as before, have narrowed. This must happen, since for any irrational point near a gap the threads will take a dense collection of values.

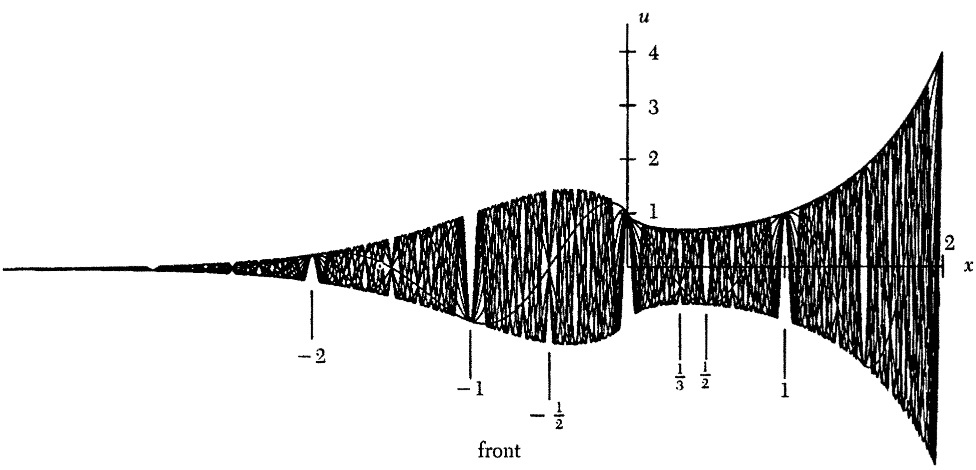


FIGURE 9

Front view of the graph of $z = x^x$ for x real and z complex. Threads t''_n are plotted, for $n = -20, -19, \dots, 19, 20$ for x in $[-4 \dots 2]$.

We list several facts about the front and top projections of the spindle.

1. In front view (FIGURES 5 and 9):
   a. Full-edge gaps at integers. (From below for $x > 0$ or for $x < 0$ and even, from above for $x < 0$ and odd.)
   b. Double half-edge gaps at negative half-integers. $(-1/2, -3/2, \dots)$
   c. Full central gaps at positive half-integers. $(1/2, 3/2, \dots)$
   d. Small-edge gaps from below for $x > 0$ and rational with odd denominator. $(1/3, 2/3, \dots)$
   e. At negative irrationals, x, the edge values $\pm|x|^x$ are never attained but are approached.
2. In the top view (FIGURE 6):
   a. Double-edge gaps at all integers and at positive half integers.
   b. All gaps are symmetric—no one-sided gaps.
   c. Full central gaps at negative half-integers.
   d. A pair of half-gaps for positive odd multiples of $1/4$.
   e. A central partial gap at negative odd multiples of $1/4$.
   f. A pair of partial gaps at integral multiples of $1/3$ that are not integers.

All of these facts and more can be understood from the analysis in Section 8.

The Limit of $x^{x^{\cdot^{\cdot^{\cdot^{x}}}}}$ as x Tends to Zero

J. MARSHALL ASH
DePaul University
Chicago, IL 60614-3504

The tower of exponents in the title poses some interesting challenges for devotees of L'Hôpital's rule. We show in this note that as x tends to 0 from the right, the expression in the title tends to 0 when the number of x's in that expression is odd and to 1 when the number of x's is even.

Let ${}^0x = 1$, and define recursively ${}^nx = x^{({}^{n-1}x)}$, $n = 1, 2, 3, \cdots$, so that, for example,

$${}^4x = x^{(x^{(x^x)})} = x^{x^{x^x}}.$$

The expression ${}^nx$ is sometimes called the n-th hyperpower of x. Notice that the nesting of the parentheses from the left is quite important. It is clear that $\lim_{x\to 0^+} {}^0x = \lim_{x\to 0^+} 1 = 1$ and $\lim_{x\to 0^+} {}^1x = \lim_{x\to 0^+} x = 0$ and it is a standard calculus exercise to show that

$$\lim_{x\to 0^+} {}^2x = \lim_{x\to 0^+} x^x = e^{(\lim_{x\to 0^+} x \ln x)} = e^0 = 1.$$

The reader should fill in the details using either L'Hôpital's rule or remark 1 below. This 1, 0, 1 pattern persists.

THEOREM 1. *For every nonnegative integer k, we have*

$$\begin{cases} \lim_{x\to 0^+} {}^{2k}x = 1 & (A, k) \\ \lim_{x\to 0^+} {}^{2k+1}x = 0. & (B, k) \end{cases}$$

Proof. Since ${}^nx = e^{({}^{n-1}x \ln x)}$, for every positive integer k the above two statements are equivalent, respectively, to

$$\begin{cases} \lim_{x\to 0^+} {}^{2k-1}x \ln x = 0 & (A', k) \\ \lim_{x\to 0^+} {}^{2k}x \ln x = -\infty. & (B', k) \end{cases}$$

We saw above that equations (A, 0), (B, 0), and (A, 1) hold. To see how the general induction proof should go, we will first establish the next two equations explicitly.

To verify equation (B, 1) we see that there holds the equivalent (B′, 1):

$$\lim_{x\to 0^+} {}^2x \ln x = -\infty,$$

since $\lim_{x\to 0^+} {}^2x = 1$ and $\lim_{x\to 0^+} \ln x = -\infty$.

To confirm equation (A, 2), note that from $\lim_{x\to 0^+} {}^2x = 1$, it follows that for x sufficiently small, ${}^2x > 1/2$. Thus for $x \in (0, 1)$ sufficiently small we have

$${}^3x = x^{({}^2x)} < x^{1/2}.$$

From this and L'Hôpital's rule (alternatively, See Remark 1 below) we have

$$\lim_{x\to 0^+} |{}^3x \ln x| \le \lim_{x\to 0^+} |x^{1/2} \ln x| = \lim_{x\to 0^+} \left|\frac{(\ln x)'}{(x^{-1/2})'}\right| = \lim_{x\to 0^+} |-2x^{1/2}| = 0,$$

which is the equivalent equation (A′, 2).

Passing to the general proof, assume now that equations (A, k) and (B, k) hold. It suffices to establish equations (A′, $k+1$) and (B′, $k+1$).

By induction hypothesis (A, k), $\lim_{x\to 0^+} {}^{2k}x = 1$. It follows that for x sufficiently small, ${}^{2k}x > 1/2$. Thus for $x \in (0,1)$ sufficiently small we have

$$x^{({}^{2k}x)} < x^{1/2}.$$

From this and L'Hôpital's rule (alternatively, see Remark 1 below) we have

$$\lim_{x\to 0^+} |{}^{2(k+1)-1}x \ln x| = \lim_{x\to 0^+} |x^{({}^{2k}x)} \ln x| \le \lim_{x\to 0^+} |x^{1/2} \ln x| = \lim_{x\to 0^+} \left|\frac{\ln x}{x^{-1/2}}\right|$$

$$= \lim_{x\to 0^+} \left|\frac{(\ln x)'}{(x^{-1/2})'}\right| = \lim_{x\to 0^+} \left|\frac{x^{-1}}{-\frac{1}{2}x^{-3/2}}\right| = \lim_{x\to 0^+} 2x^{1/2} = 0,$$

which is equation (A′, $k+1$). Of course the equivalent equation (A, $k+1$) also holds.

It is now almost immediate that

$$\lim_{x\to 0^+} {}^{2(k+1)}x \ln x = -\infty,$$

since $\lim_{x\to 0^+} {}^{2(k+1)}x = 1$ by equation (A, $k+1$) and $\lim_{x\to 0^+} \ln x = -\infty$. This is equation (B′, $k+1$). QED

Remarks.

1. The proof may be made more elementary by proving that $\lim_{x\to 0^+}|x^{1/2} \ln x| = 0$ without using L'Hôpital's rule. To do this we will need the fact that $0 < \ln y < y$ for $y \in (1, \infty)$. This is geometrically evident if $\ln y$ is defined as $\int_1^y (dt/t)$, the area under the function $1/t$ and between $t = 1$ and $t = y$. Now set $y = x^{-1/4}$. Then

$$\lim_{x\to 0^+} x^{1/2} \ln x = \lim_{y\to\infty} \frac{\ln y^{-4}}{y^2} = -4 \lim_{y\to\infty} \left(\left(\frac{1}{y}\right)\left(\frac{\ln y}{y}\right)\right).$$

   This is zero by the squeeze theorem, since $\lim_{y\to\infty} 1/y = 0$ and $0 < \ln y < y$ implies $0 < \ln y/y < 1$. One can similarly show that $\lim_{x\to 0^+} x^{\alpha} \ln x = 0$ for any $\alpha > 0$.
2. Perhaps I should have credited *Macsyma*® as coauthor. While fooling around with it, I asked it to find successively the limits of ${}^2x$ through ${}^6x$ as x tended to 0 from the right. Since I was expecting all these limits to be 1, my first reaction to the results, 1, 0, 1, 0, and 1 was to erroneously think that I had discovered a bug in *Macsyma*®.
3. Let ${}^{\infty}x = \lim_{n\to\infty} {}^n x$. One might like to study $\lim_{x\to 0^+} {}^{\infty}x$. Unfortunately, ${}^{\infty}x$ exists only when $x \in [e^{-e}, e^{1/e}]$. See the references for this. Reference **[2]** has a superb bibliography.
4. The phenomenon described by THEOREM 1 persists if we replace each x in $x^{\cdot^{\cdot^{x}}}$ by a function that is asymptotic to a power of x. More explicitly, the following generalization of THEOREM 1 has much the same proof. Let a be any real number and let $\{g_n\}$ be a sequence of functions such that there are constants $c_n > 0$ and b_n

such that

$$\lim_{x \to a^+} \frac{g_n(x)}{b_n(x-a)^{c_n}} = 1.$$

Define $\{f_n(x)\}$ recursively by $f_0(x) = g_0(x) = 1$, and, for $n \geq 1, f_n(x) = g_n(x)^{f_{n-1}(x)}$. Then, for $k = 0, 1, 2, \ldots,$

$$\begin{cases} \lim_{x \to a^+} f_{2k}(x) = 1 \\ \\ \lim_{x \to a^+} f_{2k+1}(x) = 0. \end{cases}$$

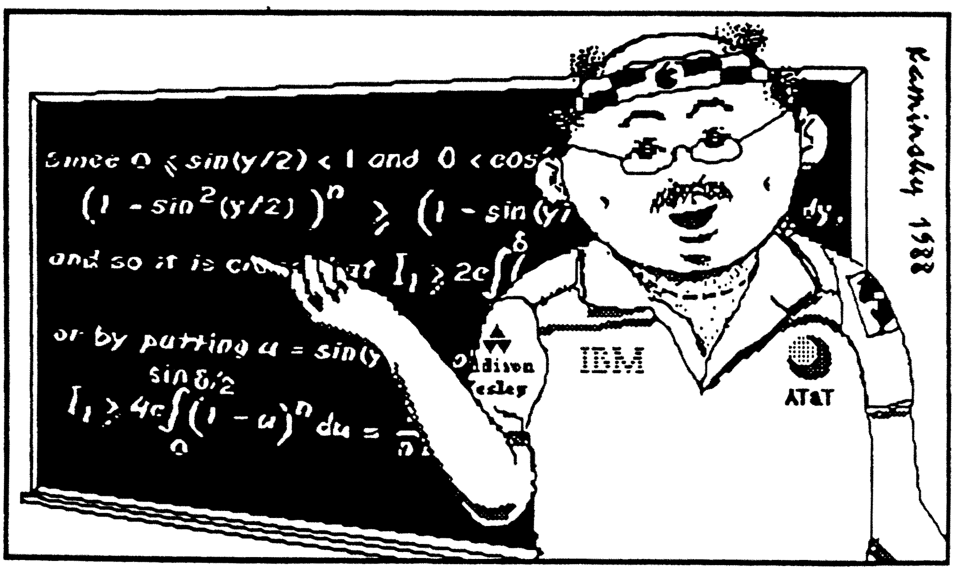
Acknowledgments. Thanks to Allan Berele, William Chin, Susanna S. Epp, Constantine Georgakis, Eric Rieders, Gang Wang, and Yuen-Fat Wong, all of DePaul University, and to Peter Ash for helpful ideas on this topic. I am also grateful to the referee for some suggestions.

REFERENCES

1. J. M. de Villiers and P. N. Robinson, The interval of convergence and limiting functions of a hyperpower sequence, *Amer. Math. Monthly*, 93(1986) 13–23.
2. R. Arthur Knoebel, Exponentials reiterated, *Amer. Math. Monthly*, 88(1981) 235–252.

Professor Fogelfroe

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



Professor Fogelfroe cashes in on the lucrative endorsement craze.

—Kenneth Kaminsky
Augsburg College
Minneapolis, MN 55454

such that

$$\lim_{x \to a^+} \frac{g_n(x)}{b_n(x-a)^{c_n}} = 1.$$

Define $\{f_n(x)\}$ recursively by $f_0(x) = g_0(x) = 1$, and, for $n \geq 1, f_n(x) = g_n(x)^{f_{n-1}(x)}$. Then, for $k = 0, 1, 2, \ldots,$

$$\begin{cases} \lim_{x \to a^+} f_{2k}(x) = 1 \\ \lim_{x \to a^+} f_{2k+1}(x) = 0. \end{cases}$$

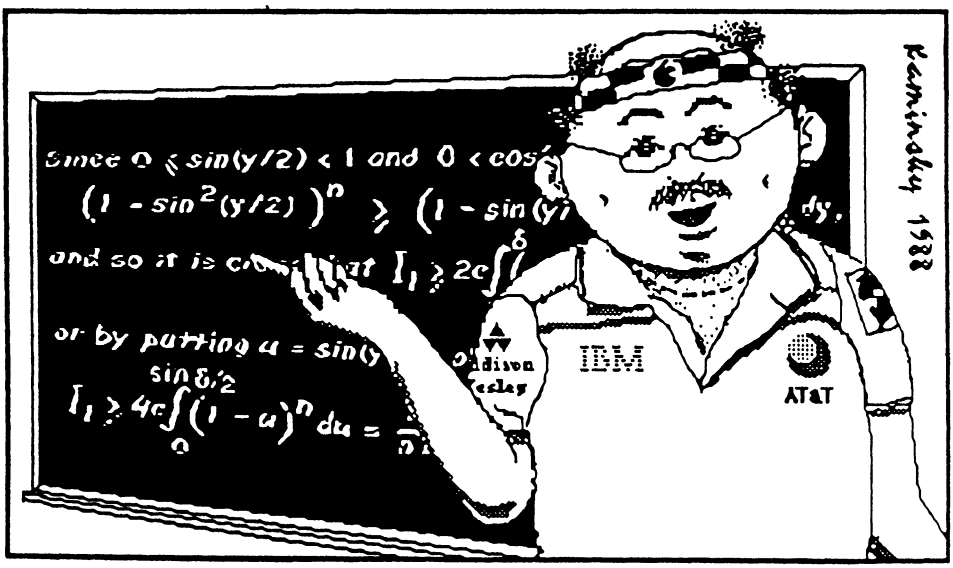
Acknowledgments. Thanks to Allan Berele, William Chin, Susanna S. Epp, Constantine Georgakis, Eric Rieders, Gang Wang, and Yuen-Fat Wong, all of DePaul University, and to Peter Ash for helpful ideas on this topic. I am also grateful to the referee for some suggestions.

REFERENCES

1. J. M. de Villiers and P. N. Robinson, The interval of convergence and limiting functions of a hyperpower sequence, *Amer. Math. Monthly*, 93(1986) 13–23.
2. R. Arthur Knoebel, Exponentials reiterated, *Amer. Math. Monthly*, 88(1981) 235–252.

Professor Fogelfroe

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



Professor Fogelfroe cashes in on the lucrative endorsement craze.

—KENNETH KAMINSKY
AUGSBURG COLLEGE
MINNEAPOLIS, MN 55454

On the Convergence of the Sequence of Powers of a 2 × 2 Matrix

ROMAN W. WONG
Washington and Jefferson College
Washington, PA 15301

Introduction The sequence $\{r^n\}$, r real, is convergent if, and only if, $|r| < 1$ or $r = 1$. The same statement is true when r is a complex number, and $|r|$ is the modulus of r. We show an analogous result for the convergence of the sequence $\{A^n\}$, where A is a 2×2 matrix over $\mathbb{R}$. The sequence $\{A^n\}$ converges to the zero matrix $\mathbb{O}$ if, and only if, $|\det(A)| < 1$ and $|1 + \det(A)| > |\mathrm{tr}(A)|$. This result is then applied to prove a well-known theorem in Markov chains for 2×2 regular stochastic matrices and to obtain an explicit formula for the stationary matrix and eigenvector.

1. Convergence of $\{A^n\}$ Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2 matrix over the complex field $\mathbb{C}$. Let α and β be the eigenvalues of A. Let $n \in \mathbb{Z}^+$. Denote the 2×2 identity matrix and the zero matrix by $\mathbb{I}$ and $\mathbb{O}$, respectively. In [4], Kenneth Williams shows how to compute the powers of A using eigenvalues:

$$A^n = \begin{cases} \alpha^n\left(\dfrac{A - \beta\mathbb{I}}{\alpha - \beta}\right) + \beta^n\left(\dfrac{A - \alpha\mathbb{I}}{\beta - \alpha}\right), & \text{if } \alpha \neq \beta, \\ \alpha^{n-1}\left(nA - (n-1)\alpha\mathbb{I}\right), & \text{if } \alpha = \beta \end{cases}$$

$$= \begin{cases} \dfrac{1}{\alpha - \beta}\left((\alpha^n - \beta^n)A - \alpha\beta(\alpha^{n-1} - \beta^{n-1})\mathbb{I}\right), & \text{if } \alpha \neq \beta, \\ n\alpha^{n-1}A - (n-1)\alpha^n\mathbb{I}, & \text{if } \alpha = \beta. \end{cases}$$

As a consequence of this formula, we see that if A has distinct eigenvalues α and β, then $\{A^n\}$ converges if, and only if, $\{\alpha^n - \beta^n\}$ converges. If $\lim_{n \mapsto \infty}(\alpha^n - \beta^n)$ exists, say the limit is k, then

$$\lim_{n \mapsto \infty} A^n = \frac{k}{\alpha - \beta}\left(A - \alpha\beta\mathbb{I}\right).$$

Since both $\{\alpha^n\}$ and $\{\beta^n\}$ must converge to 0 or 1, we see that the values of k above can only be 0, 1, or -1 and that $k = 0$ if, and only if, $|\alpha| < 1$ and $|\beta| < 1$ since $\alpha \neq \beta$.

Now suppose that $\alpha = \beta$ and $|\alpha| < 1$, then $\{A^n\}$ also converges to $\mathbb{O}$ since both $\{n\alpha^{n-1}\}$ and $\{(n-1)\alpha^n\}$ approach zero by L'Hôpital's rule. Conversely, if we write, using Williams's formula,

$$A^n = \begin{pmatrix} n\alpha^{n-1}a - (n-1)\alpha^n & n\alpha^{n-1}b \\ n\alpha^{n-1}c & n\alpha^{n-1}d - (n-1)\alpha^n \end{pmatrix}$$

$$= \begin{pmatrix} \alpha^{n-1}(n(a-\alpha) + \alpha) & n\alpha^{n-1}b \\ n\alpha^{n-1}c & \alpha^{n-1}(n(d-\alpha) + \alpha) \end{pmatrix},$$

we see that if $\{A^n\}$ converges to $\mathbb{O}$, we must have $|\alpha| < 1$.

We have proved the following proposition.

PROPOSITION. *Let A be a 2×2 matrix over $\mathbb{C}$ and let α and β be the eigenvalues of A. Then $\{A^n\}$ converges to $\mathbb{O}$ if, and only if, $|\alpha| < 1$ and $|\beta| < 1$.*

Remark. This proposition is true in general for any $n \times n$ matrix A over the complex field. Using Jordan canonical form, one can show that the sequence $\{A^n\}$ converges to the $n \times n$ zero matrix $\mathbb{O}$ if, and only if, $\max\{|\lambda| : \lambda$ is an eigenvalue of $A\} < 1$. A proof of this can be found in **[1]**.

The case when the inequalities in the proposition become equalities will be discussed later. For the rest of this paper, we restrict A to be a 2×2 matrix over the reals $\mathbb{R}$. We now present the theorem that characterizes the convergence of $\{A^n\}$ directly in terms of A. As usual, denote the determinant of A by $\det(A)$ and the trace of A by $\mathrm{tr}(A)$.

THEOREM. *Let A be a 2×2 matrix over $\mathbb{R}$. Then $\{A^n\}$ converges to $\mathbb{O}$ if, and only if, $|\det(A)| < 1$ and $|1 + \det(A)| > |\mathrm{tr}(A)|$.*

The proof of the theorem uses the result of the previous proposition and a lemma that will be presented later. The theorem is not true over the complex number $\mathbb{C}$. A counterexample will be given later.

2. Examples We now illustrate the theorem by some numerical examples. Computer software can be used to show the convergence.

A	α and β	$\det(A)$	$1+\det(A)$	$\mathrm{tr}(A)$	Convergence to $\mathbb{O}$
$\begin{pmatrix} 1/2 & 0 \\ 1 & -1/3 \end{pmatrix}$	$1/2, -1/3$	$-1/6$	$5/6$	$1/6$	yes
$\begin{pmatrix} 1 & -1 \\ 1 & -1.9 \end{pmatrix}$	$3/5, -3/2$	-0.9	0.1	-0.9	no
$\begin{pmatrix} 2 & -1 \\ 0 & -1/3 \end{pmatrix}$	$2, -1/3$	$-2/3$	$1/3$	$5/3$	no
$\begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$	$\pm 2i$	4	5	0	no
$\begin{pmatrix} 0 & 1/2 \\ -1/2 & 0 \end{pmatrix}$	$\pm \frac{i}{2}$	$1/4$	$5/4$	0	yes
$\begin{pmatrix} -5.2 & 3 \\ -10.6 & 6.1 \end{pmatrix}$	$0.1, 0.8$	0.08	1.08	0.9	yes

In each of the above examples, when $\{A^n\}$ does not converge to $\mathbb{O}$, the entries in A^n actually tend to ∞. This may not be the case when the inequalities in the theorem become equalities. Here are some examples where one or both of the boundary conditions are attained.

A	α and β	$\det(A)$	$1+\det(A)$	$\mathrm{tr}(A)$	Convergence
$\begin{pmatrix} 1 & -1/2 \\ 1 & 1/2 \end{pmatrix}$	$\frac{3 \pm \sqrt{15}\,i}{4}$	1	2	3/2	no convergence: entries tend to ∞
$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$	0, 1	0	1	1	$\{A^n\}$ converges to A since $A^2 = A$
$\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$	± 1	-1	0	0	A^n equals A and A^2 $(=\mathbb{I})$ alternately
$\begin{pmatrix} 3 & 1 \\ -4 & -1 \end{pmatrix}$	1, 1	1	2	2	$A^n = \begin{pmatrix} 2n+1 & n \\ -4n & 1-2n \end{pmatrix}$ entries tend to $\pm\infty$

Another example on the boundary is $A = \begin{pmatrix} 5/2 & 1 \\ -3 & -1 \end{pmatrix}$ with $\det(A) = 1/2$ and $\det(A) + 1 = 3/2 = \mathrm{tr}(A)$. The eigenvalues of A are 1 and 1/2. By the proposition, since $\{\alpha^n - \beta^n\}$ approaches 1 as a limit, $\{A^n\}$ is convergent. In fact, A^n converges to $\begin{pmatrix} 4 & 2 \\ -6 & -3 \end{pmatrix} \neq \mathbb{O}$.

Here is one more interesting example. Let $A = \begin{pmatrix} 1 & -1/2 \\ 1 & 1/2 \end{pmatrix}$ with $\det(A) = 1$ and $\mathrm{tr}(A) = 3/2 < 1 + \det(A)$. The eigenvalues of A are the complex conjugates $\bar{\beta} = \alpha = (3 + \sqrt{7}\,i)/4$ with modulus 1. To four decimal places,

$$A^{1000} = \begin{pmatrix} 1.0491 & -0.1264 \\ 0.2527 & 0.9228 \end{pmatrix}; \quad A^{1001} = \begin{pmatrix} 0.9228 & -0.5877 \\ 1.1755 & 0.3350 \end{pmatrix}; \quad A^{1002} = \begin{pmatrix} 0.3350 & -0.7532 \\ 1.5105 & -0.4202 \end{pmatrix}.$$

The entries of A never go to infinity, but do not stabilize at any values either. A closer look at these matrices reveals that this system is not quite chaotic. In fact, if A is considered as a linear transformation in the xy-plane, the orbit of a vector under A is shown to be elliptic. My colleague Michael Woltermann pointed out that for any vector $P = \begin{pmatrix} a \\ b \end{pmatrix}$, both P and $AP = \begin{pmatrix} a - b/2 \\ a + b/2 \end{pmatrix}$ lie on the ellipse $x^2 - \frac{1}{2}xy + \frac{1}{2}y^2 - a^2 + \frac{1}{2}ab - \frac{1}{2}b^2 = 0$ and that the equation remains unchanged when a and b are replaced by $a - b/2$ and $a + b/2$, respectively. FIGURE 1 shows the orbits of the vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 0.3 \\ 0.4 \end{pmatrix}$, and $\begin{pmatrix} 0.6 \\ -1.4 \end{pmatrix}$ under the transformation A.

3. Stochastic matrices Recall that a square matrix $A = [a_{ij}]$ is called *stochastic* if $a_{ij} \geqq 0$ for all i and j and $\Sigma_i a_{ij} = 1$ for each j. A is called *regular* if for some positive integer n, all the entries in A^n are strictly positive. The fundamental theory of Markov chains says that the powers of every regular stochastic matrix approach a stationary matrix with identical columns, and that this column vector is indeed an eigenvector for the eigenvalue 1. (See, for instance, [2].) We give a different proof here for the special case of 2×2 regular stochastic matrices and provide an explicit formula for the stationary matrix.

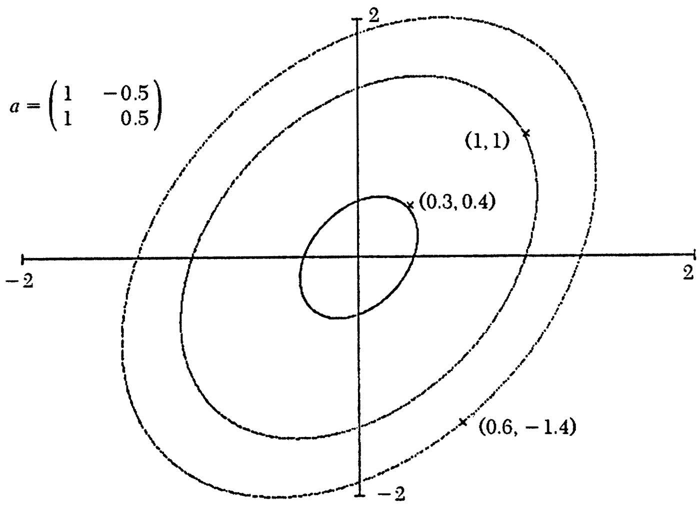


FIGURE 1

Let A be a 2×2 regular stochastic matrix with eigenvalues α and β. We can write A as

$$A = \begin{pmatrix} a & b \\ 1-a & 1-b \end{pmatrix}, \qquad 0 \le a, b \le 1.$$

We may assume that $\alpha = 1$ since one of the eigenvalues of a stochastic matrix must be 1. If β also equals to 1, we must have $\mathrm{tr}(A) = a + 1 - b = \alpha + \beta = 2$, which implies that $a = 1 + b$. Since $0 \leqq a,\ b \leqq 1$, we must have $b = 0$ and $a = 1$ and so $A = \mathbb{I}$. This is impossible since A is regular. Thus $\beta \neq 1$ and the two eigenvalues of A are distinct. Also $\beta = \alpha\beta = \det(A) = a - b$ and so $|\beta| < 1$. Thus $\lim_{n \to \infty}(\alpha^n - \beta^n) = 1$ exists and we conclude that

$$\lim_{n \mapsto \infty} A^n = \frac{1}{1-a+b}\left(A - (a-b)\mathbb{I}\right) = \frac{1}{1-a+b}\begin{pmatrix} b & b \\ 1-a & 1-a \end{pmatrix}.$$

Thus

$$\frac{1}{1-a+b}\begin{pmatrix} b & b \\ 1-a & 1-a \end{pmatrix}$$

is the stationary matrix and the vector $\begin{pmatrix} b \\ 1-a \end{pmatrix}$ is an eigenvector for the eigenvalue 1. As an example, consider the regular stochastic matrix $A = \begin{pmatrix} 1/4 & 1/3 \\ 3/4 & 2/3 \end{pmatrix}$. By the above formula, the stationary matrix for A would be

$$\frac{1}{1-1/4+1/3}\begin{pmatrix} 1/3 & 1/3 \\ 3/4 & 3/4 \end{pmatrix} = \frac{12}{13}\begin{pmatrix} 1/3 & 1/3 \\ 3/4 & 3/4 \end{pmatrix} = \frac{1}{13}\begin{pmatrix} 4 & 4 \\ 9 & 9 \end{pmatrix}$$

and the vector $\begin{pmatrix} 4 \\ 9 \end{pmatrix}$ is an eigenvector for the eigenvalue 1.

4. Proof of the theorem We first prove a lemma.

LEMMA. *Let both α and β be real numbers or be complex conjugates. Then $|\alpha| < 1$ and $|\beta| < 1$ if, and only if, $|\alpha\beta| < 1$ and $|1 + \alpha\beta| > |\alpha + \beta|$.*

Proof. First we assume that both α and β are reals. Suppose that $|\alpha| < 1$ and $|\beta| < 1$. Then $|\alpha\beta| < 1$ and

$$\begin{aligned} & 1-\alpha^2>0 \quad \text{and} \quad 1-\beta^2>0 \\ & \Rightarrow (1-\alpha^2)(1-\beta^2)>0 \\ & \Leftrightarrow 1+\alpha^2\beta^2>\alpha^2+\beta^2 \\ & \Leftrightarrow 1+2\alpha\beta+(\alpha\beta)^2>\alpha^2+2\alpha\beta+\beta^2 \\ & \Leftrightarrow (1+\alpha\beta)^2>(\alpha+\beta)^2 \\ & \Leftrightarrow |1+\alpha\beta|>|\alpha+\beta|. \end{aligned}$$

Conversely, suppose that $|\alpha\beta| < 1$ and $|1+\alpha\beta| > |\alpha+\beta|$. Reversing the above argument, we have $(1-\alpha^2)(1-\beta^2) > 0$. This implies that either both $1 > \alpha^2$ and $1 > \beta^2$ or both $1 < \alpha^2$ and $1 < \beta^2$. Since $1 < \alpha^2$ and $1 < \beta^2$ would imply that $1 < |\alpha\beta|$, which is impossible, we thus conclude that $1 > \alpha^2$ and $1 > \beta^2$, that is, $|\alpha| < 1$ and $|\beta| < 1$.

Now suppose that $\alpha = r(\cos\theta + i\sin\theta)$ and $\beta = \bar{\alpha} = r(\cos\theta - i\sin\theta)$ are complex conjugates with $r < 1$. Then

$$|\alpha+\beta| = |2r\cos\theta| \leqq 2r < 1 + r^2 = |1+\alpha\beta|$$

since $1 + r^2 - 2r = (1-r)^2 > 0$. Also $|\alpha\beta| = r^2 < 1$ implies that $|\alpha| = |\beta| = r < 1$. Thus $|\alpha\beta| < 1$ if, and only if, $|\alpha| < 1$ and $|\beta| < 1$. This completes the proof of the lemma.

We now prove the theorem.

THEOREM. *Let A be a 2×2 matrix over $\mathbb{R}$. Then $\{A^n\}$ converges to $\mathbb{O}$ if, and only if,* $|\det(A)| < 1$ *and* $|1 + \det(A)| > |\operatorname{tr}(A)|$.

Proof. The characteristic polynomial of A can be written as

$$\begin{aligned} p(x) &= x^2 - (a+d)x + (ad - bc) \\ &= x^2 - \operatorname{tr}(A)x + \det(A) \\ &= x^2 - (\alpha+\beta)x + \alpha\beta \end{aligned}$$

where α and β are the eigenvalues of A. Since $p(x)$ is a polynomial with real coefficients, α and β are either both reals or are complex conjugates. The result then follows from the lemma and the proposition.

Notes. The lemma is not true in general for any complex numbers α and β. For example, consider $\alpha = \beta = \frac{1}{2}i$. Then $|\alpha| = |\beta| = \frac{1}{2} < 1$. But $|1+\alpha\beta| = |1 - \frac{1}{4}| = \frac{3}{4} < 1 = |i/2 + i/2| = |\alpha+\beta|$. This also shows that the theorem is not true if A is over the complex numbers. Let $A = \begin{pmatrix} i/2 & 0 \\ 0 & i/2 \end{pmatrix}$. Then the same computation as above shows that $|\det(A)| = |-\frac{1}{4}| = \frac{1}{4} < 1$ whereas $|1 + \det(A)| = \frac{3}{4} < 1 = |i| = |\operatorname{tr}(A)|$. Thus A does not satisfy the conditions in the theorem. However, since $A = (\frac{1}{2}i)\mathbb{I}$ is a scalar matrix, we see that $A^n = (\frac{1}{2}i)^n\mathbb{I}$. Since $|\frac{1}{2}i| < 1$, $\{A^n\}$ converges to $\mathbb{O}$. Observe that the convergence of $\{A^n\}$ to $\mathbb{O}$ is also a consequence of the proposition since the eigenvalues of A are $\alpha = \beta = i/2$. Of course, the theorem remains true as long as the characteristic polynomial of A is a polynomial with real coefficients.

Finally, we remark that the theorem is not true for 3×3 matrices. Consider the matrix

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

Then $|\det(A)| = 0 < 1$ and $|1 + \det(A)| = 1 > |\mathrm{tr}(A)| = 0$ and A satisfies the conditions in the theorem. However, it is easy to see that

$$A^n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq \mathbb{O} \quad \text{for all } n.$$

Thus $\{A^n\}$ does not converge to $\mathbb{O}$.

Acknowledgement. The author would like to thank his colleague Michael Woltermann for finding the equation of the ellipse in one of the above examples. The author also wants to thank the referees and the former editor for their valuable suggestions.

REFERENCES

1. Charles Cullen, *Matrices and Linear Transformations*, Addison-Wesley Publishing Co., Reading, MA, 1966, p. 173.
2. John Kemeny and J. Laurie Snell, *Finite Markov Chains*, D. Van Nostrand Co., Inc., New York, NY, 1960.
3. Gareth Williams, *Linear Algebra with Applications*, 2nd edition, Wm. C. Brown Publishers, Dubuque, IA, 1991.
4. Kenneth S. Williams, The nth power of a 2×2 matrix, this MAGAZINE, **65** (1992), 336.

Power Mean for Zero Exponent

N. SCHAUMBERGER
Hofstra University
Hempstead, NY 11550

If a, b, and c are positive, then the usual proof that

$$\lim_{r \to 0} \left(\frac{a^r + b^r + c^r}{3} \right)^{1/r} = \sqrt[3]{abc}$$

uses exponentials and L'Hôpital's rule. (See, for example, [**1**].)

We use the ordinary $AM - GM$ inequality on the left, and since

$$\frac{a^r}{a^r + b^r + c^r} + \frac{b^r}{a^r + b^r + c^r} + \frac{c^r}{a^r + b^r + c^r} = 1,$$

the weighted $AM - GM$ inequality on the right, to get

$$\frac{1}{(a^r b^r c^r)^{1/3}} \geq \frac{3}{a^r + b^r + c^r} = \frac{\frac{a^r}{a^r} + \frac{b^r}{b^r} + \frac{c^r}{c^r}}{a^r + b^r + c^r}$$

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

Hence

$$(abc)^{1/3} \leq \left(\frac{a^r + b^r + c^r}{3} \right)^{1/r} \leq \left(a^{a^r} b^{b^r} c^{c^r} \right)^{\frac{1}{a^r + b^r + c^r}}.$$

As $r \to 0$, the term on the right tends to $\sqrt[3]{abc}$.

An analogous argument can be used to show that

$$\lim_{r \to 0} \left(w_1 a_1^r + w_2 a_2^r + \cdots + w_n a_n^r \right)^{1/r} = a_1^{w_1} a_2^{w_2} \ldots a_n^{w_n}$$

where $\Sigma w_i = 1$ and $w_i \geq 0$.

REFERENCE

1. G. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge University Press, New York, 1988, p. 15.

The Space of Closed Subsets of a Convergent Sequence

HAROLD REITER
University of North Carolina
Charlotte, NC 28223

ASHLEY REITER
University of Chicago
Chicago, IL 60637

Many topological spaces are simply sets of points (atoms) endowed with a topology. Some spaces, however, have elements that are functions, matrices, or other non-atomic items. Another special type of space has elements that are themselves subsets of another space; these spaces are called *hyperspaces*. Hyperspaces are metric spaces, and the metric defined on them is called the *Hausdorff* metric. Hyperspaces whose points are the closed subsets and hyperspaces whose points are the closed *connected* subsets (of metric spaces) have been extensively studied. Also, hyperspaces of closed and convex subsets of a bounded convex set in Euclidean space are of great interest in geometry. See Lay **[3]**. In recent years, geometers and topological dynamicists have explored spaces of closed and bounded subsets of the plane in connection with the study of fractals. One of the major results in the theory is that the hyperspace of closed subsets of a closed interval of real numbers is homeomorphic with the Hilbert cube, and with the space I^{∞}, the countable product of unit intervals. For more general information about the Hausdorff metric and spaces of subsets, see Devaney **[2]** and Sieradski **[6]**.

The main result of this paper is a topological characterization of the space of closed subsets of a convergent sequence of points. The proof given here provides a *homeomorphic* embedding of the space in the plane, E_2. The result was first proved by Pelczynski **[5]**. He studied arbitrary compact zero-dimensional metric spaces, so his proofs are much more widely applicable, but they are also somewhat technical. Our proof depends only on some well-known results in the theory of metric spaces, and is therefore accessible to advanced undergraduate mathematics majors. We also prove that for no convergent sequence of real numbers is there an *isometric* embedding of the hyperspace in Euclidean space, E_n, for any n. For other results and discussion, see Nadler **[4]**.

Let (X, d) denote a compact metric space. The *hyperspace* $(2^X, D)$ of X is the metric space whose points are the closed nonempty subsets of X and whose metric is the Hausdorff metric D given by

$$D(A, B) = \inf\{\epsilon > 0 : A \subset N_\epsilon(B) \text{ and } B \subset N_\epsilon(A)\},$$

where

$$N_\epsilon(A) = \{x \in X \mid d(a, x) < \epsilon \text{ for some } a \in A\}.$$

For example, let $A = \{(x, y) \mid x = 0 \text{ and } 0 \le y \le 20\}$ and $B = \{(x, y) \mid x = 10 \text{ and } 10 \le y \le 20\}$. FIGURE 1 shows that $\forall \epsilon > 0$, $N_{10+\epsilon}(A) \supset B$ and $N_{10\sqrt{2}+\epsilon}(B) \supset A$. However, $N_{10\sqrt{2}}(B) \not\supset A$, so $D(A, B) = 10\sqrt{2}$.

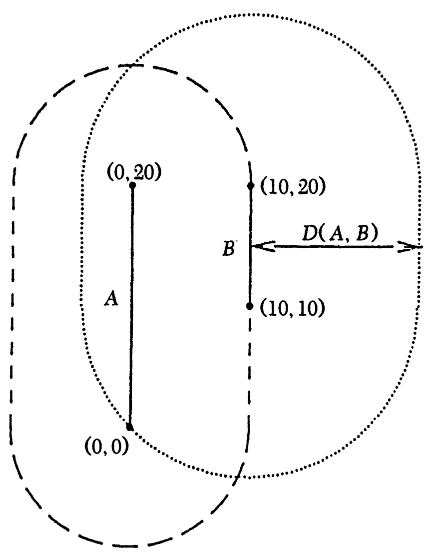


FIGURE 1
The distance function D, where A and B are vertical segments.

It is well known that 2^X is compact if and only if X is compact and that the topology of 2^X depends only on the topology of X and not on the metric for X. Also, since each point of X is a closed subset, *X is a subset of 2^X*, and the copy of X embedded in 2^X is isometric with X itself since the Hausdorff metric D for 2^X restricted to the subset of singletons is identical to the original metric d for X. See Nadler [4] for this result and more. Pelczynski investigated the hyperspace of $K = \{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\}$ with the metric it inherits from the real line. A corollary to his theorem answers the question, "topologically, what is 2^K?"

PELCZYNSKI'S THEOREM. *Let X be a zero-dimensional infinite compact metric space with a dense set of isolated points. Then the space 2^X is homeomorphic with the subset $T(C)$ of the product space $[0,1] \times C$ given by*

$$T(C) = (0, C) \cup \bigcup_{k=1}^{\infty} \bigcup_{n=1}^{N(k)} (k^{-1}, x_{k,n}),$$

where $\bigcup_{n=1}^{N(k)}\{x_{k,n}\}$ is an arbitrary fixed k^{-1}-net for C, and C is the Cantor discontinuum.[1]

Now we can state the corollary that is our first theorem.

THEOREM 1. *The space 2^K is homeomorphic with a space $Y = Y_1 \cup Y_2 \subset E_2$, where Y_1 is homeomorphic with the Cantor ternary set and Y_2 is a countable set every point of which is isolated in $Y = Y_1 \cup Y_2$ and such that every point of Y_1 is the limit of a sequence of points of Y_2.*

Pelczynski also showed that such a space is unique up to homeomorphism.

[1]By a k^{-1}-net for C, Pelczynski means a set S of points in $T(C)$ such that every point of C is within k^{-1} of some point of S.

Proof. First, consider which subsets of K are closed. Clearly, every finite subset is closed. Which infinite ones are closed? Precisely those that contain 0. It is helpful to view 2^K as the union $Z \cup F$ where Z is the family of all subsets of K that contain 0, and F is the set of nonempty finite subsets that do not contain 0. Let C denote the Cantor ternary set. That is, $C = \{x \in \mathbf{R} : 0 \le x \le 1$ and x has a ternary representation consisting only of 0's and 2's$\}$. Next we embed 2^K in the plane. We define a function $\varphi : 2^K \to E_2$ in terms of a function f from Z to sequences of 0's and 2's: For each $U \in Z$

$$(f(U))_i = \begin{cases} 0 & \text{if } \frac{1}{i} \notin U \\ 2 & \text{if } \frac{1}{i} \in U. \end{cases}$$

Then

$$\varphi(U) = (\varphi_1(U), \varphi_2(U)) = \left(\sum_{i=1}^{\infty} f(U)_i \cdot 3^{-i}, 0 \right).$$

Each finite set not containing zero is mapped as follows. If $U \in F$, order the elements of U from smallest to largest: $a_1 < a_2 < \cdots < a_n$. Then define

$$\varphi(U) = (\varphi_1(U \cup \{0\}), a_1).$$

In other words, $\varphi(U \cup \{0\})$ is a point of C and $\varphi(U)$ is a point just above $\varphi(U \cup \{0\})$ in the plane at a height equal to the smallest element of U. For example

$$\varphi(\{0, 1, \tfrac{1}{2}, \tfrac{1}{3}\}) = (\tfrac{26}{27}, 0),$$

and

$$\varphi(\{1, \tfrac{1}{2}, \tfrac{1}{3}\}) = (\tfrac{26}{27}, \tfrac{1}{3}).$$

Thus the finite subsets of K are mapped to left endpoints of C if they have 0 as a member and to points above the x-axis otherwise. Our task is to prove that this mapping φ is a homeomorphism. Since 2^K is compact, it is enough to show that φ is both one-to-one and continuous. To see that φ is one-to-one, take distinct points U and V of 2^K. We will show that $\varphi(U) \neq \varphi(V)$. If one belongs to Z and the other to F, $\varphi_2(U) \neq \varphi_2(V)$. If both belong to F, there exists k such that $1/k$ belongs to $(U \setminus V) \cup (V \setminus U)$. If k_0 is the smallest such k and $1/k_0 \in U$, then $\varphi_1(U) > \varphi_1(V)$. Similarly, if both U and V belong to Z, then $\varphi_1(U) \neq \varphi_1(V)$.

To see that φ is continuous, let $\{U_n\}_{n=1}^{\infty}$ be a convergent sequence of elements of 2^K and suppose $\lim_{n\to\infty} U_n = U_0$. To see that

$$\lim_{n\to\infty} \varphi(U_n) = \varphi(U_0),$$

we distinguish two cases:

Case 1. $0 \notin U_0$. In this case, $\{U_n\}$ is eventually constant.

Case 2. $0 \in U_0$. We show (A) $\varphi_{,1}(U_n) \to \varphi_1(U_0)$; and (B) $\varphi_2(U_n) \to \varphi_2(U_0)$.

(A) For each $k \in \mathbf{N}$, let $I_k = \{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{k}\}$. We prove that for each $k \in \mathbf{N}$ there exists $m_0 \in \mathbf{N}$ such that if $m > m_0$ then $I_k \cap U_0 = I_k \cap U_m$. Suppose not. Then there exists $j \leq k$ such that $1/j \in U_0$ and $1/j \notin U_m$ (or vice-versa), in which case

$$D(U_0, U_m) \geq \left|\frac{1}{j} - \frac{1}{j+1}\right| = \left|\frac{j+1-j}{j(j+1)}\right| = \frac{1}{j(j+1)} \geq \frac{1}{(k+1)^2}.$$

This contradicts the convergence of the sequence $\{U_n\}$. Now we prove that $\varphi_1(U_n) \to \varphi_1(U_0)$. Let $\epsilon > 0$ be given. There exists a positive integer m_0 such that $\sum_{i=m_0}^{\infty} 2\cdot 3^{-i} < \epsilon$. Now pick m_0 large enough so that if $m > m_0$ then $I_{m_0} \cap U_0 = I_{m_0} \cap U_m$. Let $D = \{i \mid 1/i \in U_m \setminus U_0 \cup U_0 \setminus U_m\}$. Then $|\varphi_1(U_m) - \varphi_1(U_0)| = \sum_{k \in D} 2\cdot 3^{-k} \leq \sum_{k=m_0}^{\infty} 2\cdot 3^{k} < \epsilon$.

(B) To see that $\varphi_2(U_n) \to \varphi_2(U_0)$, let $\epsilon > 0$ be given. We can take m_0 large enough that for any $m > m_0$, U_m contains a point $1/k$ less than ϵ. Then $\varphi_2(U_m) \leq 1/k < \epsilon$. This completes the proof.

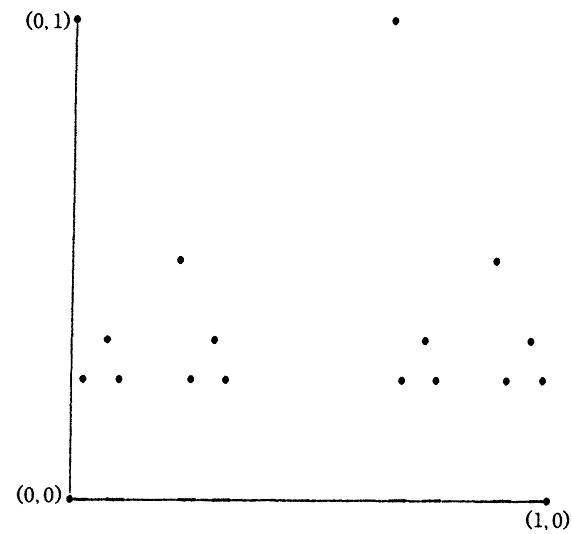


FIGURE 2
Part of the image of φ.

Rather than defining a homeomorphism into the plane, we could use a modified function to map 2^K into the line by mapping each set not containing 0 to the middle of the interval above whose right endpoint it lies. For example, $\varphi(\{1\}) = \frac{1}{2}$.

Our second theorem shows that it is not possible to improve on the homeomorphic embedding of 2^K in the plane given in the proof of Theorem 1.

THEOREM 2. *Let $\{a_i\}_{i=1}^{\infty}$ be a decreasing sequence of real numbers with $a_1 \leq 1$ and $\lim_{n\to\infty} a_i = 0$, and let $I = \{0, a_1, a_2, \dots\}$. Then, the hyperspace $(2^I, D)$ is not isometrically embeddable in Euclidean space E_n for any n.*

Proof. Suppose there is such an embedding. Let $b = a_1$ and let a be a point of the sequence satisfying $a < b/2$. Consider the six elements $\{0\}$, $\{0, b\}$, $\{b\}$, $\{a\}$, $\{a, b\}$, and $\{0, a\}$ of 2^K and denote their images in E_n under the embedding as $\{0\}'$, $\{0, b\}'$, $\{b\}'$, $\{a\}'$, $\{a, b\}'$, and $\{0, a\}'$. See FIGURE 3. If A, B, and C are three distinct points of the plane for which $D(A, C) = D(A, B) + D(B, C)$, then B lies on the segment from A to C. Using this fact, it follows that $\{a\}'$ belongs to the segment determined by $\{0\}'$

and $\{b\}'$, $\{a,b\}'$ belongs to the segment determined by $\{b\}'$ and $\{0,b\}'$, and $\{0,a\}'$ belongs to the segment determined by $\{0\}'$ and $\{0,b\}'$. Now $\{0\}'$, $\{0,b\}'$, and $\{b\}'$ are the vertices of an equilateral triangle, as are $\{b\}'$, $\{a\}'$, $\{a,b\}'$ and $\{0\}'$, $\{a\}'$, and $\{0,a\}'$. Since $D(\{a,b\},\{0,a\}) = b-a$ and $D(\{0,b\},\{0,a\}) = b-a$, we can see that $\angle\{0,b\}'\{a,b\}'\{0,a\}' = \angle\{a,b\}'\{0,b\}'\{0,a\}' = 60°$. The triangle $\{a,b\}'\{0,b\}'\{0,a\}'$ is equilateral, so $D(\{a,b\},\{0,a\}) = D(\{a,b\},\{0,b\}) = b-a$, but $D(\{a,b\},\{0,b\}) = a$, which contradicts our assumption that $a<b/2$.

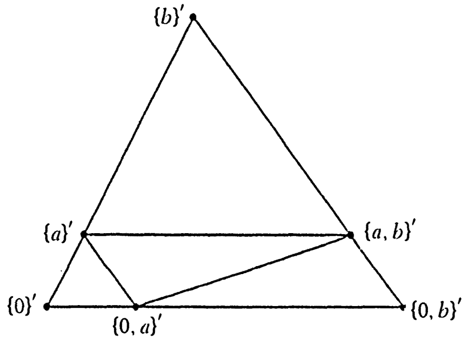


FIGURE 3
The triangle with vertices $\{0\}'$, $\{b\}'$, and $\{0,b\}'$ in the hyperspace.

Challenge 1: Find a formula for the distance between two subsets A and B based on the size of the first integer i for which $1/i$ belongs to one of the sets but not the other.
Challenge 2: Show that Theorem 1 follows no matter what metric is given for the convergent sequence.
Challenge 3: Show that Theorem 2 follows no matter what metric is given for the convergent sequence.
Challenge 4: If we use the alternative embedding of 2^K in the line, the homeomorphism establishes a linear ordering on 2^K. Describe this ordering explicitly.
Challenge 5: What goes wrong if we define φ using powers of 2 instead of powers of 3 as follows:

$$(f(U))_i = \begin{cases} 0 & \text{if } \frac{1}{i} \notin U \\ 1 & \text{if } \frac{1}{i} \in U \end{cases} \quad \text{and} \quad \varphi(U) = \left(\sum_{i=1}^{\infty} f(U)_i \cdot 2^{-i}, 0 \right)?$$

Acknowledgement. The authors gratefully acknowledge the comments of a referee that our proof of Theorem 2 holds for E_n, not just for E_2, and for other helpful remarks.

REFERENCES

1. M. F. Barnsley, *Fractals Everywhere*, Academic Press, New York, NY, 1988.
2. R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin Cummings, Menlo Park, CA, 1986.
3. S. R. Lay, *Convex Sets and their Applications*, Wiley Interscience, New York, 1982.
4. Sam B. Nadler, *Hyperspaces of Sets*, Marcel Dekker, Inc., New York, 1978.
5. A. Pelczynski, A remark on spaces 2^X for zero-dimensional X, *Bulletin of the Polish Academy of Sciences*, 13 (1965) 85–89.
6. A. J. Sieradski, *An Introduction to Topology and Homotopy*, PWS-Kent, Boston, MA, 1992.

Extending the Pythagorean Theorem to Other Geometries

PAUL FJELSTAD
32991 Dresden Avenue
Northfield, MN 55057

Many theorems of Euclidean plane geometry can be generalized to forms that also hold for some other plane geometries. This is done by generalizing the definitions for distance and angle. To do this for the Pythagorean Theorem it suffices to generalize distance, and to do this, the *odd square function* $^{\ddot{2}}$ and its inverse, the *odd square root function* $^{1/\ddot{2}}$, are defined for x real by (note—umlauts usually signify an *odd* pronounciation for a vowel)

$$x^{\ddot{2}} = (\operatorname{sgn} x)\, x^2, \qquad x^{1/\ddot{2}} = (\operatorname{sgn} x)|x|^{1/2}. \tag{1}$$

Points are represented in the usual way by vectors $\mathrm{v} = (x, y)$, the transpose being $\mathrm{v}^\dagger = \begin{pmatrix} x \\ y \end{pmatrix}$. The geometries considered are defined by the *metric* $g(t, s, r)$, for t, s, r real, and the *inner product* $\langle \mathrm{v}, \mathrm{v}' \rangle$, where

$$g(t,s,r) = \begin{pmatrix} t & s \\ s & r \end{pmatrix}, \qquad \langle \mathrm{v}, \mathrm{v}' \rangle = \mathrm{v} g \mathrm{v}'^\dagger = txx' + s(xy' + yx') + ryy'. \tag{2}$$

For $t, r = 1, s = 0$, the Euclidean case, $\langle \mathrm{v}, \mathrm{v} \rangle = x^2 + y^2$ is nonnegative and the distance of point v from the origin is $d(\mathrm{v}) = \langle \mathrm{v}, \mathrm{v} \rangle^{1/2}$. In the general case $\langle \mathrm{v}, \mathrm{v} \rangle$ can be negative, in which case this distance is defined to be negative by defining *distance* in the following way,

$$d(\mathrm{v}) = \langle \mathrm{v}, \mathrm{v} \rangle^{1/\ddot{2}}. \tag{3}$$

It is this 'trick' that allows the Pythagorean Theorem to be extended. For two points v, v', the distance between them is $d(\mathrm{v}' - \mathrm{v})$. The case $t = 1, s = 0, r = -1$ is Lorentzian (also called Minkowskian) geometry, where $\mathrm{v}' - \mathrm{v}$ is timelike or spacelike depending on $d(\mathrm{v}' - \mathrm{v})$ being positive or negative. The case $t = 1$, $s, r = 0$ is Galilean geometry, which is the space-time for Newtonian physics in the same way Lorentzian geometry is for special relativity physics. With the parameter r one can pass continuously from Euclidean through Galilean to Lorentzian.

Now consider the triangle formed by points $\mathrm{v}, \mathrm{v}', \mathrm{v}''$ with

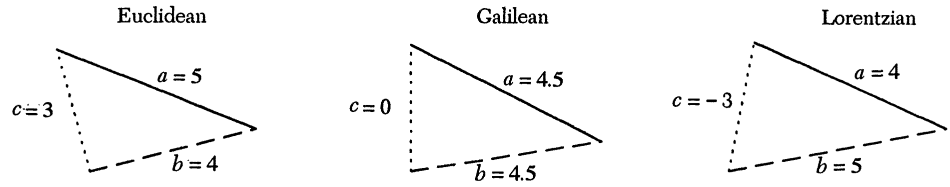
$$a = d(\mathrm{v}'' - \mathrm{v}'), \quad b = d(\mathrm{v}' - \mathrm{v}), \quad c = d(\mathrm{v}'' - \mathrm{v}), \quad \text{and} \quad \langle \mathrm{v}'' - \mathrm{v}, \mathrm{v}' - \mathrm{v} \rangle = 0. \tag{4}$$

The last equation states that $\mathrm{v}'' - \mathrm{v}$ and $\mathrm{v}' - \mathrm{v}$ are orthogonal, so we have a 'right triangle' with hypotenuse of length a and legs of length b and c. By the algebra of inner products, it follows that

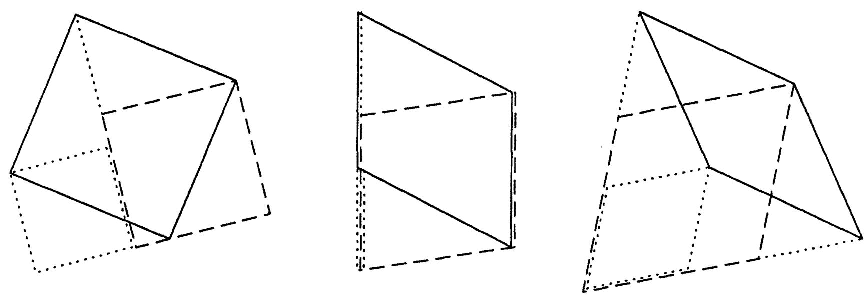
$$\begin{aligned} a^{\ddot{2}} &= \langle \mathrm{v}'' - \mathrm{v}', \mathrm{v}'' - \mathrm{v}' \rangle = \langle (\mathrm{v}'' - \mathrm{v}) - (\mathrm{v}' - \mathrm{v}), (\mathrm{v}'' - \mathrm{v}) - (\mathrm{v}' - \mathrm{v}) \rangle \\ &= \langle \mathrm{v}'' - \mathrm{v}, \mathrm{v}'' - \mathrm{v} \rangle + \langle \mathrm{v}' - \mathrm{v}, \mathrm{v}' - \mathrm{v} \rangle - 2\langle \mathrm{v}'' - \mathrm{v}, \mathrm{v}' - \mathrm{v} \rangle \\ &= b^{\ddot{2}} + c^{\ddot{2}}. \end{aligned} \tag{5}$$

This form of the Pythagorean Theorem thus holds for all the geometries considered.

Examples for three geometries are



'Proofs Without Words' abound for the Euclidean case and corresponding ones exist for the other geometries. For example (the dashed and dotted squares add up to the solid square):



PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

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

Proposals

To be considered for publication, solutions should be received by November 1, 1996.

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

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

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

Let r be a positive real number and let $\Delta A_0B_0C_0$ be equilateral. For each $n \geq 0$ let A_{n+1} and B_{n+1} divide the sides A_nB_n and A_nC_n, respectively, in the internal ratio $r:1$, and set $C_{n+1} = A_n$. If $P = \lim_{n\to\infty} \Delta A_nB_nC_n$, prove that the measures of $\angle B_0PC_0$, $\angle C_0PA_0$, and $\angle A_0PB_0$ form an arithmetic progression.

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

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

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

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

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

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

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

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

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

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

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

Quickies

Answers to the Quickies are on page 229

Q850. *Proposed by Homer White, Pikeville College, Pikeville, Kentucky.*

Do there exist complex polynomials f, g, and h such that every primitive Pythagorean triple (a, b, c) with $0 < a < b < c$ can be represented as $(f(z), g(z), h(z))$ for some $z \in \mathbb{C}$?

Q851. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Prove that among all parallelepipeds of given edge lengths, the rectangular one has the greatest sum of the lengths of the four body diagonals.

Q852. *Proposed by Mihály Bencze, Braşov, Romania.*

Let $z_1, \dots, z_n$ be complex numbers with absolute value 1. Prove that

$$\max_{|z|=1} \prod_{k=1}^{n} |z - z_k| \geq 2,$$

with equality if and only if $z_1, \dots, z_n$ form the vertices of a regular polygon.

Solutions

A Minimal Angle of a Constrained Triangle **June 1995**

1474. *Proposed by G. A. Edgar, The Ohio State University, Columbus, Ohio.*

Consider triangle ABC with side lengths $a = BC$, $b = AC$, and $c = AB$. Suppose r, r', and r'' are positive numbers satisfying

$$r' \leq a, \qquad r'' \leq b \leq r'' + r, \qquad r' \leq c \leq r' + r,$$
$$r' \geq 2r, \qquad r'' \geq 2r, \qquad r'' \leq \tfrac{4}{3}r'.$$

What is the least possible measure of angle A?

Solution by Anchorage Math Solutions Group, University of Alaska Anchorage, Anchorage, Alaska.

We may set $r=1$ (or, equivalently, divide all inequalities by r and rename variables). This gives

$$2 \le r' \le a,$$
$$2 \le r' \le c \le r' + 1,$$
$$2 \le r'' \le b \le r'' + 1 \le \tfrac{4}{3}r' + 1.$$

For given b and c, $\cos A$ is clearly maximized when $a = r'$. Coordinatize with $B=(0,0)$ and $C=(a,0)$, and assume A is in the upper half plane. The feasible region for $A=(x,y)$ is the intersection of the two semi-annuli $a^2 \le x^2+y^2 \le (a+1)^2$ and $4 \le (x-a)^2 + y^2 \le (\frac{4}{3}a+1)^2$. The level curves for $\angle A$ are arcs of circles which have BC as a chord. Since none of the boundaries of the feasible region is tangent to a member of this family, this eliminates all locations for optimal A except the "corners" corresponding to $\{b=\frac{4}{3}a+1, c=a\}$, $\{b=\frac{4}{3}a+1, c=a+1\}$, $\{b=2, c=a+1\}$, and $\{b=2, c=a\}$. In each case, $\cos A$ is a decreasing function of a for $a \ge 2$:

$$\frac{2}{3}+\frac{1}{2a}, \quad \frac{8a^2+21a+9}{12a^2+21a+9}, \quad \frac{1}{2a}+\frac{3}{4a+4}, \text{ and } \frac{1}{a},$$

respectively. The first of the four gives the greatest value at $a=2$, and so the minimal $\angle A$ is $\cos^{-1}\frac{11}{12}$.

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

An Inequality of Greatest Integer Parts of Square Roots **June 1995**

1475. *Proposed by Peter J. Ferraro, Roselle Park, New Jersey.*

Show that

$$\lfloor\sqrt{\alpha}\rfloor + \lfloor\sqrt{\beta}\rfloor + \lfloor\sqrt{\alpha+\beta}\rfloor \ge \lfloor\sqrt{2\alpha}\rfloor + \lfloor\sqrt{2\beta}\rfloor$$

for all real numbers α and β, $\alpha \ge 1$ and $\beta \ge 1$. ($\lfloor x \rfloor$ denotes the greatest integer not exceeding x.)

Solution by Robert L. Doucette, McNeese State University, Lake Charles, Louisiana.

By the concavity of the square root function

$$\sqrt{\alpha+\beta} = \sqrt{\frac{2\alpha+2\beta}{2}} \ge \frac{1}{2}\sqrt{2\alpha} + \frac{1}{2}\sqrt{2\beta} \ge \left\lfloor \frac{1}{2}\sqrt{2\alpha} \right\rfloor + \left\lfloor \frac{1}{2}\sqrt{2\beta} \right\rfloor.$$

It follows that $\lfloor\sqrt{\alpha+\beta}\rfloor \ge \lfloor\frac{1}{2}\sqrt{2\alpha}\rfloor + \lfloor\frac{1}{2}\sqrt{2\beta}\rfloor$. Therefore it suffices to show that

$$\lfloor\sqrt{x}\rfloor + \left\lfloor\frac{1}{2}\sqrt{2x}\right\rfloor \ge \lfloor\sqrt{2x}\rfloor, \text{ for } x \ge 1. \qquad (*)$$

The identity $\lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor = \lfloor 2x \rfloor$ has a straightforward proof. We use it to replace $\lfloor\frac{1}{2}\sqrt{2x}\rfloor$ with $\lfloor\sqrt{2x}\rfloor - \lfloor\frac{1}{2}\sqrt{2x}+\frac{1}{2}\rfloor$ in $(*)$. This yields

$$\lfloor\sqrt{x}\rfloor \ge \left\lfloor\frac{1}{2}\sqrt{2x}+\frac{1}{2}\right\rfloor, \text{ for } x \ge 1.$$

This last inequality follows by noting that $x \ge 4$ implies $(2-\sqrt{2})\sqrt{x} > 1$ or $\sqrt{x} > \frac{1}{2}\sqrt{2x}+\frac{1}{2}$ and $1 \le x < 4$ implies $\frac{1}{2}\sqrt{2x}+\frac{1}{2} < 2$.

Also solved by Anchorage Math Solutions Group, Con Amore Problem Group (Denmark), Qais Haider Darwish (Oman), Edward D. Onstott, Xavier Retnam, I. A. Sakmar, and the proposer.

Limit of an Arc to Chord Ratio **June 1995**

1476. *Proposed by Hugh Thurston, University of British Columbia, Vancouver, British Columbia, Canada.*

Is there a curve and a point P on the curve such that as Q approaches P on the curve, the limit $\lim_{Q \to P}$ arc PQ/chord PQ exists but does not equal 1?

Solution by Michael H. Andreoli, Miami Dade Community College (North), Miami, Florida.

The answer is yes: If the limit exists, it can be any number greater than or equal to 1. For a fixed but arbitrary real number α, define a continuous plane curve on $[0, 1]$ parametrically by

$$Q(t) = (x(t), y(t)) = e^{-1/(1-t)}\left(\cos\frac{\alpha}{(1-t)}, \sin\frac{\alpha}{(1-t)}\right) \text{ for } t \in [0,1)$$

and $Q(1) = P = (0, 0)$.

Then for $t \in [0, 1)$

$$\text{chord } PQ(t) = \sqrt{[x(t)]^2 + [y(t)]^2} = e^{-1/(1-t)}$$

and

$$\text{arc } PQ(t) = \int_t^1 \sqrt{[x'(u)]^2 + [y'(u)]^2}\, du = \sqrt{1+\alpha^2}\, e^{-1/(1-t)}.$$

Therefore,

$$\lim_{Q \to P}\left(\frac{\text{arc } PQ}{\text{chord } PQ}\right) = \lim_{t \to 1^-}\left(\frac{\text{arc } PQ(t)}{\text{chord } PQ(t)}\right) = \sqrt{1+\alpha^2}$$

for any given α.

Comment. Robert L. Doucette points out that an example of such a curve appears in the proposer's article, this MAGAZINE, October 1993, 231–232.

Also solved by Anchorage Math Solutions Group, Adam Coffman, Con Amore Problem Group (Denmark), Robert L. Doucette, Thomas Leong (student), O. P. Lossers (The Netherlands), Howard Morris, Stephen Noltie, Joseph L. Pe, Robert R. Rogers, Seth Zimmerman, Paul J. Zwier, and the proposer. There was one incorrect solution.

A Cotangent Inequality **June 1995**

1477. *Proposed by Wu Wei Chao, He Nan Normal University, Xin Xiang City, He Nan Province, China.*

Show that $\dfrac{1+\sqrt{1-x}}{2x} < \cot x$ for all x in the open interval $(0, 1)$.

Solution by David Doster, Choate Rosemary Hall, Wallingford, Connecticut.

On the interval $(0, 1)$, the inequality $\dfrac{1+\sqrt{1-x}}{2x} < \cot x$ is equivalent to $\tan x < 2 - 2\sqrt{1-x}$, and hence to

$$\tan^2 x - 4\tan x + 4x > 0.$$

Let $f(x) = \tan^2 x - 4\tan x + 4x$. Then

$$f'(x) = 2\tan x \sec^2 x - 4\sec^2 x + 4 = 2\tan x(\tan x - 1)^2 \geq 0.$$

Thus, f increases on $(0,1)$. Since $f(0)=0$, it follows that $f(x)>0$ and the proposed inequality holds.

Also solved by Anchorage Math Solutions Group, Brian D. Beasley, Marc Brodie, Con Amore Problem Group (Denmark), Robert L. Doucette, David L. Farnsworth, Arthur H. Foss, Paul M. Harms, Gerald A. Heuer, Mickey Holsey and Starlett Louis, Joe Howard and John Jeffries, Benjamin G. Klein, Kee-Wai Lau (Hong Kong), O. P. Lossers (The Netherlands), Kim McInturff, Can A. Minh (student), Edward D. Onstott and Noll N. Gurwell, Xaviér Retnam, Noah Rosenberg (student), I. A. Sakmar, Heinz-Jürgen Seiffert (Germany), Trinity University Problem Group, Dennis Walsh, WMC Problems Group, Yan-Loi Wong (Singapore), and the proposer. There were also four incorrect solutions and one incomplete solution.

Generators of an Ideal of Continuous Functions **June 1995**

1478. *Proposed by Piotr Zarzycki, University of Gdańsk, Gdańsk, Poland.*

Let $C[0,1]$ denote the ring of continuous real-valued functions on the closed interval $[0,1]$. Is the set $I=\{f\in C[0,1]\colon f(0)=0 \text{ and } f'_+(0)=0\}$ a countably generated ideal of $C[0,1]$?

I. *Solution by Donald Plank, Stockton College of New Jersey, Pomona, New Jersey.*

The ideal I is not countably generated algebraically. To see this, let $R=C[0,1]$, and let $M=\{g\in R\colon g(0)=0\}$, a maximal ideal in R.

Note first that $f\in I$ if and only if there is some $g\in M$ such that $f(x)=xg(x)$ for $0\le x\le 1$. For, if such a $g\in M$ exists, $f'_+(0)=0$ by the continuity of g at 0, and conversely, if $f\in I$, define $g(0)=0$ and $g(x)=f(x)/x$ for $x>0$.

Note next that $\{g_a(x)\colon a\in A\}$ is a set of generators for M if and only if $\{xg_a(x)\colon a\in A\}$ is a set of generators for I. It suffices, therefore, to show that M is not countably generated.

That M is not countably generated follows from a more general result of Leonard Gillman, *Proceedings of the American Mathematical Society* 11 (1960), 660–666, but we shall provide a separate proof for this case. Assume then that $M=(g_1,g_2,g_3,\dots)$. Replacing each g_n by $2^{-2n}g_n(1+g_n^2)^{-1}$ if necessary, we may assume that $|g_n|^{1/2}\le 2^{-n}$ for each n. Define $h=\sum_{n=1}^{\infty}|g_n|^{1/2}$. Since the series converges uniformly, $h\in R$. Also the zero set of h is $\{0\}$, so that $h\in M$. Therefore, $h=\sum_{n=1}^{m}c_ng_n$ for functions $c_1,c_2,\dots,c_m\in R$. Choose a real number $B>0$ so that $|c_n|\le B$ for $n=1,2,\dots,m$, and let

$$V=\left\{x\in[0,1]\colon |g_n(x)|<B^{-2},\ n=1,2,\dots,m\right\}.$$

Then V is a neighborhood of 0, so there is some $z\in V$ with $z\neq 0$. Since $0<h(z)=\sum_{n=1}^{m}c_n(z)g_n(z)$, we must have $g_n(z)\neq 0$ for some $n\le m$. This guarantees the strict inequality in

$$\begin{aligned} h(z) &\le \sum_{n=1}^{m}|c_n(z)||g_n(z)|\le \sum_{n=1}^{m}B|g_n(z)|^{1/2}|g_n(z)|^{1/2} \\ &< \sum_{n=1}^{m}BB^{-1}|g_n(z)|^{1/2}\le h(z). \end{aligned}$$

So, by contradiction, M is not countably generated.

II. *Solution by Con Amore Problem Group, The Royal Danish School of Educational Studies, Copenhagen, Denmark.*

The ideal I is countably generated topologically.

Let f be an element of I and ε, $0<\varepsilon<1$, be arbitrary. The set of polynomials in one variable with rational coefficients can be ordered into a sequence $p_1, p_2, \dots, p_i, \dots$. According to the Weierstrass approximation theorem, f can be approximated uniformly on $[0,1]$ by real polynomials, and since $\{p_i\}$ is dense in the set of real polynomials, there is then a natural number i such that, for every x in $[0,1]$,

$$|f(x)-p_i(x)|<\frac{\varepsilon}{3}.$$

For each natural number j define a function ϕ_j from $[0,1]$ to $\mathbb{R}$ by

$$\phi_j(x)=\begin{cases} j^2x^2 & \text{for } 0\le x\le 1/j,\\ 1 & \text{for } 1/j\le x\le 1.\end{cases}$$

It is easily seen that $\phi_j\in I$ and $|\phi_j(x)|\le 1$ for $j=1,2,\dots$ and $x\in[0,1]$.

Now, because $f(0)=0$, there is a j for which $0\le x\le 1/j$ implies $|f(x)|<\varepsilon/3$. For $0\le x\le 1/j$ it follows that

$$|p_i(x)|\le|p_i(x)-f(x)|+|f(x)|<\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\frac{2\varepsilon}{3}.$$

Furthermore,

$$|f(x)-\phi_j(x)p_i(x)|\le|f(x)|+|\phi_j(x)||p_i(x)|<\frac{\varepsilon}{3}+1\cdot\frac{2\varepsilon}{3}=\varepsilon.$$

On the other hand, $1/j\le x\le 1$ implies

$$|f(x)-\phi_j(x)p_i(x)|=|f(x)-p_i(x)|<\frac{\varepsilon}{3}.$$

So, for $0\le x\le 1$,

$$|f(x)-\phi_j(x)p_i(x)|<\varepsilon.$$

Since I is an ideal, and $\phi_j(x)\in I$, we have $\phi_j(x)p_i(x)\in I$. Consequently, f can be approximated as closely as one wishes with elements of the following countable subset E of I:

$$E=\{\phi_j p_i\colon j=1,2,\dots;\quad i=1,2,\dots\}.$$

Hence I is countably generated topologically in $C[0,1]$.

Also solved by the proposer.

Answers

Solutions to the Quickies on page 225

A850. No. Suppose that such polynomials f, g, and h exist. For all positive n, $(2n+1, 2n^2+2n, 2n^2+2n+1)$ is a primitive Pythagorean triple with $2n+1<2n^2+2n<2n^2+2n+1$, hence $h(z)-g(z)=1$ for infinitely many z. Thus, $h(z)-g(z)=1$ identically. But $(2n, n^2-1, n^2+1)$ is also a primitive Pythagorean triple with $2n<n^2-1<n^2+1$ for all even $n\ge 4$. Similar reasoning implies that $h(z)-g(z)=2$ identically, a contradiction. Thus the polynomials, f, g, and h do not exist.

A851. Let a, b, and c be the given edge lengths. Let vectors **A**, **B**, and **C** denote three corresponding vectors along three coterminal edges of the parallelepiped. We want to maximize

$$S = |\mathbf{A}+\mathbf{B}+\mathbf{C}| + |\mathbf{A}+\mathbf{B}-\mathbf{C}| + |\mathbf{A}-\mathbf{B}+\mathbf{C}| + |-\mathbf{A}+\mathbf{B}+\mathbf{C}|.$$

We may write

$$\mathbf{B}\cdot\mathbf{C} = bc\cos\alpha, \qquad \mathbf{C}\cdot\mathbf{A} = ca\cos\beta, \qquad \mathbf{A}\cdot\mathbf{B} = ab\cos\gamma,$$

where α, β, and γ denote the angles between the pairs of vectors. Hence,

$$|\pm\mathbf{A}\pm\mathbf{B}\pm\mathbf{C}| = \left(a^2+b^2+c^2 \pm 2bc\cos\alpha \pm 2ca\cos\beta \pm 2ab\cos\gamma\right)^{1/2},$$

where the appropriate $\pm$ signs are chosen. Since $\sqrt{x}$ is concave for $x \geq 0$,

$$S \leq 4\left(a^2+b^2+c^2\right)^{1/2},$$

with equality if and only if the four body diagonals are equal, or, equivalently, if the parallelepiped is rectangular.

A852. Let $P(z) = \prod_{k=1}^{n}(z-z_k)$. Multiplying $z_1, \ldots, z_n$ be a complex number of absolute value 1 does not change $\max_{|z|=1}|P(z)|$. Thus, there is no loss of generality in assuming $P(0)=1$. Now, let $Q(z) = P(z) - z^n - 1$, a polynomial of degree at most $n-1$, and let $\zeta = \exp(2\pi i/n)$. Because $\sum_{k=1}^{n-1}\zeta^{jk} = 0$ for $j = 1, \ldots, n-1$, we have

$$\sum_{k=0}^{n-1} P(\zeta^k) = \sum_{k=0}^{n-1}\left(2 + Q(\zeta^k)\right) = 2n.$$

It follows from the triangle inequality that $\max_{|z|=1}|P(z)| \geq 2$.

Furthermore, equality of the maximum implies that $P(\zeta^k) = 2$, hence $Q(\zeta^k) = 0$, for $k = 0, 1, \ldots, n-1$. Therefore, $Q(z)$ is identically 0. Finally, $|z^n + 1| \leq 2$ for $|z| = 1$ by the triangle inequality.

REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of mathematics literature. Readers are invited to suggest items for review to the editors.

Paulos, John Allen, Dangerous abstractions, *New York Times* (Ntl. Ed.) (7 April 1996) Section 4, Op-Ed. Pinsky, Mark, et al., Letters: In Unabomber case, don't blame mathematics, (12 April 1996) A14. Appleby, Doris, Letters: Mathematics remains remote from crimes, (17 April 1996) A14.

"The Unabomber a mathematician? It figures." That sidebar summarizes Paulos's Op-Ed column. He notes that axiomatic thinking, attention to detail, beautiful ideas blinding one to reality, and abstraction as characteristics of mathematical thinking and also of the antisocial thinking and behavior of accused Unabomber T.J. Kaczynski, a former mathematician. Several respondents find Paulos's comparison to be absurd, though one thinks it must have been a tongue-in-cheek spoof. A more fundamental question may be: Is mathematics particularly attractive to certain kinds of personalities?

Gordon, Carolyn, and David Webb, You can't hear the shape of a drum, *American Scientist* 84 (January–February 1996) 46–55.

Thirty years ago, the late Mark Kac posed the question, "Can one hear the shape of a drum?" The problem is to deduce the shape of the drumhead from the frequencies that the drum can emit. Authors Gordon and Webb discuss the one-dimensional analogue of the vibrating string, cite Milnor's 1964 negative result for Riemannian manifolds, show how group theory can be brought in, and describe their own negative solution for planar regions.

Tucker, Alan C. (ed.), *Models that Work: Case Studies in Effective Undergraduate Mathematics Programs*, MAA, 1996; x + 78 pp, $24 (P). ISBN 0-88385-096-6.

This report summarizes effective practices at some mathematics programs with widely varying missions. Common features were particular "states of mind" held by most or all faculty (respecting students: teaching the students one has, not the ones one would wish for; caring about students; and enjoying one's career as a collegiate educator), plus a curriculum geared toward student needs instead of faculty values, and an interest in a variety of pedagogical approaches.

Stewart, Ian, Mathematical recreations: How fair is Monopoly, *Scientific American* 274 (4) (April 1996) 104–105.

Stewart claims that the game of Monopoly is fair because each player's probability of landing on any square is the same for all squares in the long run. Unfortunately, his Markov-chain model does not take into account the secondary ways of moving about the board, the Chance and Community Chest cards plus the "Go Directly to Jail" square, which induce major nonuniformity. A more sophisticated analysis, including expected income per turn from the various properties, was presented by Robert B. Ash and Richard L. Bishop, "Monopoly as a Markov process," THIS MAGAZINE 45 (1972) 26–29.

Hayes, Brian, A question of numbers, *American Scientist* 84 (January-February 1996) 10–14; Shearer, James B., Letters to the editors: Periods of strings, (May-June 1996) 207.

This article notifies readers of Neil J.A. Sloane and Simon Plouffe's's *Encyclopedia of of Integer Sequences* (1995) and of accompanying electronic "oracles," computer programs that respond to email queries. Sloane's programs attempt to identify a sequence of integers from initial terms (send email to `sequences@research.att.com` with message the word `lookup` followed by the sequence elements, separated by spaces). Plouffe has developed a similar server, called the Inverse Symbolic Calculator (ISC), that recognizes real numbers: You give it the first sixteen decimal digits, and it suggests possible algebraic expressions as sources. You can find instructions on how to use it at `http://www.cecm.sfu.ca/projects/ISC.html`.

Chronis, Peter G., Math professor shares $15 million lotto jackpot. *Denver Post*, 9 April 1996, B3.

"Math professor Celestino Mendez was discussing expected value in his class at Metropolitan State College and remarked that, in a lottery, the expected winning increases when the jackpot gets higher. He told his students that they ought to buy a ticket in the current Colorado Lottery because the expected value was positive (14 cents when you buy a $1 ticket). Professor Mendez thought he should put his money where his mouth is, and so, on the way home, he stopped and bought ten tickets. One of these had the lucky numbers, and he shared the $15 million prize with one other winner."—from *Chance News* 5.05 (29 March—23 April 1996), at `http://www.geom.umn.edu/locate/chance`.

Swetz, Frank, et al., eds., *Learn from the Masters!*, MAA, 1995; x + 303 pp, $23 (P). ISBN 0–88385–703–0.

It was Abel who said, "[I]f one wants to make progress in mathematics one should study the masters." This book is the proceedings of a conference on the history of mathematics that took place in 1988 near where Abel lived briefly and was buried. The book focuses on the use of history as a pedagogical tool in teaching mathematics, with 23 essays on history in school mathematics and in higher mathematics. Everyone who reads this book will learn something that they can take to their classroom and use.

Fagin, Ronald, Moni Naor, and Peter Winkler, Comparing information without leaking it, *Communications of the Association for Computing Machinery* 39 (5) (May 1996) 77–85.

"A group of friends may wish to determine who among them is the oldest, has had the most lovers, or has the extreme value of any potentially embarrassing parameter." Or, as in the real-life situation that prompted this readable article, two managers may want to determine if the same person has complained to both of them about a sensitive matter, without either manager divulging any information if there are in fact two different complainants. The authors describe properties that a solution should offer, then describe 14 solutions of varying degrees of sophistication and satisfaction of the properties.

Wilson, Robin, A successful battle to save a math program, *Chronicle of Higher Education* (12 April 1996) A17, A24.

The Mathematics Dept. and the administration of the University of Rochester have agreed on a compromise that will result in a smaller reduction in staff and reinstate the graduate program in mathematics in 1977. In return, the mathematics faculty will work closely with other departments to improve undergraduate instruction in mathematics.

ERRATUM: In Vol. 69, No. 1 (February 1996), p. 76, the year for the article about Torricelli should be 1994 (thanks to Joe Fiedler).

NEWS AND LETTERS

USA and International Mathematical Olympiads - 1995

Twenty-Fourth Annual USA Mathematical Olympiad Problems and Solutions

1. Let p be an odd prime. The sequence $(a_n)_{n\geq 0}$ is defined as follows: $a_0 = 0$, $a_1 = 1, \ldots, a_{p-2} = p-2$ and, for all $n \geq p-1$, a_n is the least positive integer that does not form an arithmetic sequence of length p with any of the preceding terms. Prove that, for all n, a_n is the number obtained by writing n in base $p-1$ and reading the result in base p.

Solution. Our proof uses the following result.

PROPOSITION *Let $B = \{b_0, b_1, b_2, \ldots\}$, where b_n is the number obtained by writing n in base $p-1$ and reading the result in base p. Then (i) for every $a \notin B$, there exists $d > 0$ so that $a - kd \in B$ for $k = 1, 2, \ldots, p-1$, and (ii) B contains no p-term arithmetic progression.*

Proof. Note that $b \in B$ if and only if the representation of b in base p does not use the digit $p-1$.

(i) Since $a \notin B$, when a is written in base p at least one digit is $p-1$. Let d be the positive integer whose representation in base p is obtained from that of a by replacing each $p-1$ by 1 and each digit other than $p-1$ by 0. Then none of the numbers $a-d$, $a-2d, \ldots, a-(p-1)d$ has $p-1$ as a digit when written in base p, and the result follows.

(ii) Let a, $a+d$, $a+2d, \ldots, a+(p-1)d$ be an arbitrary p-term arithmetic progression of nonnegative integers. Let δ be the rightmost nonzero digit when d is written in base p, and let α be the corresponding digit in the representation of a. Then α, $\alpha+\delta, \ldots, \alpha+(p-1)\delta$ is a complete set of residues modulo p. It follows that at least one of the numbers a, $a+d, \ldots, a+(p-1)d$ has $p-1$ as a digit when written in base p. Hence at least one term of the given arithmetic progression does not belong to B.

Let $(a_n)_{n\geq 0}$ be the sequence defined in the problem. To prove that $a_n = b_n$ for all $n \geq 0$, we use mathematical induction. Clearly $a_0 = b_0 = 0$. Assume that $a_k = b_k$ for $0 \leq k \leq n-1$, where $n \geq 1$. Then a_n is the smallest integer greater than b_{n-1} such that $\{b_0, b_1, \ldots, b_{n-1}, a_n\}$ contains no p-term arithmetic progression. By part (i) of the proposition, $a_n \in B$ so $a_n \geq b_n$. By part (ii) of the proposition, the choice of $a_n = b_n$ does not yield a p-term arithmetic progression with any of the preceding terms. It follows by induction that $a_n = b_n$ for all $n \geq 0$.

2. A calculator is broken so that the only keys that still work are the sin, cos, tan, $\sin^{-1}$, $\cos^{-1}$, and $\tan^{-1}$ buttons. The display initially shows 0. Given any positive rational number q, show that pressing some finite sequence of buttons will yield q. Assume that the calculator does real number calculations with infinite precision. All functions are in terms of radians.

Solution. Since $\cos^{-1}\sin\theta = \pi/2 - \theta$ and $\tan(\pi/2-\theta) = 1/\tan\theta$ for $0 < \theta < \pi/2$, we have for any $x > 0$,

$$\tan\cos^{-1}\sin\tan^{-1}x = \tan(\pi/2 - \tan^{-1}x) = 1/x. \quad (1)$$

Also for $x \geq 0$,

$$\cos\tan^{-1}\sqrt{x} = 1/\sqrt{x+1},$$

so by (1),

$$\tan\cos^{-1}\sin\tan^{-1}\cos\tan^{-1}\sqrt{x} = \sqrt{x+1}. \qquad (2)$$

By (1) and (2), we can obtain $\sqrt{r}$ for any nonnegative rational number r that can be obtained from 0 using the operations

$$x \mapsto x+1 \qquad \text{and} \qquad x \mapsto 1/x.$$

We now prove that every nonnegative rational number r can be so obtained, by induction on the denominator of r. If the denominator is 1, we can obtain the nonnegative integer r by repeated application of $x \mapsto x+1$. Now assume we can get all r with denominator up to n. In particular, we can get any of

$$\frac{n+1}{1}, \frac{n+1}{2}, \ldots, \frac{n+1}{n},$$

so we can also get

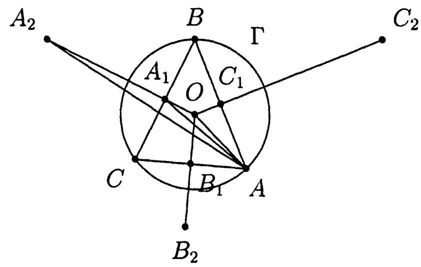
$$\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1},$$

and any positive r of exact denominator $n+1$ can be obtained by repeatedly adding 1 to one of these.

Thus for any positive rational number r, we can obtain $\sqrt{r}$. In particular, we can obtain $\sqrt{q^2} = q$.

3. Given a nonisosceles, nonright triangle ABC, let O denote the center of its circumscribed circle, and let A_1, B_1, and C_1 be the midpoints of sides BC, CA, and AB, respectively. Point A_2 is located on the ray OA_1 so that ΔOAA_1 is similar to ΔOA_2A. Points B_2 and C_2 on rays OB_1 and OC_1, respectively, are defined similarly. Prove that lines AA_2, BB_2, and CC_2 are concurrent, i.e. these three lines intersect at a point.

<u>Solution</u>. Let Γ denote the circumcircle of ΔABC.



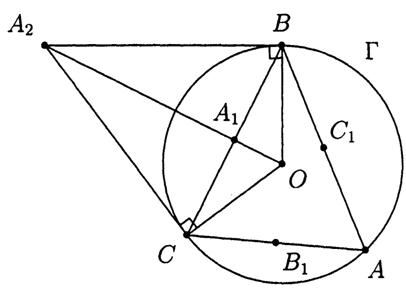
Since ΔOAA_1 and ΔOA_2A are similar, we have

$$\frac{OA_1}{OA} = \frac{OA}{OA_2}$$

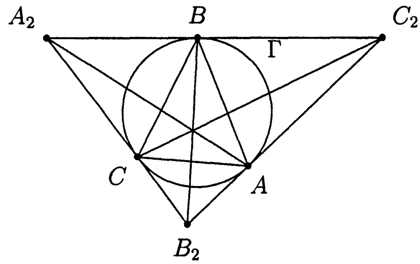
so

$$(OA_1)(OA_2) = (OA)^2 = (OB)^2 = (OC)^2.$$

(Thus A_2 is the image of A_1 with respect to inversion about Γ.) Note that OA_1 is perpendicular to chord BC, and that $OA_2/OB = OB/OA_1$, so ΔOA_2B is similar to ΔOBA_1. Hence the segment A_2B is tangent to Γ at B. Likewise, A_2C is tangent to Γ at C.



Similarly, B_2C and B_2A are tangents to Γ from B_2, and C_2A and C_2B are tangents to Γ from C_2. Thus Γ is the incircle of $\Delta A_2B_2C_2$ with A, B, C as the points of tangency. It is well known and easy to prove (using Ceva's theorem and equal tangents) that the cevians AA_2, BB_2, and CC_2 of $\Delta A_2B_2C_2$ are concurrent.



4. Suppose $q_0, q_1, q_2, \ldots$ is an infinite sequence of integers satisfying the following two conditions:

(i) $m-n$ divides $q_m - q_n$ for $m > n \geq 0$,

(ii) there is a polynomial P such that $|q_n| < P(n)$ for all n.

Prove that there is a polynomial Q such that $q_n = Q(n)$ for all n.

Solution. Let d be the degree of P. There is a polynomial $Q(x)$ of degree at most d and having rational coefficients such that $Q(i) = q_i$ for $i = 0, 1, 2, \ldots, d$. An explicit expression for this polynomial is

$$Q(x) = q_0L_0(x) + q_1L_1(x) + \cdots + q_dL_d(x),$$

where

$$L_i(x) = \prod_{\substack{0 \leq j \leq d \\ j \neq i}} \left(\frac{x-j}{i-j}\right).$$

(This is the form of the interpolating polynomial due to Lagrange.) We shall prove that $q_n = Q(n)$ for all $n \geq 0$.

Let $k \geq 1$ be a common denominator for the coefficients of Q, and for $n = 0, 1, 2, \ldots$ set $r_n = k(Q(n) - q_n)$. Then $r_i = 0$ for $i = 0, 1, \ldots, d$. It is well known that for any polynomial $\mathcal{P}$ with integer coefficients, and any distinct integers m, n,

$$m - n \text{ divides } \mathcal{P}(m) - \mathcal{P}(n).$$

The polynomial $kQ(n)$ has integer coefficients and $m-n$ divides $q_m - q_n$, so $m-n$ divides $r_m - r_n$ for all $m > n \geq 0$. Also, there is a polynomial R of degree at most d such that $|r_n| < R(n)$ for all $n \geq 0$. This follows from the fact that P and Q have degree at most d and

$$|r_n| \leq k(|Q(n)| + |q_n|) < k(|Q(n)| + P(n)).$$

Thus $|r_n| < an^d + b$ for all $n \geq 0$ if a and b are appropriately large constants. For any $n > d$ and $0 \leq i \leq d$,

$$n - i \text{ divides } r_n - r_i = r_n,$$

so $\mathrm{LCM}(n, n-1, \ldots, n-d)$ divides r_n. We claim that

$$\mathrm{LCM}(n, n-1, \ldots, n-d) > R(n)$$

for all sufficiently large n, say for all $n \geq N$. Since $\mathrm{LCM}(n, n-1, \ldots, n-d)$ divides r_n and $R(n) > |r_n|$, the truth of this claim implies that $r_n = 0$ for all $n \geq N$. Now $r_n = 0$ for $n < N$ as well, since for arbitrary large $m \geq N$,

$$m - n \text{ divides } r_m - r_n = -r_n.$$

Hence $r_n = 0$ for all $n \geq 0$. By definition of r_n, we get $q_n = Q(n)$ for all $n \geq 0$.

Proof of the claim. We use the fact that for any finite sequence of positive integers $a_1, a_2, \ldots, a_m$,

$$\mathrm{LCM}(a_1, a_2, \ldots, a_m) \geq \frac{\prod_{i=1}^{m} a_i}{\prod_{1 \leq i < j \leq m} \mathrm{GCD}(a_i, a_j)}.$$

To see this, let p be an arbitrary prime and for $i = 1, 2, \ldots, m$ let e_i be the exponent of p in the prime factorization of a_i.

Then the exponent of p in the prime factorization of $\mathrm{LCM}(a_1, \ldots, a_m)$ is $\max(e_1, e_2, \ldots, e_m)$, and this is clearly at least as large as $\sum_{i=1}^{m} e_i - \sum_{1\le i<j\le m} \min(e_i, e_j)$, which is the exponent of p in the prime factorization of the right hand side of (1). Since

$$\prod_{0\le i<j\le d} \mathrm{GCD}(n-i, n-j) = \prod_{0\le i<j\le d} \mathrm{GCD}(n-i, j-i) \le \prod_{0\le i<j\le d} (j-i),$$

the degree of R is d, and the degree of $n(n-1)\cdots(n-d)$ is $d+1$, it follows that

$$\mathrm{LCM}(n, n-1, \ldots, n-d) \ge \frac{n(n-1)\cdots(n-d)}{\prod_{0\le i<j\le d}(j-i)} > R(n)$$

for all sufficiently large n.

5. Suppose that in a certain society, each pair of persons can be classified as either *amicable* or *hostile.* We shall say that each member of an amicable pair is a *friend* of the other, and each member of a hostile pair is a *foe* of the other. Suppose that the society has n persons and q amicable pairs, and that for every set of three persons, at least one pair is hostile. Prove that there is at least one member of the society whose foes include $q(1-4q/n^2)$ or fewer amicable pairs.

Solution. Graph theory provides an appropriate model. Represent each person in the society by a vertex, and an amicable pair by an edge joining the two given vertices. The set of all vertices is denoted by V. Two vertices that are joined by an edge are said to be *adjacent* and the set of all vertices adjacent to a given vertex v is called the *neighborhood* of v, denoted by $N(v)$. The number of vertices in $N(v)$ is the *degree* of v, denoted by $\deg(v)$. It is a well-known fact that for a graph with q edges,

$$\sum_{w\in V} \deg(w) = 2q.$$

Thus by Cauchy's inequality,

$$n \sum_{w\in V} (\deg(w))^2 \ge \left(\sum_{w\in V} \deg(w)\right)^2 = 4q^2. \qquad (1)$$

In the language of graph theory, we are to prove that in any triangle-free graph with n vertices and q edges, there is at least one vertex v_0 such that the number of edges joining pairs of vertices distinct from v_0 and not in $N(v_0)$ is at most $q(1-4q^2/n)$. Since the graph is triangle-free, for any vertex v the number of edges joining pairs of vertices distinct from v and not in $N(v)$ is $q - f(v)$ where $f(v) = \sum_{w\in N(v)} \deg(w)$. Thus we need to prove that there is a vertex v_0 such that $f(v_0) \ge (2q/n)^2$. Using (1), we find that the average value of f is

$$\begin{aligned} \overline{f} &= \frac{1}{n}\sum_{v\in V} f(v) \\ &= \frac{1}{n}\sum_{v\in V}\sum_{w\in N(v)} \deg(w) \\ &= \frac{1}{n}\sum_{w\in V} \deg(w) \sum_{v\in N(w)} 1 \\ &= \frac{1}{n}\sum_{w\in V} (\deg(w))^2 \\ &\ge 4q^2/n^2. \end{aligned}$$

Thus we know that some vertex v_0 satisfies $f(v_0) \ge (2q/n)^2$.

Thirty-Sixth Annual International Mathematical Olympiad Problems

1. Let A, B, C and D be four distinct points on a line, in that order. The circles with diameters AC and BD intersect at the points X and Y. The line XY meets BC at the point Z. Let P be a point on the line XY different from Z. The line CP intersects the circle with diameter AC at the points C and M, and the line BP intersects the circle with diameter BD at the points B and N. Prove that the lines AM, DN and XY are concurrent.

2. Let a, b and c be positive real numbers such that $abc = 1$. Prove that

$$\frac{1}{a^3(b+c)} + \frac{1}{b^3(c+a)} + \frac{1}{c^3(a+b)} \geq \frac{3}{2}.$$

3. Determine all integers $n > 3$ for which there exist n points $A_1, A_2, \ldots, A_n$ in the plane, and real numbers $r_1, r_2, \ldots, r_n$ satisfying the following two conditions:

(i) no three of the points $A_1, A_2, \ldots, A_n$ lie on a line;

(ii) for each triple i, j, k ($1 \leq i < j < k \leq n$) the triangle $A_iA_jA_k$ has area equal to $r_i + r_j + r_k$.

4. Find the maximum value of x_0 for which there exists a sequence of positive real numbers $x_0, x_1, \ldots, x_{1995}$ satisfying the two conditions:

(i) $x_0 = x_{1995}$;

(ii) $x_{i-1} + 2/x_{i-1} = 2x_i + 1/x_i$ for each $i = 1, 2, \ldots, 1995$.

5. Let $ABCDEF$ be a convex hexagon with

$$AB = BC = CD, \quad DE = EF = FA,$$

and

$$\angle BCD = \angle EFA = 60°.$$

Let G and H be two points in the interior of the hexagon such that $\angle AGB = \angle DHE = 120°$. Prove that

$$AG + GB + GH + DH + HE \geq CF.$$

6. Let p be an odd prime number. Find the number of subsets A of the set $\{1, 2, \ldots, 2p\}$ such that

(i) A has exactly p elements, and

(ii) the sum of all the elements in A is divisible by p.

Notes

The 1995 USA Mathematical Olympiad was prepared by Titu Andreescu, Anne Hudson, James Propp, Cecil Rousseau (chair), Richard Stong, and Paul Zeitz.

The top eight students in the 1995 USAMO were:

Christopher Chang	Palo Alto, CA
Jay Chyung	Iowa City, IA
Samit Dasgupta	Silver City, MD
Andrei Gnepp	Orange, OH
Craig Helfgott	New York, NY
Aleksandr L. Khazanov	New York, NY
Jacob A. Lurie	Silver Spring, MD
Josh Nichols-Barrer	Newton, MA

The 36th International Mathematical Olympiad was held in Toronto, Canada during mid-July. Members of the USA team were Christopher Chang, Jay Chyung, Andrei Gnepp, Aleksandr Khazanov, Jacob Lurie, and Josh Nichols-Barrer.

The training session to prepare the USA team for the IMO was held at the Illinois Mathematics and Science Academy in Aurora, Illinois. Titu Andreescu, Elgin Johnston, and Paul Zeitz served as instructors

The booklet *Mathematical Olympiads 1995* presents additional solutions to problems on the 24th USAMO and solutions to the 36th IMO. This booklet is available from:

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

Such a booklet has been published every year since 1976. Copies are \$5.00 for each year 1976–1994.

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

This report was prepared by Cecil Rousseau, Memphis State University.

Letters to the Editor

Dear Editor:

There is a small correction to be made in Dixon J. Jones's article *Continued powers and a sufficient condition for their convergence*, this MAGAZINE, December 1995, pp. 387-92. In his Example III, the author offers an example of a continued square that converges but that doesn't satisfy the sufficient condition of the main theorem of the paper. Specifically, he says that the continued square defined by $x_n = (4^n + 1)/4^{n+1}$ does not satisfy his convergence condition for continued squares. That is, he asserts (without proof) that $4x_n^{2^n}$ is unbounded, even though the sequence is dominated by the terms of the convergent continued square in his Example II, and thus the corresponding continued square converges anyway. But the sequence defined by $(1 + 1/4^n)^{2^n}$, which he says is unbounded, is clearly bounded by $(1 + 1/2^n)^{2^n}$, which in turn is always less than e. So his example does not show that his condition is not necessary.

Of course, this doesn't contradict the major point of the paper, that being to show the sufficiency of the condition. However, whether the condition is necessary or not is still up in the air.

Josh Nichols-Barrer
24 Hazelton Rd.
Newton, MA 02159

Dear Editor:

Thanks to Martin E. Muldoon and Abraham A. Ungar (*Beyond sin and cos*, this MAGAZINE, February 1996, pp. 3–14) for their interesting review of some higher-order generalizations of circular and hyperbolic trigonometric functions, including a helpful list of references.

I point out that the special case of skew-circulant matrices has found an engineering application in the field of fluid dynamics; see P. Kelly and R. L. Panton, *Three-dimensional potential flows from functions of a 3-d complex variable*, Fluid Dynam. Res. 6 (1990), pp. 119-137, and E. Dale Martin, *Some elements of a theory of multidimensional complex variables: Parts 1 & 2*, J. Franklin Inst. 326 (1989), pp. 611–647, 649–681.

Greg Demers
1321 Jerome St.
Lansing, MI 48912

Dear Editor:

In regard to Dieter Jungnickel's article (*On the order of a product in a finite abelian group*, this MAGAZINE, February 1996, pp. 53–56), there is a sort of formula for the order which he does not write down explicitly. This "formula" surely must be well known and probably has been published before, but I have not seen it.

Suppose a and b are elements of an abelian group with finite order $o(a) = \prod_{i=1}^{r} p_i^{a_i}$ and $o(b) = \prod_{i=1}^{r} p_i^{b_i}$, where $p_1, \ldots, p_r$ are distinct primes and $a_1, \ldots, b_r$ are non-negative integers. Then the order of ab is $\prod_{i=1}^{r} p_i^{c_i}$, where $c_i = \max(a_i, b_i)$ when $a_i \neq b_i$, and $0 \leq c_i \leq a_i$ when $a_i = b_i$. Moreover, for those i for which $a_i = b_i$, any allowable c_i can occur.

This "formula" would certainly make an interesting problem for students; both proving it and finding it would be a challenge.

Allen D. Bell
University of Wisconsin-Milwaukee
Milwaukee, WI 53201

Editor's Correction

The first paragraph of Robin Chapman's note *A polynomial taking integer values*, this MAGAZINE, April 1996, p. 121, contains two serious typesetting errors, one of which involves the title polynomial. The affected sentences should read as follows:

In [2] Sury proves that for integers $a_1 < a_2 < \cdots < a_n$, the expression

$$\prod_{n \geq i > j \geq 1} \frac{a_i - a_j}{i - j}$$

is also an integer. ... Sury gives an elementary but indirect proof, based on the stronger result that

$$\prod_{n \geq i > j \geq 1} \frac{X^{a_i - a_j} - 1}{X^{i-j} - 1} \in \mathbb{Z}[x].$$

We regret these errors.

Ed.

gineering application in the field of fluid dynamics; see P. Kelly and R. L. Panton, *Three-dimensional potential flows from functions of a 3-d complex variable*, Fluid Dynam. Res. 6 (1990), pp. 119-137, and E. Dale Martin, *Some elements of a theory of multidimensional complex variables: Parts 1 & 2*, J. Franklin Inst. 326 (1989), pp. 611–647, 649–681.

Greg Demers
1321 Jerome St.
Lansing, MI 48912

Dear Editor:

In regard to Dieter Jungnickel's article (*On the order of a product in a finite abelian group*, this MAGAZINE, February 1996, pp. 53–56), there is a sort of formula for the order which he does not write down explicitly. This "formula" surely must be well known and probably has been published before, but I have not seen it.

Suppose a and b are elements of an abelian group with finite order $o(a) = \prod_{i=1}^{r} p_i^{a_i}$ and $o(b) = \prod_{i=1}^{r} p_i^{b_i}$, where $p_1, \ldots, p_r$ are distinct primes and $a_1, \ldots, b_r$ are non-negative integers. Then the order of ab is $\prod_{i=1}^{r} p_i^{c_i}$, where $c_i = \max(a_i, b_i)$ when $a_i \neq b_i$, and $0 \leq c_i \leq a_i$ when $a_i = b_i$. Moreover, for those i for which $a_i = b_i$, any allowable c_i can occur.

This "formula" would certainly make an interesting problem for students; both proving it and finding it would be a challenge.

Allen D. Bell
University of Wisconsin-Milwaukee
Milwaukee, WI 53201

Editor's Correction

The first paragraph of Robin Chapman's note *A polynomial taking integer values*, this MAGAZINE, April 1996, p. 121, contains two serious typesetting errors, one of which involves the title polynomial. The affected sentences should read as follows:

In [2] Sury proves that for integers $a_1 < a_2 < \cdots < a_n$, the expression

$$\prod_{n \geq i > j \geq 1} \frac{a_i - a_j}{i - j}$$

is also an integer. ... Sury gives an elementary but indirect proof, based on the stronger result that

$$\prod_{n \geq i > j \geq 1} \frac{X^{a_i - a_j} - 1}{X^{i-j} - 1} \in \mathbb{Z}[x].$$

We regret these errors.

Ed.

MicroCalc, ver. 7.0

Interactive Calculus Software

MicroCalc covers almost all topics needed for teaching and learning calculus: single variable, several variable, differential equations. (Programmed by Harley Flanders)

MicroCalc is menu-driven; there is no language to learn; its input is the way you write mathematics. *MicroCalc* is dedicated to calculus, and has desired calculus topics ready to run. Examples: generation of sine and tangent, graphs of Riemann sums, solids of revolution by slabs and by shells, implicit curves and surfaces, Newton steps for solving equations, graphical Lagrange multipliers. About 70 other topics, symbolic, numerical, graphical, are on *MicroCalc's* menus, plus several utilities.

MS-DOS platform only. 3.5" diskette with *install* program.

Site license fees:
First 50 workstations: $850; up to 50 additional: $10 each; still more workstations: $5 each. Upgrade from previous version: 40% off. (First class mail free. Express mail at cost.)

Sample disk:
$15 paid in advance; $25 by purchase order.

Purchase orders:
(Please include the full name of the licensing school/department.)

MathCalcEduc
1449 Covington Drive
Ann Arbor, MI 48103-5630

Federal EIN: 38-2740157
Telephone: 313 761 4666

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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