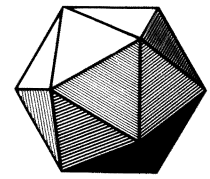
Vol. 65 No. 2, April 1992



MATHEMATICS MAGAZINE

$$\varphi(n)\ \varphi(n)\ \sigma(n)\ \sigma(n)\ \tau(n)\ \tau(n)$$

$$\sigma(n)\ \tau(n)\ \varphi(n)\ \tau(n)\ \sigma(n)\ \varphi(n)$$

$$\tau(n)\ \sigma(n)\ \tau(n)\ \varphi(n)\ \varphi(n)\ \sigma(n)$$

$$\varphi(n)\ \varphi(n)\ \sigma(n)\ \sigma(n)\ \tau(n)\ \tau(n)$$

$$\sigma(n)\ \tau(n)\ \varphi(n)\ \tau(n)\ \sigma(n)\ \varphi(n)$$

$$\tau(n)\ \sigma(n)\ \tau(n)\ \varphi(n)\ \varphi(n)\ \sigma(n)$$

$$\varphi * \tau = \sigma$$

- Number-Theoretic Functions via Convolution Rings
- From Intermediate Value Theorem to Chaos
- Tetrahedra with Integer Edges and Integer Volume

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

The aim of *Mathematics Magazine* is 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. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these are essential to the presentation of some good mathematics. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

The full statement of editorial policy appears in this *Magazine*, Vol. 64, pp. 71–72, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher.

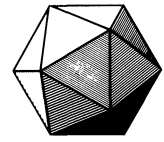
Send new manuscripts to: Martha Siegel, Editor, Mathematics Magazine, Towson State University, Towson, MD 21204. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and two copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added. Do not use staples.

AUTHORS

Sterling K. Berberian received his Ph.D. from the University of Chicago in 1955. The seed for the present article was planted in 1983 when the author received a review copy of the manuscript of Gerhard Hochschild's remarkable *Perspectives of Elementary Mathematics*. In his reply to Walter Kaufmann Bühler of Springer-Verlag, the author expressed the opinion that the clever presentation of number-theoretic functions had the makings of one of those long MAA expository articles, and, so he thought, forgot about it. Surreptitiously, the topic began to take root in his undergraduate algebra courses, a draft was written up to pass out to students, and eventually, demanded to be drafted to a wider audience. There is always a reason for love at first sight; the author's exposure to the convolution of functions in analysis probably predisposed him to be swept away by the pretty analogue for number-theoretic functions.

Xun-Cheng Huang has earned his degrees from Shanghai Jiaotong University (M.S., 1982), and Marquette University (M.S., 1987, Ph.D., 1988). He currently is an Associate Professor of Mathematics at New Jersey Institute of Technology. His research interests include biomathematical modeling, differential equations, integrable, and chaotic dynamical systems.

Vol. 65 No. 2, April 1992



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

Number-Theoretic Functions via Convolution Rings

S. K. BERBERIAN
University of Texas
Austin, TX 78712

For Peter John Durbin

We've all met and befriended at least one of the 'number-theoretic functions' τ, σ, φ: for a positive integer n,

$\tau(n)$ = the number of divisors of n,
$\sigma(n)$ = the sum of the divisors of n,
$\varphi(n)$ = the number of integers k $(1 \leqslant k \leqslant n)$ that are relatively prime to n.

Here 'divisor' means positive integral divisor; and to say that k is relatively prime to n—usually written $(k, n) = 1$—means that 1 is the only common divisor of k and n. For example,

$$\tau(6) = |\{1, 2, 3, 6\}| = 4$$

(vertical bars around a finite set count the number of elements of the set),

$$\sigma(6) = 1 + 2 + 3 + 6 = 12,$$

and $\varphi(6) = |\{1, 5\}| = 2$.

There is a beautiful formula that relates these three functions:

$$\varphi * \tau = \sigma.$$

What does it mean? The left side alludes to a way of combining two functions to form a third—a law of composition for functions. Here's how we evaluate the function $\varphi * \tau$ at a positive integer n: We take a divisor d of n, form the product $\varphi(d)\tau(n/d)$ (so to speak, n/d is the divisor of n 'complementary to d'), and we sum these products over all possible divisors d of n. For example,

$$(\varphi * \tau)(6) = \varphi(1)\tau(6) + \varphi(2)\tau(3) + \varphi(3)\tau(2) + \varphi(6)\tau(1)$$

$$= 1 \cdot 4 + 1 \cdot 2 + 2 \cdot 2 + 2 \cdot 1 = 12,$$

which equals $\sigma(6)$ (a miracle!). Try it for $n = 12$.

Why does it work? One way to see it would be to derive formulas for φ, τ, and σ and verify that the equation is true (in the last section, we give a one-line proof). Both aspects—deriving the formulas and verifying the equation—call on a property of the functions that is not readily apparent: They are multiplicative. This means that

$$\tau(mn) = \tau(m)\tau(n) \quad \text{whenever } (m, n) = 1,$$

and similarly for the functions σ and φ (all of these facts will be proved below). In view of the Fundamental Theorem of Arithmetic (factorization into powers of primes, unique apart from the order of the factors), this reduces the computation of, say, $\varphi(n)$, to the computation of $\varphi(p^k)$, where p is prime and k is a positive integer. For example,

$$\begin{aligned}\varphi(60) &= \varphi(2^2 \cdot 3 \cdot 5) = \varphi(2^2)\varphi(3 \cdot 5) \\ &= \varphi(4)\varphi(3)\varphi(5) = 2 \cdot 2 \cdot 4 = 16.\end{aligned}$$

To reduce the verification of $(\varphi * \tau)(n) = \sigma(n)$ to the case that $n = p^k$, we need to know that $\varphi * \tau$ is also multiplicative. And therein lies the tale....

The functions τ, σ, φ, and a host of other interesting functions live in a fascinating ring A. The elements of A are the functions $f\colon \mathbb{P} \to \mathbb{Z}$, where $\mathbb{P}$ is the set of positive integers and $\mathbb{Z}$ is the ring of integers:

$$\mathbb{P} = \{1,2,3,\ldots\}, \qquad \mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}.$$

Functions f, g in A are added 'pointwise',

$$(f+g)(n) = f(n) + g(n),$$

and their product $f * g$, called the *convolution* of f and g, is defined by the formula

$$(f * g)(n) = \sum_{d|n} f(d)\, g(n/d)$$

(the sum extends over all *positive* integral divisors d of n). We will see that A, equipped with the operations $f+g$ and $f * g$, is a commutative ring with a unity element u (the only computation of any substance is the associative law for convolution). The set of units (= invertible elements) of the ring A is a group, called the *group of units* of A and denoted $\mathrm{U_A}$:

$$\mathrm{U_A} = \{f \in \mathrm{A}\colon f * g = u \text{ for some } g \in \mathrm{A}\}.$$

It turns out that the units of A are the functions $f \in \mathrm{A}$ such that $f(1) = \pm 1$. A function $f \in A$ will be called *multiplicative* if $f(1) = 1$ and $f(mn) = f(m)f(n)$ whenever $(m, n) = 1$. The group of units includes all of the multiplicative functions f (simply because $f(1) = 1$). The centerpiece of the theory is the following theorem:

The multiplicative functions form a **subgroup** *of the group of units of* A.

Some of the preliminaries will be presented in slightly greater generality: We will look at convolution rings $\mathrm{A} = \mathrm{A_R(S)}$ of functions $f\colon \mathrm{S} \to \mathrm{R}$, for certain monoids S and for certain rings R.[1] The motive is twofold: to clear the air of irrelevant special properties of the monoid $\mathrm{S} = \mathbb{P}$ and the ring $\mathrm{R} = \mathbb{Z}$, and to point the way for further exploration. The reader who prefers to leave the generality for another day can substitute $\mathbb{P}$ for S, and $\mathbb{Z}$ for R; very little substance (and none of the fun) will be lost by doing so. With or without the generality, the topic is technically within the reach of (and, judging from my experience, greatly enjoyed by) an undergraduate class in abstract algebra, toward the end of the semester. This article is written in the hope of encouraging students and teachers to give it a try.

[1]The exposition is inspired by that of G. Hochschild, *Perspectives of Elementary Mathematics* (Springer-Verlag, 1983), Chapter 2.

1. Counting Divisors (The Function τ)

In the first two sections we prove a core of facts from number theory 'with our bare hands' (no fancy techniques); these pertain to the functions τ and φ described in the introduction.

LEMMA. *For each positive integer n, write D_n for the set of positive integral divisors of n:*

$$D_n = \{x \in \mathbb{P}: x|n\}.$$

For every pair of positive integers m and n, the formula $f(x,y) = xy$ defines a surjective mapping $f: D_m \times D_n \to D_{mn}$.

Proof. If x and y are divisors of m and n, respectively, then xy is a divisor of mn, thus the formula $f(x,y) = xy$ does indeed define a mapping $f: D_m \times D_n \to D_{mn}$; it remains to prove that f is surjective. Assuming $z|mn$, we seek a factorization $z = xy$ with $x|m$ and $y|n$. Say $mn = tz$. Let $x = (m,z)$ (the greatest common divisor of m and z) and write $m = xm_1$, $z = xz_1$. We have

$$(xm_1)n = mn = tz = t(xz_1),$$

so $m_1 n = tz_1$. Since $z_1|m_1 n$ and $(z_1, m_1) = 1$ we have $z_1|n$, and the desired factorization $z = xy$ is obtained by setting $y = z_1$.

THEOREM 1. *If m and n are relatively prime positive integers, then the mapping $f: D_m \times D_n \to D_{mn}$ of the lemma is bijective.*

Proof. In view of the lemma, it suffices to show that f is injective. Assuming $xy = x'y'$, where x, x' are divisors of m, and y, y' are divisors of n, we must show that $x = x'$ and $y = y'$. Since $x|m$, $y'|n$ and $(m,n) = 1$, we have also $(x,y') = 1$; but x divides $xy = x'y'$, so necessarily $x|x'$. Similarly $x'|x$, so $x = x'$; then $y = y'$ by cancellation.

Remark. The converse of Theorem 1 is true: If f is injective then $(m,n) = 1$. (Hint: Assuming $(m,n) = d > 1$, show that f is not injective. Write $m = m_1 d$, $n = n_1 d$ and look at $m_1 n = m n_1$.)

Definition 1. For every positive integer n, the *number* of (positive integral) divisors of n is denoted $\tau(n)$; thus $\tau(n) = |D_n|$, where $D_n = \{k \in \mathbb{P}: k|n\}$.

Examples. $\tau(6) = 4$ because $D_6 = \{1,2,3,6\}$; $\tau(1) = 1$; for $n > 1$, $\tau(n) \geqslant 2$ because n has at least the divisors 1 and n; $\tau(7) = 2$ because $D_7 = \{1,7\}$; $\tau(n) = 2 \Leftrightarrow n$ is a prime number.

The key property of τ:

THEOREM 2. $\tau(mn) = \tau(m)\tau(n)$ *whenever* $(m,n) = 1$.

Proof. If $(m,n) = 1$ then, by Theorem 1, $|D_{mn}| = |D_m \times D_n| = |D_m| \cdot |D_n|$.

Definition 2. A function $f: \mathbb{P} \to \mathbb{Z}$ is called a *number-theoretic function*; f is said to be *multiplicative* if $f(1) = 1$ and if $f(mn) = f(m)f(n)$ whenever $(m,n) = 1$.

Remark. If $f(mn) = f(m)f(n)$ whenever $(m,n) = 1$ and if $f(1) \neq 1$, then f is identically zero. (Look at $f(n \cdot 1) = f(n)f(1)$.)

Multiplicativity is a powerful tool for computation:

COROLLARY. *If n is a positive integer with prime-power factorization*

$$n = p_1^{k_1} \cdots p_r^{k_r}$$

(where $p_1, \ldots, p_r$ are distinct primes and $k_1, \ldots, k_r$ are positive integers) then

$$\tau(n) = \prod_{i=1}^{r} (k_i + 1).$$

Proof. By Theorem 2, it suffices to show that $\tau(p^k) = k + 1$ when p is prime and k is a positive integer. The set of divisors of p^k is $\{1, p, p^2, \ldots, p^k\}$.

2. Euler's φ-function

According to the definition in the introduction,

$$\varphi(n) = |\{k \in \mathbb{P}: 1 \leqslant k \leqslant n, (k, n) = 1\}|.$$

The derivation of the properties of φ is expedited by finding another formula for it.

LEMMA 1. *If n is an integer greater than 1, $\mathbb{Z}_n = \mathbb{Z}/(n)$ is the ring of integers modulo n and $\mathrm{U}_{\mathbb{Z}_n}$ is the group of units of $\mathbb{Z}_n$, then the order of $\mathrm{U}_{\mathbb{Z}_n}$ is $\varphi(n)$.*

Proof. For any integer k, write $\bar{k} = k + (n)$ for the equivalence class of k modulo n, where $(n) = \mathbb{Z}n$ is the set of all integral multiples of n. By Euclid's algorithm for calculating the greatest common divisor, k and n are relatively prime if and only if $1 = rk + sn$ for suitable integers r and s. On passing to quotients modulo n, this means that $\bar{1} = \bar{r}\bar{k}$ for some integer r, that is, $\bar{k}$ is invertible in the quotient ring $\mathbb{Z}_n$. Thus, as k runs over the integers from 1 to $n - 1$ that are relatively prime to n, $\bar{k}$ runs over the group of units of $\mathbb{Z}_n$.

LEMMA 2. *If m and n are relatively prime integers greater than 1, then $\mathbb{Z}_{mn}$ and $\mathbb{Z}_m \times \mathbb{Z}_n$ are isomorphic as rings.*

Proof. Define a mapping $f: \mathbb{Z} \to \mathbb{Z}_m \times \mathbb{Z}_n$ by the formula $f(k) = (k + (m), k + (n))$; this is easily seen to be a ring homomorphism, with kernel $(m) \cap (n) = ([m, n])$, where $[m, n]$ is the least common multiple of m and n. Since $mn = (m, n)[m, n]$ and $(m, n) = 1$ we have $mn = [m, n]$, thus the kernel of f is (mn). By the First Isomorphism Theorem, passage to quotients modulo the kernel yields a monomorphism $\mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$. But

$$|\mathbb{Z}_{mn}| = mn = |\mathbb{Z}_m| \cdot |\mathbb{Z}_n| = |\mathbb{Z}_m \times \mathbb{Z}_n|,$$

so the mapping is necessarily surjective.

Remark. With notations as in Lemma 2, if a and b are any two integers then there exists an integer k such that $k + (m) = a + (m)$ and $k + (n) = b + (n)$ (that's the meaning of surjectivity in the preceding proof); in other words $k \equiv a \pmod{m}$ and $k \equiv b \pmod{n}$. In this form, the lemma is known as the Chinese Remainder Theorem.

LEMMA 3. *If R and S are rings with unity and $\mathrm{R} \times \mathrm{S}$ is the product ring, then $\mathrm{U}_{\mathrm{R}\times\mathrm{S}} = \mathrm{U}_{\mathrm{R}} \times \mathrm{U}_{\mathrm{S}}$.*

Proof. Products in $\mathrm{R} \times \mathrm{S}$ are defined by the formula $(x, y)(x', y') = (xx', yy')$; such a product is equal to the unity element $(1, 1)$ if and only if $xx' = 1$ and $yy' = 1$.

THEOREM 3. *If m and n are relatively prime positive integers, then $\varphi(mn) = \varphi(m)\varphi(n)$.*

Proof. If $m = 1$ or $n = 1$ the equation is trivial. Suppose m and n are $\geqslant 2$. In view of Lemmas 1–3, the asserted equation is immediate from the formula for the number of elements in a product set.

On the way to a formula for actually computing $\varphi(n)$:

LEMMA 4. *For integers $a, b_1, \ldots, b_r$,*

$$(a, b_1 \cdots b_r) = 1 \quad \Leftrightarrow \quad (a, b_i) = 1 \quad \textit{for all } i.$$

Proof. In words, an integer is relatively prime to each of a list of integers if and only if it is relatively prime to their product. For $a = 0$ or $a = \pm 1$ the asserted equivalence is trivial. Assuming $a \geqslant 2$, write $u_i = b_i + (a)$ for the class of b_i modulo a. By the proof of Lemma 1, the assertion is that

$$u_1 \cdots u_r \text{ is a unit of } \mathbb{Z}_a \quad \Leftrightarrow \quad u_1, \ldots, u_r \text{ are units of } \mathbb{Z}_a,$$

which is obviously true.

THEOREM 4. *If n is a positive integer with prime-power factorization $n = p_1^{k_1} \cdots p_r^{k_r}$, then*

$$\varphi(n) = \prod_{i=1}^{r} \left(p_i^{k_i} - p_i^{k_i - 1}\right).$$

Proof. By Theorem 3 we need only show that $\varphi(p^k) = p^k - p^{k-1}$ for p prime and k a positive integer. Let $1 \leqslant a \leqslant p^k$. By Lemma 4,

$$(a, p^k) = 1 \quad \Leftrightarrow \quad (a, p) = 1;$$

since p is prime, this means that

$$(a, p^k) > 1 \quad \Leftrightarrow \quad (a, p) = p \quad \Leftrightarrow \quad a \text{ is a multiple of } p.$$

Let $\mathrm{X} = \{a \in \mathbb{P}\colon\ 1 \leqslant a \leqslant p^k\}$, a set with p^k elements. The elements of X that are multiples of p form a subset $\mathrm{Y} = \{p, 2p, 3p, \ldots, p^{k-1}p\}$ with p^{k-1} elements. Thus, for $a \in \mathrm{X}$,

$$(a, p^k) = 1 \quad \Leftrightarrow \quad a \in \mathrm{X} - \mathrm{Y},$$

where $\mathrm{X} - \mathrm{Y}$ is a set with $p^k - p^{k-1}$ elements.

3. Monoids of Finite Type

A *semigroup* (written multiplicatively) is a nonempty set S with a binary operation $x \cdot y$ (or simply xy) that is associative ($xy \cdot z = x \cdot yz$ for all x, y, z). A *monoid* is a semigroup having an element 1 that is neutral for the operation ($1x = x1 = x$ for all x). The data for a monoid is conveniently expressed as a triple $(\mathrm{S}, \cdot, 1)$. A monoid is *cancellative* if $xy = xz$ (or $yx = zx$) implies $y = z$, and *commutative* if $xy = yx$ for all x, y.

Example. The monoid $(\mathbb{P}, \cdot, 1)$ of positive integers under multiplication is both cancellative and commutative. It also has the following property, which is crucial for our discussion.

Definition 3. A monoid $(S, \cdot, 1)$ is said to be of *finite type* if each element of S has only finitely many factorizations in S; in other words, for each $x \in S$, the set of ordered pairs $\{(y, z) \in S \times S: x = yz\}$ is finite. (This set contains at least the ordered pairs $(1, x)$ and $(x, 1)$.)

Example 1. In the monoid $(\mathbb{P}, \cdot, 1)$, $ab = n \Leftrightarrow a|n$ and $b = n/a$, so

$$\{(a, b): ab = n\} = \{(a, n/a): a|n\},$$

a set with $\tau(n)$ elements.

Example 2. Let $S = F_1[t]$ be the set of all *monic* (leading coefficient 1) polynomials with coefficients in a field F. With ordinary polynomial multiplication as the law of composition and the constant polynomial 1 as neutral element, S is a commutative and cancellative monoid. It is of finite type because each monic polynomial is factorable uniquely (apart from order) as a product of powers of a finite list of irreducible monic polynomials. (If p, q are irreducible monic polynomials and $p|q$, then $p = q$.)

Example 3. Let $S = \{1, t, t^2, t^3, \ldots\}$, t an indeterminate, with $t^i t^j = t^{i+j}$ as the operation (where $t^0 = 1$). This is essentially the monoid $(\mathbb{N}, +, 0)$ in multiplicative disguise, where $\mathbb{N} = \{0, 1, 2, 3 \ldots\}$.

Definition 4. An element x of a monoid S is said to be a *unit* if there is an element $x' \in S$ (necessarily unique) such that $xx' = x'x = 1$. The set U_S of all units of S is a group (called, naturally, the group of units of S).

Examples. For the monoid $(\mathbb{P}, \cdot, 1)$, the group of units is $\{1\}$; for $(\mathbb{Z}, \cdot, 1)$ it is $\{1, -1\}$; for $(\mathbb{Z}, +, 0)$ it is $\mathbb{Z}$ (here the notation is 'additive'); for $(F, \cdot, 1)$, where F is a field, it is $F - \{0\}$; for the monoid $F_1[t]$ of Example 2, it is $\{1\}$; for $(R, \cdot, 1)$, where R is a ring with unity, the group of units is the set of invertible elements of R. A group is a monoid all of whose elements are units.

Remark. If $(S, \cdot, 1)$ is a monoid of finite type, then its group of units is finite. (Proof: The set of all ordered pairs (x, y) with $xy = 1$ is finite; the units of S are among the first coordinates of such pairs.) It follows that an infinite group—for example $(\mathbb{Z}, +, 0)$—is not a monoid of finite type.

4. Rings of Functions on Monoids of Finite Type

In this section $(S, \cdot, 1)$ is a monoid of finite type and R is any ring with unity. (In Section 5 we specialize to $S = \mathbb{P}$; in Sections 6 and 7, $S = \mathbb{P}$ and the ring R is required to be commutative; in the final Section 8, we throw in the towel and assume that $S = \mathbb{P}$ and $R = \mathbb{Z}$.)

Definition 5. The set of all functions $f: S \to R$ will be denoted $A_R(S)$, briefly A. For $f, g \in A$, $f = g$ means that $f(x) = g(x)$ for all $x \in S$.

Two operations must be defined to make A a ring: *addition* (along with concepts of 'zero' and 'negatives') and *multiplication*. The definitions pertaining to addition are as follows:

Definition 6. For $f, g \in A$ the *sum* $f+g$ of f and g, and the *negative* $-f$ of f, are defined 'pointwise':

$$(f+g)(x) = f(x) + g(x), \qquad (-f)(x) = -f(x)$$

for all $x \in S$. The *zero* element of A is the constant function $0\colon S \to R$ defined by $0(x) = 0$ for all $x \in S$, where the 0 on the right side is the zero element of the ring R.

LEMMA 1. *For the operations of Definition* 6, $(A, +, 0)$ *is an abelian group.*

The proof of the lemma is straightforward and elementary. The multiplicative structure of A is more subtle.

Definition 7. For $f, g \in A$ we define a function $f * g \in A$, called the *convolution* of f and g, by the formula

$$(f * g)(x) = \sum_{yz=x} f(y)g(z)$$

for all $x \in S$. (The sum on the right is finite because only finitely many ordered pairs (y, z) qualify.)

Example. If $S = \mathbb{P}$ (under multiplication) then

$$(f * g)(n) = \sum_{d|n} f(d)g(n/d),$$

where d runs over all positive integral divisors of n (the sum has $\tau(n)$ terms). When n is a prime p, there are only two terms: $(f * g)(p) = f(1)g(p) + f(p)g(1)$.

The proof that Definition 7 makes A a ring with unity (commutative when S and R are commutative) is arranged in a series of lemmas.

LEMMA 2. *For all* $f, g, h \in A$, $f*(g+h) = f*g + f*h$ *and* $(f+g)*h = f*h + g*h$.

Proof. This is a straightforward consequence of the definitions and the distributive laws in the ring R.

LEMMA 3. *For all* $f, g, h \in A$, $(f*g)*h = f*(g*h)$.

Proof. For all $x \in S$,

$$\begin{aligned}[f*(g*h)](x) &= \sum_{yz=x} f(y)(g*h)(z) \\ &= \sum_{yz=x} f(y) \sum_{st=z} g(s)h(t) \\ &= \sum_{y\cdot st=x} f(y)\cdot g(s)h(t),\end{aligned}$$

summed over all triples (y, s, t) for which $y \cdot st = x$; similarly,

$$[(f*g)*h](x) = \sum_{ys\cdot t=x} f(y)g(s)\cdot h(t),$$

summed over all triples (y, s, t) for which $ys \cdot t = x$. Since $y \cdot st = ys \cdot t$ and $f(y)\cdot g(s)h(t) = f(y)g(s)\cdot h(t)$ (because the operations in S and R are associative), we see that $[f*(g*h)](x) = [(f*g)*h](x)$ for all $x \in S$.

From Lemmas 1–3 we know that A is a ring. The unity element u of A is defined as follows:

Definition 8. The function u: $S \to R$ is defined by $u(1_S) = 1_R$ and $u(x) = 0$ for $x \neq 1_S$. (Writing 1 for the identity element of either S or R, the definition can be expressed in 'Kronecker delta' notation: $u(x) = \delta_{x,1}$.)

LEMMA 4. *With u as in Definition* 8, $f * u = u * f = f$ *for all* $f \in A$.

Proof. Let $x \in S$. In the formula

$$(f * u)(x) = \sum_{yz=x} f(y)u(z),$$

$u(z)$ is 0 unless $z = 1$. The only possible nonzero term on the right side corresponds to the factorization $x1 = x$, so $(f * u)(x) = f(x)u(1) = f(x)1 = f(x)$. This shows that $f * u = f$; similarly $u * f = f$.

LEMMA 5. *If* S *and* R *are commutative then* $f * g = g * f$ *for all* $f, g \in A$.

Proof. In the formula

$$(f * g)(x) = \sum_{yz=x} f(y)g(z),$$

interchanging y and z—then $f(z)$ and $g(y)$—yields the formula for $(g * f)(x)$. The interchanges are permissible by the assumed commutativity.

Summarizing Lemmas 1–5:

THEOREM 5. *If* S *is a monoid of finite type and* R *is a ring with unity, then* $A = A_R(S)$ *is a ring with unity for the pointwise sum and the convolution product, where the unity element* u *is the function defined by* $u(x) = \delta_{x,1}$. *If, moreover,* S *and* R *are commutative, then* A *is a commutative ring.*

5. The Group of Units in a Convolution Ring

In this section, $A = A_R(\mathbb{P})$; that is, the monoid is specialized to $(\mathbb{P}, \cdot, 1)$ but R can still be any ring with unity.

THEOREM 6. *A function* $f \in A$ *is a unit of* A *if and only if* $f(1)$ *is a unit of* R, *that is,*

$$U_A = \{f \in A: f(1) \in U_R\}.$$

Proof. Suppose f is a unit of A, say $f * g = g * f = u$. Since $1 = 1 \cdot 1$ is the only factorization of 1, we have $1 = u(1) = (f * g)(1) = f(1)g(1)$. Similarly $g(1)f(1) = 1$, so $f(1)$ is a unit of R (with inverse $g(1)$).

Conversely, assuming $f \in A$ with $f(1)$ invertible in R, we must show that f has a convolution inverse. This will be done by constructing functions g and h in A such that $g * f = u$ and $f * h = u$ (then $g = h$ by the associative law for convolution). The functions g and h will be constructed recursively (whence the restriction to the monoid of positive integers).

We seek a function $g \in A$ satisfying the relation

$$(*) \qquad \sum_{d|k} g(d)f(k/d) = u(k)$$

for all $k \in \mathbb{P}$. For $k = 1$ this calls for $g(1)f(1) = 1$. Define $g(1) = f(1)^{-1}$. Let $n > 1$ and assume inductively that $g(k)$ has been defined for all $k < n$ in such a way that $(*)$ holds for $1 \leqslant k < n$. The relation $(*)$ calls for $g(n)$ to satisfy the condition

$$\sum_{d|n,\, d<n} g(d)f(n/d) + g(n)f(1) = 0,$$

where, by the induction hypothesis, all terms of the summation on the left have already been defined. We take this as a cue to *define* $g(n)$ by the formula

$$g(n) = \Big(-\sum_{d|n,\, d<n} g(d)f(n/d)\Big)f(1)^{-1}.$$

This completes the construction of $g \in \mathrm{A}$ such that $g * f = u$.

Similarly, if $h \in \mathrm{A}$ is defined recursively by the formulas $h(1) = f(1)^{-1}$ and

$$h(n) = f(1)^{-1}\Big(-\sum_{d|n,\, d>1} f(d)h(n/d)\Big),$$

then $f * h = u$.

Remark 1. With notations as in the theorem, define $\Phi\colon \mathrm{A} \to \mathrm{R}$ by $\Phi(f) = f(1)$, that is, Φ is 'evaluation at 1.' From $(f * g)(1) = f(1)g(1)$ we see readily that Φ is a ring homomorphism of A onto R. The assertion of the theorem is that $\mathrm{U}_{\mathrm{A}} = \Phi^{-1}(\mathrm{U}_{\mathrm{R}})$. Incidentally, $\operatorname{Ker}\Phi = \{f \in \mathrm{A}\colon f(1) = 0\}$ and $\mathrm{A}/\operatorname{Ker}\Phi \cong \mathrm{R}$. (Remember polynomial rings and the evaluation map $p \mapsto p(0)$?)[2]

Remark 2. When $f(1) = 1$ the recursive formula for the convolution inverse f^{-1} of f simplifies to

$$f^{-1}(n) = -\sum_{d|n,\, d<n} f^{-1}(d)f(n/d).$$

If $n = p^k$ (p prime, k a positive integer), this can be written

$$f^{-1}(p^k) = -\sum_{i=0}^{k-1} f^{-1}(p^i)f(p^{k-i});$$

in particular, $f^{-1}(p) = -f(p)$, $f^{-1}(p^2) = -f(p^2) + f(p)^2$ and $f^{-1}(p^3) = -f(p^3) + f(p)f(p^2) + f(p^2)f(p) - f(p)^3$.

6. The Subgroup of Multiplicative Functions

We now specialize to the case that $\mathrm{A} = \mathrm{A}_{\mathrm{R}}(\mathbb{P})$, where R is a *commutative* ring with unity. The classical terminology in Section 1 is extended to cover R-valued functions:

Definition 9. A function $f\colon \mathbb{P} \to \mathrm{R}$ is said to be *multiplicative* if $f(1) = 1$ and $f(mn) = f(m)f(n)$ whenever m and n are relatively prime.

[2]However, the formula of Theorem 6 is false for rings of polynomial functions. The more pertinent ring is the 'ring of formal power series' (G. Hochschild, op. cit., p. 22), for which Theorem 6 *is* true (with the multiplicative monoid $\mathbb{P}$ replaced by the additive monoid $\mathbb{N}$—and $f(1)$ by $f(0)$).

Some easy examples: $f =$ the constant function 1; $f = u$ the unity element of A. The next theorem is a powerful tool for generating and analyzing further examples.

THEOREM 7. *The multiplicative functions $f \in A$ form a subgroup of the group of units of* A.

Proof. If $f \in A$ is multiplicative then $f(1) = 1$, so $f \in U_A$ by Theorem 6. Assuming $f, g \in A$ are multiplicative, we have to show that $f * g$ and f^{-1} are also multiplicative. At any rate,

$$(f * g)(1) = f(1)g(1) = 1 \quad \text{and} \quad f^{-1}(1) = f(1)^{-1} = 1.$$

Assuming $(m, n) = 1$ we must show that

$$(f * g)(mn) = (f * g)(m)(f * g)(n) \qquad 1°$$

$$f^{-1}(mn) = f^{-1}(m)f^{-1}(n). \qquad 2°$$

The proof of 1° is straightforward; notation is 90% of the battle. The proof of 2°, based on 1°, is devious.

Proof of 1°. We note for later use that every divisor of m is relatively prime to every divisor of n. By definition,

$$(*) \qquad (f * g)(mn) = \sum_{yz = mn} f(y)g(z).$$

Also,

$$(f * g)(m) = \sum_{rs = m} f(r)g(s),$$

$$(f * g)(n) = \sum_{ab = n} f(a)g(b),$$

so

$$(f * g)(m)(f * g)(n) = \sum_{rs = m} \sum_{ab = n} f(r)g(s) \cdot f(a)g(b)$$

$$= \sum_{rs = m} \sum_{ab = n} f(r)f(a) \cdot g(s)g(b)$$

(by the commutativity of R). Since r, s are divisors of m, and a, b are divisors of n, we have

$$f(r)f(a) = f(ra) \quad \text{and} \quad g(s)g(b) = g(sb)$$

by the multiplicativity of f and g, therefore

$$(**) \qquad (f * g)(m)(f * g)(n) = \sum_{rs = m,\, ab = n} f(ra)g(sb).$$

The strategy of the proof is to set up a one-to-one correspondence between the terms of $(*)$ and $(**)$ in such a way that corresponding terms are equal. Let

$$V = \{(y, z) \in \mathbb{P}^2 : yz = mn\} = \{(y, mn/y) : y | mn\}$$

(a set with $\tau(mn)$ elements) and let

$$W = \{(r, s, a, b) \in \mathbb{P}^4 : rs = m \text{ and } ab = n\}$$

$$= \{(r, m/r, a, n/a) : r|m \text{ and } a|n\}$$

(a set with $\tau(m)\tau(n)$ elements). Since $\tau(mn) = \tau(m)\tau(n)$ (Theorem 2), V and W have the same number of elements. The earlier equations can be written

$$(*) \qquad (f * g)(mn) = \sum_{(y,z)\in V} f(y)g(z),$$

$$(**) \qquad (f * g)(m)(f * g)(n) = \sum_{(r,s,a,b)\in W} f(ra)g(sb).$$

If $(r, s, a, b) \in W$ then $ra \cdot sb = rs \cdot ab = mn$, therefore $(ra, sb) \in V$; thus, the formula

$$\theta(r, s, a, b) = (ra, sb)$$

defines a mapping θ: $W \to V$. Moreover, if $w = (r, s, a, b) \in W$ and $v = \theta(w) = (ra, sb)$, then the term $f(ra)g(sb)$ of $(**)$ corresponding to w is equal to the term of $(*)$ corresponding to $v = \theta(w)$. It remains to show that the mapping θ is bijective, and since $|W| = |V|$ we need only show that it is injective.

Suppose $\theta(r, s, a, b) = \theta(r', s', a', b')$, that is, $(ra, sb) = (r'a', s'b')$; then

$$ra = r'a' \quad \text{and} \quad sb = s'b',$$

where r, r', s, s' are divisors of m, and a, a', b, b' are divisors of n (thus every element of the first list is relatively prime to every element of the second). From $r|r'a'$ and $(r, a') = 1$ we conclude that $r|r'$; similarly $r'|r$, so $r = r'$ and then $a = a'$ by cancellation. Similarly $s = s'$ and $b = b'$, thus $(r, s, a, b) = (r', s', a', b')$. This completes the proof that $f * g$ is multiplicative.

Proof of 2°. The idea of the proof is to define a (suitable) function $h \in A$ that is visibly multiplicative and then show (using 1°) that $h = f^{-1}$.

Define $h(1) = 1$. If $n > 1$ and

$$n = p_1^{k_1} \cdots p_r^{k_r}$$

is its prime-power factorization, define

$$h(n) = f^{-1}(p_1^{k_1}) \cdots f^{-1}(p_r^{k_r});$$

h is defined unambiguously because, by the commutativity of R, the expression on the right is invariant under any permutation of the primes $p_1, \ldots, p_r$.

We assert that h is multiplicative. Suppose $m > 1$, $n > 1$ with $(m, n) = 1$. The primes $q_1, \ldots, q_t$ occurring in m are different from the primes $p_1, \ldots, p_r$ occurring in n; thus if

$$m = q_1^{j_1} \cdots q_t^{j_t} \quad \text{and} \quad n = p_1^{k_1} \cdots p_r^{k_r}$$

then

$$mn = q_1^{j_1} \cdots q_t^{j_t} \cdot p_1^{k_1} \cdots p_r^{k_r}$$

is the prime-power factorization of mn. Therefore

$$\begin{aligned} h(mn) &= f^{-1}(q_1^{j_1}) \cdots f^{-1}(q_t^{j_t}) \cdot f^{-1}(p_1^{k_1}) \cdots f^{-1}(p_r^{k_r}) \\ &= h(m)h(n) \end{aligned}$$

by the definition of h.

Since f and h are multiplicative, $f*h$ is multiplicative by the first part of the proof. We assert that $f*h=u$. Indeed, $(f*h)(1)=f(1)h(1)=1=u(1)$ and, for p prime and $k\in\mathbb{P}$,

$$(f*h)(p^k)=\sum_{i+j=k} f(p^i)h(p^j)=\sum_{i+j=k} f(p^i)f^{-1}(p^j)$$

$$=(f*f^{-1})(p^k)=u(p^k) \qquad (=0),$$

therefore $(f*h)(n)=u(n)$ for all $n>1$ by the multiplicativity of $f*h$ and u.

But $f*h=u$ implies $h=f^{-1}$, thus f^{-1} has been identified with a function h known to be multiplicative.

Remarks (optional). A moment's thought about the construction of h in the proof of 2° persuades us that the values of a multiplicative function can be specified arbitrarily on the set of prime-powers p^k. Let's capture this idea in a convenient notation. Let Λ be the set of all prime numbers, $\mathbb{N}$ the set of all nonnegative integers. For every function $\beta\colon \Lambda\times\mathbb{N}\to \mathrm{R}$ such that $\beta(p,0)=1$ for all $p\in\Lambda$, there exists a unique multiplicative function $f_\beta\in \mathrm{A}$ such that $f_\beta(p^k)=\beta(p,k)$ for all p and k.

Write $\mathscr{B}$ for the set of all such functions β, and $\mathrm{M_A}$ for the set of all multiplicative functions $f\in \mathrm{A}$; the correspondence $\beta\mapsto f_\beta$ defines a bijection $\mathscr{B}\to \mathrm{M_A}$. (Since $\mathrm{M_A}$ is a subgroup of $\mathrm{U_A}$ (Theorem 7), $\mathscr{B}$ acquires a group structure via the bijection ('transport of structure'). Problem: Find a formula for the group law of $\mathscr{B}$.) One can identify $\mathscr{B}$ with the set $\mathscr{F}(\Lambda\times\mathbb{P},\mathrm{R})$ of all functions $\Lambda\times\mathbb{P}\to \mathrm{R}$ (a set sometimes denoted $\mathrm{R}^{\Lambda\times\mathbb{P}}$). This answers, in a nebulous way, the question of 'how many' multiplicative functions there are.

7. The Möbius Function μ

Let's revert for a moment to the general case $\mathrm{A}=\mathrm{A_R(S)}$, where $(\mathrm{S},\cdot,1)$ is any monoid of finite type and R is any ring with unity. The element $u\in \mathrm{A}$ is the neutral element for the convolution product (Lemma 4 of §4). There is a second 'product' on A that makes it a ring: the pointwise product fg, where $(fg)(x)=f(x)g(x)$ for all $x\in \mathrm{S}$. For the pointwise product, the neutral element is the 'constant function 1':

Definition 10. The function $\gamma\colon \mathrm{S}\to \mathrm{R}$ is defined by $\gamma(x)=1$ for all $x\in \mathrm{S}$.

When $\mathrm{S}=\mathbb{P}$, γ is a unit of A for the convolution product (Theorem 6) and γ is multiplicative in the sense of Definition 9. If, moreover, R is commutative then, although γ is itself a boring function, its convolution inverse is interesting:

THEOREM 8. *If* $\mathrm{A}=\mathrm{A_R}(\mathbb{P})$, *where* R *is a commutative ring with unity, and if* $\mu=\gamma^{-1}$, *then*

$$\mu(n)=\begin{cases}1 & \text{if } n=1,\\ (-1)^r & \text{if } n \text{ is the product of } r \text{ distinct primes},\\ 0 & \text{otherwise}.\end{cases}$$

Proof. Since $\mu(1)=\gamma(1)^{-1}=1$, by Remark 2 of §5 we have, for $n>1$,

$$\mu(n)=-\sum_{d\mid n,\, d<n}\mu(d)\gamma(n/d)=-\sum_{d\mid n,\, d<n}\mu(d).$$

In particular, for p prime and $k \in \mathbb{P}$,

$$\mu(p^k) = -\sum_{i=0}^{k-1} \mu(p^i).$$

Thus $\mu(p) = -\mu(1) = -1$ and

$$\mu(p^{k+1}) = -\sum_{i=0}^{k} \mu(p^i) = -\sum_{i=0}^{k-1} \mu(p^i) - \mu(p^k)$$
$$= \mu(p^k) - \mu(p^k) = 0.$$

Summarizing: $\mu(1) = 1$, and if p is a prime then $\mu(p) = -1$ and $\mu(p^k) = 0$ for $k \geqslant 2$. Since μ is multiplicative (Theorem 7) the formula for $\mu(n)$ in the statement of the theorem follows at once from the prime-power factorization of n.

The function μ of Theorem 8 is called the *Möbius function*. Its role in number theory rests on two mappings $A \to A$ defined as follows:

Definition 11. For $f \in A = A_R(\mathbb{P})$, R a commutative ring with unity, the functions $f', f^\circ \in A$ are defined by the formulas

$$f'(n) = \sum_{d|n} f(d)\mu(n/d), \qquad f^\circ(n) = \sum_{d|n} f(d).$$

Thus $f^\circ(n)$ is the sum of $f(d)$ over all divisors d of n. In view of Theorem 8, $f'(n)$ is a 'weighted sum' (weights ± 1) of the $f(d)$ for which n/d is either 1 or a product of distinct primes.

The mappings $f \mapsto f'$ and $f \mapsto f^\circ$ are 'linear' in an appropriate sense. For example, $(f+g)' = f' + g'$ and $(rf)' = rf'$ for $r \in R$, where, by definition, $(rf)(x) = rf(x)$ for all $x \in S$. Looking at the formulas in Definition 11, it seems a miracle that these mappings are mutually inverse.

THEOREM 9 (Möbius inversion formulas). *Let* $A = A_R(\mathbb{P})$, *where* R *is a commutative ring with unity. For all* $f \in A$, $f'^\circ = f$ *and* $f^{\circ\prime} = f$.

Proof. From Definition 11 it is clear that $f' = f * \mu$ and $f^\circ = f * \gamma$, so the asserted formulas are immediate from $\mu = \gamma^{-1}$.

8. The Classical Case

Having indulged in some harmless and possibly helpful generality (it made the proofs no harder) we settle down to the classical case that motivated it all: $A = A_{\mathbb{Z}}(\mathbb{P})$, the convolution ring of integer-valued functions on the monoid $(\mathbb{P}, \cdot, 1)$ of positive integers under multiplication.

The fact that $(\mathbb{P}, \cdot, 1)$ is a submonoid of $(\mathbb{Z}, \cdot, 1)$ yields an important dividend: The insertion mapping

$$\varepsilon: \mathbb{P} \to \mathbb{Z}, \qquad \varepsilon(n) = n \quad \text{for all } n \in \mathbb{P},$$

is an element of the ring A and is trivially multiplicative. This completes the cast of characters for our discussion:

$$u, \gamma, \varepsilon, \tau, \sigma, \varphi.$$

A common thread of neutrality runs through the first three. Thus u is neutral for the convolution product of A, γ is neutral for the pointwise product, and ε is as neutral as it can be for composition: For every $f \in A$ the composition $f \circ \varepsilon$ makes sense and $f \circ \varepsilon = f$.

THEOREM 10. *$\tau = \gamma * \gamma$ and $\sigma = \varepsilon * \gamma$.*

Proof. For all positive integers n,

$$(\gamma * \gamma)(n) = \sum_{d|n} \gamma(d)\gamma(n/d) = \sum_{d|n} 1 = \tau(n)$$

and

$$(\varepsilon * \gamma)(n) = \sum_{d|n} \varepsilon(d)\gamma(n/d) = \sum_{d|n} d \cdot 1 = \sigma(n).$$

An interesting by-product of the formula $\sigma = \varepsilon * \gamma$ is that ε and γ are multiplicative (obvious), therefore σ is multiplicative (not obvious!) by Theorem 7.

THEOREM 11. *$\varphi * \gamma = \varepsilon$.*

Proof. All functions in sight are multiplicative, so we need only show that $(\varphi * \gamma)(n) = \varepsilon(n)$ for $n = p^k$ (p prime, $k \in \mathbb{P}$). Citing the formula for φ (Theorem 4) at the appropriate step, we have

$$(\varphi * \gamma)(p^k) = \sum_{i=0}^{k} \varphi(p^i)\gamma(p^{k-i}) = \sum_{i=0}^{k} \varphi(p^i)$$

$$= \varphi(1) + \sum_{i=1}^{k} (p^i - p^{i-1})$$

$$= 1 + (p^k - 1) = p^k = \varepsilon(p^k).$$

THEOREM 12. *$\varphi * \tau = \sigma$.*

Proof.[3] Citing Theorems 10 and 11, we have

$$\varphi * \tau = (\varepsilon * \gamma^{-1}) * (\gamma * \gamma) = \varepsilon * \gamma = \sigma.$$

From the formulas $\tau = \gamma * \gamma$, $\sigma = \varepsilon * \gamma$, $\varphi = \varepsilon * \gamma^{-1}$, we see that all of the functions under discussion belong to the subgroup $\langle \gamma, \varepsilon \rangle$ of U_A generated by γ and ε (where U_Λ is the group of units of A). This subgroup is worth dwelling on.

In general, if G is a group and a, b are elements of G such that $ab = ba$, then the mapping $\mathbb{Z}^2 \to G$ defined by $(m, n) \mapsto a^m b^n$ is a homomorphism of groups. Its range is therefore the subgroup $\langle a, b \rangle$ generated by a and b, thus $\langle a, b \rangle = \{a^m b^n : m, n \in \mathbb{Z}\}$.

Let's apply this to a pair of elements f, g of the abelian group U_A. But first we need a distinctive notation for 'convolution powers': Let's write $f^{(n)}$ for the convolution nth power of f ($n \in \mathbb{Z}$), that is,

$$f^{(0)} = u,\ f^{(1)} = f,\ f^{(n+1)} = f^{(n)} * f \text{ for } n \in \mathbb{P},$$

[3] It is striking that a single formula links these functions. Doesn't it remind you of $e^{\pi i} + 1 = 0$?

and $f^{(-n)} = (f^{-1})^{(n)}$ for $n \in \mathbb{P}$. The subgroup of U_A generated by f and g is

$$\langle f, g \rangle = \{f^{(m)} * g^{(n)} : m, n \in \mathbb{Z}\},$$

and the mapping $(m, n) \mapsto f^{(m)} * g^{(n)}$ is a group homomorphism of $\mathbb{Z}^2$ onto $\langle f, g \rangle$. We are going to show that this is an isomorphism when $f = \gamma$ and $g = \varepsilon$.

LEMMA. *If $f \in U_A$ and $f(1) = 1$ then $f^{(k)}(p) = kf(p)$ for all primes p and all $k \in \mathbb{Z}$.*

Proof. Fix a prime p and consider the mapping $\lambda: \mathbb{Z} \to \mathbb{Z}$ defined by $\lambda(k) = f^{(k)}(p)$. In particular, $\lambda(1) = f(p)$. For all integers j and k,

$$\begin{aligned}\lambda(j+k) &= f^{(j+k)}(p) = (f^{(k)} * f^{(j)})(p) \\ &= f^{(k)}(1) f^{(j)}(p) + f^{(k)}(p) f^{(j)}(1) \\ &= 1 \cdot \lambda(j) + \lambda(k) \cdot 1 = \lambda(j) + \lambda(k);\end{aligned}$$

this shows that λ is a group homomorphism, so

$$\lambda(k) = k\lambda(1) \quad \text{for all } k \in \mathbb{Z},$$

in other words $f^{(k)}(p) = kf(p)$ for all $k \in \mathbb{Z}$.

THEOREM 13. *The mapping $\mathbb{Z}^2 \to \langle \gamma, \varepsilon \rangle$ defined by $(m, n) \mapsto \gamma^{(m)} * \varepsilon^{(n)}$ is an isomorphism of groups, thus $\langle \gamma, \varepsilon \rangle \cong \mathbb{Z}^2$.*

Proof. It remains only to prove injectivity of the mapping. Assuming $\gamma^{(m)} * \varepsilon^{(n)} = u$, we have to show that $m = n = 0$. For *every* prime p, we have

$$\begin{aligned}0 = u(p) &= (\gamma^{(m)} * \varepsilon^{(n)})(p) \\ &= \gamma^{(m)}(1)\varepsilon^{(n)}(p) + \gamma^{(m)}(p)\varepsilon^{(n)}(1) \\ &= 1 \cdot \varepsilon^{(n)}(p) + \gamma^{(m)}(p) \cdot 1 \\ &= n\varepsilon(p) + m\gamma(p) \quad \text{(by the lemma)} \\ &= np + m,\end{aligned}$$

whence it is obvious that $m = n = 0$.

In conclusion, here are some prospects for further exploration.

1. The functions γ and ε are, so to speak, 'independent'; they generate a subgroup of U_A (more precisely, of M_A) of 'rank 2'. That's an open invitation to find an interesting multiplicative function f such that γ, ε, f generate a subgroup of rank 3 (and why stop at 3?).[4]

2. The constant function γ is boring, but its convolution inverse $\mu = \gamma^{-1}$ proved to be interesting. How about the inverses of the other functions? For example, $\tau^{-1} = \mu * \mu$ by Theorem 10. From the formula for μ (Theorem 8) it is easy to derive the formula (for p prime)

$$\tau^{-1}(p^k) = \begin{cases} 0 & \text{if } k \geqslant 3, \\ 1 & \text{if } k = 2, \\ -2 & \text{if } k = 1. \end{cases}$$

[4]A good place to start would be to browse through P. J. McCarthy's *Introduction to Arithmetical Functions* (Springer-Verlag, 1986), where a host of intriguing special functions are discussed.

Since τ^{-1} is multiplicative (Theorem 7), it follows that if $n = p_1^{k_1} \cdots p_r^{k_r}$ is the prime-power factorization of n, then $\tau^{-1}(n) = 0$ if $k_i \geq 3$ for some i, otherwise $\tau^{-1}(n) = (-2)^m$, where $m \geq 0$ is the number of indices i for which $k_i = 1$.

The formulas for ε^{-1}, φ^{-1} and σ^{-1} are equally accessible (use Remark 2 following Theorem 6).

3. Theorems 10 and 11 invite exploration of convolution formulas for other pairs of functions, in particular convolution squares. One of them is nice—$\varepsilon * \varepsilon = \tau \cdot \varepsilon$ (the pointwise product!)—but $\tau * \tau$, $\varphi * \varphi$ and $\sigma * \sigma$ seem to be messy.

4. From Theorem 13 we know the 'structure' of the subgroup $\langle \gamma, \varepsilon \rangle$ of the group M_A of multiplicative functions. What is the 'structure' of M_A? The remarks at the end of Section 6 are a start, but not totally satisfying. When that's settled, the groups U_A and U_A/M_A beckon—and what more can we say about the ring A?

5. Abandoned in Section 3 but not forgotten is the monoid $S = F_1[t]$ of monic polynomials with coefficients in a field F (§3, Example 2). Its group of units is $\{1\}$. Could the proof of Theorem 6 be adapted to S by inducting on degree? The ring $A_R(S)$ with $R = F[t]$ invites exploration (in particular, S is a submonoid of R).

6. What about monoids S that are not necessarily of finite type? They can be admitted, at the cost of considering only functions $f: S \to R$ of 'finite support' (i.e., vanishing at all but finitely many points of S) so as to assure the existence of the sums defining convolutions.[5] At first glance, we seem to have lost the functions $\tau, \varphi, \ldots$ that motivated it all; however, thinking of τ as an infinite sequence

$$(\tau(1), \tau(2), \tau(3), \ldots),$$

we surely recapture its essence by considering the totality of all of its 'finite truncates'

$$(\tau(1), \ldots, \tau(n), 0, 0, 0, \ldots).$$

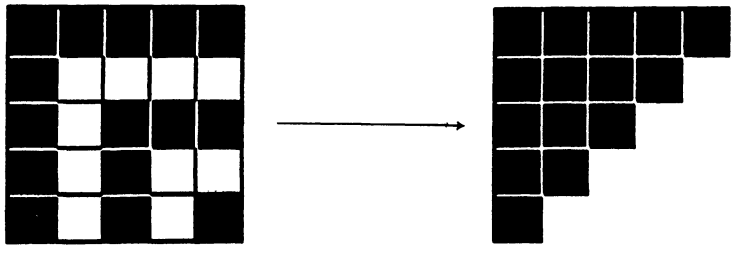
Should we have started with such rings in the first place?

7. An integral domain is a commutative ring with unity having no divisors of zero (if $x \neq 0$ and $y \neq 0$ then $xy \neq 0$). Exercise: If R is an integral domain, then so is the convolution ring $A_R(\mathbb{P})$.

[5]Such convolution rings are called 'monoidal algebras', generalizing the more familiar 'group algebras.

Proof without Words: Alternating Sum of Squares = Triangular Number

$$n^2 - (n-1)^2 + \cdots + (-1)^{n-1}(1)^2 = \sum_{k=0}^{n} (-1)^k (n-k)^2 = \frac{(n)(n+1)}{2}$$



—Stephen L. Snover
University of Hartford
W. Hartford, CT 06117

Since τ^{-1} is multiplicative (Theorem 7), it follows that if $n = p_1^{k_1} \cdots p_r^{k_r}$ is the prime-power factorization of n, then $\tau^{-1}(n) = 0$ if $k_i \geqslant 3$ for some i, otherwise $\tau^{-1}(n) = (-2)^m$, where $m \geqslant 0$ is the number of indices i for which $k_i = 1$.

The formulas for ε^{-1}, φ^{-1} and σ^{-1} are equally accessible (use Remark 2 following Theorem 6).

3. Theorems 10 and 11 invite exploration of convolution formulas for other pairs of functions, in particular convolution squares. One of them is nice—$\varepsilon * \varepsilon = \tau \cdot \varepsilon$ (the pointwise product!)—but $\tau * \tau$, $\varphi * \varphi$ and $\sigma * \sigma$ seem to be messy.

4. From Theorem 13 we know the 'structure' of the subgroup $\langle \gamma, \varepsilon \rangle$ of the group M_A of multiplicative functions. What is the 'structure' of M_A? The remarks at the end of Section 6 are a start, but not totally satisfying. When that's settled, the groups U_A and U_A/M_A beckon—and what more can we say about the ring A?

5. Abandoned in Section 3 but not forgotten is the monoid $S = F_1[t]$ of monic polynomials with coefficients in a field F (§3, Example 2). Its group of units is $\{1\}$. Could the proof of Theorem 6 be adapted to S by inducting on degree? The ring $A_R(S)$ with $R = F[t]$ invites exploration (in particular, S is a submonoid of R).

6. What about monoids S that are not necessarily of finite type? They can be admitted, at the cost of considering only functions $f\colon S \to R$ of 'finite support' (i.e., vanishing at all but finitely many points of S) so as to assure the existence of the sums defining convolutions.[5] At first glance, we seem to have lost the functions $\tau, \varphi, \ldots$ that motivated it all; however, thinking of τ as an infinite sequence

$$(\tau(1), \tau(2), \tau(3), \ldots),$$

we surely recapture its essence by considering the totality of all of its 'finite truncates'

$$(\tau(1), \ldots, \tau(n), 0, 0, 0, \ldots).$$

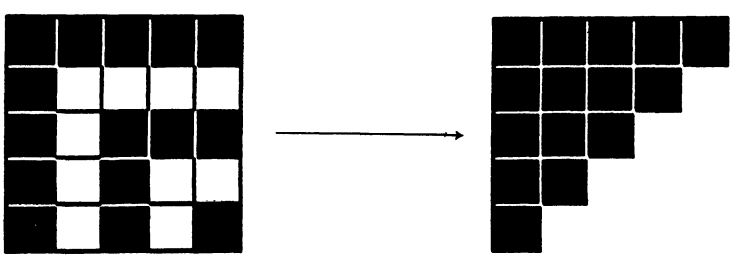
Should we have started with such rings in the first place?

7. An integral domain is a commutative ring with unity having no divisors of zero (if $x \neq 0$ and $y \neq 0$ then $xy \neq 0$). Exercise: If R is an integral domain, then so is the convolution ring $A_R(\mathbb{P})$.

[5] Such convolution rings are called 'monoidal algebras', generalizing the more familiar 'group algebras.'

Proof without Words: Alternating Sum of Squares = Triangular Number

$$n^2 - (n-1)^2 + \cdots + (-1)^{n-1}(1)^2 = \sum_{k=0}^{n} (-1)^k (n-k)^2 = \frac{(n)(n+1)}{2}$$



—Stephen L. Snover
University of Hartford
W. Hartford, CT 06117

From Intermediate Value Theorem To Chaos

XUN-CHENG HUANG
New Jersey Institute of Technology
Newark, NJ 07102

1. Introduction

Continuous functions of a single variable have been studied extensively for over 200 years. Great mathematicians such as Newton (1642–1727), Leibniz (1646–1716), and Euler (1707–1783) have left enduring monuments in this field. Their rich achievements (for example, Euler published 886 papers and books in his 76 years of life) now comprise the major part of the calculus with which every student of science and engineering is familiar. It is hard to believe that in such a field, ploughed and cultivated repeatedly by so many great masters, there is still some virgin land.

In 1975, the article: "Period three implies chaos", was published in the *American Mathematical Monthly* by Li and Yorke. (Here, the word "period" is used in a different way from that seen in elementary mathematics. For example, period three means that there is a point x_0 such that $f^3(x_0) = f(f(f(x_0))) = x_0$, $f^k(x_0) \neq x_0$ for $k = 1, 2$; in other words, the image of x_0 comes back to x_0 after three iterations.) In that article, Li and Yorke announced that a new theorem for continuous functions of a single variable was discovered. The theorem states that if a continuous function has period three, it must have period n for every positive integer n. Soon afterwards, it was found that Li and Yorke's theorem is only a special case of a remarkable theorem published a decade earlier by Soviet mathematician A. N. Sarkovskii, in a Ukrainian journal. Sarkovskii reordered the natural numbers and proved that if $l \lhd m$ (which means l is "less than" m in Sarkovskii's ordering) and if a function has period l then it must have period m. The number 3 is the "smallest" in Sarkovskii's ordering. So, obviously, period 3 implies all the other periods, and Li-Yorke's theorem was not a new one. However, it was in Li-Yorke's article that the new concept of chaos was first introduced into mathematics. People were surprised that iterations of even a very simple continuous function of a single variable can display extremely complicated chaotic behavior.

The original proof of Sarkovskii's theorem is quite difficult. More recently, several authors have simplified the proof (see for example, [3, 5]), however, their proofs are still overly complicated. In this article, we introduce a proof that is based on the intermediate value theorem, accessible to readers with some knowledge of calculus.

2. Common Facts and the Intermediate Value Theorem

Without any special knowledge of mathematics, one can understand the following common facts:

Two trains that depart at the same time from New York and Chicago, destined for Chicago and New York, respectively, must meet each other along their trips.

In a marathon race, a contestant who is lagging behind at first and wants to win must catch up and pass all the other contestants.

Now, let's play a little trick on these common facts. Is it still so obvious?

Suppose Robert starts to climb up a mountain at 8 a.m. and reaches the top at 6 p.m., and then at the same time next day begins his return using the same route. Is there any place on his way up and down the mountain where his watch indicates the same time?

The answer is yes. We can imagine two Roberts start at the same time, one climbing up and the other climbing down by the same route. If their watches are adjusted before starting, of course, they will show exactly the same time where the two Roberts meet on their ways.

The above idea can be summarized in mathematics as the following theorem:

INTERMEDIATE VALUE THEOREM. *If f is continuous on $[a, b]$ and N is any number between $f(a)$ and $f(b)$, then there exists at least one x_0 between a and b such that $f(x_0) = N$.*

This is one of the most fundamental theorems in calculus. Although it is very simple and ordinary, people still pay great attention to it and use it as a test question. For example, the following question is often asked: Prove

PROPOSITION 2.1. *Let f be continuous on $[a, b]$. If the range of f contains $[a, b]$, then equation*

$$f(x) = x$$

has at least one solution in $[a, b]$.

The solution is straightforward. Since the range of f contains $[a, b]$, there must be some $x_1, x_2 \in [a, b]$ such that (see FIGURE 1)

$$f(x_1) \leqslant a \leqslant x_1, \qquad f(x_2) \geqslant b \geqslant x_2.$$

Let $g(x) = f(x) - x$. The result follows from applying the intermediate value theorem with $N = 0$.

By the way, if the assumption "the range of f contains $[a, b]$" is replaced by "the range of f is contained in $[a, b]$," Proposition 2.1 is still valid (see FIGURE 2).

A point x_0 satisfying equation (2.1) is called a *fixed point*. A natural generalization of fixed point is *periodic point*.

Assume $R\{f\} \subset D\{f\}$, the range of f is contained in the domain of f. Denote $f^0(x) = x$, $f^1(x) = f(x)$, $f^2(x) = f(f(x))$, $f^3(x) = f(f^2(x)), \ldots, f^n(x) = f(f^{n-1}(x))$. If x_0 satisfies

$$\begin{cases} f^n(x_0) = x_0 \\ f^k(x_0) \neq x_0, \qquad k = 1, 2, \ldots, n-1, \end{cases} \tag{2.1}$$

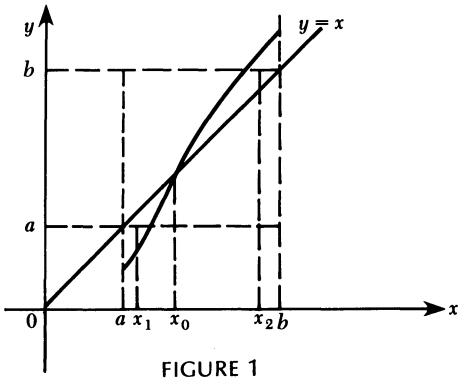


FIGURE 1

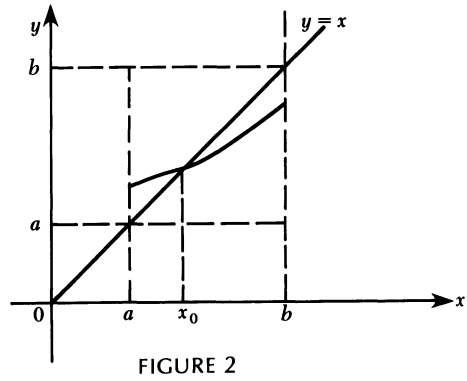


FIGURE 2

then x_0 is called an *n-periodic point with period n*. Clearly, a fixed point is a 1-periodic point.

If x_0 is an n-periodic point of f, then $x_0, f(x_0), \dots, f^{n-1}(x_0)$ are distinct and the set $\{x_0, f(x_0), \dots, f^{n-1}(x_0)\}$ is called a *periodic orbit* of f.

If f has an n-periodic point, we say that *f has period n*.

The existence of a fixed point of a function is generally clear by the inspection of its graph. But the existence of an n-periodic point is not so easy to see even if n is a small integer. As an example, let us consider the function

$$\psi(x) = \begin{cases} x + \frac{1}{2}, & 0 \leqslant x \leqslant \frac{1}{2} \\ 2 - 2x, & \frac{1}{2} < x \leqslant 1, \end{cases} \tag{2.2}$$

whose graph is in FIGURE 3.

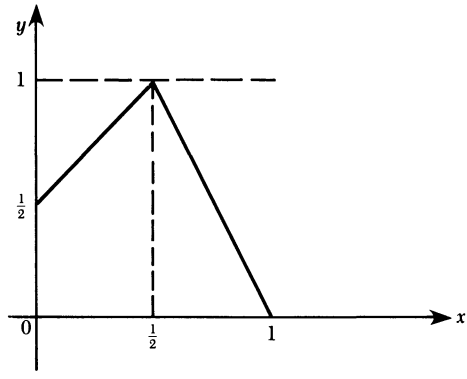


FIGURE 3

As you can see, $\psi(0) = \frac{1}{2}$, $\psi^2(0) = \psi(\frac{1}{2}) = 1$, $\psi^3(0) = \psi^2(\frac{1}{2}) = \psi(1) = 0$; that is ψ has a 3-periodic point 0. Does it have a 5-periodic point? A 7-periodic point? It is hard to ascertain this by just looking at the graph. We need to do some deeper analysis.

The following is a generalized version of the Intermediate Value Theorem.

PROPOSITION 2.2. *Let f be continuous on $[a, b]$, and let $I_0, I_1, \dots, I_{n-1}$ be closed subintervals of $[a, b]$. If*

$$\begin{aligned} f(I_k) &\supset I_{k+1}, \qquad k = 0, 1, \dots, n-2, \\ f(I_{n-1}) &\supset I_0, \end{aligned} \tag{2.3}$$

then, the equation

$$f^n(x) = x \tag{2.4}$$

has at least one solution $x = x_0 \in I_0$ such that

$$f^k(x_0) \in I_k, \qquad k = 0, 1, \dots, n-1. \tag{2.5}$$

In the proposition, $f(I_k) \supset I_{k+1}$ means that the range of f on I_k contains I_{k+1}. We will use the notation

$$I_i \to I_j \qquad \text{or} \qquad I_j \leftarrow I_i \tag{2.6}$$

if $f(I_i) \supset I_j$ ($f(I_i)$ "covers" I_j). The condition (2.3) can be written as

$$I_0 \to I_1 \to I_2 \to \cdots \to I_{n-1} \to I_0.$$

Clearly, if $n = 1$, Proposition 2.2 is reduced to Proposition 2.1.

The proof of Proposition 2.2 is based on the following fact:

If $I_1 \to I_2$, then there exists a subinterval $I_1^* \subset I_1$ such that $f(I_1^*) = I_2$.

This is true, since if $I_2 = [c, d]$, there exist x_1 and x_2 in I_1 such that $f(x_1) = c$ and $f(x_2) = d$. Let $I_1^* = [x_1, x_2]$. Then $I_1^* \subset I_1$, and by the intermediate value theorem, $f(I_1^*) = I_2$.

This fact implies that there exist $I_{n-1}^* \subset I_{n-1}$ such that $f(I_{n-1}^*) = I_0$, $I_{n-2}^* \subset I_{n-2}$ such that $f(I_{n-2}^*) = I_{n-1}^*, \ldots$, and $I_0^* \subset I_0$ such that $f(I_0^*) = I_1^*$. In other words, there exist $I_k^* \subset I_k$ such that

$$f(I_k^*) = I_{k+1}^* \subset I_{k+1} \quad \text{for } k = 0, 1, \ldots, n-2$$

and (2.7)

$$f(I_{n-1}^*) = I_0 \supset I_0^*.$$

From (2.7), $f^k(I_0^*) = I_k^*$, for $k = 0, 1, \ldots, n-2$, and $f^n(I_0^*) \supset I_0^*$. Thus, by Proposition 2.1, equation (2.4) has a solution $x_0 \in I_0^* \subset I_0$ such that (2.5) holds.

Note: In geometry, (2.5) means that mapped successively by f, x_0 visits $I_1, I_2, \ldots, I_{n-1}$ and finally comes back to where it was.

Proposition 2.2 itself is not a remarkable result. But it is the only calculus we need for the proof of our main result.

PROPOSITION 2.3. *Let f: $I \to I$ be continuous, and let f have a $(2n+1)$-periodic orbit $\{x_k = f^k(x_0), k = 0, 1, \ldots, 2n\}$, but no $(2m+1)$-periodic orbit for $1 \leqslant m < n$. Suppose x_0 is in the middle of all the x_i's. Then one of the two permutations*

$$\text{(i)} \quad x_{2n} < x_{2n-2} < \cdots < x_2 < x_0 < x_1 < \cdots < x_{2n-3} < x_{2n-1}$$

$$\text{(ii)} \quad x_{2n-1} < x_{2n-3} < \cdots < x_1 < x_0 < x_2 < \cdots < x_{2n-2} < x_{2n} \tag{2.8}$$

is valid. (FIGURE 4 *is for the case* $n = 3$).

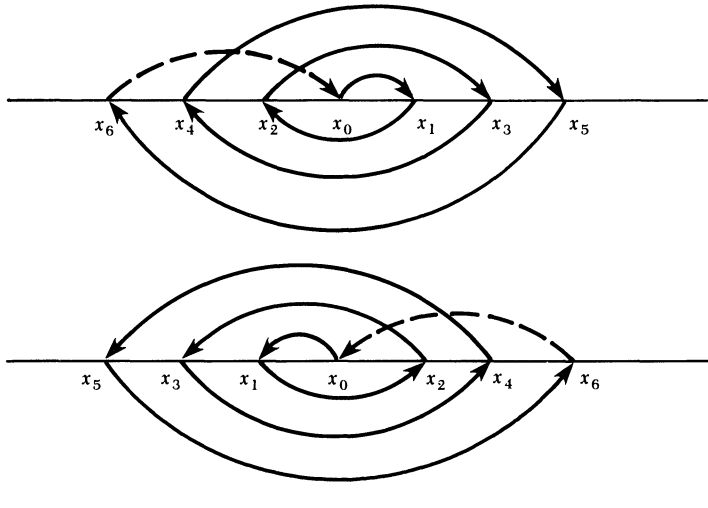


FIGURE 4

Proof. Suppose $n > 1$. Reorder $\{x_i, i = 0, 1, \ldots, 2n\}$ as $\{z_i, i = 1, 2, \ldots, 2n+1\}$ such that

$$z_1 < z_2 < \cdots < z_{2n+1}.$$

Let S_{kl} be the set $\{z_i, k \leqslant i \leqslant l\}$. Assume

$$\min\{f(z): z \in S_{kl}\} = z_i$$
$$\max\{f(z): z \in S_{kl}\} = z_j,$$

and define the set function f^* as

$$f^*(S_{kl}) = S_{ij}. \tag{2.9}$$

We use the notation $S_{kl} \to S_{ij}$, to denote $f^*(S_{kl}) \supset S_{ij}$. Our proof is based on the following claim:

There exist

(i) positive integers m, l $(m, l < 2n+1)$,

(ii) a family of sets $S_i = S_{k_i l_i}$ $(i = 1, 2, \ldots, 2n)$ such that

$$S_1 = \{z_m, z_{m+1}\}, \qquad S_{2n} = \{z_l, z_{l+1}\},$$

S_i has only one common point with S_{i+1}, and

$$\begin{aligned} &S_1 \to S_2 \to \cdots \to S_{2n} \to S_1 \\ &S_1 \subset S_2 \subset \cdots \subset S_{2n-1} \not\supset S_{2n}. \end{aligned} \tag{2.10}$$

In fact, since $f(z_1) > z_1$ and $f(z_{2n+1}) < z_{2n+1}$, we can choose the largest i, say m, such that $f(z_m) > z_m$. Clearly, $m \leqslant 2n$.

Let $S_1 = \{z_m, z_{m+1}\}$, $S_2 = f^*(S_1), \ldots, S_{i+1} = f^*(S_i), \ldots$. Since x_0 is not a 2-periodic point, we have

$$S_{i+1} \supset S_i,\ i = 1, 2, 3, \ldots, \quad \text{and}$$
$$S_i \not\supset S_{i+1} \text{ if } S_i \text{ is not the set } \{z_1, z_2, \ldots, z_{2n+1}\}.$$

Suppose that $i = 1, 2, \ldots, t-1$. If $t = 2n$, then (2.10) is valid. We only need to prove that $t = 2n$.

Since the number of points in $S_{1m} = \{z_1, \ldots, z_m\}$ differs from that in $S_{m+1,2n+1} = \{z_{m+1}, \ldots, z_{2n+1}\}$, there exists $l \neq m$ such that $f(z_l)$ and $f(z_{l+1})$ are on different sides of $[z_m, z_{m+1}]$. Thus, $[z_l, z_{l+1}] \to [z_m, z_{m+1}]$.

Let S_t be $\{z_l, z_{l+1}\}$. Here $t-1$ is chosen as the smallest i such that $S_i \to \{z_l, z_{l+1}\}$. This is possible because $S_1 \subset S_2 \subset S_3 \subset \ldots$.

As an illustration, in the first case of FIGURE 5, $S_1 = \{z_m, z_{m+1}\} = \{z_4, z_5\}$, $S_2 = \{z_3, z_5\}$, $S_3 = \{z_3, z_6\}$, $S_4 = \{z_2, z_6\}$, $S_5 = \{z_2, z_7\}$, $S_t = \{z_l, z_{l+1}\} = \{z_1, z_2\}$, $S_{1m} = \{z_1, z_2, z_3, z_4\}$, and $S_{m+1,2n+1} = \{z_5, z_6, z_7\}$.

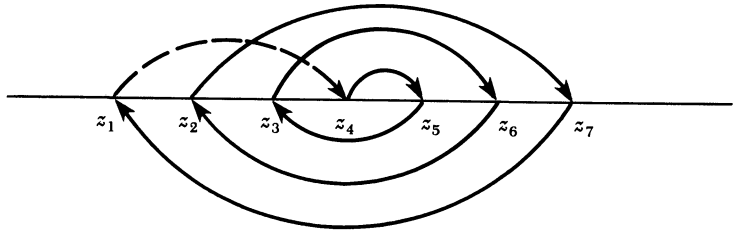


FIGURE 5

Assume that I_i is the smallest closed interval that contains S_i. Then

$$\begin{aligned} &I_1 \subset I_2 \subset \cdots \subset I_{t-1} \not\supset I_t \\ &I_1 \to I_2 \to \cdots \to I_t \to I_1. \end{aligned} \tag{2.11}$$

Since $S_{t-1} \not\supset S_t = \{z_l, z_{l+1}\}$, and S_{t-1} contains at least t points, we have $t \leqslant 2n$.

Assume that $t < 2n$. Let $J_0 = J_1 = \cdots J_{2n-t-1} = I_1$, $J_{2n-t} = I_2$, $J_{2n-t+1} = I_3, \ldots$ $J_{2n-2} = I_t$. We have

$$J_0 \to J_1 \to J_2 \to \cdots J_{2n-2} \to J_0. \tag{2.12}$$

By Proposition 2.2, there exists $x_0^* \in J_1$, such that

$$\begin{aligned} &f^{2n-1}(x_0^*) = x_0^* \\ \text{and} \qquad &f^k(x_0^*) \in J_k \quad \text{for } k = 0, 1, \ldots, 2n-2. \end{aligned} \tag{2.13}$$

It is easy to see that $x_0^*, f(x_0^*), \ldots, f^{2n-2}(x_0^*)$ are distinct. For otherwise, the period of x_0^* is less than $2n-1$, and consequently $f^{2n-2}(x_0^*)$ would then be equal to one of $x_0^*, f(x_0^*), \ldots, f^{2n-3}(x_0^*)$. Then by (2.13)

$$f^{2n-2}(x_0^*) \in J_0 \cap J_{2n-2} = I_1 \cap I_t. \tag{2.14}$$

That is impossible because, by the construction of I_i, $I_1 \cap I_t$ is empty for $t > 2$. Thus, $t = 2n$, and $S_{i+1} \setminus S_i$ is a singleton for each $i = 1, 2, \ldots 2n-2$.

For Proposition 2.3, it is sufficient to show that for each $i = 1, 2, \ldots 2n-1$, f always maps one end point of S_i, say A, to the other, say B, and A is always between B and $f(B)$. Let $S_i \setminus S_{i-1} = \{A_i\}$. If it is not the case for some $k < 2n$, then we have

$$\curvearrowright [A_{k-2}, A_k] \leftrightarrows [A_{k-1}, A_{k-3}] \tag{2.15}$$

(See FIGURE 6).

Clearly, these two cases will lead to the fact that f has period 3 by Proposition 2.2.

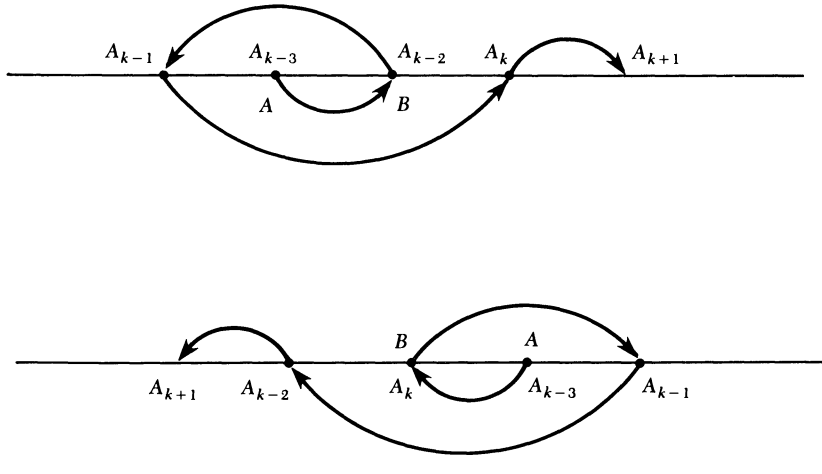


FIGURE 6

3. Period Three and Chaos

The famous Li-Yorke theorem is the following:

THEOREM 3.1. *Let f be continuous on $[a, b]$, its range contained in $[a, b]$. If f has a 3-periodic point, then f has n-periodic points for all positive integers n.*

Proof. Let $x_0 < x_1 < x_2$ be a 3-periodic orbit of f. Then either $f(x_1) = x_0$ or $f(x_1) = x_2$. Without loss of generality, suppose $f(x_1) = x_0$. Then $f(x_0) = x_2$, $f(x_2) = x_1$. Let $\tilde{I}_0 = [x_0, x_1]$, $\tilde{I}_1 = [x_1, x_2]$. By the intermediate value theorem

$$\tilde{I}_0 \leftrightarrows \tilde{I}_1. \tag{3.1}$$

Let

$$\begin{aligned} I_0 = I_1 = \cdots I_{n-2} = \tilde{I}_0 \\ I_{n-1} = \tilde{I}_1. \end{aligned} \tag{3.2}$$

Proposition 2.2 implies there exists $x_0^* \in \tilde{I}_0$ such that $f^n(x_0^*) = x_0^*$ and

$$\begin{aligned} f^k(x_0^*) &\in \tilde{I}_0, \qquad k = 0, 1, \ldots, n-2 \\ f^{n-1}(x_0^*) &\in \tilde{I}_1. \end{aligned} \tag{3.3}$$

By the same argument as for (2.14), if $x_0^*, f(x_0^*), \ldots, f^{n-1}(x_0^*)$ is not an n-periodic point of f, then $f^{n-1}(x_0^*)$ would be one of $f^k(x_0^*)$, $k = 0, 1, \ldots, n-2$. Thus,

$$f^{n-1}(x_0^*) \in \tilde{I}_0 \cap \tilde{I}_1 = x_1,$$

and

$$\begin{aligned} x_0^* &= f^n(x_0^*) = x_0 \\ f(x_0^*) &= f(x_0) = x_2 \notin \tilde{I}_0, \end{aligned}$$

which is a contradiction to $f(x_0^*) \in \tilde{I}_0$.

The proof of Theorem 3.1 is completed.

Theorem 3.1 tells us that the function whose graph is shown in FIGURE 3 has period n for each n. It is beyond one's imagination that such a simple function is such a complicated phenomenon.

In [2] the new concept "chaos" was first introduced. The meaning of chaos in mathematics is that if f has a 3-periodic point in I then there exists an uncountable set $S \subset I$ such that for any two points $x_0, y_0 \in S$, the distance between the two iterative series $x_n = f^n(x_0)$, $y_n = f^n(y_0)$, $n = 1, 2 \ldots$, has the property that, as $n \to \infty$, the limit inferior equals zero while the limit superior is greater than zero.

Clearly the points in S have very interesting properties under successive mapping by f. That the limit inferior equals zero means that there are infinitely many n such that $\{f^n(x)\}$ and $\{f^n(y)\}$ are as close as you like, and that the limit superior is greater than zero means that there are infinitely many n such that the distance between $\{f^n(x)\}$ and $\{f^n(y)\}$ is always positive. In other words, under the successive iteration of f, different points of S are sometimes close, sometimes separated, and none of them is periodic.

4. Sarkovskii's Theorem

The order of natural numbers is

$$1, 2, 3, 4, 5, \ldots .$$

But not many people know that they can also be reordered as the following:

$$3, 5, 7, \ldots, 2\cdot 3, 2\cdot 5, 2\cdot 7, \ldots, 2^2\cdot 3, 2^2\cdot 5, 2^2\cdot 7, \ldots, 16, 8, 4, 2, 1. \tag{4.1}$$

That is, first list all odd numbers except 1, followed by 2 times the odds, 2^2 times the odds, 2^3 times the odds, etc. This exhausts all the natural numbers with the exception of the powers of two that are listed at the end in decreasing order. The number 1 is last.

The ordering (4.1) is now known as *Sarkovskii's ordering* of natural numbers, and is denoted as

$$\begin{aligned} &3 \lhd 5 \lhd 7 \lhd \cdots \lhd 2\cdot 3 \lhd 2\cdot 5 \lhd 2\cdot 7 \lhd \cdots \lhd \\ &2^2\cdot 3 \lhd 2^2\cdot 5 \lhd 2^2\cdot 7 \lhd \cdots \lhd 16 \lhd 8 \lhd 4 \lhd 2 \lhd 1. \end{aligned} \tag{4.2}$$

Is there any application of Sarkovskii's ordering? Let us consider the following theorem.

THEOREM 4.1 (Sarkovskii, 1964). *Let $f\colon I \to I$ be continuous and let f have an l-periodic point. If $l \lhd m$, then f has an m-periodic point, too.*

Here, I can be any interval, finite or infinite, open or closed, semi-open or semi-closed. This theorem tells us that the periods of a continuous function show a wonderful regularity. Before proving the theorem, we make several remarks:

1. If f has a periodic point whose period is not a power of two, then f must have infinitely many periodic points. Conversely, if f has only finitely many periodic points, then each period must be a power of two. Here, the number 1 is considered as 2^0.
2. Period 3 is the least period in the Sarkovskii ordering and therefore implies the existence of all other periods as we saw in Theorem 3.1.
3. The converse of Sarkovskii's Theorem is also true. There exist functions that have p-periodic points and no "higher" periodic points in the sense of Sarkovskii's ordering.

Proof of Theorem 4.1.

Case 1. If f has period 2^m, then f has period 2^l for each $l < m$.

We just need to prove that period 2^m implies period 2^{m-1}.

If $m = 1$, then f has a 2-periodic point. Let x_1, x_2 $(x_1 < x_2)$ be a 2-periodic orbit of f. That is, $f(x_1) = x_2$, $f(x_2) = x_1$, or $f([x_1, x_2]) \supset [x_1, x_2]$. By Proposition 2.1, f has a fixed point (i.e. 2^0-periodic point).

Suppose the conclusion is valid for $m = k$. We want to show it is also valid for $m = k + 1$.

Let $g = f^2$. Then f has period 2^{k+1} implies that g has period 2^k. Now by the induction hypothesis, g has period 2^{k-1}. That is, there exists an $x_0 \in I$ such that

$$g^{2^{k-1}}(x_0) = x_0.$$

$$g^t(x_0) \neq x_0 \quad \text{for } t = 1, 2, \ldots, 2^{k-1},$$

which is equivalent to

$$f^{2^k}(x_0) = x_0,$$
$$f^{2t}(x_0) \neq x_0 \quad \text{for } t = 1, 2, \ldots, 2^{k-1}.$$

Suppose that x_0 is not a 2^k-periodic point of f. Then there must be some $s \in \{1, 3, 5, \ldots, 2^k - 1\}$ such that

$$f^s(x_0) = x_0.$$

But, it is impossible because this implies that

$$f^{2s_0}(x_0) = x_0$$

for some $s_0 \in \{1, 2, 3, \ldots, 2^{k-1} - 1\}$.

Therefore, the induction completes the proof.

Case 2. If f has period $2m + 1$, $(m > 1)$, then f has period k for all $k > 2m + 1$.

By Proposition 2.3, letting $I_1 = [x_0, x_1]$, $I_2 = [x_2, x_0], \ldots, I_{2n-1} = [x_{2n-3}, x_{2n-1}]$, $I_{2n} = [x_{2n}, x_{2n-2}]$, we have the following diagram (FIGURE 7)

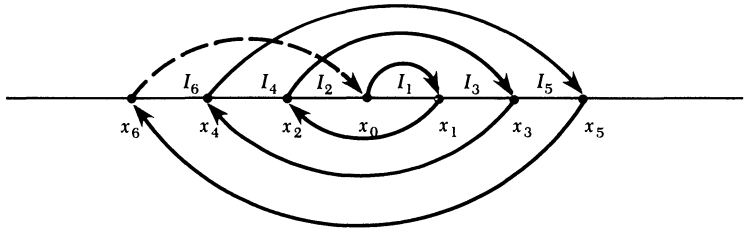


FIGURE 7

and

$$I_1 \to I_2 \to I_3 \to I_4 \to I_5 \to \cdots \to I_{2m-1} \to I_{2m}. \tag{4.3}$$

Suppose there is no $(2n + 1)$-periodic point for $1 \leqslant n < m$. Then, for $k > 2m + 1$

$$\underbrace{I_1 \to I_1 \to \cdots \to I_1}_{k-(2m-1)} \to I_2 \to \cdots I_{2m} \to I_1. \tag{4.4}$$

Since there is no common point in I_1 and I_{2n-1} Proposition 2.2 will result in f having a k-periodic point.

Case 3. If f has period $2m + 1$, $m > 1$, then f has period $2k$ for any positive integer k.

We only need to prove the case where $2k \leqslant 2m$. From (4.3), we have

$$I_{2(m-k)+1} \to I_{2(m-k)+2} \to \cdots I_{2m} \to I_{2(m-k)+1}.$$

By the same argument as in case 2, f has a $2k$-periodic point.

Case 4. Let $m \lhd n$, where $m = 2^k p$, $n = 2^t q$, p, q are odd numbers, $p > 1$, $k \geq 1$. Then period m implies period n. Without loss of generality, suppose that for $l \lhd m$ there is no l-periodic point of f.

According to the Sarkovskii ordering, we need to consider the following possibilities:

$$\text{(i)} \quad t > k, \ q \geq 1, \quad \text{and}$$
$$\text{(ii)} \quad t = k, \ q > p.$$

Let $g(x) = f^{2^k}(x)$. Then f has period $2^k p$ implies g has period p.

By Case 3, g has period $2^{t-k}q$ for $t > k$ and $q \geq 1$. Therefore f has period $2^t q$ for $t > k$, $q \geq 1$, and (i) is valid.

By Case 2, g having period p implies that g has period q. Then f has period $2^t q$, and (ii) is valid.

The proof of Theorem 4.1 is complete.

The following example shows that period 5 does not imply period 3.

Let f be the piecewise linear function defined on $[1,5]$ with $f(1) = 3$, $f(2) = 5$, $f(3) = 4$, $f(4) = 2$ and $f(5) = 1$, whose graph is shown in FIGURE 8.

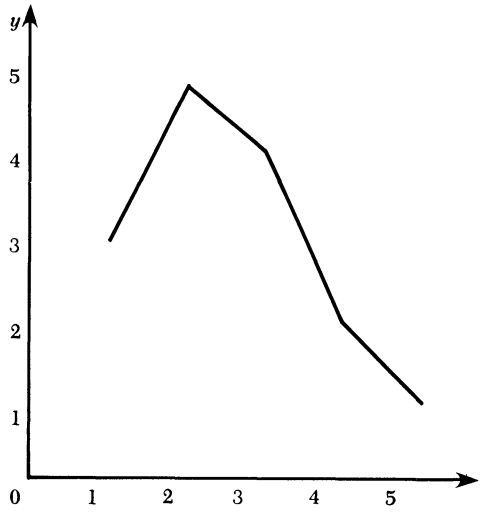


FIGURE 8

It is easy to check that

(i) $10/3$ is a fixed (or 1-periodic) point;

(ii) $5/3$ is a 2-periodic point;

(iii) $1, 2, 3, 4, 5$ are 5-periodic points.

We can also prove that f has no 3-periodic point. Since

$$f^3[1,2] = [2,5], \ f^3[2,3] = [3,5], \ f^3[4,5] = [1,4],$$

f^3 has no periodic point in any of these intervals. Also, since $f^3[3,4] = [1,5]$ and f^3 is monotonically decreasing on $[3,4]$, there exists a unique $x_0 \in [3,4]$ such that

$$f^3(x_0) = x_0.$$

Since $f(x) = 10 - 2x$ on $[3,4]$, $f(x)$ has a unique fixed point $\bar{x} = 10/3$ on $[3,4]$. Since,

$$f^3(\bar{x}) = f^2(\bar{x}) = f(\bar{x}) = \bar{x} = x_0,$$

x_0 is not a 3-periodic point. Hence, f has no 3-periodic point.

5. Conclusion and Discussion

Sarkovskii's theorem has not ended the discussion of the periodic orbits of continuous functions. On the contrary, it created a new direction for studying the problem. Many articles and books have been appearing. The question is why so many great classical analysts didn't discover such an important theorem. The reason is that the classical analysts concentrated on the local properties of functions. Essentially, continuity, differentiability, and integrability are determined by the local properties of functions. Although some global properties such as uniform continuity had been obtained, these global properties can be derived simply from the local properties.

A remarkable advance in modern analysis is viewing the situation as a whole in the study of functions. Actually, the concepts such as iteration and periodic orbits have inseparable relations to the global properties. For example, $f(x)$ can be iterated on $[a,b]$ but may not be iterated on any subinterval of $[a,b]$.

Studying the global properties of functions or mappings now forms a new branch of mathematics called global analysis, which includes differential dynamical systems, global differential geometry, qualitative theory of differential equations, differential topology, etc. It is one of the main directions in modern mathematics.

Now we use another example to show that even in fundamental problems one can benefit from taking into account the global structure.

Example. Monkeys' Apples

There was a pile of apples on the beach that belonged to five monkeys and was to be distributed equally among them. The first monkey came and waited for a while but no others followed. He divided those apples into five piles each of which had the same number of apples. But one was left and he threw it into the sea and went away with his own pile of apples. The second monkey came and divided the rest of the apples into five piles equally, too. Again, one was left and was thrown into the sea. Then he went away with his own apples, too. Later, one by one, each monkey did the same as the first two did.

What is the least number of apples on the beach in the beginning? What is the least number of apples left after all the monkeys take away theirs?

The problem is not easy to solve if you use the usual equations. So the famous physicist Dirac suggested doing it as follows.

Let N be the number of apples in the beginning, and A_1, A_2, A_3, A_4, A_5 be numbers of apples taken by the monkeys. Then, we will have a system of linear equations

$$\begin{cases} N - 5A_1 = 1 \\ 4A_1 - 5A_2 = 1 \\ 4A_2 - 5A_3 = 1 \\ 4A_3 - 5A_4 = 1 \\ 4A_4 - 5A_5 = 1 \end{cases}, \tag{5.1}$$

which possesses a particular solution

$$(N, A_1, A_2, A_3, A_4, A_5) = (-4, -1, -1, -1, -1, -1). \tag{5.2}$$

The corresponding homogenous equations of (5.1) have a general solution

$$\left(5\left(\frac{5}{4}\right)^4 k, \left(\frac{5}{4}\right)^4 k, \left(\frac{5}{4}\right)^3 k, \left(\frac{5}{4}\right)^2 k, \frac{5}{4}k, k\right) \tag{5.3}$$

where k is any constant. Therefore, the general solution of (5.1) is

$$\left(5\left(\frac{5}{4}\right)^4 k - 4, \left(\frac{5}{4}\right)^4 k - 1, \left(\frac{5}{4}\right)^3 k - 1, \left(\frac{5}{4}\right)^2 k - 1, \frac{5}{4}k - 1, k - 1\right). \tag{5.4}$$

From (5.4), we can determine that the least positive integer for N is $5^5 - 4 = 3121$ when $k = 4^4 = 256$; and the number of apples left is $4A_5 = 4(k - 1) = 1020$.

As you can see this solution is based on the structure of solutions of linear equations. If you don't know the theory, it is really hard to find it.

The method we used for this problem is fundamental and very simple. Suppose x is the number of apples before a monkey came and y the number after he left. Clearly, y is determined by x, say $y = f(x)$, and

$$f(x) = \frac{4}{5}(x - 1). \tag{5.5}$$

If there were N apples at first and M apples at last, then

$$M = f(f(f(f(f(N))))) = f^5(N). \tag{5.6}$$

Now, we consider how to get a formula for $f^5(x)$. We rewrite $f(x)$ as

$$f(x) = \frac{4}{5}(x + 4) - 4 \tag{5.7}$$

where, as you can see, -4 is a fixed point of $f(x)$.

Obviously,

$$\begin{aligned} f^2(x) &= \left(\frac{4}{5}\right)^2 (x + 4) - 4 \\ f^3(x) &= \left(\frac{4}{5}\right)^3 (x + 4) - 4 \\ f^4(x) &= \left(\frac{4}{5}\right)^4 (x + 4) - 4 \\ f^5(x) &= \left(\frac{4}{5}\right)^5 (x + 4) - 4 \end{aligned} \tag{5.8}$$

and hence

$$M = \left(\frac{4}{5}\right)^5 (N + 4) - 4. \tag{5.9}$$

In order to have a positive integer M, $N + 4$ must be a multiple of 5^5. So the least positive integer value of N is

$$N = 5^5 - 4 = 3121,$$

and consequently

$$M = 4^5 - 4 = 1020.$$

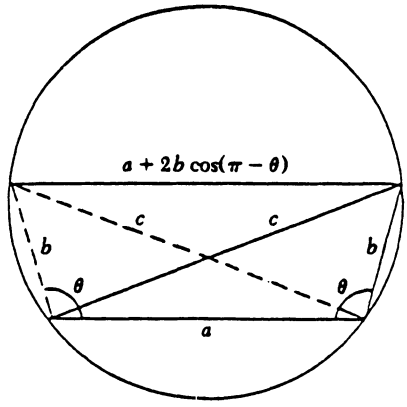
Before ending the article, we mention again that the proof of Sarkovskii's beautiful theorem is far from advanced mathematics. This big surprise shows that people need not have advanced knowledge to establish mathematics if, when opportunity arrives, it is recognized.

Acknowledgements. The author would like to thank Professors Stephen Merrill and Roman Voronka for their useful comments.

REFERENCES

1. A. N. Sarkovskii, Coexistence of cycles of a continuous map of a line into itself, *Ukr. Math. Z.* 16 (1964), 61–71.
2. T. Li and J. A. Yorke, Period three implies chaos, *Amer. Math. Monthly* 82 (1975), 985–992.
3. P. Stefen, A theorem of Sarkovskii on the existence of periodic orbits of continuous endomorphisms of the real line, *Commun. Math. Phys.* 54 (1977), 237–248.
4. Z. Jing, From ordinary fact to surprising theorem, *Nature J.* 8–7 (1985), 532–536.
5. R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin/Cummings Publishing Co., Menlo Park, CA, 1986.
6. M. Liang, Tidbits of chaos theory and homotopy method, *Nature J.* 9:2 (1986), 139–142.

Proof without Words: The law of cosines via Ptolemy's Theorem



$$c \cdot c = b \cdot b + (a + 2b\cos(\pi - \theta)) \cdot a$$
$$c^2 = a^2 + b^2 - 2ab \cdot \cos\theta$$

—Sidney H. Kung
Jacksonville University
Jacksonville, FL 32211

$$M = 4^5 - 4 = 1020.$$

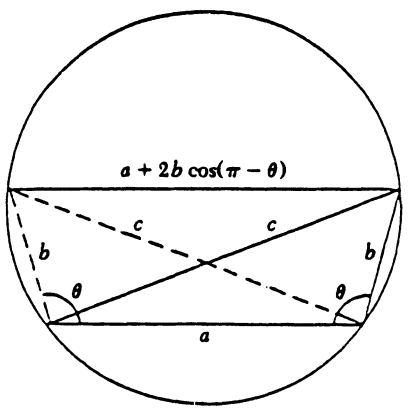
Before ending the article, we mention again that the proof of Sarkovskii's beautiful theorem is far from advanced mathematics. This big surprise shows that people need not have advanced knowledge to establish mathematics if, when opportunity arrives, it is recognized.

Acknowledgements. The author would like to thank Professors Stephen Merrill and Roman Voronka for their useful comments.

REFERENCES

1. A. N. Sarkovskii, Coexistence of cycles of a continuous map of a line into itself, *Ukr. Math. Z.* 16 (1964), 61–71.
2. T. Li and J. A. Yorke, Period three implies chaos, *Amer. Math. Monthly* 82 (1975), 985–992.
3. P. Stefen, A theorem of Sarkovskii on the existence of periodic orbits of continuous endomorphisms of the real line, *Commun. Math. Phys.* 54 (1977), 237–248.
4. Z. Jing, From ordinary fact to surprising theorem, *Nature J.* 8–7 (1985), 532–536.
5. R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin/Cummings Publishing Co., Menlo Park, CA, 1986.
6. M. Liang, Tidbits of chaos theory and homotopy method, *Nature J.* 9:2 (1986), 139–142.

Proof without Words: The law of cosines via Ptolemy's Theorem



$$c \cdot c = b \cdot b + (a + 2b\cos(\pi - \theta)) \cdot a$$
$$c^2 = a^2 + b^2 - 2ab \cdot \cos\theta$$

—Sidney H. Kung
Jacksonville University
Jacksonville, FL 32211

NOTES

Tetrahedra with Integer Edges and Integer Volume

KEVIN L. DOVE
JOHN S. SUMNER
University of Tampa
Tampa, FL 33606

Among the many problems that have been proposed in *Mathematics Magazine*, we found the following problem [1] from the February 1987 issue an especially interesting one.

Problem 1261b. Determine the volume of the smallest tetrahedron with integer edges and integer volume.

This problem was the second part of a two-part problem. The first part asked the reader to determine the smallest area of a triangle with integer edges and integer area. The smallest area can be shown to be 6 (the 3-4-5 right triangle is the only such triangle). The proof is not too difficult and makes for a challenging problem. As late as the October 1991 issue of *Mathematics Magazine*, Problem 1261b had not been resolved. The purpose of this note is to solve Problem 1261b and to establish in the process several interesting facts concerning the existence of tetrahedra with integer edges and integer volume.

Early attempts When we began work on this problem our knowledge of tetrahedra was limited to the definition of a tetrahedron and the well-known volume formula $V = (1/3)Bh$, where B is the area of a base of the tetrahedron and h is the height of the tetrahedron from the base. In other words, we knew very little.

As a prelude to solving Problem 1261b, it seemed proper to look for at least one tetrahedron with integer edges and integer volume. (For all we knew there might not be one!) This in itself turned out to be a challenging problem. (The reader should try finding one!) One simple way that we attempted to construct a tetrahedron with integer edges was to try to build the tetrahedron so that it would fit into the corner of a room. That is, we wanted to construct a tetrahedron three of whose coincident edges meet in right angles. See FIGURE 1.

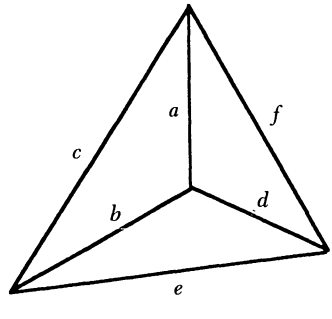


FIGURE 1

Here it was required that faces a-b-c, a-d-f, and b-d-e be right triangles. Of course this can only happen if a^2+b^2, a^2+d^2, and b^2+d^2 are all squares of integers. The reader is urged to try various small values of a, b, and d and satisfy himself that there is no "obvious solution." Encountering the same problem, we resorted to a computer program to search for solutions. In our program we varied the values of a, b, and d from 3 to 300 assuming $a \geqslant b \geqslant d$. Amazingly, the first solution that was generated occurred with $a = 240$, $b = 117$, and $d = 44$. In this case $c = 267$, $e = 125$, $f = 244$, and the volume is equal to $V = (1/6)abd = 205{,}920$. For a general solution to this number theoretical problem see [2].

Although a volume of 205,920 left us a long way from finding the smallest volume, we were pleased to find at least one tetrahedron with integer edges and integer volume.

With a little thought we constructed another tetrahedron with integer edges and a much smaller integer volume. This tetrahedron contains exactly two right triangles and was formed in the following way. We formed the base of the tetrahedron by glueing two 3-4-5 right triangles together along the edge of length 4. This formed a triangular base with edges of lengths 5, 5, and 6, and area 12. Next we attached two 5-12-13 right triangles to the base. See FIGURE 2.

It is clear that the tetrahedron in FIGURE 2 has volume $V = 48$. This was certainly an improvement over 205,920 and lowered the smallest volume by a large amount. However, at this point we ran out of ideas. Since the volume formula $V = (1/3)Bh$ is useless if one does not know the height h, we felt that Problem 1261b could only be solved if a volume formula could be developed (or found) that depends only upon the edges of the tetrahedron. We did not know of one at the time and a brief search of the literature did not indicate the existence of such a formula. Fortunately, we were able to construct a formula ourselves.

An alternative volume formula Consider a tetrahedron with integer edges a, b, c, d, e, and f as in FIGURE 3. We have arranged the tetrahedron so that the base a-b-c sits in the first quadrant. Of course it is possible that the vertex d-e-f does not sit over the interior of the base a-b-c. However, the derivation of the volume formula does not depend upon the location of vertex d-e-f.

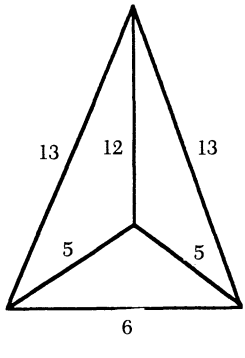


FIGURE 2

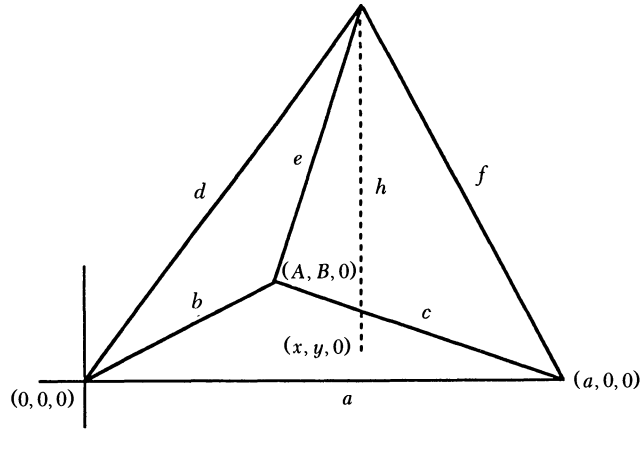


FIGURE 3

We clearly have the equations:

$$x^2 + y^2 = d^2 - h^2 \tag{1}$$

$$(x - A)^2 + (y - B)^2 = e^2 - h^2 \tag{2}$$

$$(x - a)^2 + y^2 = f^2 - h^2 \tag{3}$$

$$A^2 + B^2 = b^2 \tag{4}$$

$$(a - A)^2 + B^2 = c^2. \tag{5}$$

Using equations (1), (2), and (3) to solve for x and y we find that:

$$x = \frac{a^2 + d^2 - f^2}{2a} \quad \text{and} \quad y = \frac{a(A^2 + B^2 + d^2 - e^2) - A(a^2 + d^2 - f^2)}{2Ba}.$$

Substituting these expressions for x and y in equation (1) we have

$$h^2 = d^2 - \left[\frac{a^2 + d^2 - f^2}{2a}\right]^2 - \left[\frac{a(A^2 + B^2 + d^2 - e^2)}{2Ba} - \frac{A(a^2 + d^2 - f^2)}{2Ba}\right]^2.$$

If we multiply both sides of this last equation by $4a^2B^2$ and use the fact that $144V^2 = 4a^2B^2h^2$, where V is the volume of the tetrahedron, we find that:

$$144V^2 = 4a^2B^2d^2 - B^2(a^2 + d^2 - f^2)^2 - [a(A^2 + B^2 + d^2 - e^2) - A(a^2 + d^2 - f^2)]^2.$$

Use equations (4) and (5) to solve for A and B^2. The results are

$$A = \frac{a^2 + b^2 - c^2}{2a} \quad \text{and} \quad B^2 = \frac{4a^2b^2 - (a^2 + b^2 - c^2)^2}{4a^2}.$$

After substituting these expressions for A and B^2 and solving for V, we find that the volume of the tetrahedron is given by

$$V = \frac{1}{12}\Big\{4a^2b^2d^2 - a^2(b^2 + d^2 - e^2)^2 - b^2(a^2 + d^2 - f^2)^2$$

$$-d^2(a^2 + b^2 - c^2)^2 + (b^2 + d^2 - e^2)(a^2 + d^2 - f^2)(a^2 + b^2 - c^2)\Big\}^{1/2}. \tag{6}$$

We later discovered that Euler had found this formula in the 18th century. See [3], [6].

Care must be taken when applying (6). Unfortunately, it is possible for the integer values a, b, c, d, e, and f to form triangular faces a-b-c, a-d-f, b-d-e, and c-e-f without the radical in (6) being real. For example if $a = b = c = 14$ and $d = e = f = 8$, then the faces a-b-c, a-d-f, b-d-e, and c-e-f are all triangles. However, it is a simple matter to check that the value of the radicand in (6) is equal to $-153{,}664$. Thus no tetrahedron exists with these faces. To see why, let α be the included angle between the edges a and b, let β be the included angle between the edges b and d, and let γ be the included angle between the edges a and d. Then it is clear that a tetrahedron exists if and only if $\alpha + \beta > \gamma$, $\alpha + \gamma > \beta$, and $\beta + \gamma > \alpha$. This is precisely what goes wrong with the previous example. Clearly $\alpha = 60°$. A simple calculation shows that $\beta < 29°$. Since $\gamma = \beta$, we have $\beta + \gamma < \alpha$. Using the law of cosines and simple trigonometric identities, we can write (6) as:

$$V = \frac{abd}{3}\left\{\sin\left(\frac{\alpha+\beta+\gamma}{2}\right)\sin\left(\frac{\alpha+\beta-\gamma}{2}\right)\sin\left(\frac{\alpha-\beta+\gamma}{2}\right)\sin\left(\frac{-\alpha+\beta+\gamma}{2}\right)\right\}^{1/2} \tag{7}$$

Note the resemblance of (7) to Heron's Formula for the area of a triangle. It is a simple matter to show that the radicand of (7) is positive if and only if $\alpha+\beta>\gamma$, $\alpha+\gamma>\beta$, and $\beta+\gamma>\alpha$. Therefore we can say the following concerning the use of (6). If a tetrahedron exists and has edges a, b, c, d, e, and f and triangular faces a-b-c, a-d-f, b-d-e, and c-e-f, then the volume is given by (6). Conversely if the real numbers a, b, c, d, e, and f form triangular faces a-b-c, a-d-f, b-d-e, and c-e-f and if the radicand of (6) is positive, then a tetrahedron exists and has volume given by (6).

One interesting consequence of (6) is given by the following proposition.

PROPOSITION 1. *Suppose a tetrahedron has integer edges and integer volume. Then no edge has length* 1.

Proof. Suppose there exists a tetrahedron with integer edges, integer volume, and an edge of length 1. Without loss of generality we may assume that $c=1$. In the triangular faces a-b-c and c-e-f we must have $a=b$ and $e=f$. It follows from (6) that

$$\begin{aligned}144V^2 &= 4a^4d^2 - 2a^2(a^2+d^2-e^2)^2 - d^2(2a^2-1)^2 \\ &\quad + (a^2+d^2-e^2)^2(2a^2-1) \\ &= 4a^4d^2 - d^2(2a^2-1)^2 - (a^2+d^2-e^2)^2 \\ &= d^2(4a^2-1) - (a^2+d^2-e^2)^2.\end{aligned}$$

Thus $(12V)^2 + (a^2+d^2-e^2)^2 = d^2(4a^2-1)$. This means that the integer $d^2(4a^2-1)$ is a sum of two squares. It is well known [4] that an integer n is representable as a sum of two squares if and only if each prime factor of n of the form $4k-1$ occurs to an even power in the prime factorization of n. This is clearly true for the integer d^2. However, it is never true for the integer $4a^2-1$. Then $d^2(4a^2-1)$ is not representable as a sum of two squares and we have a contradiction. Thus no tetrahedron with integer edges and integer volume can have an edge of length 1.

Using (6) and Proposition 1 we again resorted to a computer program and generated a list of all tetrahedra having integer edges between 2 and 20 and integer volume. We had hoped that a tetrahedron of volume 1 would be generated and in the process solve Problem 1261b. Unfortunately, the program verified that the smallest tetrahedron with integer edges between 2 and 20 and integer volume has volume equal to 6 (e.g., $a=b=4$, $c=2$, $d=7$, $e=6$, $f=5$). However, several good things came out of this list. Of course it lowered the smallest known volume to 6. More importantly, though, it was observed that the volume of each tetrahedron in the list was a multiple of 3. That this must be the case is given by the following proposition.

The volume is a multiple of 3

PROPOSITION 2. *Suppose a tetrahedron has integer edges and integer volume. Then the volume is a multiple of* 3.

Proof. We consider cases depending on the number of edges whose lengths are multiples of 3.

First we note that perfect squares are congruent to either 0 or 1 (modulo 3). Since the volume is an integer, the polynomial inside the radical of (6), call it U, must be a perfect square. In fact, $U = (12V)^2$ and hence $U \equiv 0 \pmod 3$. If we multiply out the polynomial U we find that:

$$\begin{aligned} U = {} & a^2b^2e^2 + a^2b^2f^2 + a^2c^2d^2 + a^2c^2e^2 + a^2d^2e^2 + a^2e^2f^2 \\ & + b^2c^2d^2 + b^2c^2f^2 + b^2d^2f^2 + b^2e^2f^2 + c^2d^2e^2 + c^2d^2f^2 \\ & - a^2b^2c^2 - a^2d^2f^2 - b^2d^2e^2 - c^2e^2f^2 - a^2e^4 - a^4e^2 \\ & - b^2f^4 - b^4f^2 - c^2d^4 - c^4d^2. \end{aligned}$$

From this expression it can be seen that the volume formula (6) does not depend upon which vertex of the tetrahedron is labeled as the *a-b-d* vertex.

Case 1. If no edge is a multiple of 3, then the square of each edge is congruent to 1 (mod 3). Then computing U we find that $U \equiv 2 \pmod 3$, which is not possible. Thus case 1 cannot occur.

Case 2. Suppose exactly one edge is a multiple of 3. Without loss of generality assume $a \equiv 0 \pmod 3$ and all other edges are not. Then

$$\begin{aligned} U = {} & \{-a^4e^2\} + \{a^2(b^2e^2 + b^2f^2 + c^2d^2 + c^2e^2 + d^2e^2 + e^2f^2 - b^2c^2 - d^2f^2 - e^4)\} \\ & + \{b^2c^2d^2 + b^2c^2f^2 + b^2d^2f^2 + b^2e^2f^2 + c^2d^2e^2 + c^2d^2f^2 - b^2d^2e^2 - c^2e^2f^2 \\ & - c^4d^2 - c^2d^4 - b^4f^2 - b^2f^4\}. \end{aligned}$$

The first expression in braces is a multiple of 81 since $a^4 \equiv 0 \pmod{81}$, and the second is a multiple of 27 since $a^2 \equiv 0 \pmod 9$ and what remains in parentheses is congruent to 0 (mod 3). So if we can show that the last brace is a multiple of 27, then U is a multiple of 27.

Replacing each square by a multiple of 3 plus one (i.e., $b^2 = 3B + 1$, $c^2 = 3C + 1$, etc.), the last brace becomes

$$\begin{aligned} & \{27(BCD + BCF + BDF + BEF + CDE + CDF - BDE - CEF - C^2D - CD^2 \\ & \qquad - B^2F - BF^2) \\ & + 9(2BC + BD - CD - BF + CF + 2DF - C^2 - D^2 - F^2 - B^2)\} \\ & = 27(*) - 9(B - C + D - F)^2 \end{aligned}$$

where $(*)$ is a polynomial in B, C, D, E, and F.

Now since $(B - C + D - F)^2 \equiv 0$ or 1 (mod 3), we have $(4V)^2 = U/9 \equiv 0$ or 2 (mod 3). Since $(4V)^2$ is a perfect square, it must be congruent to either 0 or 1 (mod 3). Therefore $(B - C + D - F)^2 \equiv 0 \pmod 3$ and U is a multiple of 27. Therefore $144V^2 = U \equiv 0 \pmod{27}$ and V is a multiple of 3.

Case 3. If exactly two edges are multiples of 3, either they meet or they do not. In the first instance let $a \equiv b \equiv 0 \pmod 3$ and all other edges not. Then as in case 1, $U \equiv 2 \pmod 3$, a contradiction.

If the two edges don't meet they must be opposite. For example suppose $a \equiv e \equiv 0 \pmod 3$ and all other edges not. As in Case 2, we notice that $U = 27(*) - 9(B - C + D - F)^2$ although the two asterisks do not represent the same polynomial since $e^2 \neq 3E + 1$. At any rate, we conclude that V is a multiple of 3 as before.

Case 4. Suppose at least three edges are multiples of 3. Several cases are disposed of as in case 1. If exactly three edges are multiples of 3 but these edges do not form a face of the tetrahedron, then either the edges form a vertex (e.g. *a-b-d*) or a spine (e.g. *a-b-e*). In either case we have $U \equiv 2 \pmod 3$. If exactly four edges are multiples of 3 and these do not form a face then the four edges must form two pairs of opposite edges (e.g. *a-e* and *b-f*). In this case we have $U \equiv 1 \pmod 3$. But we must have $U \equiv 0 \pmod 3$. So the only remaining possibility is that a face of the tetrahedron is made up of edges that are all multiples of 3.

Suppose that $a = 3A$, $b = 3B$ and $c = 3C$. Then as before

$$U = 27(*) + 9\{A^2(be^2 - f^2)(d^2 - e^2) + B^2(e^2 - f^2)(f^2 - d^2) + C^2(d^2 - e^2)(f^2 - d^2)\}.$$

Since at least two of the squares d^2, e^2, and f^2 must be congruent modulo 3, all but at most one term of the expression in the braces must be congruent to 0 (mod 3). Without loss of generality assume that the first term $A^2(e^2 - f^2)(d^2 - e^2)$ is not congruent to 0 (mod 3). A quick check of the possible values for A^2, d^2, e^2, and f^2 shows that $(e^2 - f^2)(d^2 - e^2) \equiv 2 \pmod 3$ and $A^2 \equiv 1 \pmod 3$. Thus $(4V)^2 = U/9 \equiv 2 \pmod 3$, a contradiction. Thus all three of the terms inside the braces are multiples of 3 and U again is a multiple of 27.

In view of Proposition 2, our search for the smallest such volume became a matter of checking whether a tetrahedron of volume 3 exists or not. The program that we originally used to find the volume 6 example was extremely slow when the edges got larger than 20—there were essentially six nested loops running from $i = 2$ to $i = a$ (the largest edge). This meant that if we had any hope of finding a tetrahedron of volume 3, the computer program we were using would have to be greatly modified.

A tetrahedron of volume 3 We noticed that if we fix the base *a-b-c* of the tetrahedron, then the polynomial U is a degree 4 polynomial in the variables d, e, and f. Since *a-d-f* is a face of the tetrahedron we know that $f - a < d < f + a$ so we can let $d = f + i$ where i takes on values between $-a$ and a. Similarly *c-e-f* is a face so we can let $e = f + j$ where j takes on values between $-c$ and c. Making these replacements into the polynomial U, we are guaranteed a polynomial of degree four or less in the variable f (where a, b, c, i, and j are held constant). Fortunately (and surprisingly) for us, the polynomial U actually becomes degree two since all higher order terms cancel. That is,

$$144V^2 = U = (r)f^2 + (s)f + t$$

where r, s, and t are polynomials in the variables a, b, c, i, and j. By setting the volume equal to 3, we get a quadratic equation in f for each choice of the other variables. Thus for a fixed base *a-b-c* and for a fixed i and j we check to see whether the solutions of this equation are integers. If so, then we have found our tetrahedron of volume 3.

This method is considerably faster in locating a tetrahedron (assuming one exists) because the index on the loops is smaller for most values of a, and more importantly, because the edge a need not be the largest edge of the tetrahedron—only the largest edge of the face *a-b-c*.

After running such a program we found two tetrahedra of volume 3 as follows:

$$a = 35 \quad b = 33 \quad c = 32 \quad d = 76 \quad e = 44 \quad f = 70$$

$$a = 47 \quad b = 32 \quad c = 21 \quad d = 58 \quad e = 56 \quad f = 76$$

Note that the largest edge of each is 76, so in our previous program we would have had to wait until $a = 76$ to get a solution—a period of approximately three weeks on our available system.

If we consider the shape of these tetrahedra, the first thing that springs to mind is that one of the faces of each is a fairly large triangle. The height of each tetrahedron corresponding to that largest face must therefore be quite small. In the second case if we measure the edges in yards (so that the base 47-58-76 covers much of a football field) and compute the height using the formula $V = (1/3)Bh$, we find that the height is less than one-quarter inch. Thus a model of these tetrahedra is probably not physically possible—they are much too flat. Their shapes are shown in FIGURE 4 as viewed from directly above when the largest face is in the plane:

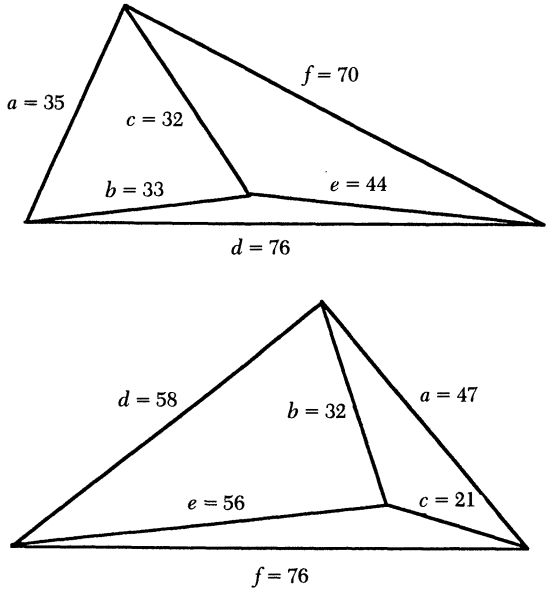


FIGURE 4

Some questions In our searches, some questions have arisen concerning the distribution of volumes of tetrahedra. It would appear that some volumes correspond to a large number of tetrahedra. For instance we have many examples with volume 42 (possibly not surprising since 42 is the ultimate hitchhiker answer to the universe), whereas we struggled to find a tetrahedron of volume 39. Thus the following questions:

1. Are there an infinite number of tetrahedra with integral edges and the same integral volume? We have two or more for most volumes less than 100, with notable exceptions. The same question for triangles yields the answer "no."

2. Is there a tetrahedron with volume any given multiple of 3? We have examples for every volume up to 99 except for $V = 87$. We know that any tetrahedron with this volume has an edge of at least 90 on every face. So far all attempts to discover this tetrahedron have been fruitless. Most of the volumes less than 99 correspond to tetrahedra with small edges (less than 50), but a few were troublesome. Volume 39 for instance corresponds to an integral tetrahedron whose smallest edge is 48 and whose largest edge is 308.

3. Is there a tetrahedron with integer edges, integer volume and all of whose faces have integer area? This problem was solved in 1985 (see [5]).

REFERENCES

1. Problem 1261, this MAGAZINE, 60 (1987), 40.
2. L. E. Dickson, *History of the Theory of Numbers*, Vol. 2, Chelsea Publishing Co., New York, pp. 497–502.
3. H. Dörrie, *100 Great Problems of Elementary Mathematics*, Dover Publications, Inc., New York, 1965, pp. 285–289.
4. B. M. Stewart, *Theory of Numbers*, MacMillan Publishing Co., New York, 1952, p. 183.
5. Problem 930, *Crux Mathematicorum* 11 (1985), 162–166.
6. D. M. Y. Sommerville, *An Introduction to the Geometry of N Dimensions*, Dover Publications, Inc., New York, 1958, pp. 124–125.

Alternate Solutions to Putnam Competition Problems

L.-S. HAHN
University of New Mexico
Albuquerque, NM 87131

In a recent article, Gary A. Martin [3] generalized Problem B-1 of the 1973 Putnam Competition problem by replacing the set of integers with abelian groups that have no nontrivial element of odd order. Some years ago, this author [1] found a totally different solution to the Putnam problem that also discards the condition that the numbers be integers. Our solution translates the given conditions into a system of linear equations and solves the system with basic linear algebra. Thus our solution provides an interesting exercise for linear algebra students.

The lemma we present below also solves Problem B-5 of the 1988 Putnam Competition.

Problem. Let $x_1, x_2, \ldots, x_{2n+1}$ be a set of integers such that, if any one of them is removed, the remaining ones can be divided into two sets of n integers with equal sums. Prove $x_1 = x_2 = \cdots = x_{2n+1}$.

We could rephrase it in a more appealing way: There are $2n + 1$ balls. If any one of these balls is removed, then the remaining $2n$ balls can be balanced n versus n in a pan balance. Is it necessary that all of them have the same weight?

Solution. Suppose the ball with weight x_1 is removed, then, by assumption, the remaining index set $\{2, 3, \ldots, 2n + 1\}$ can be divided into two disjoint sets E and F, each containing n elements such that

$$\sum_{j \in E} x_j = \sum_{j \in F} x_j.$$

REFERENCES

1. Problem 1261, this MAGAZINE, 60 (1987), 40.
2. L. E. Dickson, *History of the Theory of Numbers*, Vol. 2, Chelsea Publishing Co., New York, pp. 497–502.
3. H. Dörrie, *100 Great Problems of Elementary Mathematics*, Dover Publications, Inc., New York, 1965, pp. 285–289.
4. B. M. Stewart, *Theory of Numbers*, MacMillan Publishing Co., New York, 1952, p. 183.
5. Problem 930, *Crux Mathematicorum* 11 (1985), 162–166.
6. D. M. Y. Sommerville, *An Introduction to the Geometry of N Dimensions*, Dover Publications, Inc., New York, 1958, pp. 124–125.

Alternate Solutions to Putnam Competition Problems

L.-S. HAHN
University of New Mexico
Albuquerque, NM 87131

In a recent article, Gary A. Martin [3] generalized Problem B-1 of the 1973 Putnam Competition problem by replacing the set of integers with abelian groups that have no nontrivial element of odd order. Some years ago, this author [1] found a totally different solution to the Putnam problem that also discards the condition that the numbers be integers. Our solution translates the given conditions into a system of linear equations and solves the system with basic linear algebra. Thus our solution provides an interesting exercise for linear algebra students.

The lemma we present below also solves Problem B-5 of the 1988 Putnam Competition.

Problem. Let $x_1, x_2, \ldots, x_{2n+1}$ be a set of integers such that, if any one of them is removed, the remaining ones can be divided into two sets of n integers with equal sums. Prove $x_1 = x_2 = \cdots = x_{2n+1}$.

We could rephrase it in a more appealing way: There are $2n + 1$ balls. If any one of these balls is removed, then the remaining $2n$ balls can be balanced n versus n in a pan balance. Is it necessary that all of them have the same weight?

Solution. Suppose the ball with weight x_1 is removed, then, by assumption, the remaining index set $\{2, 3, \ldots, 2n + 1\}$ can be divided into two disjoint sets E and F, each containing n elements such that

$$\sum_{j \in E} x_j = \sum_{j \in F} x_j.$$

Therefore, we have

$$\sum_{j=2}^{2n+1} \mathscr{E}_{1j} x_j = 0,$$

where half of the $\mathscr{E}_{1j}$'s are $+1$ and the other half are -1 $(j = 2, \ldots, 2n+1)$. Continuing in the same manner, we obtain a system of linear equations

$$\sum_{j=1}^{2n+1} \mathscr{E}_{ij} x_j = 0 \qquad (i = 1, 2, \ldots, 2n+1),$$

where

$$\mathscr{E}_{ii} = 0 \qquad (i = 1, 2, \ldots, 2n+1),$$
$$\mathscr{E}_{ij} = \pm 1 \qquad (i \neq j),$$

and

$$\sum_{j=1}^{2n+1} \mathscr{E}_{ij} = 0 \qquad (i = 1, 2, \ldots, 2n+1).$$

We know that $x_1 = x_2 = \cdots = x_{2n+1}$ describes a one-dimensional family of solutions of this system of linear homogeneous equations, so the $(2n+1) \times (2n+1)$ coefficient matrix has nullity at least one. Equivalently, the rank of the coefficient matrix is at most $2n$, and our problem becomes whether the rank is actually equal to $2n$. The following lemma shows that the $(2n) \times (2n)$ submatrix in the upper left corner of the coefficient matrix has rank $2n$. From this we conclude that the only solutions satisfy $x_1 = x_2 = \cdots = x_{2n+1}$.

LEMMA. *Suppose an $m \times m$ square matrix A has the property that all the entries on the main diagonal are even numbers, while all the rest of the entries are odd numbers. Then the rank of A is m (i.e., A is invertible) provided that m is even.*

Proof. If m is even, then

$$A^2 \equiv I_m \qquad (\mathrm{mod}\, 2),$$

where I_m is the identity matrix of order m. Therefore

$$\det A \not\equiv 0 \qquad (\mathrm{mod}\, 2).$$

We remark that our proof is valid even if the "weights" $x_1, x_2, \ldots, x_{2n+1}$ are complex numbers.

REFERENCES

1. L.-S. Hahn, Solution to Problem 4304, *Mathmedia*, 16 (1981), 122–123, Inst. Math, Acad. Sinica, Taipei, Taiwan (in Chinese).
2. A. P. Hillman, The William Lowell Putnam Mathematics Competition, *Amer. Math. Monthly*, 81 (1974).
3. G. A. Martin, A class of Abelian groups arising from an analysis of a proof, *Amer. Math. Monthly* 95 (1988).

On Two Classes of Extremum Problems without Calculus

MURRAY S. KLAMKIN
University of Alberta
Edmonton, Alberta, Canada T6G 2G1

Calculus provides us with systematic, general, and powerful methods for attacking extremum problems. Because of its generality, there are often simpler methods for solving specific problems. These usually involve inequalities, e.g., Arithmetic-Geometric Mean (A.M.-G.M.), Jensen, Hölder, Minkowski, or a discriminant. If one surveys the maximum and minimum problems in our calculus texts, one will find that a great many of them can be done much more efficiently just using the A.M.-G.M. inequality in two and three variables, i.e.,

$$a^2 + b^2 \geqslant 2ab \qquad \text{and} \qquad a^3 + b^3 + c^3 \geqslant 3abc.$$

In this note, we will first consider a two-dimensional problem of Ogilvy [1], which was solved using Lagrange multipliers. We redo the problem using Hölder's inequality [2] plus some other quite elementary inequalities. We also point out an oversight regarding the maximum values and then generalize the problem to n dimensions. Secondly, we consider two problems that can be solved using an appropriate discriminant. The first one is due to Langley [3]. Again, we point out an oversight and then extend the result.

PROBLEM 1. Ogilvy generalizes the problem "If a point $P(h,k)$ is a point in the first quadrant, what line through P, together with the positive coordinate axes, forms the right triangle that minimizes (I) the sum of the base (a) and altitude (b); (II) the hypotenuse?" by minimizing $f(a,b) = a^n + b^n$, $n > 0$, subject to the constraint $h/a + k/b = 1$. He obtains

$$\min f(a,b) = \{h^{n/(n+1)} + k^{n/(n+1)}\}^{n+1}$$

by the usual Lagrange multiplier rule. He also notes that "If $n < 0$, the problem becomes that of maximizing the function $f(a,b)$ instead of minimizing it. All the calculations made for $n > 0$ hold for $n < 0$, except for the adjustments in wording required by the fact that the extremal is now a maximum instead of a minimum." There is an oversight in the latter statement that we will correct.

Our solution is via Hölder's inequality

$$(u+v)^{1/p}(r+s)^{1/q} \geqslant u^{1/p}r^{1/q} + v^{1/p}s^{1/q}$$

where $u, v, r, s \geqslant 0$, $1/p + 1/q = 1$ and $p, q > 1$. There is equality *iff* $u/r = v/s$.

Case 1. $n > 0$.

$$(a^n + b^n)^{1/(n+1)}(h/a + k/b)^{n/(n+1)} \geqslant h^{n/(n+1)} + k^{n/(n+1)}.$$

Thus,

$$\min(a^n + b^n) = \{h^{n/(n+1)} + k^{n/(n+1)}\}^{n+1}$$

and with equality *iff* $h/a^{n+1} = k/b^{n+1}$.

For the remaining cases with $n < 0$, we will change n to $-n$ and consider $1/a^n + 1/b^n$.

Case 2. $1 > n > 0$.

$$(h/a + k/b)^n (h^{n/(n-1)} + k^{n/(n-1)})^{1-n} \geqslant 1/a^n + 1/b^n.$$

Thus,

$$\max(1/a^n + 1/b^n) = (h^{n/(n-1)} + k^{n/(n-1)})^{1-n}.$$

For this case, there is also a lim inf and it is determined by

$$1/a^n + 1/b^n \geqslant \{(h/a)^n + (k/b)^n\}/\max(h^n, k^n) > (h/a + k/b)^n/\max(h^n, k^n).$$

This yields

$$\liminf(1/a^n + 1/b^n) = 1/\max(h^n, k^n).$$

(Note that $\{(h/a)^n + (k/b)^n\} = (h/a + k/b)^n$ only if one of $h/a, k/b$ vanishes.) This lim inf is not achieved as the minimum value since in order to obtain this value, the line must be parallel to one of the coordinate axes. But this is ruled out since a and b are finite. However, we can get arbitrarily close to $1/\max(h^n, k^n)$ by taking the line arbitrarily close to being parallel to one of the axes, i.e., letting a or $b \to \infty$.

Case 3. $n = 1$.

$$(h/a + k/b)/\max(h, k) < 1/a + 1/b < (h/a + k/b)/\min(h, k).$$

We are assuming that $h \neq k$, otherwise there is no problem here. This yields

$$\liminf(1/a + 1/b) = 1/\max(h, k) \text{ and } \limsup(1/a + 1/b) = 1/\min(h, k).$$

Case 4. $n > 1$.

$$(1/a^n + 1/b^n)^{1/n} (h^{n/(n-1)} + k^{n/(n-1)})^{(n-1)/n} \geqslant h/a + k/b = 1.$$

Thus,

$$\min(1/a^n + 1/b^n) = (h^{n/(n-1)} + k^{n/(n-1)})^{1-n}$$

and is achieved *iff* $(a, b) = \lambda(h^{1/(1-n)}, k^{1/(1-n)})$ for some real number λ. There is also a lim sup and it is determined by

$$(1/a^n + 1/b^n) \leqslant \{(h/a)^n + (k/b)^n\}/\min(h^n, k^n) < (h/a + k/b)^n/\min(h^n, k^n).$$

Whence,

$$\limsup(1/a^n + 1/b^n) = 1/\min(h^n, k^n).$$

The special case here of $n = 2$ corresponds geometrically to finding the extreme values of the perpendicular from the origin onto the line.

The above shows that Ogilvy's oversight is in the maximum values for the cases $n \leqslant -1$. Most likely the error crept in since there was no check on "sufficiency," which is usual when one uses Lagrange multipliers.

For more than a two variable case, as here, the Lagrange multiplier method usually leads to some algebraic difficulties. One has to determine all the real solutions of the

resulting set of simultaneous equations that are usually non-linear. This only takes care of the necessary conditions for the extrema. Then one has to check out end point extrema as well as sufficiency. For example, just consider the problem of finding the maximum of $F(x, y, z)$ in the cube $a \geqslant x, y, z \geqslant 0$. First one has to solve the 3-dimensional problem $F_x = F_y = F_z = 0$ in the interior of the cube. Then one has to consider the six 2-dimensional cases on the faces of the cube. Next comes the 12 1-dimensional cases on the edges of the cube. Finally, one has to check the values at the eight vertices of the cube. One should always be on the lookout for some method (like an appropriate inequality) to avoid all this work. As a further example of this, we show that the following generalization of Ogilvy's problem to n dimensions is hardly more work than before.

Here, $P(x_1, x_2, \ldots, x_n)$ is a point in the first orthant and we want to pass a hyperplane through it which minimizes or maximizes the sum of the mth powers of the coordinate intercepts. Our problem is to find the extrema of

$$F \equiv a_1^m + a_2^m + \cdots + a_n^m$$

subject to the constraint

$$x_1/a_1 + x_2/a_2 + \cdots + x_n/a_n = 1.$$

Consider the minimum for $m > 0$. By Hölder's inequality for n variables,

$$\begin{aligned}(a_1^m + a_2^m + \cdots + a_n^m)^{1/(m+1)}&(x_1/a_1 + x_2/a_2 + \cdots + x_n/a_n)^{m/(m+1)} \\ &\geqslant x_1^{m/(m+1)} + x_2^{m/(m+1)} + \cdots + x_n^{m/(m+1)}\end{aligned}$$

and with equality if, and only if,

$$x_1/a_1^{(m+1)} = x_2/a_2^{(m+1)} = \cdots = x_n/a_n^{(m+1)}.$$

Thus,

$$\min(a_1^m + a_2^m + \cdots + a_n^m) = \left\{x_1^{m/(m+1)} + x_2^{m/(m+1)} + \cdots + x_n^{m/(m+1)}\right\}^{m+1}.$$

The rest of the cases go through as before. For some extra geometric dividends associated with this extended problem see the analogous two-dimensional results in [1].

Also, it should be noted that there are many times when a non-calculus solution of an extremum problem is discovered only after first obtaining the necessary conditions from the calculus methods.

Our second class of extrema problems are those solvable by means of the discriminant of a quadratic equation.

PROBLEM 2. Here we want to determine the maximum and minimum values of

$$\left\{\sin\theta + \sqrt{\sin^2\theta + \sin^2\alpha}\right\}\cos\theta, \tag{1}$$

where α is fixed and θ is the variable. This problem was solved nicely and without the calculus by E. M. Langley [3] and illustrates still another way of determining maxima and minima. However, there is a further remark to add to complete the solution. Also, we then give extensions to more general functions.

Langley's elegant solution goes as follows:

Let

$$u = \{\sin\theta + \sqrt{\sin^2\theta + \sin^2\alpha}\}\cos\theta,$$

then

$$(u - \sin\theta\cos\theta)^2 = (\sin^2\theta + \sin^2\alpha)\cos^2\theta,$$

or

$$u^2 - 2u\sin\theta\cos\theta = \sin^2\alpha\cos^2\theta,$$

or

$$u^2\tan^2\theta - 2u\tan\theta + u^2 - \sin^2\alpha = 0.$$

Since $\tan\theta$ must be real, the discriminant of the latter quadratic equation in $\tan\theta$ must be nonnegative. Thus,

$$u^2 \geqslant u^2(u^2 - \sin^2\alpha)$$

or

$$u^2 \leqslant 1 + \sin^2\alpha.$$

Finally,

$$-\sqrt{1 + \sin^2\alpha} \leqslant u \leqslant \sqrt{1 + \sin^2\alpha}\,.$$

To complete Langley's solution, one should show that the latter extreme values of u are actually achieved and give the corresponding θ values. To find θ, we solve

$$\sqrt{1 + \sin^2\alpha} = \{\sin\theta + \sqrt{\sin^2\theta + \sin^2\alpha}\}\cos\theta.$$

By rearranging and squaring, we get

$$(\sin^2\theta + \sin^2\alpha)\cos^2\theta = 1 + \sin^2\alpha - 2(\sin\theta\cos\theta)\sqrt{1 + \sin^2\alpha} + \sin^2\theta\cos^2\theta$$

or

$$2(\sin\theta\cos\theta)\sqrt{1 + \sin^2\alpha} = 1 + \sin^2\alpha\sin^2\theta.$$

Squaring again, replacing $\cos^2\theta$ by $1 - \sin^2\theta$, and rearranging gives

$$\{(\sin^2\alpha + 2)\sin^2\theta - 1\}^2 = 0.$$

Hence, $\sin^2\theta = 1/(\sin^2\alpha + 2)$ and then $\cos^2\theta = (\sin^2\alpha + 1)/(\sin^2\alpha + 2)$. Substituting these values back in (1) gives the previously obtained extremum values. For the maximum, $\sin\theta$ and $\cos\theta$ are both positive while for the minimum value, $\sin\theta$ is positive and $\cos\theta$ is negative.

A calculus solution here is direct but the solution of the derivative equation involves considerably more work than before. Incidentally, the method of obtaining extreme values by appealing to the discriminant of a quadratic equation appears in quite a few old English texts on mechanics. In fact, the Langley problem is equivalent physically to obtaining the maximum range of a shot-put released at a height h above the ground with an initial velocity V_0 and assuming no drag [4]. Here, $\sin^2\alpha$ corresponds to $2hg/V_0^2$ where g is the gravitational constant.

One way to generalize the function in (1) is to start with the equation

$$u^2 \tan^2 \theta - 2u \tan \theta + u^2 - \sin^2 \alpha = 0 \tag{2}$$

and replace $\tan \theta$ by a more general function $F(\theta)$. Retracing our steps, the function corresponding to (1) is

$$\left\{F + \sqrt{F^2 + (F^2+1)\sin^2 \alpha}\right\} \Big/ (F^2+1). \tag{3}$$

The extremum values of (3) are the same as for (1). Again, one should check whether these extreme values can actually be achieved. For further generalizations, we can not only replace $\tan \theta$ by $F(\theta)$ in (2) but we can also change the functional appearance of the u. However, the resulting equation in u should be solvable.

Our last problem is a well-known one and can be solved in quite a variety of different ways. The discriminant method here is quite elegant.

PROBLEM 3. Find the shortest distance in E^n from a given point $P(p_1, p_2, \ldots, p_n)$ to a given line l whose parametric representation is $x_i = a_i + h_i t$, $i = 1, 2, \ldots, n$.

Geometrically, the desired distance is given by the radius r of a sphere centered at the given point P and that is tangent to the given line l. Consequently, the intersection of the sphere with the line must consist of a single point. Since the equation of the sphere is $\Sigma(x_i - p_i)^2 = r^2$ (here and subsequently, the sums are over $i = 1, 2, \ldots, n$), the parameter t must satisfy

$$\sum (h_i t + a_i - p_i)^2 = r^2$$

or

$$t^2 \sum h_i^2 + 2t \sum h_i (a_i - p_i) + \sum (a_i - p_i)^2 - r^2 = 0.$$

Since the discriminant of this quadratic in t must be zero, the square of the shortest distance is given by

$$r^2 = \sum (a_i - p_i)^2 - \left\{\sum h_i (a_i - p_i)\right\}^2 \Big/ \sum h_i^2.$$

For other examples of solutions of extrema problems without the calculus, see [5, 6, 7, 8].

Acknowledgement. I am grateful to the referees for a number of helpful suggestions.

REFERENCES

1. C. S. Ogilvy, Extra dividends from a calculus problem, this MAGAZINE 41 (1968), 280–281.
2. D. S. Mitrinovic, *Analytic Inequalities*, Springer-Verlag, Heidelberg, 1970, pp. 50–52.
3. E. M. Langley, Problem #38, *Math. Gazette* 1 (1896), 18.
4. M. S. Klamkin, Dynamics: throwing the javelin, putting the shot, *UMAP Journal* 6 (1985), 3–18.
5. G. Pólya, *Induction and Analogy in Mathematics*, Princeton University Press, Princeton, NJ, 1954, pp. 121–141.
6. I. Niven, *Maxima and Minima without Calculus*, MAA, Washington, DC, 1981.
7. R. P. Boas and M. S. Klamkin, Extrema of polynomials, this MAGAZINE 50 (1977), 75–78.
8. M. S. Klamkin, On the teaching of mathematics so as to be useful, *Educ. Studies in Math.* 1 (1968), 126–160.

Confidence Intervals from Groups

GARY J. SHERMAN
Rose-Hulman Institute of Technology
Terre Haute, IN 47803

Suppose you want a confidence interval for the mean, μ, of a continuous random variable $\mathbf{Y}$. What do you do? More than likely, you take a random sample $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ from $\mathbf{Y}$, compute the sample mean $\bar{\mathbf{Y}} = (1/n)\Sigma_{i=1}^{n}\mathbf{Y}_i$, compute the sample variance

$$\mathbf{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right)^2,$$

and then ponder

$$\mathbf{Z} = \frac{\sqrt{n}}{\sigma}\left(\bar{\mathbf{Y}} - \mu\right) \quad \text{and} \quad \mathbf{T} = \frac{\sqrt{n}}{\mathbf{S}}\left(\bar{\mathbf{Y}} - \mu\right),$$

where σ is the standard deviation of $\mathbf{Y}$. If σ is known (or n is large enough, say $n \geqslant 30$, to replace σ with $\mathbf{S}$), then the Central Limit Theorem implies that $\mathbf{Z}$ is approximately standard normal. If σ is unknown (or n is too small, say, $n < 30$, to replace σ with $\mathbf{S}$) and $\mathbf{Y}$ is normal (or at least "mound shaped") then $\mathbf{T}$ is Student t (approximately Student t) with $n - 1$ degrees of freedom. The appropriate choice of $\mathbf{Z}$ or $\mathbf{T}$—only your random sample knows for sure—provides an exact or approximate $\alpha \times 100\%$ confidence interval for μ:

$$\left[\bar{\mathbf{Y}} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \bar{\mathbf{Y}} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right] \quad \text{or} \quad \left[\bar{\mathbf{Y}} - t_{\alpha/2}\cdot\frac{\mathbf{S}}{\sqrt{n}}, \bar{\mathbf{Y}} + t_{\alpha/2}\cdot\frac{\mathbf{S}}{\sqrt{n}}\right].$$

Here's a nonparametric approach based on order statistics that you might try: Compute the sample mean for some subsamples of your random sample and list these subsample means in ascending order. If the subsamples you have chosen form a group with respect to symmetric difference, then the intervals formed by successive (ordered) subsample means are equally likely to contain μ. This method of producing confidence intervals makes use of the notion of "resampling" (see [1]), which has been attributed in [3] to Tukey [6]. The connection with group theory is due to a theorem of John Hartigan [3].

I first became aware of Hartigan's result while auditing Robert Ling's course on data analysis at Clemson University a few years ago and immediately found the theorem appealing because it establishes a specific, elementary, and, in my opinion, aesthetic connection between statistics and group theory (or, depending on your tastes, vector spaces over $\mathbf{Z}_2$ or algebraic coding theory). The goal of this note is to convince you that Hartigan's result is accessible to undergraduates.

A group theoretic approach Normality, σ and S will not be mentioned (again) in this section. Given the form of the two confidence intervals constructed above you may find this a bit surprising. The price of this generality and of group theory's contribution is that $\mathbf{Y}$ will be assumed to be symmetrically distributed about its location parameter. (If the mean does not exist, then take μ to be the median.) Let's take $\mu = 0$ and $n = 3$ for transient notational convenience and permanent pedagogical clarity, respectively.

Again, suppose you want a confidence interval for the mean (median), 0, of a continuous symmetric random variable $\mathbf{Y}$. Compute $\overline{\mathbf{Y}}$. The symmetry of $\mathbf{Y}$ guarantees $\overline{\mathbf{Y}}$ and $-\overline{\mathbf{Y}}$ have the same distribution. So, $\text{Prob}(\overline{\mathbf{Y}} < 0) = \text{Prob}(-\overline{\mathbf{Y}} > 0) = \text{Prob}(\overline{\mathbf{Y}} > 0)$ which means that $(-\infty, \overline{\mathbf{Y}})$ and $(\overline{\mathbf{Y}}, \infty)$ are 50% confidence intervals for $\mu = 0$. I'll anticipate your objections before you stop reading in protest:

i) Where's the group?
ii) The sample of size three is unnecessary; one observation of $\mathbf{Y}$ would have served the same purpose.
iii) The confidence intervals are not finite.
iv) The confidence level is only 50%.

Here's the group $\mathbf{Z}_2^3 = \mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus \mathbf{Z}_2$. (Recall that addition is done componentwise modulo two.) To each nonzero element abc of $\mathbf{Z}_2^3$ we associate the subsample mean, denoted by $\overline{\mathbf{X}}_{abc}$, of the sample elements identified by a 1 in the corresponding position in abc. Thus $\overline{\mathbf{X}}_{111} = \overline{\mathbf{Y}}$ while $\overline{\mathbf{X}}_{000}$ is defined to be μ.

Computing several subsample means will enable us to extract more information from the random sample and better illustrate the role of $\mathbf{Z}_2^3$. Specifically, associate a subsample mean with each element of the subgroup

$$\begin{aligned} C &= \{000, 011, 101, 110\}, \\ \overline{\mathbf{X}}_{000} &= 0 \\ \overline{\mathbf{X}}_{011} &= (\mathbf{Y}_2 + \mathbf{Y}_3)/2 \\ \overline{\mathbf{X}}_{101} &= (\mathbf{Y}_1 + \mathbf{Y}_3)/2 \\ \overline{\mathbf{X}}_{110} &= (\mathbf{Y}_1 + \mathbf{Y}_2)/2, \end{aligned}$$

and then denote the ordered non-trivial subsample means by $\overline{\mathbf{X}}_1$, $\overline{\mathbf{X}}_2$, and $\overline{\mathbf{X}}_3$.

Finite intervals? Sure:

$$\left(\overline{\mathbf{X}}_1, \overline{\mathbf{X}}_2\right), \left(\overline{\mathbf{X}}_2, \overline{\mathbf{X}}_3\right), \quad \text{and} \quad \left(\overline{\mathbf{X}}_1, \overline{\mathbf{X}}_3\right).$$

Confidence intervals? That's a bit more difficult but the idea is the same as above. There we observed that 0 or $\overline{\mathbf{Y}}$ was the smaller of $\{0, \overline{\mathbf{Y}}\}$ with equal probability. Here we will see that, for each $abc \in C$, $\overline{\mathbf{X}}_{abc}$ is the kth smallest element of $\{\overline{\mathbf{X}}_{000}, \overline{\mathbf{X}}_{011}, \overline{\mathbf{X}}_{101}, \overline{\mathbf{X}}_{110}\}$ with probability $1/4$.

Let's be specific and compute the probability that $\overline{\mathbf{X}}_{011}$ is second smallest. The following array is helpful in understanding the computation. Read " $\sim$ " as "has the same distribution as."

$$\begin{aligned} \overline{\mathbf{X}}_{011} - \overline{\mathbf{X}}_{000} &= (\mathbf{Y}_2 + \mathbf{Y}_3)/2 \sim (\mathbf{Y}_2 + \mathbf{Y}_3)/2 = \overline{\mathbf{X}}_{011+000} = \overline{\mathbf{X}}_{011} \\ \overline{\mathbf{X}}_{011} - \overline{\mathbf{X}}_{011} &= 0 \sim 0 = \overline{\mathbf{X}}_{011+011} = \overline{\mathbf{X}}_{000} \\ \overline{\mathbf{X}}_{011} - \overline{\mathbf{X}}_{101} &= (-\mathbf{Y}_1 + \mathbf{Y}_2)/2 \sim (\mathbf{Y}_1 + \mathbf{Y}_2)/2 = \overline{\mathbf{X}}_{011+101} = \overline{\mathbf{X}}_{110} \\ \overline{\mathbf{X}}_{011} - \overline{\mathbf{X}}_{110} &= (-\mathbf{Y}_1 + \mathbf{Y}_3)/2 \sim (\mathbf{Y}_1 + \mathbf{Y}_3)/2 = \overline{\mathbf{X}}_{011+110} = \overline{\mathbf{X}}_{101} \end{aligned}$$

Notice that $\overline{\mathbf{X}}_{011}$ is second smallest if and only if exactly two of the differences $\overline{\mathbf{X}}_{011} - \overline{\mathbf{X}}_{abc}$ are negative. But the differences in question have the same distribution as a permutation of the original set of four subsample means because $\mathbf{Y}$ is symmetric (column two to column three) and C is a subgroup of $\mathbf{Z}_2^3$ (column four to column five). Therefore, the probability, say p, that $\overline{\mathbf{X}}_{011}$ is second smallest is the probability

that two of $\bar{\mathbf{X}}_{000}$, $\bar{\mathbf{X}}_{011}$, $\bar{\mathbf{X}}_{101}$, and $\bar{\mathbf{X}}_{110}$ are negative. The utility of C's group structure strikes again since the role of $\bar{\mathbf{X}}_{011}$ in this paragraph may be played by each $\bar{\mathbf{X}}_{abc}$. Thus each $\bar{\mathbf{X}}_{abc}$ is second smallest with probability p, so $p = 1/4$.

Rereading the previous paragraph with kth smallest in place of second smallest should convince you that 0 is the kth smallest element of $\{\bar{\mathbf{X}}_{000}, \bar{\mathbf{X}}_{011}, \bar{\mathbf{X}}_{101}, \bar{\mathbf{X}}_{110}\}$ with probability $1/4$ as claimed. This means

$$P(0 < \bar{\mathbf{X}}_1) = P(\bar{\mathbf{X}}_1 < 0 < \bar{\mathbf{X}}_2) = P(\bar{\mathbf{X}}_2 < 0 < \bar{\mathbf{X}}_3)$$
$$= P(\bar{\mathbf{X}}_3 < 0) = 1/4.$$

It follows that $(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_3)$ is a finite 50% confidence interval for 0. The $\bar{\mathbf{X}}_i$'s are called **typical values for 0**.

The journey from column one to column four in the array went smoothly because of the symmetry of $\mathbf{Y}$, the group structure of C *and* the fact that each non-trivial subsample had the same number of observations. For example, if C were chosen to be $\{000, 011, 100, 111\}$, then

$$\bar{\mathbf{X}}_{011} - \bar{\mathbf{X}}_{100} = (-2\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3)/2 \sim (2\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3)/2 \neq \bar{\mathbf{X}}_{011+100} = \bar{\mathbf{X}}_{111}.$$

If, for the time being, we use subsample sums ($\mathbf{X}_{abc}$) to construct the intervals rather than subsample means ($\bar{\mathbf{X}}_{abc}$), this difficulty can be avoided. For example,

$$\mathbf{X}_{011} - \mathbf{X}_{100} = (-\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3) \sim (\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3) = \mathbf{X}_{011+100} = \mathbf{X}_{111}.$$

At the risk of feeling as if you are caught in a nested loop: Reread the paragraph containing the array again with $\mathbf{X}_{abc}$ in place of $\bar{\mathbf{X}}_{abc}$ and a subgroup of your choice to see that our conclusions remain valid for subsample sums regardless of the sizes of the subsamples. The temporary cost of this change is that the confidence intervals constructed for 0 from subsample sums will be longer than those constructed from subsample means.

The lemma lurking in the preceding discussion is that *if* $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ *is a random sample from* $\mathbf{Y}$, *which is continuous and symmetric about* 0, *and* C *is a subgroup of* $\mathbf{Z}_2^n$, *then the ordered* $\mathbf{X}_{ab\cdots c}$*'s,* $ab \cdots c \in C - \{00\ldots0\}$, *are typical values for* 0; *i.e., each of the* $|C|$ *intervals* $(-\infty, \mathbf{X}_1), (\mathbf{X}_1, \mathbf{X}_2), \ldots, (\mathbf{X}_{|C|-1}, \infty)$ *contain* 0 *with probability* $1/|C|$.

THEOREM. *If* $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ *is a random sample from* $\mathbf{Y}$, *which is continuous and symmetric about* μ, *and* C *is a subgroup of* $\mathbf{Z}_2^n$, *then the ordered* $\bar{\mathbf{X}}_{ab\cdots c}$*'s,* $ab \cdots c \in C - \{00 \cdots 0\}$, *are typical values for* μ; *i.e., each of the* $|C|$ *intervals* $(-\infty, \bar{\mathbf{X}}_1), (\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2), \ldots, (\bar{\mathbf{X}}_{|C|-1}, \infty)$ *contains* μ *with probability* $1/|C|$.

Proof. Notice that $\bar{\mathbf{X}}_{ab\cdots c} < \mu$ is equivalent to $\mathbf{X}_{ab\cdots c} - w \cdot \mu < 0$, where w is the number of ones in $ab \cdots c$. Thus the probability that μ is the kth smallest of $\{\bar{\mathbf{X}}_{ab\cdots c}\}$ is the probability that 0 is the kth smallest of $\{\mathbf{X}_{ab\cdots c} - w \cdot \mu\}$. Applying the lemma to the random sample $\mathbf{Y}_1 - \mu, \ldots, \mathbf{Y}_n - \mu$ of $\mathbf{Y} - \mu$ implies that this probability is $1/|C|$.

An example Suppose we have a random sample of size 31 (see FIGURE 1) from a continuous symmetric random variable $\mathbf{Y}$ ($\mathbf{Y}$ is uniform on $[0, 1]$) and we would like a 90% (at least) confidence interval for μ. Taking $(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_{|C|-1})$ as the confidence interval requires $(|C| - 2)/|C| \geq .90$; i.e., $|C| \geq 20$. Since the order of each subgroup of $\mathbf{Z}_2^{31}$ is a divisor of 2^{31}, choosing a subgroup, C, of order 2^5 will suffice. There are a lot of subgroups of $\mathbf{Z}_2^{31}$ of order 2^5. How do we get one? By taking the linear algebra

1) 0.181916	2) 0.765071	3) 0.686501
4) 0.935966	5) 0.245895	6) 0.731319
7) 0.499784	8) 0.612373	9) 0.008250
10) 0.844945	11) 0.503263	12) 0.900799
13) 0.314173	14) 0.586790	15) 0.997342
16) 0.426548	17) 0.218642	18) 0.373209
19) 0.198238	20) 0.133826	21) 0.217568
22) 0.174174	23) 0.995979	24) 0.284221
25) 0.856858	26) 0.356958	27) 0.707513
28) 0.204196	29) 0.641297	30) 0.739791
31) 0.621815		

FIGURE 1

point of view and thinking of C as the row space of a 5×31 binary matrix of rank five.

$$M = \begin{bmatrix} 1010111011000111110011010010000 \\ 0101011101100011111001101001000 \\ 0010101110110001111100110100100 \\ 0001010111011000111110011010010 \\ 0000101011101100011111001101001 \end{bmatrix}$$

is such a matrix. The matrix product $[11111]M$ produces the element 11001101001000010101111011000111 of C and the associated subsample mean 0.469372.

Turning your computer loose yields the 93.75% confidence interval we were after: [0.383746, 0.618873]. Normal theory and MINITAB* yield [0.415, 0.615] if σ is estimated from the sample, and [0.418, 0.612] if σ is taken to be $\sqrt{1/12}$.

This example illustrates, as you may have suspected, that confidence intervals based on standard normal asymptotic methods (when applicable) are more efficient than those based on groups—but not nearly as much fun. The statistical package IDA [4] contains a subsampling procedure (SAMP) for producing confidence intervals that is even less efficient because it uses only the spirit of the group approach rather than group structure itself. The rationale of the approach stems from [2] and [3] and goes something like this: Empirical analysis suggests that when groups that do nearly as well as normal theory are found they tend to consist mostly of subsamples using about one half of the observations on Y; such groups are difficult to construct so why not select so called random half-samples by performing Bernoulli trials (with $p = 1/2$) on each observation of Y and hope for the best. (For the random sample above, 31 random half-samples yielded the approximate 93.75% confidence interval [0.404216, 0.613835].) A report of a simulation study comparing the efficiency of groups with random subsamples appears in [2]. The message is that groups are "sufficiently superior" to random subsamples "to offset the tricky generation procedure" for groups, particularly for small sets (order less than 31) of random subsamples.

The development of algebraic coding theory, of which the authors of [2], [3], and [4] were apparently unaware, renders the rationalization mentioned in the preceding paragraph less compelling and makes the group generation procedure less "tricky" than the ad hoc techniques of [2] and [3]. Take another look at the matrix M. Viewing C as a code means that M is its generator matrix [5]; i.e., each codeword in C is a linear combination of the rows of M. Since rows two, three, four and five of M are cyclic permutations of row one, virtually all of the information in M is in row one.

*MINITAB is a registered trademark of Minitab, Inc.

Indeed, thinking of row one as the polynomial $g(x) = 1 + x^2 + x^4 + x^5 + x^6 + x^8 + x^9 + x^{13} + x^{14} + x^{15} + x^{16} + x^{17} + x^{20} + x^{21} + x^{23} + x^{26} \in Z_2[x]$, C may be viewed as the principal ideal of $Z_2[x]/(x^{31} - 1)$ generated by $g(x)$. Thus to obtain a 93.75% confidence level for a random sample of size 31, we need only $g(x)$ and an understanding of the algebra of $\mathbf{Z}_2[x]/(x^{31} - 1)$.

A problem To my knowledge, the connection between confidence intervals and algebraic coding theory has been exploited by neither statisticians nor coding theorists. While pursuing this connection is beyond the scope of this paper, the general problem is clear: Characterize codes using statistical criteria rather than criteria with their genesis in information transmission.

We close with a less ambitious problem. Hartigan's original result is for any set of continuous random variables (not just a random sample) symmetric about a common location parameter and shows, in this more general case, that group structure is both necessary and sufficient to produce typical values. Since we have been restricting our attention to random samples one should ask if group structure is really necessary. I asked Erich Friedman, a former student of mine, who is now a graduate student at Cornell. He said "hmmmmm", went away for a couple of days and returned to say "No, just look at this example".

Let $\mathbf{Y}_1$, $\mathbf{Y}_2$, and $\mathbf{Y}_3$ be a random sample from the continuous random variable $\mathbf{Y}$ that is symmetric about 0. The subset $\{000, 110, 101\}$ of $\mathbf{Z}_2^3$ is not a subgroup yet it produces a typical set for 0.

To see that $\mathbf{X}_{110} = \mathbf{Y}_1 + \mathbf{Y}_2$ and $\mathbf{X}_{101} = \mathbf{Y}_1 + \mathbf{Y}_3$ comprise a typical set for 0 consider the eight possibilities for the signs of $\mathbf{Y}_1$, $\mathbf{Y}_2$, and $\mathbf{Y}_3$.

SIGN($\mathbf{Y}_1$)	+	+	+	+	−	−	−	−
SIGN($\mathbf{Y}_2$)	+	+	−	−	+	+	−	−
SIGN($\mathbf{Y}_3$)	+	−	+	−	+	−	+	−

Since the distributions of the $\mathbf{Y}_i$'s are symmetric, the eight sign patterns occur with equal probability. Thus the probability that 0 occurs in the interval $(\mathbf{X}_2, \infty)$, which is the probability that both $\mathbf{X}_{110}$ and $\mathbf{X}_{101}$ are negative, is $(1/8)(0 + 0 + 0 + 1/3 + 1/3 + 1/2 + 1/2 + 1) = 1/3$. A similar argument shows the probability that 0 occurs in the interval $(-\infty, \mathbf{X}_1)$ also to be $1/3$. This means $\mathbf{X}_1$, and $\mathbf{X}_2$ are typical values for 0. In a similar manner one can show that if the random sample is of size n, then $\{0 \cdots 0, 110 \cdots 0, 1010 \cdots 0, \ldots, 10 \cdots 01\}$ produces a typical set for 0.

The problem: Characterize those subsets of $\mathbf{Z}_2^n$ that produce a typical set for a continuous random variable symmetric about 0.

Acknowledgment. I wish to thank the referees for their helpful suggestions.

REFERENCES

1. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia, 1982.
2. A. Forsythe and J. A. Hartigan, Efficiency of confidence intervals generated by repeated subsample calculations, *Biometrika* 57 (1970), 629–639.
3. J. A. Hartigan, Using subsample values as typical values, *J. Amer. Stat. Assoc.* 64 (1969), 1303–1317.
4. R. F. Ling and H. V. Roberts, *Users Manual for IDA*, The Scientific Press, S. San Francisco, 1980.
5. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, New York, 1977.
6. J. W. Tukey, Bias and confidence in not quite large samples, *Annals of Math. Stat.* 29 (1958), 614.

A Mean-Value Property of Cubic Polynomials—without Mean Values

D. F. BAILEY
Trinity University
San Antonio, TX 78212

It is well known that for quadratic polynomials the Mean Value Theorem takes the form

$$\frac{f(x)-f(y)}{x-y}=f'\left(\frac{x+y}{2}\right). \tag{1}$$

Several authors [3], [4], [5] have likewise shown that if condition (1) holds then $f(x)=ax^2+bx+c$. Aczél [1] shows (a bit more than) if f satisfies

$$f(x)-f(y)=(x-y)h(x+y), \tag{2}$$

then f is a quadratic. Surprisingly Aczél's result assumes neither differentiability nor continuity of f and there is no mention of the mean of x and y.

An easy computation will show that if f is a cubic polynomial, then

$$f[x,y,z]=\frac{1}{2}f''\left(\frac{x+y+z}{3}\right), \tag{3}$$

where $f[x,y,z]$ denotes the divided difference on $x, y,$ and z. The divided difference on n points is defined inductively as follows. $f[x_0]=f(x_0)$ and

$$f[x_0,x_1,\ldots,x_k]=\frac{f[x_0,x_1,\ldots,x_{k-1}]-f[x_1,x_2,\ldots,x_k]}{x_0-x_k}.$$

One can, likewise, show that any function satisfying condition (3) is a cubic. This leads one to wonder if the result of Aczél can be generalized. That this is true is the subject of this note.

THEOREM. *If f is a differentiable function satisfying*

$$f[x,y,z]=h(x+y+z)$$

then f is a cubic.

Proof. The condition of the theorem implies that

$$f(x)(y-z)+f(y)(z-x)+f(z)(x-y)=(x-y)(y-z)(x-z)h(x+y+z). \tag{4}$$

Now if we replace f by g where $g(x)=f(x)-f(0)$ it is easily verified that g satisfies condition (4). Thus we may assume that $f(0)=0$. Under this assumption we set $z=0$ in (4) and obtain

$$yf(x)-xf(y)=xy(x-y)h(x+y).$$

This of course yields

$$\frac{f(y)}{y}-\frac{f(x)}{x}=(y-x)h(x+y). \tag{5}$$

Now under the assumption that f is differentiable, h is continuous and thus if we allow y to approach 0 on each side of equation (5), we obtain

$$f'(0) - \frac{f(x)}{x} = -xh(x).$$

Therefore, if we define

$$q(x) = \frac{f(x)}{x} \quad \text{when } x \neq 0 \text{ and } q(0) = f'(0),$$

we have $f(x) = xq(x)$ for all x and

$$q(y) - q(x) = (y - x)h(x + y),$$

which is condition (2). Thus by Aczél's result we have $q(x) = ax^2 + bx + c$ so that $f(x) = ax^3 + bx^2 + cx$. Removing the assumption that $f(0) = 0$ we have $f(x) = ax^3 + bx^2 + cx + d$ and the proof is complete.

This rather modest generalization of Aczél's result of course suggests several other possibilities of generalization. One obvious possibility concerns making our result look more like that of Aczél by trying to remove the differentiability condition on f. It is not clear that the requirement can be completely removed; however, one of the referees has shown how we can obtain our result while requiring only that f be continuous. One is also led to wonder if

$$f[x_1, x_2, \ldots, x_n] = h(x_1 + x_2 + \cdots + x_n)$$

and f continuous (or perhaps differentiable) will imply that f is a polynomial of degree no more than n. At this point we have no answer. A closely related question however can be answered. In [2] it is shown that if f is a polynomial of degree not more than $n \geq 2$, then

$$f[x_1, x_2, \ldots, x_n] = \frac{1}{(n-1)!} f^{(n-1)}\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right),$$

and conversely.

REFERENCES

1. J. Aczél, A mean value property of the derivative of quadratic polynomials—without mean values and derivatives, this MAGAZINE, 58 (1985), 42–45.
2. D. F. Bailey and G. F. Fix, A generalization of the mean value theorem, *Appl. Math. Lett.* (1988), 327–330.
3. J. R. Boone and V. P. Schielack, private correspondence.
4. F. Charlton, A fixed feature of the mean value theorem, *The Mathematical Gazette*, 67 (1983), 49–50.
5. T. L. Saaty, *Modern Nonlinear Equations*, Dover Publications, Inc., New York, 1981, p. 122.

An Elementary Proof that Schoenberg's Space-Filling Curve Is Nowhere Differentiable

HANS SAGAN
North Carolina State University
Raleigh, NC 27695

Up to the end of the nineteenth century (and in six, randomly selected, contemporary calculus books) a plane curve was (and is) defined as the "graph of a pair of parametric equations

$$\left.\begin{aligned} x &= f(t) \\ y &= g(t) \end{aligned}\right\} t \in I \tag{1}$$

in which the functions f, g are continuous on the interval I", or by words to that effect. This "conventional wisdom" was shattered in 1890 when G. Peano [1] demonstrated that this definition embraces what are now called "space-filling curves," that is, curves that pass through every point of a two-dimensional region with positive content such as a square. Since one does not ordinarily call a square a curve, one has to place restrictions on f and g for (1) to produce what one might conventionally call a curve. For example, if one assumes that the mapping $(f, g): I \to E^2$ is continuous and injective, one obtains what is generally referred to as a *Jordan arc* that is, by a theorem of Netto ("a bi-jective map from a line onto a surface is, by necessity, discontinuous" [2]) not space filling.

Peano's pioneering work spawned numerous other examples of space-filling curves, one of which, namely Schoenberg's, will be the object of this discussion.

To produce Schoenberg's curve, let $I = [0, 1]$ and let f, g in (1) be defined by

$$f(t) = \frac{1}{2}\sum_{k=0}^{\infty} p(3^{2k}t)/2^k, \qquad g(t) = \frac{1}{2}\sum_{k=0}^{\infty} p(3^{2k+1}t)/2^k \tag{2}$$

where the 2-periodic even function p is defined as in FIGURE 1. (See also [3].)

It is not difficult to demonstrate that f, g are continuous on $[0, 1]$ and that (1), with f, g defined as in (2), represents a space-filling curve (see [3] or [4, p. 365] or [5, p. 438]).

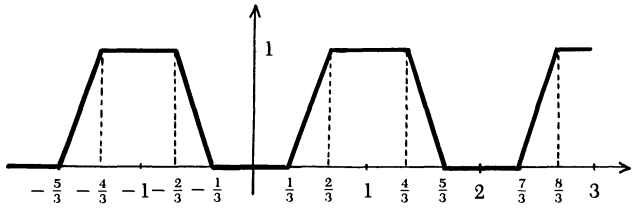


FIGURE 1

Some space-filling curves (but not all) are nowhere differentiable and Schoenberg's is one of them. This was not proved until 1981 when J. Alsina published a fairly complicated and laborious proof [6]. (By contrast, Lebesgue's space-filling curve is a.e. differentiable [4, p. 365].) Schoenberg himself proved in [7] that for $0 < a < 1$, b odd, and $ab > 4$, the function

$$\sum_{k=0}^{\infty} a^k E(b^k t)$$

is nowhere differentiable, where E is linear between consecutive integers and $E(n) = (-1)^n$ for all integers n. He did this by a clever adaptation of Rudin's proof of the nondifferentiability of Weierstrass' first example of such a function ([8, pp. 125–127]). He used this result, in turn, to establish the nowhere differentiability of his curve.

Our proof that Schoenberg's curve is nowhere differentiable is straightforward and does not utilize any advanced notions and techniques. It represents a suitable modification of a proof that the function

$$\sum_{k=1}^{\infty} s_k(t)$$

with $s_1, s_2, s_3, \ldots$ defined as in FIGURE 2, is nowhere differentiable ([9]) and is based on the following.

LEMMA. *If $f:[0,1] \to R$ is differentiable at $t \in (0,1)$, then, for any two sequences $\{a_n\} \to t$, $\{b_n\} \to t$ with $0 < a_n < t < b_n < 1$, by necessity,*

$$\lim_{n\to\infty} \frac{f(b_n) - f(a_n)}{b_n - a_n} = f'(t) \tag{3}$$

exists.

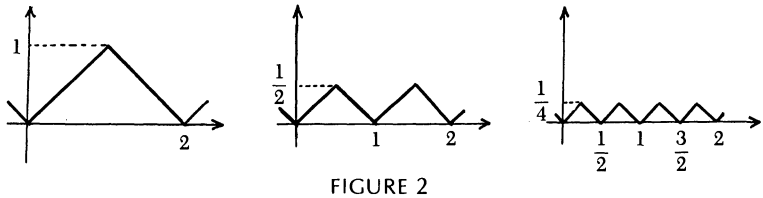


FIGURE 2

The proof of this lemma consists of the observation that

$$\frac{f(b_n) - f(a_n)}{b_n - a_n} - f'(t) = \frac{b_n - t}{b_n - a_n}\left(\frac{f(b_n) - f(t)}{b_n - t} - f'(t)\right) + \left(1 - \frac{b_n - t}{b_n - a_n}\right)\left(\frac{f(a_n) - f(t)}{a_n - t} - f'(t)\right)$$

and that

$$\left|\frac{b_n - t}{b_n - a_n}\right|, \left|1 - \frac{b_n - t}{b_n - a_n}\right|$$

are bounded.

We are now ready to establish our main result:

THEOREM. *The functions f, g as defined in (2) are nowhere differentiable.*

Proof. (i) First, let $t = 0$. Choose $b_n = 1/9^n$ and consider

$$f(0) = 0$$

$$f(1/9^n) = \frac{1}{2}\sum_{k=0}^{\infty} p(9^k/9^n)/2^k.$$

Since

$$p(9^k/9^n) = \begin{cases} 0 & \text{for } k < n \\ 1 & \text{for } k \geqslant n, \end{cases}$$

we have

$$f(1/9^n) = \frac{1}{2} \sum_{k=n}^{\infty} 1/2^k = 1/2^n,$$

and hence

$$\frac{f(b_n) - f(0)}{b_n} = \left(\frac{9}{2}\right)^n \to \infty$$

as $n \to \infty$, i.e., $f'(0)$ does not exist.

(ii) Next, let $t = 1$ and choose $a_n = 1 - 1/9^n$ to obtain

$$\frac{f(1) - f(a_n)}{1 - a_n} = 9^n - \left(\frac{9}{2}\right)^n \to \infty$$

as $n \to \infty$, i.e., $f'(1)$ does not exist.

Observe, that we did not need the lemma for these two cases but only the definition of the derivative as the limit of an average rate of change.

(iii) Finally, let $t \in (0, 1)$. To achieve our objective, we have to find for every such t two sequences $\{a_n\}, \{b_n\}$ that satisfy the requirements of the lemma so that the limit in (3) does not exist. We will do this for the function f.

Let

$$k_n = [9^n t], \tag{4}$$

where $[x]$ denotes the largest integer that is less than or equal to x, and let

$$a_n = k_n/9^n, \qquad b_n = k_n/9^n + 1/9^n.$$

It is a simple matter to show that, for sufficiently large n,

$$0 < a_n < t < b_n < 1$$

and that $\{a_n\} \to t$, $\{b_n\} \to t$.

From (4), infinitely many k_n are even or infinitely many k_n are odd, or both. We will assume for the sequel that infinitely many of them are *even* and denote the corresponding subsequence again by k_n in order to avoid double subscripts.

From (2) and (4),

$$f(a_n) = \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{2^k} p\left(\frac{9^k}{9^n} k_n\right), \qquad f(b_n) = \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{2^k} p\left(\frac{9^k}{9^n} k_n + \frac{9^k}{9^n}\right).$$

Hence,

$$f(b_n) - f(a_n) = \frac{1}{2} \sum_{k=0}^{n-1} \frac{1}{2^k} \left[p\left(\frac{9^k}{9^n} k_n + \frac{9^k}{9^n}\right) - p\left(\frac{9^k}{9^n} k_n\right) \right]$$

$$+ \frac{1}{2} \sum_{k=n}^{\infty} \frac{1}{2^k} \left[p\left(\frac{9^k}{9^n} k_n + \frac{9^k}{9^n}\right) - p\left(\frac{9^k}{9^n} k_n\right) \right] \triangleq S_1 + S_2.$$

If $k < n$, then $9^k/9^n \leq 1/9$ and, in order to obtain a lower estimate for S_1, we assume the worst possible situation where $(9^k/9^n)k_n + (9^k/9^n)$, as well as $(9^k/9^n)k_n$ both lie in an interval where p descends with slope -3 (see FIGURE 1). Then,

$$p\left(\frac{9^k}{9^n}k_n + \frac{9^k}{9^n}\right) - p\left(\frac{9^k}{9^n}k_n\right) \geq -3(9^k/9^n).$$

Hence,

$$S_1 \geq -\frac{1}{2}\sum_{k=0}^{n-1}\frac{1}{2^k}3\frac{9^k}{9^n} = -\frac{3}{9^n \cdot 2}\sum_{k=0}^{n-1}(9/2)^k = -\frac{3}{7 \cdot 9^n}\left[(9/2)^n - 1\right]. \quad (5)$$

If $k \geq n$, then $(9^k/9^n) \geq 1$, odd. Hence, $(9^k/9^n)k_n$ is even and $(9^k/9^n)k_n + (9^k/9^n)$ = even + odd = odd. Therefore,

$$S_2 = \frac{1}{2}\sum_{k=n}^{\infty}\frac{1}{2^k}\left[p(\text{odd}) - p(\text{even})\right] = \frac{1}{2}\sum_{k=n}^{\infty}\frac{1}{2^k} = 1/2^n. \quad (6)$$

From (5) and (6),

$$\frac{f(b_n) - f(a_n)}{b_n - a_n} = 9^n(S_1 + S_2) \geq \frac{4}{7}\left(\frac{9}{2}\right)^n + \frac{3}{7} \to \infty$$

as $n \to \infty$, i.e., $f'(t)$ does not exist.

We assumed for the preceding argument that infinitely many of the k_n are even. If infinitely many of them are odd, we reverse our strategy and seek an upper bound for the difference quotient. We obtain by analogous reasoning that

$$\frac{f(b_n) - f(a_n)}{b_n - a_n} \leq -\frac{4}{7}\left(\frac{9}{2}\right)^n - \frac{3}{7} \to -\infty.$$

Since $g(t) = f(3t)$ by (2), g is also nowhere differentiable.

That the Schoenberg curve is nowhere differentiable as opposed to the Lebesgue curve that is differentiable almost everywhere finds a very graphic expression in the behavior of their approximating polygons [**4**, FIGURES 2 and 8].

REFERENCES

1. G. Peano, Sur une courbe qui remplit toute une aire plane, *Math. Ann.* 36 (1890), 157–160.
2. E. Netto, Beitrag zur Mannigfaltigkeitslehre, *Crelle J.* 86 (1879), 263.
3. I. J. Schoenberg, The Peano-Curve of Lebesgue, *Bull. Amer. Math. Soc.* 44 (1938), 519.
4. H. Sagan, Approximating polygons for Lebesgue's and Schoenberg's space-filling curves, *Amer. Math. Monthly*, 93 (1986), 361–368.
5. H. Sagan, *Advanced Calculus*, Houghton-Mifflin, Boston, 1974.
6. J. Alsina, The Peano Curve of Schoenberg is nowhere differentiable, *J. Approx. Theory* 33 (1981), 28–42.
7. I. J. Schoenberg, *Mathematical Time Exposures*, MAA, Washington, DC, 1982, pp. 135–148.
8. W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, 1953.
9. J. F. Randolph, *Basic Real and Abstract Analysis*, Academic Press, New York, 1968, pp. 365–367.

PROBLEMS

LOREN C. LARSON, *editor*
St. Olaf College

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by September 1, 1992.

1393. *Proposed by Florin S. Pîrvănescu, Slatina, Romania.*

The sequence (a_n) of real numbers is defined inductively by

$$a_1 = a \quad \text{and} \quad a_{n+1} = a_n^2 - 2 \quad \text{for } n \geqslant 1.$$

Compute the product $a_1 a_2 \cdots a_n$.

1394. *Proposed by David Callan, University of Wisconsin, Madison, Wisconsin.*

Prove the identity

$$\sum_{k=0}^{n} \frac{1}{2k+1}\binom{2n+1}{n-k} = \sum_{k=0}^{n} \frac{4^k}{2k+1}\binom{2n-2k}{n-k}.$$

1395. *Proposed by Ioan Sadoveanu, Ellensburg, Washington.*

Let $A_1A_2 \cdots A_n$ be an n-gon circumscribing a circle, and let $B_1, B_2, \ldots, B_n$ denote the points of tangency of the sides. Let M be a point on the circumference of the incircle. Show that

$$\prod_{i=1}^{n} d\left(M, B_{\sigma(i)}B_{\sigma_{(i+1)}}\right) = \prod_{i=1}^{n} d(M, A_iA_{i+1})$$

for any permutation σ of $\{1, 2, \ldots, n\}$. (Here, $B_{\sigma(n+1)} = B_{\sigma(1)}$ and $A_{n+1} = A_1$.)

ASSISTANT EDITORS: CLIFTON CORZAT, BRUCE HANSON, RICHARD KLEBER, KAY SMITH, *and* THEODORE VESSEY, *St. Olaf College and* MARK KRUSEMEYER, *Carleton College. We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) next to a problem number indicates that neither the proposer nor the editors supplied a solution.*

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

Solutions and new proposals should be mailed in duplicate to Loren Larson, Department of Mathematics, St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057-1098 or mailed electronically via fax: (507) 663-3549 or e-mail: larson@stolaf.edu.

1396. *Proposed by Jiro Fukuta, Gifu-ken, Japan.*

Let ABC be an arbitrary triangle, let L_1 and L_2 be the trisection points of BC, arranged in order from B to C. Describe a method for dissecting triangle ABL_1 into four parts, each of which is a triangle or a quadrilateral, so that the parts can be reassembled to form a triangle congruent to triangle AL_2C.

1397. *Proposed by John O. Kiltinen, Northern Michigan University, Marquette, Michigan.*

It is well known that every permutation on a finite set can be expressed as a "product" of transpositions. For each permutation σ of $\{1, 2, \ldots, n\}$, let $F(\sigma)$ denote the minimal number of transpositions needed to represent σ as a product. Find the average value of F over the set of all permutations of $\{1, 2, \ldots, n\}$.

Quickies

Answers to the Quickies are on page 136.

Q788. *Proposed by Anthony Malebranche and Michael Handelsman, Erasmus Hall High School, Brooklyn, New York.*

A circle intersects the parabola $y = x^2$ in points (a, a^2) and (b, b^2) non-tangentially, and (t, t^2) tangentially. Find t.

Q789. *Proposed by Norman Schaumberger, Hofstra University, Hempstead, New York.*

Show that there exist two consecutive integer squares such that there are at least 1000 primes between them.

Q790. *Proposed by Robert S. Lubarsky, Franklin & Marshall College, Lancaster, Pennsylvania.*

Show that there is no real-valued differentiable function f in a neighborhood of 0 such that $f(0) = 0 = f'(0)$ and $\lim_{x \to 0} \dfrac{f(x)}{f'(x)} = 1$.

Solutions

Cup Pouring **April 1991**

1368. *Proposed by Allen J. Schwenk, Western Michigan University, Kalamazoo, Michigan.*

In my kitchen I have a set of three measuring cups with capacities c, $1 - c$, and 1 where $0 < c < 1/2$. Starting with the largest cup full, I can measure several additional fractions of a cup by sequentially pouring from one cup to another. I always pour until either the receiving cup is full or until the dispensing cup is empty. I never spill, waste, or consume any of the liquid. Define the accuracy of this set of cups to be the length ε of the longest interval $(a, a + \varepsilon)$ in $[0, 1]$ for which no $x \in (a, a + \varepsilon)$ can be measured. For example, when $c = 1/n$, I can measure $2/n$ by filling the smallest

cup, transferring this $1/n$ to the middle cup, refilling the smallest, and transferring again. Clearly each fraction i/n can be measured in this way, so for $c = 1/n$ the accuracy is also $1/n$.

For each value of c, find the accuracy of the set.

Solution by The WMC Problems Group, Western Maryland College, Westminster, Maryland.

If $c = m/n$ is rational and reduced, then the accuracy is $1/n$. If c is irrational, the accuracy will always be zero.

Call the cups with capacities c, $1 - c$, and 1 the small, medium, and large cups, respectively. Start by pouring $1 - c$ into the medium cup, leaving c in the large cup. We argue now that in two pourings we can always increase the contents of the large cup by c or decrease its contents by $1 - c$. In other words, we can consistently increase the contents by c (mod 1).

To see this, observe that there are two possibilities:

(i) We have enough in the medium cup to pour c into the small cup and then transfer this to the large cup, thus increasing the contents of the large cup by c; or

(ii) We can pour all the medium cup into the small one, then pour $1 - c$ from the large one into the medium one, thus decreasing the contents of the large cup by $1 - c$. If $c = m/n$ is a reduced fraction, then m is a generator for the cyclic group $\mathbf{Z}_n$ and so we obtain all fractions k/n by repeatedly adding c (mod 1).

If c is irrational, then the numbers kc (mod 1), $k = 1, 2, \ldots$ are dense in the unit interval, and consequently no interval can be free of measurable quantities.

Also solved by Seung-Jin Bang (Korea), David Callan, Douglas E. Jackson, Jiro Fukuta (Japan), Stephen Noltie, Roger Williams College Number Theory Class, and the proposer.

Diophantine Analysis **April 1991**

1369. *Proposed by Mihály Bencze, Braşov, Romania.*

a. Find all natural numbers x, y, z such that $3^x + 4^y = 5^z$.

b. * Given natural numbers $A > B$ and C, find all natural numbers x, y, z such that

$$\left((A^2 - B^2)C\right)^x + (2ABC)^y = \left((A^2 + B^2)C\right)^z.$$

Solution by Harvey Schmidt, Jr., Lewis and Clark College, Portland, Oregon.

a. The only (natural number) solution is the well-known $x = y = z = 2$.

Assume x, y, and z are natural numbers. We first show that x, y, and z are all even. If $3^x + 4^y = 5^z$, then $1 \equiv (-1)^z$ (mod 3) and $(-1)^x \equiv 1$ (mod 4), so both x and z are even, say $x = 2m$ and $z = 2n$. Thus $4^y = 5^z - 3^x = 5^{2n} - 3^{2m} = (5^n - 3^m)(5^n + 3^m)$, so $5^n + 3^m = 2^t$ for some non-negative integer t. Since $m, n > 0$ it follows that $t \geq 3$. Now $0 \equiv 2^t = 5^n + 3^m \equiv 1 + (-1)^m$ (mod 4), so m is odd. Finally, $0 \equiv 5^{2n} = 3^{2m} + 4^y \equiv (-1)^m + (-1)^y$ (mod 5), so m and y have opposite parity. Consequently, y is even, say $y = 2t$.

Since $3^{2m} + 4^{2t} = 5^{2n}$ if and only if $(3^m)^2 + (4^t)^2 = (5^n)^2$, it is well known from the theorem on Pythagorean triples that there exist relatively prime, positive integers u and v of opposite parity such that

$$3^m = u^2 - v^2, \qquad 4^t = 2uv, \quad \text{and} \quad 5^n = u^2 + v^2.$$

Because $u > v$ it is clear that $v = 1$ and $u = 2^{2t-1}$. Thus $3^m = u^2 - v^2 = (2^{2t-1})2 - 1 = (2^{2t-1} - 1)(2^{2t-1} + 1)$, and therefore $2^{2t-1} - 1 = 3^a$ and $2^{2t-1} + 1 = 3^b$ for some nonnegative integers a and b. Since $2^{2t-1} - 1$ and $2^{2t-1} + 1$ are relatively prime, $2^{2t-1} - 1 = 1$ and $2^{2t-1} + 1 = 3^m$. But now $t = 1$, $u = 2$, $m = 1$, and $n = 1$. Hence, $x = y = z = 2$.

Part a was also solved by Jiro Fukuta (Japan), Peter W. Lindstrom, John S. Sumner, and the proposer. There was one incorrect solution.

Schmidt goes on to prove other instances when $x = y = z = 2$ is the only solution, namely $(A, B, C) = (3, 2, 1), (4, 3, 1), (5, 3, 1)$, or $(7, 4, 3)$. *M. J. DeLeon*, Florida Atlantic University, cites Sierpinski, *Elementary Theory of Numbers* (1988), p. 40, where the problem is stated and references are given for the proofs of (a) and special cases of (b). In particular, there are infinitely many triples for which the only solution is $x = y = z = 2$.

Sums of Constant Powers: A Comparison **April 1991**

1370. *Proposed by R. S. Luthar, University of Wisconsin Center, Janesville, Wisconsin.*

a. If x, y, and z are positive real numbers with $x + y + z = 1$, prove that

$$\frac{2}{3} \le \frac{\ln(x^5 + y^5 + z^5)}{\ln(x^7 + y^7 + z^7)} \le \frac{5}{7}.$$

b. Generalize the result of part a.

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

We will prove the following: If $\alpha, \beta, x_1, \ldots, x_n$ are positive real numbers such that

$$\alpha < \beta \quad \text{and} \quad \sum_{i=1}^{n} x_i \le 1 \quad (\text{with } n > 1),$$

then

$$\frac{\alpha}{\beta} \le \frac{\ln(\Sigma x_i^{\alpha+1})}{\ln(\Sigma x_i^{\beta+1})} < \frac{\alpha + 1}{\beta + 1}. \tag{1}$$

Proof. With $n > 1$, each $x_i < 1$, so that

$$\Sigma x_i^{\beta+1} < \Sigma x_i \le 1. \tag{2}$$

We have

$$\Sigma x_i^{\alpha+1} = \Sigma\left(x_i^{\beta+1}\right)^{\alpha/\beta} x_i^{1-\alpha/\beta} \le \left(\Sigma x_i^{\beta+1}\right)^{\alpha/\beta} \left(\Sigma x_i\right)^{1-\alpha/\beta} \le \left(\Sigma x_i^{\beta+1}\right)^{\alpha/\beta},$$

where the first inequality follows from Hölder's inequality. Hence,

$$\ln(\Sigma x_i^{\alpha+1}) \le \frac{\alpha}{\beta} \ln(\Sigma x_i^{\beta+1}).$$

By (2) we see that the inequality on the left side of (1) holds.

By Jensen's inequality,

$$\left(\Sigma x_i^{\beta+1}\right)^{1/(\beta+1)} \le \left(\Sigma x_i^{\alpha+1}\right)^{1/(\alpha+1)},$$

with equality if and only if at most one x_i is nonzero. Taking logs and using (2) again, we establish

$$\frac{\ln(\Sigma x_i^{\alpha+1})}{\ln(\Sigma x_i^{\beta+1})} < \frac{\alpha+1}{\beta+1}.$$

Letting x_1 approach 1 shows the inequality is sharp.

Also solved by M. Reza Akhlaghi, Seung-Jin Bang (Korea), Jiro Fukuta (Japan), H. Guggenheimer, M. S. Klamkin (Canada), Peter W. Lindstrom, Heinz-Jürgen Seiffert (Germany), Dick A. Wood, Michael Vowe (Switzerland), and the proposer.

A Triangle Invariant **April 1991**

1371. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

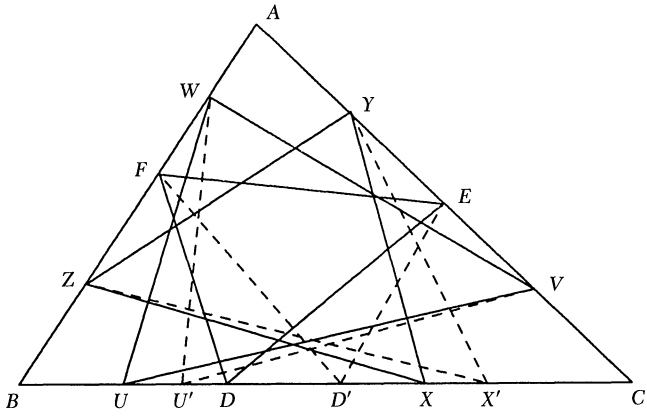
Let A, B, and C be vertices of a triangle and let D, E, and F be points on the sides of BC, AC, and AB, respectively. Let U, X, V, Y, W, Z be the midpoints of, respectively, BD, DC, CE, EA, AF, FB. Prove that

$$\text{Area}(\triangle UVW) + \text{Area}(\triangle XYZ) - \tfrac{1}{2}\,\text{Area}(\triangle DEF)$$

is a constant independent of D, E, and F.

I. *Solution by Jordi Dou, Barcelona, Spain; submitted on the occasion of his 80th birthday.*

First, let D' be any point between C and D and take U', X' to be the midpoints of BD', $D'C$. Then $XX' = UU' = \frac{1}{2}DD'$.



Let $h_F, h_A, h_W, \ldots$ denote the distances from $F, A, W, \ldots$ to BC respectively. Clearly $h_Z = \frac{1}{2}h_F$, and $h_W = \frac{1}{2}(h_F + h_A)$, and therefore by addition, $h_Z + h_W - h_F = \frac{1}{2}h_A = h_Y + h_V - h_E$. We let $[PQR]$ denote the area of triangle PQR, and set $S = [ABC]$, $\sigma = [DEF]$, $\sigma_1 = [XYZ]$, $\sigma_2 = [UVW]$, $\sigma' = [D'EF]$, $\sigma_1' = [X'YZ]$, and $\sigma_2' = [U'VW]$.

Using these identities, we find that

$$\sigma' - \sigma = \tfrac{1}{2}DD'(h_E - h_F),$$

$$\sigma_1' - \sigma_1 = \tfrac{1}{2}XX'(h_Y - h_Z) = \tfrac{1}{4}DD'(h_Y - h_Z)$$

and

$$\sigma_2' - \sigma_2 = \tfrac{1}{4}DD'(h_V - h_W).$$

It follows that

$$\begin{aligned}(\sigma_1' + \sigma_2' - \tfrac{1}{2}\sigma') - (\sigma_1 + \sigma_2 - \tfrac{1}{2}\sigma) &= \tfrac{1}{4}DD'(h_Y - h_Z + h_V - h_W - h_E + h_F) \\ &= \tfrac{1}{4}DD'((h_Y + h_V - h_E) - (h_Z + h_W - h_F)) \\ &= \tfrac{1}{4}DD'(\tfrac{1}{2}h_A - \tfrac{1}{2}h_A) \\ &= 0.\end{aligned}$$

By symmetry, it is clear that the preceding is also 0 when D' is between B and D.

In exactly the same way, taking F' on AB instead of F and triangle $D'EF'$ for $D'EF$, and after this, taking E' on AC instead of E and triangle $D'E'F'$ for $D'EF'$, we find that $\sigma_1 + \sigma_2 - \frac{1}{2}\sigma$ is invariant with respect to DEF.

We obtain the value of $\sigma_1 + \sigma_2 - \frac{1}{2}\sigma$ by putting $E = A$, $F = B$, $D = C$. Then $X = C$, $Y = A$, $Z = B$, U is the midpoint of BC, V is the midpoint of CA, W is the midpoint of AB. Also, $\sigma = S$, $\sigma_1 = S$, $\sigma_2 = (1/4)S$, and therefore, $\sigma_1 + \sigma_2 - \frac{1}{2}\sigma = \frac{3}{4}S$.

II. *Solution by László Szücs, Fort Lewis College, Durango, Colorado.*

We shall use the notation $[ABC] = \text{Area}(\triangle ABC)$. The given expression can be written as

$$\begin{aligned}&([ABC] - ([AWV] + [BUW] + [CVU])) \\ &\quad + ([ABC] - ([AZY] + [BXZ] + [CYX])) \\ &\quad - (1/2)([ABC] - ([AFE] + [BDF] + [CED])).\end{aligned}$$

Using the relations $[AFE] = 4[AWY]$, $[BDF] = 4[BUZ]$, and $[CED] = 4[CVX]$, the expression becomes

$$\begin{aligned}\tfrac{3}{2}[ABC] - ([AWV] - [AWY] + [AZY] - [AWY] \\ + [BUW] - [BUZ] + [BXZ] - [BUZ] \\ + [CVU] - [CVX] + [CYX] - [CVX]).\end{aligned}$$

We now use the relation $[AWV] - [AWY] = [VYW] = \frac{1}{4}[CAF]$ and its five analogues to obtain

$$\tfrac{3}{2}[ABC] - \tfrac{1}{4}([CAF] + [EAB] + [ABD] + [FBC] + [BCE] + [DCA]),$$

which is easily seen to equal

$$\tfrac{3}{2}[ABC] - \tfrac{3}{4}[ABC] = \tfrac{3}{4}[ABC].$$

Also solved by Larry E. Askins, Eynshteyn Averbukh, Seung-Jin Bang (Korea), Karen Benbury, Francisco Bellot Rosado (Spain), Scott D. Cohen (student), C. Patrick Collier, Miquel Amengual Covas (Spain), Jordi Dou (Spain), Ragnar Dybvik (Norway), Kao H. and Irene C. Sze, Milton P. Eisner, Jiro Fukuta (Japan), Thomas E. Gantner, John F. Goehl, Jr, Cornelius Groenewoud, H. Guggenheimer, Francis M. Henderson, Ralph P. Grimaldi, Russell Jay Hendel, Paul Irwin, Geoffrey A. Kandall, Václav Konečný, Philip Lau, Eugene Lee, Peter W. Lindstrom, James Pfaendtner, Richard E. Pfiefer, Rolf Rosenkranz (Germany), Ioan Sadoveanu, Jyotirmoy Sarkar, Volkhard Schindler (Germany), Mohammad Parvez Shaikh (student), Ching-Kuang Shene, John S. Sumner, Jordan Tabov (Bulgaria), Michael Vowe, and the proposer.

Tabov proved the more general result. Consider a triangle $A_1A_2A_3$, a real number α different from 0 and 1, and real numbers λ and μ. For arbitrary points X_1, X_2, and X_3 respectively on the lines A_2A_3, A_3A_1, and A_1A_2, define points C_{ij}, $(i, j = 1, 2, 3;\ i \neq j)$ by $\overrightarrow{OC_{ij}} = \alpha\overrightarrow{OA_i} + \beta\overrightarrow{OX_j}$, where O is any point outside the plane of the triangle $A_1A_2A_3$ and $\beta = 1 - \alpha$. Let $F(X_1, X_2, X_3) = \lambda[C_{23}C_{31}C_{12}] + \mu[C_{13}C_{21}C_{32}] - [X_1X_2X_3]$, where the square brackets denote signed area, and X_1, X_2, and X_3 describe independently respectively the lines A_2A_3, A_3A_1 and A_1A_2. Then the function $F(X_1, X_2, X_3)$ is constant if and only if $\lambda = \mu = \frac{1}{2}(1 - \alpha)^{-2}$. (The given problem corresponds to the case $\alpha = 1/2$.)

A Rational Trigonometric Sum **April 1991**

1372. *Proposed by Nick Lord, Tonbridge School, Kent, England.*

For which angles θ, a rational number of degrees, is it the case that $\tan^2\theta + \tan^2 2\theta$ is rational?

Solution by Diane and Roy Dowling, University of Manitoba, Winnipeg, Manitoba.

Suppose $2\theta = 360k/n$ degrees where k and n are relatively prime positive integers and $\tan^2\theta + \tan^2 2\theta$ is rational. Then

$$\frac{1}{4\cos^2\theta} + \frac{1}{4\cos^2 2\theta} = r$$

for some rational r. Letting $x = e^{2i\theta}$ this equation may be written

$$\frac{1}{x + \frac{1}{x} + 2} + \frac{1}{x^2 + \frac{1}{x^2} + 2} = r.$$

Each solution of this equation must be a zero of the polynomial

$$rx^6 + (2r-1)x^5 + (3r-1)x^4 + (4r-4)x^3 + (3r-1)x^2 + (2r-1)x + r. \quad (1)$$

However, $e^{2i\theta}$ (that is, $e^{360^\circ ki/n}$) is a primitive root of unity. It is therefore a zero of some cyclotomic polynomial. This cyclotomic polynomial must be a divisor of (1), since it is irreducible over the rationals and shares a zero with (1). Its degree is therefore 6 or less. The cyclotomic polynomials of degree 6 or less are

$$x-1,\ x+1,\ x^2+x+1,\ x^2+1,\ x^4+x^3+x^2+x+1,$$
$$x^2-x+1,\ x^6+x^5+x^4+x^3+x^2+x+1,\ x^4+1,\ x^6+x^3+1,$$
$$x^4-x^3+x^2-x+1,\ x^4-x^2+1,$$
$$x^6-x^5+x^4-x^3+x^2-x+1,\ x^6-x^3+1.$$

Upon dividing (1) by each of these cyclotomic polynomials it is found that with four exceptions there is no value of r for which the remainder is zero. The four exceptions are:

The remainder on division by $x-1$ is zero when r is $1/2$.

The remainder on division by x^2+x+1 is zero when r is 2.

The remainder on division by $x^4+x^3+x^2+x+1$ is zero when r is 3.

The remainder on division by x^2-x+1 is zero when r is $4/3$.

The zero of $x-1$ is 1. The zeros of x^2+x+1 are the primitive cube roots of unity. The zeros of $x^4+x^3+x^2+x+1$ are the primitive fifth roots of unity. The zeros of x^2-x+1 are the primitive sixth roots of unity.

It follows that the only possible values of 2θ in the interval $[0, 360^\circ)$ are

$$0^\circ,\ 120^\circ,\ 240^\circ,\ 72^\circ,\ 144^\circ,\ 216^\circ,\ 288^\circ,\ 60^\circ,\ \text{and}\ 300^\circ.$$

Thus the only possible values of θ are

$$0^\circ,\ 30^\circ,\ 36^\circ,\ 60^\circ,\ 72^\circ,\ 108^\circ,\ 120^\circ,\ 144^\circ,\ 150^\circ$$

and those angles that differ by a multiple of 180° from one of the preceding.

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

Answers

Solutions to the Quickies on page 130.

A788. With the usual designations for the center and radius, the equation that describes the intersection of the conics is

$$(x-h)^2+(x^2-k)^2=r^2.$$

Expanding gives a fourth degree polynomial of which the cubic term has coefficient zero. It follows that the four roots, a, b, t, t have the sum of zero, whence we obtain $t=-(a+b)/2$.

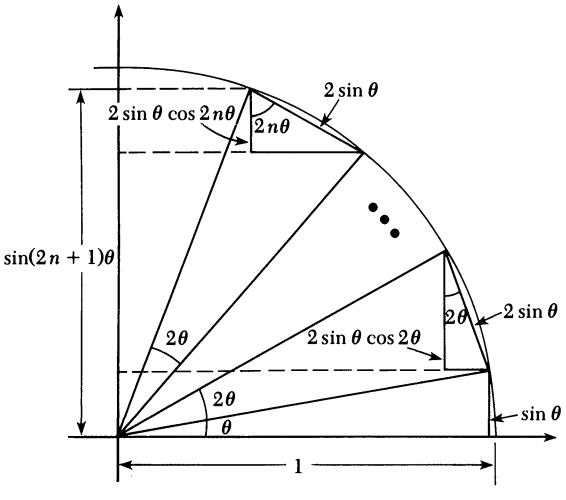
A789. Assume that for $n=1,2,\ldots,$ the interval $[n^2,(n+1)^2]$ contains fewer than 1000 primes. It follows that the sum of the reciprocals of the primes in this interval is less than $1000/n^2$. Thus, the sum of the reciprocals of all the primes is less than $1000\sum_{n=1}^{\infty}1/n^2$. Since this latter sum converges and the sum of the reciprocals of the primes diverges, we have a contradiction.

A790. If the zeros of $f(x)$ had an accumulation point at 0, so would those of $f'(x)$, a contradiction. If not, L'Hôpital's Rule implies

$$0=\lim_{x\to 0}\frac{xf(x)}{f(x)}=\lim_{x\to 0}\frac{f(x)+xf'(x)}{f'(x)}=1.$$

Proof without Words

$$\sin(2n+1)\theta=\sin\theta+2\sin\theta\sum_{k=1}^{n}\cos 2k\theta$$



—J. Chris Fisher and E. L. Koh
University of Regina
Regina, Saskatchewan, Canada S4S 0A2

Answers

Solutions to the Quickies on page 130.

A788. With the usual designations for the center and radius, the equation that describes the intersection of the conics is

$$(x-h)^2+(x^2-k)^2=r^2.$$

Expanding gives a fourth degree polynomial of which the cubic term has coefficient zero. It follows that the four roots, a, b, t, t have the sum of zero, whence we obtain $t=-(a+b)/2$.

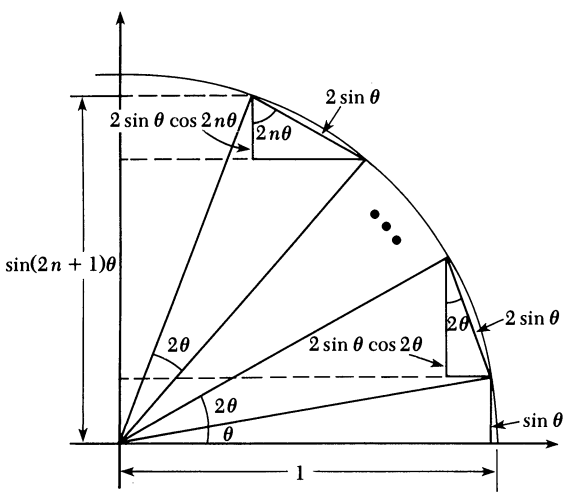
A789. Assume that for $n=1,2,\ldots,$ the interval $[n^2,(n+1)^2]$ contains fewer than 1000 primes. It follows that the sum of the reciprocals of the primes in this interval is less than $1000/n^2$. Thus, the sum of the reciprocals of all the primes is less than $1000\sum_{n=1}^{\infty}1/n^2$. Since this latter sum converges and the sum of the reciprocals of the primes diverges, we have a contradiction.

A790. If the zeros of $f(x)$ had an accumulation point at 0, so would those of $f'(x)$, a contradiction. If not, L'Hôpital's Rule implies

$$0=\lim_{x\to 0}\frac{xf(x)}{f(x)}=\lim_{x\to 0}\frac{f(x)+xf'(x)}{f'(x)}=1.$$

Proof without Words

$$\sin(2n+1)\theta=\sin\theta+2\sin\theta\sum_{k=1}^{n}\cos 2k\theta$$



—J. Chris Fisher and E. L. Koh
University of Regina
Regina, Saskatchewan, Canada S4S 0A2

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 the mathematics literature. Readers are invited to suggest items for review to the editors.

Max, Nelson, Another harmony of the spheres, *Nature* 355 (9 January 1992) 115–116.
Stewart, Ian, Mathematical recreations: The kissing number, *Scientific American* (February 1992) 112–115.

In 1990, Wu-Yi Hsiang (University of California—Berkeley) announced a proof that face-centered cubic packing is the densest packing of spheres in three dimensions, thereby settling a claim of Kepler's. Or is it settled? Hsiang's original announcement was rejected by the *Bulletin* of the AMS because the details of the proof had not yet been written down. Hsiang's argument depends on the classification and analysis of a large number of configurations, many of which he described originally only in qualitative terms. Now there is a draft of all the details. But will anyone "have the patience to repeat Hsiang's year-long verification that all possibilities are covered?" Hsiang's cases are not easily amenable to verification by computer, despite the fact that he used "only high-school algebra, vector identities, a little calculus and a lot of spherical geometry and spherical trigonometry," plus his TI-35+ calculator. Finally, Hsiang has recently announced that he has solved the 'kissing" problem in four dimensions, determining that the greatest number of spheres that can surround and touch another is 24.

vos Savant, Marilyn, Ask Marilyn, *Parade* (5 January 1992) 22–23; (26 January 1992) 10.

Marilyn continues baiting arrogant Ph.D.'s with probability paradoxes, reopening another old chestnut and writing a letter to herself (no reader wrote in) to correct the reasoning behind her solution of a previous problem. The new consternation is over what Martin Gardner calls the "paradox of the second child": "Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?" *The Second Scientific American Book of Mathematical Puzzles and Diversions*, Simon and Schuster, 1961, p. 152). With vos Savant, it's beagles instead of children. Gardner notes that such problems are "ill-defined unless stated with great precision," particularly regarding the randomizing procedure and the way in which the information "at least one is a boy" is obtained (p. 226). The letters vos Savant quotes from Ph.D.'s—none listed in the mathematical societies' *Combined Membership List*, thank goodness!—don't echo Gardner, they just make fools of their authors by slamming vos Savant.

Stewart, Ian, Numerical methods: Warning—handle with care!, *Nature* 355 (2 January 1992) 16–17.

Are you using a tool whose workings you don't understand? H.C. Lee et al. take continuous models (e.g., differential equations) with known explicit solutions, discretize them, and analyze the behavior of the resulting discrete dynamics. The shocking fact is that well-known methods (e.g., Euler, Runge-Kutta) can introduce spurious periodic points and steady states, even for an equation as simple as the logistic equation. "The main problems are not so much numerical as dynamical: the actual behaviour of the continuum models, and their relation to discretizations, must be [investigated]."

Hayman, d'Arcy, *The Calculus Virgin: An Artist's View of the Language of Calculus*, with introduction and notes by Louis Leithold, Tortue Publications (2917 Santa Monica Blvd., Santa Monica, CA 90404); vii + 118 pp, $14.85 (P).

What could an artist with no mathematical background get out of a seminar on the teaching of calculus? This book stems from just such a "virginal encounter." The result is an artist's whimsical reflection, in 53 drawings and accompanying annotations, on the *language* of calculus. The book presents a fresh insight for us mathematicians. We may have become so accustomed to the narrow and precise senses in which we use mathematical terminology that we miss non-mathematical chords that our words may strike in the minds of mathematical neophytes. Think again, then, of what other meanings our students could attach to "triangle inequality," "chain rule," "root test," or "closed ball."

Steen, Lynn Arthur (ed.), *Library Recommendations for Undergraduate Mathematics*, MAA, 1992; xi + 194 pp, $ 15 (P). ISBN 0–88385–076–1. Two-Year College Mathematics Library Recommendations MAA, 1992; xi + 76 pp, $10 (P). ISBN 0–88385–077–X.

It has been 15 years since the last versions of the *Basic Library Lists* for four-year colleges and for two-year colleges. Meanwhile, your library may have missed acquiring some of the Essential or Highly Recommended books in these newly-revised lists. It's worth having your library or a student worker in your department check your holdings against the appropriate list, which also includes books noted as Recommended or just Listed. The numbers of titles in the categories are in the rough proportions of 1:2:4:8. The four-year list includes 3,000 volumes, the two-year list 1,200, from 15,000 initial nominations. The books are arranged under 25 major areas, broken into subareas. Each volume contains an author index, plus a short list of appropriate journals and periodicals.

Silverman, Robert D., Massively distributed computing and factoring large integers, *Communications of the Association for Computing Machinery* 34 (11) (November 1991) 95–103.

This is a succinct summary, without too many deep details, of progress in factoring. Methods for numbers of a special form, specially-designed computer processors (e.g., with word length of 256 bits), and general-purpose methods are all described briefly, including elliptic curve method, the continued fraction algorithm, the quadratic sieve, and the new number field sieve. The recent "distributed" factoring of a 116-digit number, by computers across the world working on assigned subproblems, took 275 MIP-years (1 MIP year = the work done by a machine executing one million instructions per second for a year). A 129-digit number would take 1,000 to 2,000 times as much computation; a 200-digit number, 10^9 times as much. The author concludes: "No foreseeable increase in the speed or the number of computers is going to overcome a factor of 10^9," so that 120 digits is more or less the limit of practical factoring of numbers that are not of a special form.

Gardner, Martin, *Fractal Music, Hypercards and More ...: Mathematical Recreations from Scientific American Magazine*, Freeman, 1992; ix + 327 pp (P). ISBN 0–7167–2189–9

Martin Gardner will be 78 years old this year; this book reprints his Mathematical Games columns in *Scientific American* from the years 1978 and 1979. "It is the fourteenth such collection, and I have one more to go before running out of columns." The topics are timeless (despite addenda based on later information): Egyptian fractions, Bell numbers, a mathematical zoo, minimal sculpture, pi and poetry, Chaitin's Omega, "and more."

Stewart, Ian, Trees telephones and tiles, *New Scientist* (16 November 1991) 26–29.

Well-written and well-illustrated elementary exposition of the 1990 Du-Hwang result on just how much shorter a shortest Steiner tree can be than the shortest spanning tree.

Gunn, Charlie, and Delle Maxwell (directors), *Not Knot*, 20-min. VHS videotape; and Epstein, David, and Charlie Gunn, *Supplement to Not Knot*, 48 pp, (P). Jones and Bartlett, 1991. ISBN 0–867620–240–8 (video + book), 0–86720–297–1 (book only)

This "guided tour into computer-animated hyperbolic space" features marvelous computer graphics. The film and supplement deftly exposit and illustrate that the complements of different knots are different (C. Gordon and J. Luecke, 1988) and the failure of the corresponding statement for links. With some exceptions, every complement of a knot or link admits hyperbolic structure (Thurston, 1970's); the film illustrates this structure for the Borromean rings link (so perhaps the title really should be *Not Link*). The supplement contains the complete script of the film, plus questions and answers, sidebars, student activities, and bibliography. This splendid video and its accompanying text were produced at the Geometry Center at the University of Minnesota, supported in part by the National Science Foundation.

Albers, Donald J., Gerald L. Alexanderson, and Constance Reid, *More Mathematical People: Contemporary Conversations*, Harcourt Brace Jovanovich, 1990; xviii + 375 pp, $29.95.

Delightful, sensitive, and informative interviews, with photographs, of 18 contemporary mathematicians: Bers, Boas, Cohen, Dantzig, Gleason, Gosper, Kaplansky, Lax, Le Cam, Lewy, Mac Lane, Morawetz, Mosteller, J. Robinson, M.E. Rudin, Smale, Thurston, and R. Wilson. Every mathematics departmental common room should have this book (and its predecessor, *Mathematical People*). Concludes Constance Reid: "[I]n spite of the habit of mind that all mathematicians share and the common joy that they take in their subject, there is no such person as the typical mathematician."

Soifer, Alexander, *How Does One Cut a Triangle?*, Center for Excellence in Mathematical Education (885 Red Mesa Dr., Colorado Springs, CO 80906); xiii + 139 pp, $18.95 (P). ISBN 0–940263–01–7

Formulates and explores ways to solve new and engaging problems about cutting a triangle into congruent or similar triangles, gradually refining the problems and drawing the reader into the solutions: "every solved problem gives birth to a myriad of unsolved ones." The author's goal is to show students what it is that mathematicians do; except for a brief excursion into eigenvalues, only high-school mathematics is used. Several solutions are by high-school students, and one chapter is devoted to solved and unsolved problems co-created by the author with P. Erdös.

Austin, Joe Dan, *Applications of Secondary School Mathematics: Readings from the Mathematics Teacher*, NCTM, 1991; vi + 339 pp, $21.50 (P). ISBN 0–87353–336–4

Early NCTM sourcebooks of applications appeared in 1979 and 1980; this one reprints articles from the *Mathematics Teacher* over the past 15 years. Articles are grouped under the chapters Using Applications in Teaching, Arithmetic, Geometry, Algebra, Trigonometry and Elementary Analysis, Calculus, and Probability and Statistics.

Goldenberg, Stuart, and Harvey Greenwald, *Calculus Applications in Engineering and Science*, D.C. Heath, 1990; viii + 228 pp, (P). ISBN 0–669–21676–3

Although keyed to *Calculus*, 4^{th} ed., by Larson et al., this supplement can be used with any calculus text. The applications range from the fairly standard to water rationing, relativity, Egyptian water clocks, and boomerangs. Sometimes the application is just a come-on for a routine oldie: ocean oil spills (emptying a vertical barrel), global warming (melting a spherical snowball). Each application features exercises, with answers.

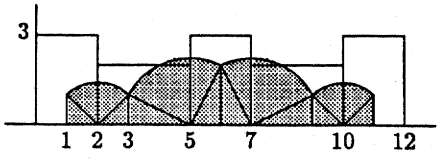
NEWS AND LETTERS

52nd ANNUAL WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

These solutions have been compiled and prepared by Loren Larson, St. Olaf College.

A-1. A 2×3 rectangle has vertices at $(0,0), (2,0), (0,3)$, and $(2,3)$. It rotates 90° clockwise about the point $(2,0)$. It then rotates 90° clockwise about the point $(5,0)$, then 90° clockwise about the point $(7,0)$, and finally, 90° clockwise about the point $(10,0)$. (The side originally on the x-axis is now back on the x-axis.) Find the area of the region above the x-axis and below the curve traced out by the point whose initial position is $(1,1)$.

Solution. As indicated in the following diagram, the area consists of four 1×1 right triangles of area $1/2$, four 1×2 right triangles of area 1, two quarter-circles of area $(\pi/4)(\sqrt{2})^2 = \pi/2$, and two quarter-circles of area $(\pi/4)(\sqrt{5})^2 = 5\pi/4$. Hence the total area is $7\pi/2 + 6$.



A-2. Let $\mathbf{A}$ and $\mathbf{B}$ be different $n \times n$ matrices with real entries. If $\mathbf{A}^3 = \mathbf{B}^3$ and $\mathbf{A}^2\mathbf{B} = \mathbf{B}^2\mathbf{A}$, can $\mathbf{A}^2 + \mathbf{B}^2$ be invertible?

Solution. No. If so, then $\mathbf{A} - \mathbf{B} = (\mathbf{A}^2+\mathbf{B}^2)^{-1}(\mathbf{A}^2+\mathbf{B}^2)(\mathbf{A}-\mathbf{B}) = (\mathbf{A}^2+\mathbf{B}^2)^{-1}(\mathbf{A}^3+\mathbf{B}^2\mathbf{A}-\mathbf{A}^2\mathbf{B}-\mathbf{B}^3) = (\mathbf{A}^2+\mathbf{B}^2)^{-1}\mathbf{0} = \mathbf{0}$, so $\mathbf{A} = \mathbf{B}$, a contradiction.

A-3. Find all real polynomials $p(x)$ of degree $n \geq 2$ for which there exist real numbers $r_1 < r_2 < \ldots < r_n$ such that
(*i*) $p(r_i) = 0, \quad i = 1, 2, \ldots, n,$
and
(*ii*) $p'\left(\frac{r_i + r_{i+1}}{2}\right) = 0, \; i = 1, 2, \ldots, n-1,$
where $p'(x)$ denotes the derivative of $p(x)$.

Solution. The set of polynomials is

$$\{ax^2 + bx + c : a \neq 0,\ b^2 - 4ac > 0\}.$$

First, if $p(x)$ is such a polynomial, it must have two distinct real roots, say r_1, r_2, with $r_1 < r_2$. It is easy to check that such polynomials meet the condition. To show nothing else does, write

$$p(x) = a(x - r_1)(x - r_2)\cdots(x - r_n)$$

where $r_1 < r_2 < \cdots < r_n$ and $n \geq 3$. Then $p'(x) = a(2x - (r_1 + r_2))\,q(x) + a(x - r_1)(x - r_2)q'(x)$, where $q(x) = (x - r_3)\cdots(x - r_n)$. By Rolle's Theorem, all the zeros of $q'(x)$ lie between r_3 and r_n. Hence $(r_1 + r_2)/2$ is not a zero of $q'(x)$, showing that $p(x)$ does not meet the condition.

A-4. Does there exist an infinite sequence of closed discs $D_1, D_2, D_3, \ldots$ in the plane, with centers $c_1, c_2, c_3, \ldots$, respectively, such that
(i) the c_i have no limit point in the finite plane,
(ii) the sum of the areas of the D_i is finite, and
(iii) every line in the plane intersects at least one of the D_i?

Solution. Yes. Let (a_i) be a decreasing sequence of positive numbers such that $\sum a_i = \infty$, and $\sum a_i^2 < \infty$ (for example, $a_i = 1/i$). Let X_i^+ be the closed disc of radius a_i centered at $\left(\sum_{k=1}^{i} a_k, 0\right)$, for all integers $i > 0$. Clearly the collection of discs X_i^+ covers the positive x-axis including the origin. Similarly define Y_i^+ along the positive y-axis, X_i^- along the negative x-axis, and Y_i^- along the negative y-axis. Then set $D_{4i} = X_i^+$, $D_{4i-1} = Y_i^+$, $D_{4i-2} = X_i^-$, and $D_{4i-3} = Y_i^-$ for all integers $i > 0$. The D_i cover both the x-axis and the y-axis, so they intersect any given line in the plane. Their total area is finite, and the centers have no limit point since any given disc contains only finitely many of the centers.

A-5. Find the maximum value of

$$\int_0^y \sqrt{x^4 + (y - y^2)^2}\, dx$$

for $0 \le y \le 1$.

Solution. The maximum value is $1/3$, attained when $y = 1$.

When $y = 1$, we have $\int_0^1 \sqrt{x^4}\, dx = 1/3$. On the other hand, $\int_0^y \sqrt{x^4 + (y - y^2)^2}\, dx \le \int_0^y \left(x^2 + (y - y^2)\right) dx = y^2 - 2y^3/3$. Setting $g(y) = y^2 - 2y^3/3$, we have $g'(y) = 2y - 2y^2 = 2y(1 - y) > 0$ on $(0, 1)$. Thus, g is strictly increasing on $[0, 1]$.

We conclude $\int_0^y \sqrt{x^4 + (y - y^2)^2}\, dx \le g(y) \le g(1) = 1/3$, with equality if and only if $y = 1$.

A-6. Let $A(n)$ denote the number of sums of positive integers $a_1 + a_2 + \cdots + a_r$ that add up to n with $a_1 > a_2 + a_3$, $a_2 > a_3 + a_4, \ldots, a_{r-2} > a_{r-1} + a_r$, $a_{r-1} > a_r$.

Let $B(n)$ denote the number of $b_1 + b_2 + \cdots + b_s$ that add up to n, with

(i) $b_1 \ge b_2 \ge \ldots \ge b_s$,

(ii) each b_i is in the sequence $1, 2, 4, \ldots, g_j, \ldots$ defined by $g_1 = 1$, $g_2 = 2$, and $g_j = g_{j-1} + g_{j-2} + 1$, and

(iii) if $b_1 = g_k$ then every element in $\{1, 2, 4, \ldots, g_k\}$ appears at least once as a b_i.

Prove that $A(n) = B(n)$ for each $n \ge 1$.

(For example, $A(7) = 5$ because the relevant sums are 7, $6+1$, $5+2$, $4+3$, $4+2+1$, and $B(7) = 5$ because the relevant sums are $4+2+1$, $2+2+2+1$, $2+2+1+1+1$, $2+1+1+1+1+1$, $1+1+1+1+1+1+1$.)

Solution. Given $a_1 + a_2 + \cdots + a_r = n$, $a_1 > a_2 + a_3, \ldots, a_{r-1} > a_r$, let the corresponding b's consist of a_r g_r's, $(a_{r-1} - a_r)$ g_{r-1}'s, $(a_{r-2} - a_{r-1} - a_r)$ g_{r-2}'s, $\ldots, (a_1 - a_2 - a_3)$ g_1's.

Then

$$\begin{aligned} n &= a_r + a_{r-1} + \cdots + a_1 \\ &= a_r(g_r - g_{r-1} - g_{r-2}) + \\ &\quad a_{r-1}(g_{r-1} - g_{r-2} - g_{r-3}) + \ldots + \\ &\quad a_2(g_2 - g_1) + a_1 g_1 \\ &= a_r g_r + (a_{r-1} - a_r) g_{r-1} + \cdots + \\ &\quad (a_1 - a_2 - a_3) g_1, \end{aligned}$$

hence the corresponding b-sum is n as well.

The map is clearly one-to-one. Set

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} a_r \\ a_{r-1} \\ a_{r-2} \\ a_{r-3} \\ \vdots \\ a_1 \end{pmatrix}, \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} \#g_r\text{'s} \\ \#g_{r-1}\text{'s} \\ \#g_{r-2}\text{'s} \\ \#g_{r-3}\text{'s} \\ \vdots \\ \#g_1\text{'s} \end{pmatrix}.$$

Since the system $\mathbf{C}\,\mathbf{A} = \mathbf{G}$ has a solution in integers, the mapping is also onto.

B-1. For each integer $n \ge 0$, let $S(n) = n - m^2$, where m is the greatest integer with $m^2 \le n$. Define a sequence $(a_k)_{k=0}^{\infty}$ by $a_0 = A$ and $a_{k+1} = a_k + S(a_k)$ for $k \ge 0$. For what positive integers A is this sequence eventually constant?

Solution. If A is a perfect square, the sequence is eventually constant, since it is identically A. Clearly the sequence diverges to infinity if it never contains a perfect square. So, say a_n is not a perfect square, but $a_{n+1} = (r+1)^2$. If $a_n \ge r^2$ then

$$\begin{aligned} a_{n+1} &= a_n + S(a_n), \\ (r+1)^2 &= a_n + (a_n - r^2), \\ r^2 + (r+1)^2 &= 2a_n, \end{aligned}$$

a contradiction because the left side is odd but the right side is even. On the other hand, if $a_n < r^2$ we have

$$(r+1)^2 = a_n + S(a_n) < r^2 + (r^2 - 1 - (r-1)^2) = r^2 + 2r - 2,$$

again a contradiction. Hence if A is not a perfect square, no a_n is a perfect square.

B-2. Suppose f and g are nonconstant, differentiable, real-valued functions on $\mathbf{R}$. Furthermore, suppose that for each pair of real numbers x and y,

$$\begin{aligned} f(x+y) &= f(x)f(y) - g(x)g(y), \\ g(x+y) &= f(x)g(y) + g(x)f(y). \end{aligned}$$

If $f'(0) = 0$, prove that $(f(x))^2 + (g(x))^2 = 1$ for all x.

Solution. Differentiate both sides of the two equations with respect to y, obtaining

$$f'(x+y) = f(x)f'(y) - g(x)g'(y),$$
$$g'(x+y) = f(x)g'(y) + g(x)f'(y).$$

Setting $y = 0$ yields

$$f'(x) = -g'(0)g(x) \text{ and } g'(x) = g'(0)f(x).$$

Multiply the first of these by $f(x)$ and the second by $g(x)$ and add to get

$$2f(x)f'(x) + 2g(x)g'(x) = 0.$$

From this it follows that $(f(x))^2 + (g(x))^2$ is identically equal to a constant.

The given functional equations imply that $f(0) = (f(0))^2 - (g(0))^2$ and $g(0) = 2f(0)g(0)$. If $g(0) \neq 0$ then $f(0) = 1/2$ and $(g(0))^2 = (f(0))^2 - f(0) < 0$, a contradiction, since g is real-valued. Therefore $g(0) = 0$, and $f(0) = (f(0))^2$. If $f(0) = 0$ then $f(x) = f(x+0) = f(x)f(0) - g(x)g(0) = 0$, a contradiction, since f is nonconstant. Therefore $f(0) = 1$, and $(f(x))^2 + (g(x))^2 = (f(0))^2 + (g(0))^2 = 1$.

B-3. Does there exist a real number L such that, if m and n are integers greater than L, then an $m \times n$ rectangle may be expressed as a union of 4×6 and 5×7 rectangles, any two of which intersect at most along their boundaries?

Solution. Yes.

Claim: If a and b are positive integers, then there exists a number L_0 so that every multiple of (a,b) (the greatest common divisor of a and b) greater than L_0 may be written in the form $ra + sb$, where r and s are nonnegative integers.

Proof of Claim: Suppose first that $(a,b) = 1$. Then $0, a, 2a, \ldots, (b-1)a$ is a complete set of residues modulo b. Thus, for any integer k greater than $(b-1)a - 1$, $k - qb = ja$ for some $q \geq 0, j = 0, 1, 2, \ldots, b-1$, hence the claim for this special case.

In general, since $a/(a,b)$ and $b/(a,b)$ are relatively prime, we make use of the above to see that for some L_1, every integer greater than L_1 can be written in the form $ra/(a,b) + sb/(a,b)$. Multiplying through by (a,b) yields the claim.

To answer the question, we begin by forming 20×6 and 20×7 rectangles. From the claim, we may form $20 \times n$ rectangles for n sufficiently large. We may also form 35×5 and 35×7 rectangles, hence $35 \times n$ rectangles for n sufficiently large. We may further form 42×4 and 42×5 rectangles, hence $42 \times n$ rectangles for n sufficiently large.

Since $(20, 35) = 5$, there exists a multiple m_0 of 5, relatively prime to 42 and independent of sufficiently large n, for which we may form an $m_0 \times n$ rectangle. Finally, since $(m_0, 42) = 1$, we may form all $m \times n$ rectangles for m and n sufficiently large.

B-4. Suppose p is an odd prime. Prove that

$$\sum_{j=0}^{p} \binom{p}{j}\binom{p+j}{j} \equiv 2^p + 1 \pmod{p^2}.$$

Solution. We have

$$\sum_{j=0}^{p} \binom{p}{j}\binom{p+j}{j} = \sum_{j=0}^{p} \frac{(p+j)!}{(j!)^2(p-j)!}$$
$$= 1 + \sum_{j=1}^{p-1} \frac{(p+j)!}{(j!)^2(p-j)!} + \frac{(2p)!}{(p!)^2}.$$

Now $\dfrac{(2p)!}{(p!)^2} = 2\dfrac{(p+1)(p+2)\cdots(p+p-1)}{1 \cdot 2 \cdots (p-1)}$.

From the expansion

$$(x+1)(x+2)\cdots(x+p-1) =$$
$$= \prod_{j=1}^{(p-1)/2} (x+j)(x+p-j)$$
$$= \prod_{j=1}^{(p-1)/2} (x^2 + px + j(p-j))$$
$$= (p-1)! + \lambda px + x^2 f(x)$$

for some integer λ and some integer polynomial $f(x)$, we see that

$$(p+1)(p+2)\cdots(p+p-1) \equiv (p-1)! \pmod{p^2}.$$

Thus, since $\dfrac{(2p)!}{(p!)^2}$ is an integer, we have

$$\frac{(2p)!}{(p!)^2} \equiv 2 \pmod{p^2}.$$

Note that since $p + i \equiv i \pmod{p}$ and $\binom{p+j}{j}$ is an integer, $\dfrac{(p+1)\cdots(p+j)}{j!} = 1 + \lambda_j p$, where λ_j is an integer. It follows that

$$\sum_{j=1}^{p-1} \frac{(p+j)!}{(j!)^2(p-j)!} =$$
$$= \sum_{j=1}^{p-1} \frac{p!}{j!(p-j)!} \frac{(p+1)\cdots(p+j)}{j!}$$

$$= \sum_{j=1}^{p-1} \frac{p!}{j!(p-j)!}(1+\lambda_j p)$$

$$\equiv \sum_{j=1}^{p-1} \frac{p!}{j!(p-j)!} \pmod{p^2}$$

$$= 2^p - 2.$$

Combining these, we get the result.

B-5. Let p be an odd prime and let $\mathbf{Z}_p$ denote (the field of) the integers modulo p. How many elements are in the set

$$\{x^2 : x \in \mathbf{Z}_p\} \cap \{y^2+1 : y \in \mathbf{Z}_p\} \ ?$$

Solution. There are $\lfloor (p+3)/4 \rfloor$ elements in the intersection.

Consider first the set of solutions to

$$x^2 = y^2 + 1. \qquad (*)$$

Rewriting this as $(x+y)(x-y) = 1$, we see that for each nonzero element r of $\mathbf{Z}_p$, there is exactly one solution to the above, namely, $x+y = r$, $x-y = r^{-1}$, or

$$x = \left(\frac{p+1}{2}\right)(r+r^{-1}),$$

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

Thus, there are $p-1$ solutions to $(*)$.

On the other hand, the element $x^2 = y^2+1$ in the intersection also arises from the pairs $(x,-y),(-x,y)$, and $(-x,-y)$ as well as (x,y). These four pairs are distinct unless $x = 0$ or $y = 0$, in which case there are just two distinct pairs. Note that the element 1 arises from $(1,0)$ and $(-1,0)$. Let $c = 1$ if 0 is also in the intersection, and let $c = 0$ if not. Then the intersection has $1+c+d$ elements, where, from the above, $p-1 = 2+2c+4d$.

We see that $c = 1$ if and only if $p-1$ is divisible by 4. Solving for d in each case, we find that $1+c+d = \lfloor (p+3)/4 \rfloor$.

Note: This proof describes when -1 is a square modulo p without recourse to Fermat's Little Theorem. Also, *Ian Richards, University of Minnesota*, points out that this problem follows from a special case of the following result: If χ is the quadratic character mod p, then $\sum_{n=0}^{p-1} \chi(n)\chi(n+k) = -1$, independent of k. This result is provable from the theory of Jacobi or Gauss sums.

B-6. Let a and b be positive numbers. Find the largest number c, in terms of a and b, such that

$$a^x b^{1-x} \le a\frac{\sinh ux}{\sinh u} + b\frac{\sinh u(1-x)}{\sinh u}$$

for all u with $0 < |u| \le c$ and for all x, $0 < x < 1$. $\left(\text{Note : } \sinh u = (e^u - e^{-u})/2\right)$.

Solution. The inequality is satisfied if and only if $0 < |u| \le |\ln(a/b)|$.

The right-hand side is an even function of u; hence it suffices to consider $u > 0$. Replacing x by $1-x$ and interchanging a and b preserves the inequality, hence we may assume $a \ge b$. Set

$$F(u) = a\frac{\sinh ux}{\sinh u} + b\frac{\sinh u(1-x)}{\sinh u} - a^x b^{1-x}.$$

By differentiating

$$f(u) = \frac{\sinh ux}{\sinh u}$$

we find that $f'(u) < 0$ if and only if $g(u) = x\tanh u - \tanh xu < 0$. This latter inequality holds because $g(0) = 0$ and $g'(u) < 0$ for $u > 0$. Thus $f(u)$ is strictly decreasing in u, and therefore, so is $F(u)$. If $a > b$ then $F(\ln(a/b)) = 0$, whereas if $a = b$ then $\lim_{u\to 0+} F(u) = 0$, and the proof is complete.

CARL B. ALLENDOERFER AWARD

1990

The recipient of the Carl B. Allendoerfer Award for mathematical exposition in the 1990 *Mathematics Magazine*, announced at the summer 1991 meetings of the MAA, was:

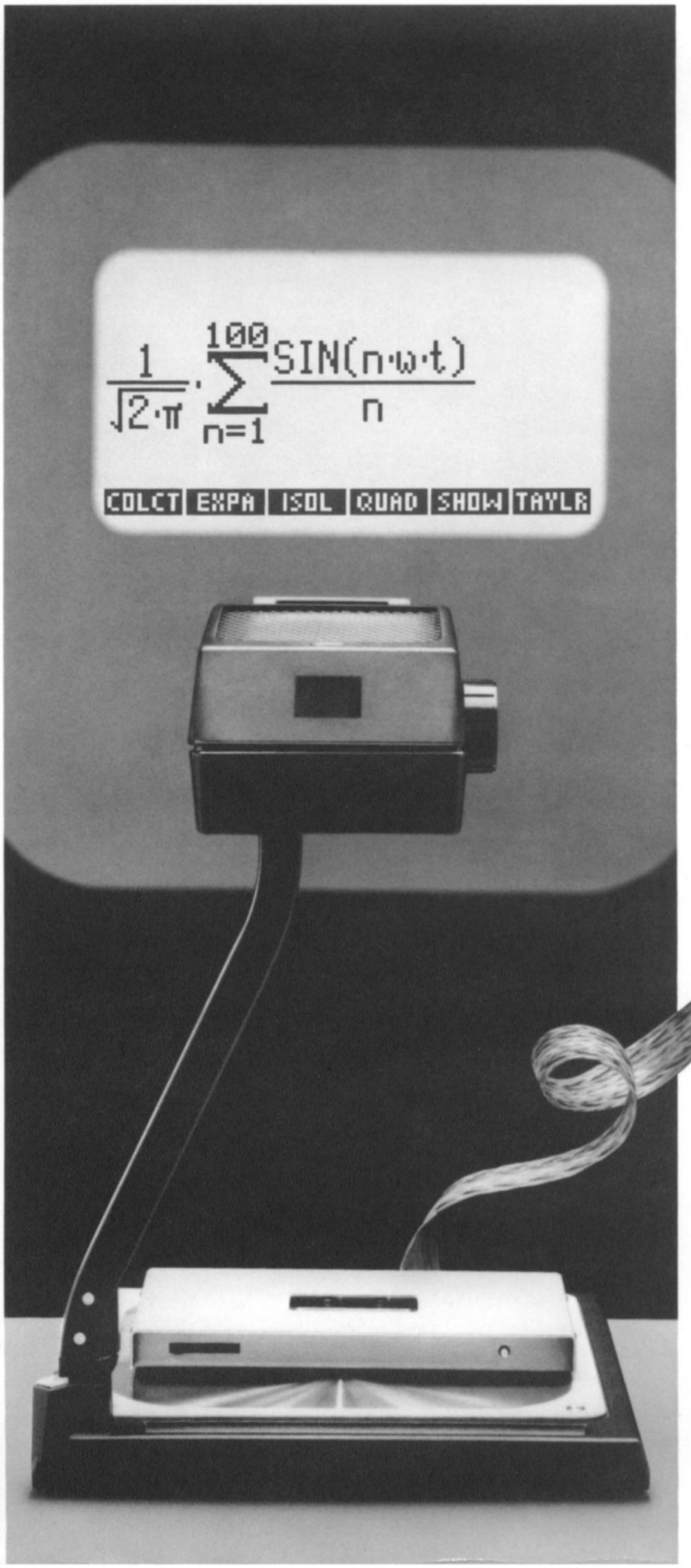
Ranjan Roy

for

The Discovery of the Series Formula for π by Leibniz, Gregory and Nilakantha,

this MAGAZINE, 63 (1990), 291-306.

Help your students discover more meaningful relationships.



Again in '92: a free classroom display device with purchase of 30 calculators.

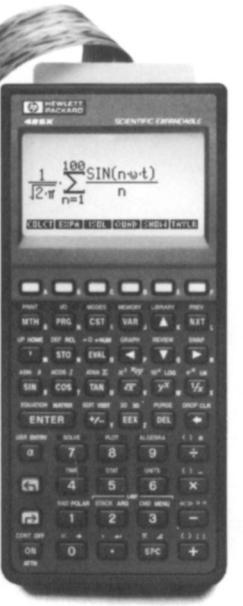
Showing is much more powerful than telling. So we've developed special classroom displays for our most advanced calculators.

The HP 48SX scientific expandable calculator and the cost-effective HP 48S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculators.

If your department or students purchase 30 HP 48SX or HP 48S calculators (or a mix of both), we'll give you free an HP 48SX and plug-in classroom display (a $900 retail value).

So call **(503) 757-2004** from 8am to 3pm PDT for details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends December 31, 1992, and applies only to college and high school instructors.

HEWLETT PACKARD



CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, D.C. 20036