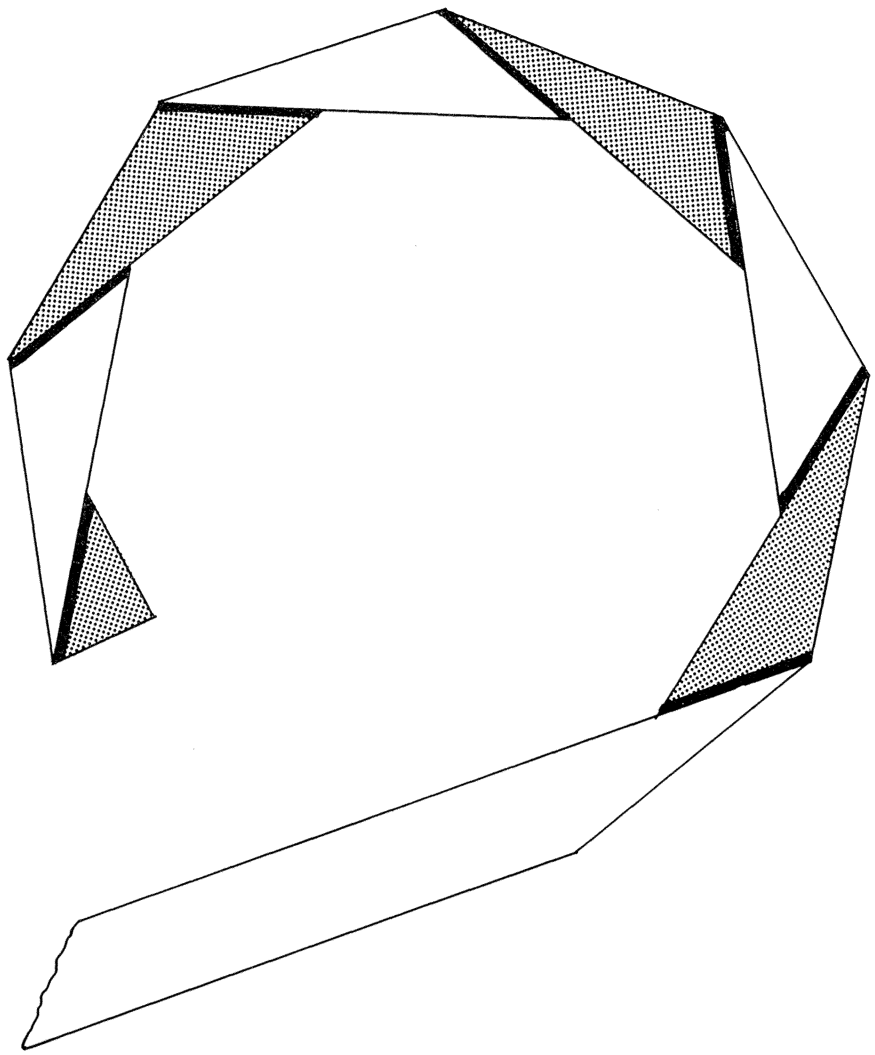
MATHEMATICS MAGAZINE



Vol. 56 No. 3
May, 1983

GAPS IN INTEGER SEQUENCES • ALGORITHM FOR $C_A(X)$
FOLDING POLYGONS • VISUALIZING MATRIX DECOMPOSITION

The RAYMOND W. BRINK SELECTED MATHEMATICAL PAPERS Series.

This series is a collection of papers on single topics selected and reprinted from the journals of the Mathematical Association of America. The papers are grouped by subject within the topic of the volume and are indexed by author. Each volume is a rich source of mathematical stimulation for students and teachers.

Four volumes are available in this series:

Volume 1.
SELECTED PAPERS ON PRECALCULUS

Edited by Tom M. Apostol, Gulbank D. Chakerian, Geraldine C. Darden, and John D. Neff. xvii + 469 pages. Hardbound.

List: $21.00; MAA Members *(personal use)* $16.00

Volume 2.
SELECTED PAPERS ON CALCULUS

Edited by Tom M. Apostol, Hubert E. Chrestenson, C. Stanley Ogilvy, Donald E. Richmond, and N. James Schoonmaker. xv + 397 pages. Hardbound.

List: $21.00; MAA Members *(personal use)* $16.00

Volume 3.
SELECTED PAPERS ON ALGEBRA

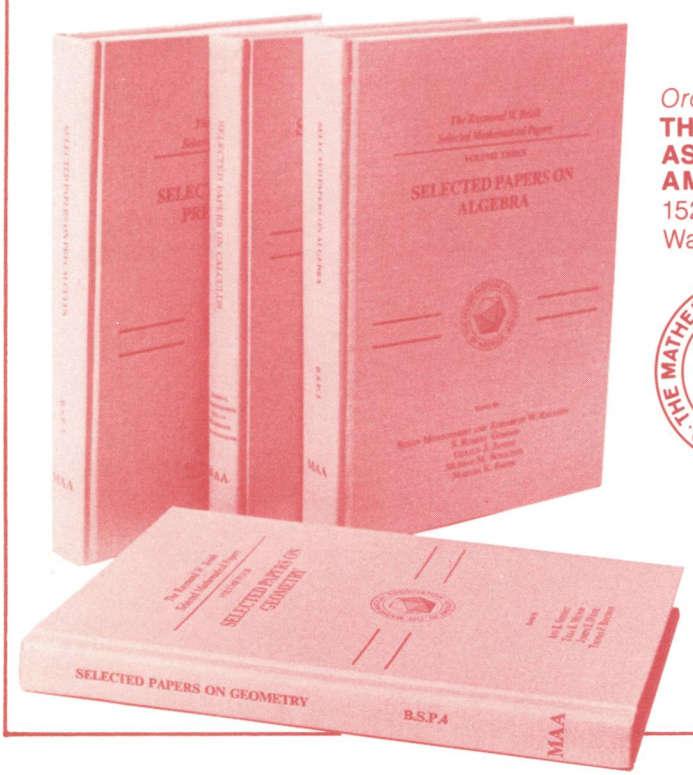
Edited by Susan Montgomery, Elizabeth W. Ralston, S. Robert Gordon, Gerald J. Janusz, Murry M. Schacher, and Martha K. Smith. xx + 537 pages. Hardbound.

List: $21.00; MAA Members *(personal use)* $16.00

Volume 4.
SELECTED PAPERS ON GEOMETRY

Edited by Ann K. Stehney, Tilla K. Milnor, Joseph D'Atri, and Thomas F. Banchoff. ix + 338 pages. Hardbound.

List: $21.00; MAA Members *(personal use)* $16.00



Order From:
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

MATHEMATICS MAGAZINE

THE MATHEMATICAL ASSOCIATION OF AMERICA

Volume 56, Number 3, May 1983
(ISSN—0025-570X)

COVER: Folding a regular 9-gon. See pp. 142–143.

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

EDITORIAL POLICY

Mathematics Magazine is a journal which aims to provide inviting, informal mathematical exposition. Manuscripts accepted for publication in the *Magazine* should be written in a clear and lively expository style and stocked with appropriate examples and graphics. Our advice to authors is: say something new in an appealing way or say something old in a refreshing way. The *Magazine* is not a research journal and so the style, quality, and level of articles submitted for publication should realistically permit their use to supplement undergraduate courses. The editor invites manuscripts that provide insight into the history and application of mathematics, that point out interrelationships between several branches of mathematics and that illustrate the fun of doing mathematics.

Authors planning to submit manuscripts should read the full statement of editorial policy which appears in this *Magazine*, Vol. 54, pp. 44–45, and is also 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.

New manuscripts should be sent to: Doris Schattschneider, Editor, Mathematics Magazine, Moravian College, Bethlehem, PA 18018. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine*. They should be typewritten and double spaced on $8\frac{1}{2}$ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added.

AUTHORS

Heini Halberstam ("Gaps in Integer Sequences") is professor and head of the Mathematics Department at the University of Illinois, Urbana. He received his education and Ph.D. at University College, London University, and taught mathematics at the University of Nottingham, England. He specializes in the theory of numbers, and has interests in the history of mathematics, and in mathematics education.

Peter Hilton ("Approximating Any Regular Polygon by Folding Paper") received a D.Phil. from Oxford University and a Ph.D. from Cambridge University. He is currently Professor of Mathematics at the State University of New York (Binghamton campus). His research interests are in algebraic topology, homological algebra and mathematical education, and he has several publications in these fields. He is currently secretary/treasurer of the International Commission on Mathematics Instruction.

Jean Pedersen received her master's degree from the University of Utah and taught there before joining the mathematics department at the University of Santa Clara. Her principal research interests are in polyhedral geometry and mathematics education, in which fields she has published several articles and books. Her most recent book is *Fear No More: An Adult Approach to Mathematics* (Addison-Wesley), co-authored with Peter Hilton. She is a past coordinator of the Women and Mathematics lectureship program for the San Francisco Bay area.

BUSINESS INFORMATION. Mathematics Magazine is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. The annual subscription price for Mathematics Magazine to an individual member of the Association is $10, included as part of the annual dues of $30. Students receive a 50% discount. Bulk subscriptions (5 or more copies to a single address) are available to colleges and universities for distribution to undergraduate students at a 35% discount. The library subscription price is $25.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N. W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.

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

Gaps in Integer Sequences

"All nature is but art unknown to thee...
All discord, harmony not understood."
—**Alexander Pope,** *An Essay on Man*

Heini Halberstam
University of Illinois
Urbana, IL 61801

Let $\mathcal{A}$: $a_1 < a_2 < \cdots$ be an infinite sequence of positive integers, and let $d_n = a_{n+1} - a_n$ denote the nth **gap**. For our present purpose we regard any sequence $\mathcal{A}$ as interesting if its elements are so irregularly distributed that the gaps do not conform to any apparent pattern. Thus an arithmetic progression $\{a + bn: n = 0, 1, 2, \dots\}$ is uninteresting in our sense because all its gaps are equal to the common difference b, and so too is the sequence of squares $\{n^2: n = 0, 1, 2, \dots\}$ uninteresting because here the gaps form the arithmetic progression $\{1 + 2n: n = 0, 1, 2, \dots\}$. Indeed, no sequence is interesting whose nth term a_n is given precisely by some algebraic formula, for the same is then true of the gaps d_n.

Given an interesting sequence $\mathcal{A}$, there are various questions that one may ask about the fluctuation in size of its gaps, and the purpose of this article is to consider some of these questions in the context of three well-known sequences: (i) the **primes**, (ii) **integers that are sums of two squares**, and (iii) the **squarefree numbers** (integers that are products of distinct primes). I shall restrict myself to results that can be proved by elementary and simple arguments, and I shall only mention some of the deeper known results. You will get some impression of the difficulty of the subject if I add that even the deepest results now available fall far short of the likely ultimate truth.

The primes

I begin with the sequence of primes. Tables of the first so many primes are readily available (see, for example, [**21**]) and study of any such table makes it seem extremely unlikely that the distribution of primes can be described in terms of simple rules. The gap $d_n = 1$ occurs only once, when $n = 1$ and $a_2 - a_1 = 3 - 2 = 1$. But $d_n = 2$ when $n = 2, 3, 5, 7, 10, 13$ and seems to recur however far the table extends, and, indeed, beyond; for computer searches throw up such prime "twins" well beyond the limits of systematic tabulation of primes (see Brent [**2**]). For example, $10^{12} + 61$ and $10^{12} + 63$ are a pair of consecutive primes. A huge pair, noted in [**20**], is $1,159,142,985 \cdot 2^{2304} \pm 1$ which have 703 digits each. Nevertheless, no one has been able to prove so far that $d_n = 2$ for infinitely many values of n. (A computer software firm has just offered $25,000 for a proof!) On the other hand, there exist long runs of consecutive composite integers and therefore d_n assumes large values; for example, there are no primes between 113 and 127, between 839 and 853, or between 20,831,323 and 20,831,533, this last a gap of length 210. Indeed, the simple example of the sequence

$$m! + 2, m! + 3, \dots, m! + m \qquad (m \geqslant 3) \tag{1}$$

of $m-1$ consecutive composite integers shows that d_n can assume arbitrarily large values. We can extract more precise information from (1). Let a_n be the largest prime not exceeding $m!+1$, so that $d_n = a_{n+1} - a_n \geqslant m-1$. Then $a_n \leqslant m!+1 < m^m (m \geqslant 2)$; taking logarithms,

$$m \ln m > \ln a_n, \tag{2}$$

and from this we deduce that $m > (\ln a_n)/(\ln \ln a_n)$. For, either $m \geqslant \ln a_n$ which is even better, or $m < \ln a_n$, in which case, by (2), $m \ln(\ln a_n) > \ln a_n$, the claimed result. Hence

$$d_n \geqslant m-1 > \frac{1}{2} \frac{\ln a_n}{\ln \ln a_n} \tag{3}$$

for infinitely many values of n. This confirms the fact that as one runs along the sequence of primes, increasingly large gaps occur. We shall see later that, on average, d_n is about $\ln a_n$, so that (3) is not so very wide off the mark. A deeper investigation (see Rankin [**14**]) shows that, actually, d_n is often significantly larger than the average: more precisely, there exists a constant $c > 0$ such that

$$d_n > c \frac{\ln a_n (\ln_2 a_n)(\ln_4 a_n)}{(\ln_3 a_n)^2}$$

for infinitely many values of n (where $\ln_2 = \ln \ln, \ln_r = \ln(\ln_{r-1})$).

In the opposite direction we may ask whether there is an upper bound to the size of d_n in terms of the nth prime a_n. More precisely, the problem is to *determine a positive increasing function $h(x)$ such that*

$$d_n < h(a_n) \text{ for all } n.$$

In our discussion, we will call such a function h **admissible**. Naturally we should like to find among all admissible functions h the one that grows the least rapidly, and a classical probabilistic argument of Cramér [**4**] suggests that $h(x) = c(\ln x)^2$ may be close to the truth. The present state of knowledge falls far short of a result of this equality, even though there is currently much successful activity in this field and the "world record" changes by the month. The best result published to date is $h(x) = x^{11/20}$ (Heath-Brown & Iwaniec [**8**]). To the innocent eye this seems rather close to $h(x) = x^{1/2}$; however, $x^{1/2}$ is already slightly better than what would follow from the truth of the famous Riemann Hypothesis!

Actually this question was first raised early in the nineteenth century by Bertrand, who, while investigating permutation groups, found that he needed to know that the choice $h(x) = x$ is admissible. He checked the truth of his "hypothesis" for all primes less than 3×10^6, but the first proof was not found until 1850, by Čebyčev. There have been several other proofs since then, all of them simpler than the original, and I shall present yet another proof in the next section. This latest proof is due to R. R. Hall (unpublished) and is based on the ideas of M. Nair [**12**]. It is, in my opinion, the simplest proof to date, and actually leads to a better result than the one I prove in the next section. The interested reader may care to find for himself the best result that the Nair-Hall method yields.

Before we prove Bertrand's "hypothesis," it is worth recalling Čebyčev's approach [**3**]. Let $A(x)$ denote the **number of elements of** $\mathcal{A}$ (of primes, in the present discussion) **that do not exceed** x. To prove the admissibility of h is equivalent to showing that

$$A(x + h(x)) - A(x) > 0 \tag{4}$$

for all x from some point onward; for

$$A(a_n + h(a_n)) - A(a_n) > 0$$

implies that $a_{n+1} < a_n + h(a_n)$, that is, $d_n < h(a_n)$. So, if we know enough about $A(x)$—for example, if we have an asymptotic formula for $A(x)$ of the type

$$\lim_{x\to\infty} A(x)\Big/\left(\frac{x}{a(\ln x)^b}\right)=1 \qquad (a>0,\ b\geqslant 0 \text{ are constants}) \tag{5}$$

or even if we know only that $x(\ln x)^{-b}$ is the right order of magnitude of $A(x)$—then we are able to prove (4) for some choice of h. Čebyčev succeeded in proving for the primes that

$$0.921\frac{x}{\ln x}<A(x)<1.106\frac{x}{\ln x}, \qquad x\geqslant x_0, \tag{6}$$

and from his work he was able to deduce that

$$A(2x)-A(x)>0.6\frac{x}{\ln 2x} \tag{7}$$

if x is large enough (in fact, (7) is true for $x\geqslant 20.5$; see Rosser and Schoenfeld [**16**]). Inequality (7) tells us even that the interval $(x,2x)$ contains about the right proportion of primes. The famous prime number theorem, proved simultaneously in 1896 by Hadamard and de la Vallée Poussin, states that (5) holds with $a=b=1$; and from it we can derive an even better gap theorem. (In the case of primes, the universal notation for $A(x)$ is $\pi(x)$.) The prime number theorem tells us also that the gaps d_n are on average about $\ln a_n$. However, to this day the prime number theorem is still hard to prove; and even the proof of (6), though elementary, is rather complicated.

Technical difficulties apart, there is a deeper reason why one should not approach gap problems about $\mathcal{A}$ by means of information about $A(x)$. In most interesting cases, there is an inherent limitation to the precision with which $A(x)$ can be approximated by a known function. For example, we know that

$$\operatorname{li} x=\int_2^x \frac{dt}{\ln t}$$

is a better approximation to $\pi(x)$ than $x/\ln x$. Nevertheless, Littlewood proved that $|\pi(x)-\operatorname{li} x|$ is larger than $\frac{1}{3}x^{1/2}\frac{\ln_3 x}{\ln x}$ infinitely often, which is vastly larger than Cramér's conjectured $(\ln x)^2$. This suggests that one should devise special methods to tackle gap problems rather than just rely on information about $A(x)$.

More about the primes

In this section, p, with or without suffix, always denotes a prime. For each integer $n\geqslant 2$ define

$$D_n=\text{L.C.M.}\ \{2,3,4,\ldots,n\}$$

and

$$\theta(n)=\sum_{p\leqslant n}\ln p.$$

Let $[y]$ denote, as usual, the integer part of y, that is, the largest integer $\leqslant y$. It is not hard to prove (see Hardy and Wright [7], sections 22.1 and 22.2) that

$$\ln D_n=\sum_{p\leqslant n}\left[\frac{\ln n}{\ln p}\right]\ln p \tag{8}$$

and that

$$\theta(n)\leqslant n\ln 4, \qquad n=2,3,4,\ldots. \tag{9}$$

(In the literature the notation for $\ln D_n$ is usually $\psi(n)$.)

Exercise. Deduce from (9) that, for any $y<n$,

$$\pi(n)<\pi(y)+\frac{n\ln 4}{\ln y}<y+\frac{n\ln 4}{\ln y}.$$

Take $y = n^{9/10}$, say, to show that $\pi(n) < 2\dfrac{n}{\ln n}$ for $n > 5^{20}$, a weaker version of the right-hand inequality in (6).

LEMMA. *For all positive integers* n,

$$\theta(n) \leqslant \ln D_n \leqslant \theta(n) + 6n^{1/2}.$$

Proof. The left-hand inequality is obvious from (8), since $[y] \geqslant 1$ if $y \geqslant 1$. In fact, for each $k = 1, 2, 3, \ldots$,

$$\left[\frac{\ln n}{\ln p}\right] = k \text{ when } n^{1/(k+1)} < p \leqslant n^{1/k},$$

so that

$$\ln D_n = \sum_{k \leqslant \ln n/\ln 2} k \sum_{n^{1/k+1} < p < n^{1/k}} \ln p = \theta(n) + \theta(n^{1/2}) + \theta(n^{1/3}) + \cdots + \theta(n^{1/s}),$$

where

$$s = \left[\frac{\ln n}{\ln 2}\right].$$

Hence, by (9),

$$\begin{aligned} \ln D_n &\leqslant \theta(n) + (n^{1/2} + n^{1/3} + \cdots + n^{1/s})\ln 4 \\ &\leqslant \theta(n) + (n^{1/2} + (s-2)n^{1/3})\ln 4 \leqslant \theta(n) + \left(1 + \frac{\ln n}{\ln 2} n^{-1/6}\right) n^{1/2} \ln 4. \end{aligned}$$

But $x^{-1/6}\ln x = 6x^{-1/6}\ln x^{1/6}$ assumes its maximum value $6/e$ at $x = e^6$, whence

$$\ln D_n \leqslant \theta(n) + \left(1 + \frac{6}{e \ln 2}\right) n^{1/2} \ln 4 \leqslant \theta(n) + 6n^{1/2}$$

since $\ln 4 + 12/e < 5.9$.

After these preliminaries we are ready for the main argument. (It would be useful but not essential to look first at the proof of Theorem 1 of Nair [**12**].) For any positive integers m, n define

$$I_{m,n} = \int_0^1 (2x-1)^{2m} x^n (1-x)^n \, dx. \tag{10}$$

(Nair works with $I_{0,n}$. The effect of the additional factor $(2x-1)^{2m}$ is to introduce another zero in the integrand, at $x = \frac{1}{2}$, and so to reduce the size of the integral.) Obviously $I_{m,n}$ is positive. The integrand of $I_{m,n}$ has the form

$$4^{-n} y^m (1-y)^n, \quad y = (2x-1)^2,$$

and therefore, by elementary calculus, takes its maximum in the range of integration at $y = (2x-1)^2 = m/(m+n)$. Hence

$$0 < I_{m,n} \leqslant \frac{1}{4^n} \frac{n^n m^m}{(m+n)^{m+n}}. \tag{11}$$

On the other hand, the integrand of $I_{m,n}$ is a polynomial of degree $2m + 2n$, with integer coefficients; if we write it as

$$\sum_{r=0}^{2m+2n} c_r x^r, \; c_r \in \mathbb{Z}, \; r = 0, 1, \ldots, 2m+2n,$$

then

$$I_{m,n} = \sum_{r=0}^{2m+2n} \frac{c_r}{r+1}.$$

Hence $D_{2m+2n+1}I_{m,n}$ is a positive integer, and therefore it is at least as large as 1. It follows from (11) that

$$1 \leqslant D_{2m+2n+1}\frac{1}{4^n}\frac{n^n m^m}{(m+n)^{m+n}},$$

and, taking logarithms, this becomes

$$\ln D_{2m+2n+1} \geqslant (m+n)\ln 4 + n\ln\left(1+\frac{m}{n}\right) + m\ln\left(\frac{m+n}{4m}\right). \tag{12}$$

Now choose $n = 2m$, when the inequality becomes

$$\ln D_{6m+1} \geqslant (3m+3)\ln 4 + m\ln\frac{27}{16} - 3\ln 4$$

$$> (3m+3)\ln 4 + \frac{m}{10}, \quad \text{provided that } m \geqslant 10.$$

By (9) and the Lemma, it follows that

$$\theta(6m+1) + 6(6m+1)^{1/2} > \theta(3m+3) + \frac{m}{10}, \quad m \geqslant 10.$$

But if all the integers $3m+4,\ 3m+5,\ldots,6m+1$ are composite, $\theta(6m+1) = \theta(3m+3)$ and therefore

$$6(6m+1)^{1/2} > \frac{m}{10},$$

a statement that is obviously false for all sufficiently large m. Indeed, $6(6m+1)^{1/2} \leqslant 6(6.1)^{1/2}m^{1/2}$ if $m \geqslant 10$, and $6(6.1)^{1/2}m^{1/2} < \frac{m}{10}$ as soon as $m > (60\sqrt{6.1})^2 = 21{,}960$. Hence, provided only that $m \geqslant 21{,}961$, the interval $(3m+3, 6m+2)$ contains at least one prime. Finally, every integer is of the form $3m+1$, $3m+2$ or $3m+3$. Since $(3m+i, 6m+3i)$ contains the interval $(3m+3, 6m+2)$ for each of $i = 1,2,3$, we have proved that $(N, 2N)$ *always contains a prime provided that* $N \geqslant 66{,}000$. The cases $N < 66{,}000$ are amply covered by existing tables, or by Bertrand's original calculations.

Exercise. Take $n = km$ in (12) and check that, for $k \geqslant 3$,

$$\ln D_{(2k+2)m+1} \geqslant (k+r)(m+1)\ln 4 + \frac{2m}{3}, \quad r = \left[\frac{\ln(k+1)}{\ln 4}\right],$$

provided $m \geqslant 6(k+r)\ln 4$. Hence derive a better result than the one obtained above.

Sums of two squares

Integers that are sums of two squares afford another instance of an interesting sequence. Now let $\mathcal{A}$ denote this sequence, which begins $1,2,4,5,8,9,10,13,16,17,18,20,25,\ldots,74,80,\ldots$. Fermat was probably the first person who knew how to prove (using his famous method of infinite descent) that every prime congruent to 1 modulo 4 is the sum of two integer squares, and Euler was the first to publish a proof. These primes and $2 = 1^2 + 1^2$ are the only primes with this remarkable property; but the famous identity

$$(x^2+y^2)(m^2+n^2) = (xm-yn)^2 + (xn+ym)^2$$

shows that if a and a' belong to $\mathcal{A}$ then so does the product aa', so that $\mathcal{A}$ obviously contains many other integers. (Note that if $a \in \mathcal{A}$ and a is odd, then $a \equiv 1 \bmod 4$.) In fact, it is not too hard to prove that if

$$n = \prod_p p^{\alpha_p(n)} \tag{13}$$

is the canonical prime decomposition of n, then n belongs to $\mathcal{A}$ if and only if $\alpha_p(n)$ is *even* whenever $p \equiv -1 \bmod 4$. This characterization of the elements of $\mathcal{A}$ enables us to show that there

exist arbitrarily large gaps in $\mathcal{Q}$. We shall need the Chinese Remainder Theorem. This tells us that the simultaneous congruences

$$N+1 \equiv b_1 \bmod m_1, N+2 \equiv b_2 \bmod m_2, \ldots, \quad N+k \equiv b_k \bmod m_k$$

have a solution N unique modulo $m_1 m_2 \cdots m_k$ provided the moduli $m_1, \ldots, m_k$ are pairwise coprime. Now take $q_1 = 3, q_2 = 7, \ldots, q_k$ to be the first k primes congruent to -1 modulo 4, and let $m_i = q_i^2, b_i = q_i$ $(i = 1, \ldots, k)$. If N is an associated solution of the k simultaneous congruences, then, for each $i, 1 \leqslant i \leqslant k$, $N+i$ is divisible by q_i but not by q_i^2, and therefore cannot be the sum of two squares. In this way we obtain k consecutive integers that do not lie in $\mathcal{Q}$; moreover, we may carry out the construction for any positive integer k, however large.

We can use this calculation to show that

$$a_{n+1} - a_n > \left(\frac{1}{2} - \varepsilon\right) \frac{\ln a_n}{\ln \ln a_n}, \qquad n \geqslant n_0(\varepsilon),$$

as we did in the case of primes; but the argument requires some information about the way in which primes congruent to -1 modulo 4 are distributed, and therefore I omit it. Quite recently Ian Richards [**15**] has shown in a surprisingly simple way that

$$a_{n+1} - a_n > (\tfrac{1}{4} - \varepsilon) \ln a_n, \quad n \geqslant n_1(\varepsilon).$$

Exercise. Show that $\mathcal{Q}$ cannot contain four consecutive integers.

Although the supply of integers that are sums of two squares seems to be quite ample, it transpires that $\mathcal{Q}$ is only slightly more plentiful than the primes. Landau proved in 1908 that for this sequence (cf (2)),

$$A(x) \sim \frac{\pi}{\sqrt{12}} \frac{x}{\sqrt{\ln x}} \text{ as } x \to \infty, \tag{14}$$

so that both the primes and sums of two squares share the property

$$\lim_{x \to \infty} \frac{A(x)}{x} = 0; \tag{15}$$

therefore they may be said to have zero density; see Iwaniec [**9**] for a different proof. Formula (14), like the prime number theorem, is hard to prove. Indeed, while the latter can be proved in a technically elementary way, (14) still requires substantial analytic tools (see LeVeque [**10**]). (In contrast to the primes, one can show that there exist infinitely many n such that both n and $n+1$ are both in $\mathcal{Q}$; but this also is too complicated to prove here.) On the other hand, there is in this case a much simpler alternative to $A(x)$ to work with. Let $r(n)$ denote the **number of ways of representing n as the sum of two integer squares;** then (cf (13)), if $r(n) \neq 0$, $r(n)$ is given by

$$r(n) = 4 \prod_{p \equiv 1 \bmod 4} \left(\alpha_p(n) + 1\right)$$

where the factor 4 derives from counting the pairs l, m; $-l, m$; m, l; $m, -l$ as giving different representations $n = l^2 + m^2$. (If n is a prime congruent to $1 \bmod 4$ then $r(n) = 8$; the other four representations derive from $l, -m$; $-l, -m$; $-m, l$; $-m, -l$. Thus such a prime has only one genuinely distinct representation as the sum of two squares.) Writing

$$R(x) = \sum_{n \leqslant x} r(n) = \sum_{\substack{l, m \\ l^2 + m^2 \leqslant x}} 1,$$

we see that $R(x)$ counts the same numbers as $A(x)$ does, but counts each element of $\mathcal{Q}$ as often as it is representable as the sum of two squares. Thus, whereas $A(x)$ is small (see (14)), $R(x)$ is roughly equal to πx, for $R(x)$ clearly counts the number of points with integer coordinates that lie within or on the circle with center at the origin and radius $x^{1/2}$. It follows that

$$\pi\left(x^{1/2} - \sqrt{2}\right)^2 \leqslant R(x) \leqslant \pi\left(x^{1/2} + \sqrt{2}\right)^2$$

and therefore

$$|R(x)-\pi x| \leqslant (2\pi\sqrt{2})x^{1/2} + 2\pi < 9x^{1/2} + 2\pi < 10x^{1/2} \quad (x \geqslant 40). \tag{16}$$

By (16),

$$\begin{aligned} R(x+10x^{1/2}) - R(x) &\geqslant \pi(x+10x^{1/2}) - 10(x+10x^{1/2})^{1/2} - \pi x - 10x^{1/2} \\ &= 10(\pi-1)x^{1/2} - 10(x+10x^{1/2})^{1/2} > 10(\pi-1)x^{1/2} - 10(x^{1/2}+5) \\ &= 10(\pi-2)x^{1/2} - 50 > 0 \end{aligned}$$

for all $x \geqslant 20$. Hence for all $x \geqslant 40$ there is an element of $\mathfrak{A}$ between x and $x + 10x^{1/2}$, and $h(x) = 10x^{1/2}$ is therefore an admissible choice here. Actually it is known, but very much harder to prove (Sierpinski [**18**]) that

$$|R(x) - \pi x| \leqslant cx^{\theta} \tag{17}$$

is true with $\theta = \frac{1}{3}$ and for some constant c; and that with $\theta = \frac{1}{4}$ relation (17) is false for every constant $c > 0$ (Hardy [**6**]). Hence a choice $h(x) = c_0 x^{1/3}$ is admissible, but no argument of *this* kind can give $h(x) = c_1 x^{1/4}$. So once again we have come up against a limitation principle of the kind mentioned in our first section. (For a very general limitation principle of almost the same quality, and with a very elegant proof, see Erdős and Fuchs [**5**].) It is the more remarkable that we can show nevertheless, in quite a different way, that *the interval* $(x, x + 3x^{1/4})$ *does contain a sum of two squares provided only* $x \geqslant 1154$. Here is the simple argument, due to Bambah & Chowla [**1**].

Let m be the largest integer not exceeding $x^{1/2}$ (see FIGURE 1), so that

$$m < x^{1/2} < m + 1.$$

(We may suppose that x is not a perfect square; otherwise $x + 1$ is available!) Let λ be the positive real number such that $m^2 + \lambda^2 = x$, i.e., such that (m, λ) lies on the circle having centre 0 and radius $x^{1/2}$. Let n be the least positive integer greater than λ, so that

$$n - 1 \leqslant \lambda < n.$$

Then $m^2 + n^2 > m^2 + \lambda^2 = x$; on the other hand,

$$m^2 + n^2 \leqslant m^2 + (\lambda + 1)^2 = x + 2\lambda + 1.$$

But

$$\lambda^2 = x - m^2 < x - (x^{1/2} - 1)^2 = 2x^{1/2} - 1 < 2x^{1/2},$$

so that $\lambda < 2^{1/2}x^{1/4}$ and

$$2\lambda + 1 < 2^{3/2}x^{1/4} + 1 < 3x^{1/4} \quad \text{if} \quad x \geqslant 1154.$$

Hence $m^2 + n^2 < x + 3x^{1/4}$ provided $x \geqslant 1154$ and so there is a number between x and $x + 3x^{1/4}$ that is the sum of two integer squares.

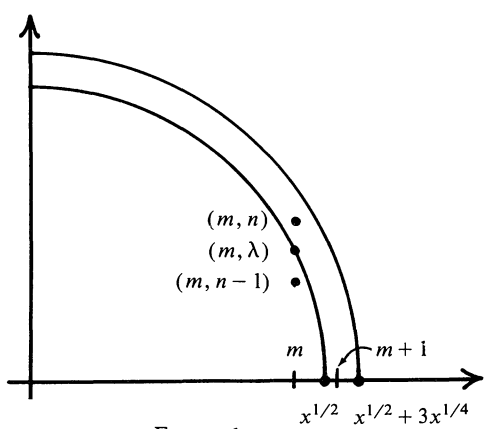


FIGURE 1

One would expect there to be more sophisticated ideas that lead to something better; almost certainly something much better is true. However, the above result is essentially the best available today.

Squarefree numbers

Now let $\mathcal{Q}$ denote the sequence of squarefree numbers, which begins 2, 3, 5, 6, 7, 10, 11, 13, 14, 15, 17,.... Here it is possible to prove (Mirsky [**11**]; but the details are too complicated to present here) that there exist infinitely many pairs and even triplets of consecutive squarefree numbers. On the other hand, we need only the Chinese Remainder Theorem again to show that arbitrarily large gaps in $\mathcal{Q}$ do occur. Solve $N+i \equiv 0 \bmod p_i^2, i=1,2,\ldots,k$, where p_i denotes the ith prime. Then, if $N \leqslant (2\cdot 3\cdots p_k)^2$ is the unique solution, none of $N+1,\ldots,N+k$ is squarefree.

Exercise. Check that $N \equiv 11069 \bmod 210^2$, $N \equiv -2 \bmod 11^2$, $N \equiv -4 \bmod 13^2$, $N \equiv -8 \bmod 17^2$, $N \equiv -9 \bmod 19^2$ gives 11 consecutive nonsquarefree numbers.

If a_n is the largest squarefree number $\leqslant N$, we have by the preceding argument that $d_n \geqslant k$. On the other hand

$$N \leqslant \exp(2\theta(p_k)) \leqslant \exp(2p_k \ln 4) = 16^{p_k}$$

by (9) so that

$$p_k \geqslant \frac{\ln N}{\ln 16} \geqslant \frac{\ln a_n}{\ln 16}.$$

Also

$$k = \pi(p_k) \geqslant \frac{p_k \ln 2}{\ln p_k}$$

for $k \geqslant 3$ (see, for example, Theorem 2 of Nair [**12**]—a weaker version of the left hand inequality in (6) that is much easier to prove). Hence if k is large enough,

$$d_n \geqslant \frac{1}{4}\frac{\ln a_n}{\ln(\ln a_n/\ln 16)} > \frac{1}{5}\frac{\ln a_n}{\ln\ln a_n} \tag{18}$$

for infinitely many values of n. I know of no result qualitatively superior to (18). It is rather easy to show (Hardy and Wright [**7**], theorem 333) that for the sequence of squarefree numbers,

$$A(x) \sim \frac{6}{\pi^2}x \quad \text{as} \quad x \to \infty, \tag{19}$$

and that

$$\left|A(x) - \frac{6}{\pi^2}x\right| \leqslant 2x^{1/2}. \tag{20}$$

On the other hand, to improve on (20) in quality, however slightly, is very hard indeed, and nothing significantly better than $x^{1/4}$ on the right of (20) can be true.

In marked contrast, I shall give a simple argument, due to Davenport (see Roth [**17**]), to show that for the squarefree numbers,

$$h(x) = 117x^{1/3}(\ln x)^{-2/3}$$

is admissible. There is a more complicated but beautiful, and still elementary, argument due to Roth [**17**] which (when combined with a modification of Nair [**13**]) shows that a certain constant multiple of $x^{1/4}$ is admissible. Roth [**17**] showed also that his method, in conjunction with other ideas, leads to exponents smaller than $\frac{1}{4}$.

Let $m = a_n$ be the nth squarefree number and suppose that $m+1, m+2,\ldots,m+h$ are all non-squarefree. Since, as we have seen, there is a prime between m and $2m$, we have $h < m$. By hypothesis there are primes $p_1,\ldots,p_h$ such that

$$m+1 = p_1^2 k_1, \ldots, m+h = p_h^2 k_h. \tag{21}$$

Let X be a parameter, to be chosen later, such that $X > \frac{1}{2}h$. We may suppose both m and h to be large.

The number of numbers in the sequence (21) with $p_i \leqslant h^{1/2}$ is at most

$$\sum_{\substack{m<p^2k\leqslant m+h \\ p<h^{1/2}}} 1 = \sum_{p<h^{1/2}} \left(\left[\frac{m+h}{p^2} \right] - \left[\frac{m}{p^2} \right] \right) \leqslant \sum_{p\leqslant h^{1/2}} \left(\frac{h}{p^2} + 1 \right)$$

$$\leqslant h \sum_{l=2}^{\infty} \frac{1}{l^2} + h^{1/2} = h\left(\frac{\pi^2}{6} - 1 \right) + h^{1/2} < \frac{2}{3}h + h^{1/2} < \frac{3}{4}h$$

since h is large ($h > 144$).

Next, consider the numbers (21) with $p_i > h^{1/2}$. No two of these numbers can have $p_i = p_j (i \neq j)$. Otherwise, if $p_i^2 k_i$ and $p_i^2 k_j$ are two of the numbers (21) with $k_i \neq k_j$ and $p_i^2 > h$, we have

$$h < |p_i^2(k_i - k_j)| = |p_i^2 k_i - p_j^2 k_j| < h,$$

a contradiction. Hence the number of numbers (21) with $h^{1/2} < p_i \leqslant X$ is no larger than the number of primes between $h^{1/2}$ and X, that is, at most $\pi(X)$.

Finally consider the numbers (21) with $p_i > X$. For these, the corresponding k_i's must be distinct. For $k_i = k_j$ $(i \neq j)$ would imply

$$h > |p_i^2 k_i - p_j^2 k_j| = |p_i^2 - p_j^2| k_i \geqslant (p_i + p_j)|p_i - p_j| \geqslant 2X,$$

contrary to the choice of X. Since each k_i associated with a $p_i > X$ is less than $(m+h)/X^2 < 2m/X^2$, it follows that the number of these numbers in (21) is at most $2mX^{-2}$.

Altogether, then, we have

$$h < \frac{3}{4}h + \pi(X) + \frac{2m}{X^2},$$

or

$$X^2 h < 4X^2 \pi(X) + 8m.$$

Now take

$$X = \frac{1}{100} h \ln h.$$

Since $\pi(x) < 2x/\ln x$ for all large enough x (this is an easy consequence of (9)), we have

$$4X^2 \pi(X) < \frac{8X^3}{\ln X} < \frac{8}{10^6} h^3 \ln^2 h,$$

provided $\ln h > 100$, so that

$$\frac{1}{10^4} h^3 \ln^2 h < \frac{8}{10^6} h^3 \ln^2 h + 8m$$

or

$$h^3 \ln^2 h < \frac{8 \times 10^6}{92} m < 10^5 m.$$

Now either $h \leqslant m^{1/4}$, an even better result than claimed; or $h > m^{1/4}$, in which case $\ln^2 h > \frac{1}{16} \ln^2 m$, and so

$$h^3 < 16 \times 10^5 \frac{m}{\ln^2 m}.$$

Hence

$$h < 117 m^{1/3} (\ln m)^{-2/3}.$$

$\mathfrak{B}$-free numbers

One last comment. Erdős proposed the following generalization of the question studied in the previous section. Squarefree numbers are integers having no integer square divisors. Replace the role of the sequence of squares of primes $\{b_i = p_i^2;\ i = 1,2,3,\dots\}$ by any integer sequence

$$\mathfrak{B}: b_1 < b_2 < b_3 < \cdots,$$

with $1 < b_1$ and satisfying only

$$\text{(i)} \quad \sum_{i=1}^{\infty} \frac{1}{b_i} \text{ converges,}$$

$$\text{(ii)} \quad \gcd(b_i, b_j) = 1, i \neq j.$$

The sequence $\mathfrak{A}$ of $\mathfrak{B}$-free integers is then the sequence of all natural numbers none of which is divisible by an element of $\mathfrak{B}$. In 1973 Szemeredi [**19**] established a gap theorem for $\mathfrak{A}$ with $h(x) = x^{1/2+\varepsilon}$, $x \geqslant x_0(\varepsilon)$, by an argument which, while much more complicated, resembles in structure Davenport's proof.

This article is based on an invited lecture given at the MAA meeting in Pittsburgh, August 1981.

I am very grateful to the many helpful comments by referees on an earlier draft. Most of these have been incorporated in the text.

References

For general reading and background, as well as bibliography, see the book of Hardy and Wright cited below and also the following:

R. K. Guy, Unsolved Problems in Number Theory (especially Chapter A), Springer, 1982.

H. Halberstam and K. F. Roth, Sequences, Oxford, 1966, Springer, 1983.

D. Shanks, Solved and Unsolved Problems in Number Theory, Chelsea, 2nd ed., 1978.

[1] R. P. Bambah and S. D. Chowla, On numbers which can be expressed as sums of two squares, Proc. Nat. Inst. Sci. India, 13 (1947) 101–103.

[2] R. P. Brent, The first occurrence of large gaps between consecutive primes, Math. Comp., 27 (1973) 959–963.

[3] P. L. Čebyčev, Mémoire sur les nombres premiers, Oeuvres I, 1952, pp. 51–70.

[4] H. Cramér, On the order of magnitude of the difference between consecutive prime numbers, Acta Arith., 2 (1957) 23–46.

[5] P. Erdős and W. H. I. Fuchs, On a problem of additive number theory, J. London Math. Soc., 31 (1956) 67–73.

[6] G. H. Hardy, On Dirichlet's divisor problem, Proc. London Math. Soc., 15 (1916) 1–25.

[7] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 5th ed., Oxford, 1979.

[8] D. R. Heath-Brown and H. Iwaniec, On the difference between consecutive primes, Invent. Math., 55 (1979) 49–69.

[9] H. Iwaniec, The half-dimensional sieve, Acta Arith., 29 (1976) 69–95.

[10] W. J. LeVeque, Topics in Number Theory, vol. 2, chapter 7–5, Addison-Wesley, 1956.

[11] L. Mirsky, Note on an asymptotic formula connected with r-free numbers, Quart. J. Math. Oxford Ser. 18 (1947) 178–182.

[12] M. Nair, On Chebyshev-type inequalities for primes, Amer. Math. Monthly, 89 (1982) 126–129.

[13] _____, Power-free values of polynomials, II, Proc. London Math. Soc., (3) 38 (1979) 353–368.

[14] R. A. Rankin, The difference between consecutive primes, J. London Math. Soc., 13 (1938) 242–247.

[15] I. Richards, On the gaps between numbers that are sums of two squares, Adv. in Math., 46 (1982) 1–2.

[16] J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, Illinois J. Math., 6 (1962) 64–94.

[17] K. F. Roth, On the gaps between squarefree numbers, J. London Math. Soc., 26 (1951) 263–268.

[18] W. Sierpinski, O pewnem zagadnienise z rachunku funkcyj asymptotycznych, Prace Mat. Fiz., 17 (1906) 77–118.

[19] E. Szemeredi, On the difference of consecutive terms of sequences defined by divisibility properties, II, Acta Arith., 23 (1973), 359–361.

[20] C. W. Trigg, J. Recreational Math., 14 (1981–82) 204.

Tables of Primes

[21] D. N. Lehmer, List of prime numbers from 1 to 10,006,721, Carnegie Institute, Washington, Pub. 165 (1914).

Approximating Any Regular Polygon by Folding Paper

An interplay of geometry, analysis and number theory.

Peter Hilton
ETH
Zürich, Switzerland

Jean Pedersen
University of Santa Clara
Santa Clara, CA 95053

The ancient question as to which regular polygons are constructible with a ruler and compass was answered by Gauss, but the search for actual constructions has continued to consume the energy of mathematicians. From *The World of Mathematics* [**1**, p. 502] we read:

> By 350 B.C. the Greeks knew Euclidean constructions for the regular polygons of 4, 8, 16,... sides and for those of 3 and 5 sides—the equilateral triangle and the regular pentagon. From these it was easy to construct regular polygons of $2^c \times 3$, $2^c \times 5$, $2^c \times 3 \times 5$ sides, where c is any positive integer, and the Greeks in effect showed how. They got no farther. Young Gauss proved that if N is of the form 2^c or 2^c times a product of *different Fermat primes* F_n, then there is a Euclidean construction for a regular polygon of N sides. This form of N is both necessary and sufficient for the possibility of a Euclidean construction. ...
>
> ... Simple Euclidean constructions for the regular polygons of 17 and 257 sides are available, and an industrious algebraist expended the better part of his years and a mass of paper in attempting to construct the F_4 regular polygon of 65,537 sides. The unfinished outcome of all this grueling labor was piously deposited in the library of a German university. ...

In sum, there is a Euclidean construction of the regular N-gon for very few values of N, and, even for these N, we do not in all cases know explicit constructions. In this article we plan to show how one can construct an *approximation* (to any desired degree of accuracy) to a regular N-gon for *any* value of N; and we will, moreover, give explicit and uncomplicated constructions involving only the folding of a straight strip of paper (like adding machine tape), in a prescribed manner. Since we wish to keep the construction as practical as possible, we also look at the question of how to use as few folds as possible.

We first give some simple examples of the folding processes we use. These will illustrate how to approximate regular polygons having either $2^n + 1$ or $2^n - 1$ sides. If the algebraist referred to in the quotation were still alive he could, by our method, obtain an approximation of his long sought 65,537-gon (since $F_4 = 2^{2^4} + 1 = 2^{16} + 1 = 65{,}537$). It would still require a huge mass of paper and a good deal of time but, in theory, it could be done.

The general discussion will lead quite naturally to the introduction of a set of natural numbers we call the **folding numbers**. As we will show, these numbers have quite remarkable and special properties, and are easily recognized by their base 2 representation. In our last section we describe some of these properties. These purely number-theoretic results enable us to prove our main result about approximating regular N-gons. In addition, the last section includes a few results on folding numbers and their generalizations not strictly relevant to our main theorem. Thus we hope to illustrate, through this work, how various branches of mathematics are interwoven, and how concrete geometrical considerations may lead to theoretical mathematics of an apparently very different nature, as well as the interest to be derived from relating theoretical facts to physical interpretations.

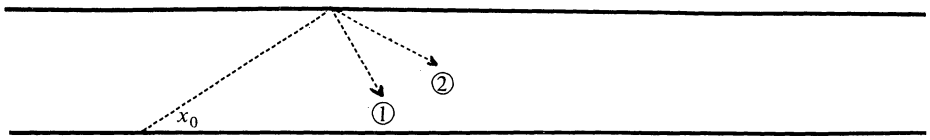


FIGURE 1

The folding procedure

Let us now turn to our two special kinds of examples. In both cases we begin with a strip of paper with parallel edges and one fold line producing an arbitrary acute angle x_0 (see FIGURE 1). In the first case we successively fold twice at each new obtuse angle created in the right-hand direction of the tape. Here, and in all subsequent primary folding processes, each new fold must *bisect* the angle formed by the last fold line and an edge of the tape. After several of these folds have been made, the tape should appear as shown in FIGURE 2(a), where the numbers along the transversals indicate the order in which the fold lines were made. One may easily see that the smallest angle on this tape is rapidly approaching $\pi/5$ (for example, call the initial angle $(\pi/5)+\varepsilon$ and solve for the next acute angle that is produced on the tape by the folding). If we cut off the first irregular pieces, say the first ten sections, we can then use the folded tape to produce the pentagons shown in FIGURE 2, (b) and (c). The solid pentagon is formed by folding the tape only on the short transversals, leaving the long transversals flat; and the "hollow" pentagon is formed by folding the tape only on the long transversals and leaving the short transversals flat.

If we had made three successive folds at each vertex, the tape would appear as shown in FIGURE 3(a). The smallest angle on this tape approaches $\pi/(2^3+1)=\pi/9$, and after cutting off the first irregular part of the strip we could construct a regular 9-gon (FIGURE 3(b)) by folding on all the lines that make an angle of $\pi/9$ with an edge of the tape. This idea generalizes; so that, if we fold n times at each successive vertex, on both the top and bottom edge of the tape, the smallest angle rapidly approaches $\pi/(2^n+1)$ (a general proof appears in [2]). We may always build the (2^n+1)-gon from this tape by throwing away the first irregular part of the strip and then folding only on the transversals that make an angle of $\pi/(2^n+1)$ with an edge of the tape. These folding processes will be denoted $\{d^n u^n\}$, meaning that we always fold n times, *d*own at the top of the tape and *u*p at the bottom.

We have seen that when $n=3$ in the folding $\{d^n u^n\}$ we can approximate a regular 9-gon (which cannot be constructed with a ruler and compass). When $n=4$ we can approximate the regular 17-gon (which would surely have been of interest to Gauss, whose sensational discovery of

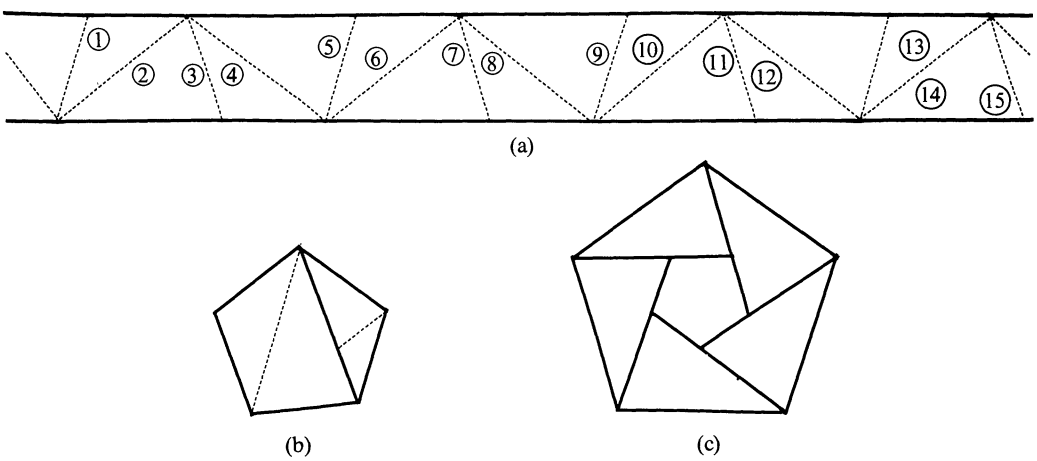


FIGURE 2

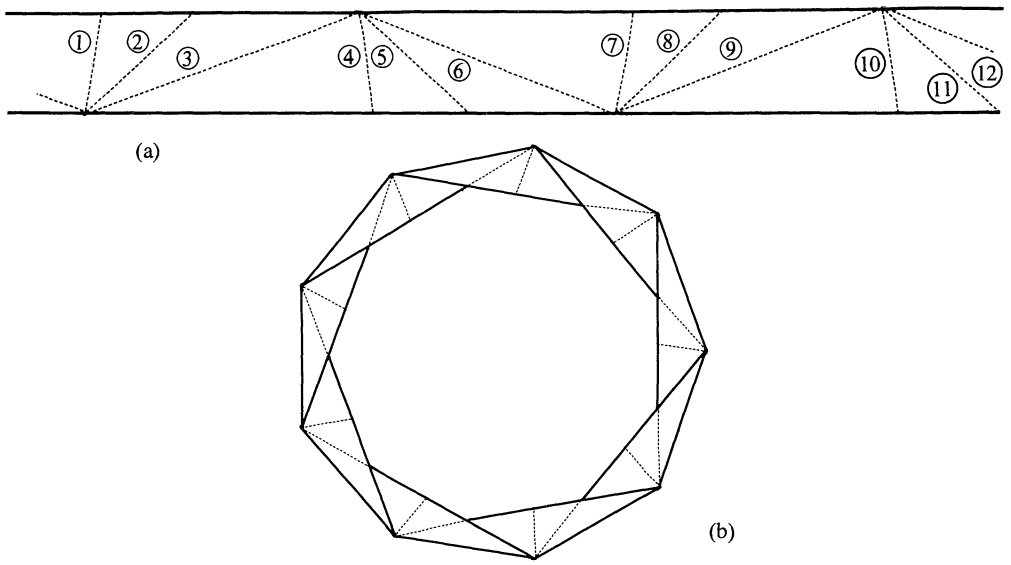


FIGURE 3

the ruler and compass construction of this figure is legendary); and when $n = 16$ we get our German algebraist's elusive 65,537-gon. But here we should note that, though the theory is fine, as a practical matter if we used tape 5 cm wide, each side of the 65,537-gon would be 1043m long, giving a perimeter of 68,358.6 kilometers! Thus we do not claim that our construction is in all cases practicable—but this is scarcely surprising in a construction claimed to be valid for *all* natural numbers N. In this respect, we are no worse off than with ruler and compass constructions; though, of course, where such a construction is possible, it has the aesthetic advantage that the error in executing the construction is entirely experimental.

Now we consider the second special type of folding. We begin, as before, with an initial angle of x_0, but this time we repeatedly fold down 2 times and up once. This is denoted $\{d^2u\}$. The folded tape is shown in FIGURE 4(a). The smallest angle on this tape can be shown to approach $\pi/(2^3 - 1) = \pi/7$ and we can throw away the first irregular part of this tape and use the subsequent portion to construct an approximation to a regular 7-gon. A way to do this is to fold

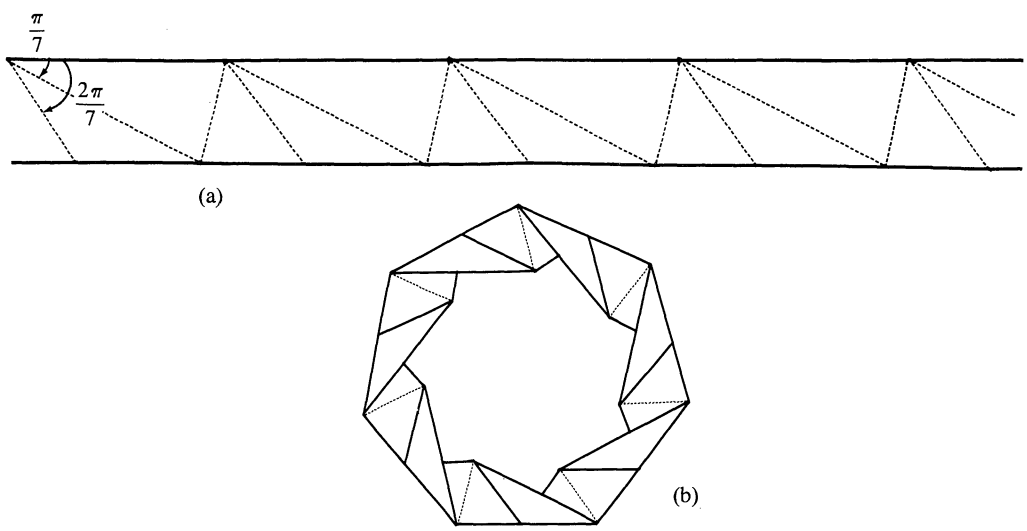


FIGURE 4

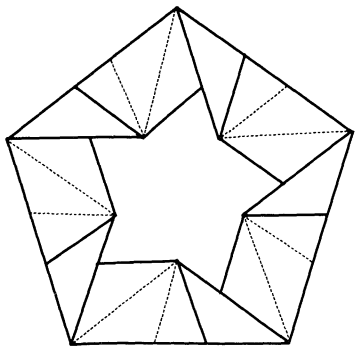


FIGURE 5

only on the transversals that make an angle of $\pi/7$ and $2\pi/7$ with the *top* edge of the tape. FIGURE 4(b) shows how the resulting 7-gon looks. One may experiment with this tape and discover that there are *other* ways of folding it so as to locate the vertices of a regular 7-gon (some arrangements give lovely star polygons)—but these other ways do not generalize as easily as this particular method.

For such a generalization, we repeatedly fold down m times and up once, denoted $\{d^m u\}$. The smallest angle on this tape will approach $\pi/(2^{m+1}-1)$ and we can use this tape to construct regular $(2^{m+1}-1)$-gons by throwing away the first irregular part of the strip and then folding only on the transversals that make an angle with the top edge of the tape of $\pi/(2^{m+1}-1)$ and $2\pi/(2^{m+1}-1)$.

It is interesting to note that this folding procedure may also be applied to the tapes we discussed earlier. If, for example, it is applied to the tape we used to construct pentagons, that is, if we fold only on the transversals that make an angle of $\pi/5$ and $2\pi/5$ with the *top* edge of the tape, we produce the regular 5-gon shown in FIGURE 5.

These fairly simple constructions, and the knowledge of how difficult ruler and compass constructions are for many regular N-gons, have motivated the following question: *Is there a reasonably simple common generalization of these two constructions which will enable us to approximate ANY regular N-gon*?

We show that the answer to this question is YES! The analytical part of the answer appears in [2]. The number-theoretical contribution is contained in this paper.

Let $\{d^m u^n\}$ designate the procedure of successively folding the tape down m times and up n times, $m \geq n \geq 1$. Thus $\{d^m u^n\}$ is a common generalization of the two procedures described earlier. By a folding procedure we will henceforth mean a $\{d^m u^n\}$-procedure. It is shown in [2] that the smaller angle u_k (which is the small angle at the top edge of the tape after carrying out the folding procedure k times) tends to the measure

$$\frac{2^n-1}{2^{m+n}-1}\pi.$$

Indeed, if the initial error for the small angle at the top was ε, that is,

$$\left|u_0 - \frac{2^n-1}{2^{m+n}-1}\pi\right| = \varepsilon,$$

then the error at the kth stage is $\varepsilon 2^{-(m+n)k}$, that is,

$$\left|u_k - \frac{2^n-1}{2^{m+n}-1}\pi\right| = \frac{\varepsilon}{2^{(m+n)k}}.$$

A similar statement holds for v_k, the small angle at the bottom of the tape at the kth stage. In other words, each stage reduces the error by a factor of 2^{m+n}.

A remark is appropriate on the different natures of the ruler and compass constructions on the one hand and our approximate constructions on the other. As everyone who has carried out ruler and compass construction knows, these constructions are only "perfect in the mind." Although the folding sequences mentioned here can never guarantee perfection, they are *convergent* sequences, and as long as we fold *approximately correctly*, each fold will produce a better approximation to the limit angle than its predecessor. On the other hand, as we all know from experience, ruler and compass constructions frequently *diverge*, the accuracy of the final result often being a function of how recently we've sharpened our pencil!

Folding numbers and the main theorem

We say that s is a **folding number** if s is a positive integer and if there is a folding procedure $\{d^m u^n\}$ such that the smaller angle u_k tends to the measure π/s. We denote by $\mathfrak{F}$ the set of all folding numbers. Thus $\mathfrak{F}$ is the set of all integers which can be expressed in the form $(2^{m+n}-1)/(2^n-1)$ with $m \geqslant n \geqslant 1$. We will study the properties of $\mathfrak{F}$ in our last section and merely refer to these as needed for the present discussion.

We begin by looking at a particularly useful way to construct a regular polygon from a straight strip of paper (with parallel edges). Since the interior angle at any vertex of a regular P-gon is $\pi - (2\pi/P)$, a straight piece of paper can be folded to produce a regular P-gon if, through vertices (located equal distances from each other along the top edge of the tape) there are fold lines sloping, say, downwards to the right, which make angles of π/P and $2\pi/P$ with the top edge of the tape. This statement may be verified by looking at FIGURE 6, where part (a) shows a general piece of tape on which only the fold lines making angles of π/P and $2\pi/P$ with the top edge of the tape are shown. The other primary and secondary fold lines that may have been made on the tape to reach this point are ignored at this stage in the construction—so we omit drawing them. The P-gon is constructed by simply folding on each of the fold lines that are shown. The effect of this folding at any particular vertex, say A_1, is to reduce the angle at that vertex by $2\pi/P$ (thus making the angle at A_1 equal to $\pi - (2\pi/P)$) and, simultaneously, to bring the line segment $\overline{A_1A_2}$ into a position where it forms the next side of the polygon. FIGURE 6(b) shows a portion of the resulting regular P-gon. Notice that it is not necessary to fold on *all* of the lines shown. For example, a larger regular P-gon with sides of length $2A_1A_2$ can be produced by folding only on the fold lines that pass through the A's with odd subscripts.

This general method of constructing a regular polygon from a straight strip is employed in the following theorem. Theorems about the set $\mathfrak{F}$ are referred to in the proof; these can be found in our next section.

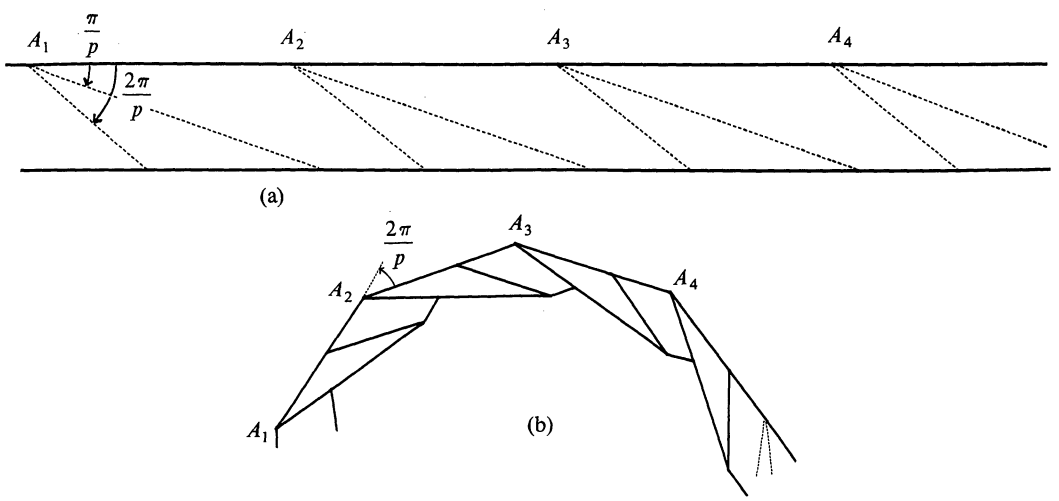


FIGURE 6

PAPER-FOLDING THEOREM ON APPROXIMATING REGULAR N-GONS. *A straight strip of paper (with parallel edges) may be folded to approximate, to any desired degree of accuracy, a regular N-gon, for any $N \geq 3$. Specifically, N will belong to at least one of the following categories:*

(i) $N \in \mathcal{F}$; *in which case a regular N-gon may be approximated from $\{d^{x(y-1)}u^x\}$-folded tape, where (x, y) are the coordinates of N as in* TABLE 1.

(ii) *There exists $s_N \in \mathcal{F}$ such that $N = 2^k s_N$, $k > 0$; in which case a regular N-gon may be approximated from tape folded first by a primary folding $\{d^{x(y-1)}u^x\}$, where (x, y) are the coordinates of s_N as in* TABLE 1, *followed by secondary folds that divide the smallest angle at the top of the $\{d^{x(y-1)}u^x\}$-folded tape into 2^k equal parts.*

(iii) *There exists $s_N \in \mathcal{F}$ such that $2^k s_N = qN$, $k \geq 0$, $q > 1$; in which case a regular qN-gon may be approximated, using* (i) *or* (ii), *according as $k = 0$ or $k > 0$. This construction is completed by gluing the qN-gon to another piece of paper and folding along lines that connect every qth vertex.*

(iv) $N = 2^k$, *where $k \geq 2$; in which case the N-gon may be constructed by an exact folding process.**

Proof. We first argue that every $N \geq 3$ is included in at least one of the four cases. Then we give specific instructions for how each case may be executed.

Suppose $N \geq 3$ and N is odd. Then by Theorem 7 we know that N is a factor of some element of $\mathcal{F}$. Thus case (i) or case (iii) applies.

If $N > 3$ and N is an even number which is not a power of 2, then N may be written in the form $2^k a$, $k \geq 1$, $a > 1$, a odd; then, again by Theorem 7 we know that a is a factor of some element of $\mathcal{F}$. So case (ii) or case (iii) applies here.

If $N > 3$ and is of the form $N = 2^k$, with $k \geq 2$, case (iv) applies.

Now, since each $N \geq 3$ is included in at least one of the four cases listed, the theorem will be proved if we can show that each case yields tape from which we can construct the required polygon.

A case (i) construction involves the special case when we know that $N \in \mathcal{F}$. The construction is carried out by first simply folding the tape as indicated in the Theorem to produce crease lines. To complete the construction lay the tape flat and then fold on the longest and second longest transversals (crease lines) that make angles of π/N and $2\pi/N$ respectively with the *top* edge of the tape. For example, $N = 5$ has coordinates (2,2) in TABLE 1 so we first fold the $\{d^{2(2-1)}u^2\}$ or $\{d^2u^2\}$-folded tape, producing the crease lines shown in FIGURE 2(a). The construction is completed by folding this strip of paper on the longest and second longest transversals that make angles of $\pi/5$ or $2\pi/5$ with the top edge of the tape. The resulting regular 5-gon is shown in FIGURE 5.

A case (ii) construction involves first finding the coordinates (x, y) of $s_N \in \mathcal{F}$ and folding the tape by means of the $\{d^{x(y-1)}u^x\}$ procedure. Then that tape must be folded again so that each of the angles between the longest transversal and the *top* edge of the tape (which will be approaching π/s_N) is divided, by folding, into 2^k equal parts. The final result is a strip of tape on which the longest and second longest transversals from the top edge of the tape make angles of $\pi/2^k s_N$ and $\pi/2^{k-1}s_N$ respectively with the top edge of the tape. FIGURE 7(a) illustrates how the fold lines look on the tape when $N = 10$; so that $s_N = 5$ and $k = 1$. The primary folds for $\{d^2u^2\}$ are dotted lines (-------) and the secondary folds producing the desired angle of $\pi/10$ with the top edge of the tape are dot-dashed lines (·-·-·-·-). FIGURE 7(b) shows a portion of the resulting regular 10-gon.

Once we know how to do case (i) and case (ii) constructions, the case (iii) construction is described by the Theorem (FIGURE 8 illustrates the construction). In FIGURE 8, we ignore the fact that $5 \in \mathcal{F}$ and instead exploit the fact that $15 = 3 \times 5$, with $15 \in \mathcal{F}$. Thus 15 has coordinates (1,4) so that the $\{d^3u\}$-folding will produce tape that can be used to approximate a 15-gon. This may be glued to a piece of paper so that folding through every third vertex (say, through

*An exact construction is considered to be an especially good approximation!

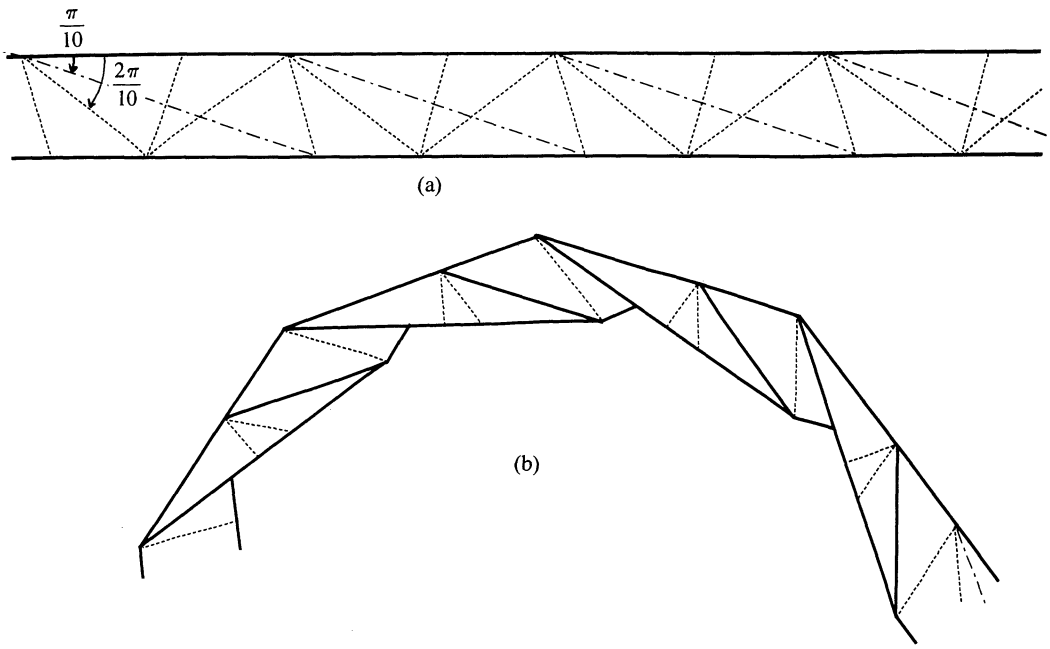


FIGURE 7. (a) $\{d^2u^2\}$-tape with a secondary fold. (b) Folding the tape shown in (a) on *all* the lines that make an angle of $\frac{\pi}{10}$ or $\frac{2\pi}{10}$ with the *top edge* of the tape produces a regular 10-gon. This figure shows four sides of the resulting polygon.

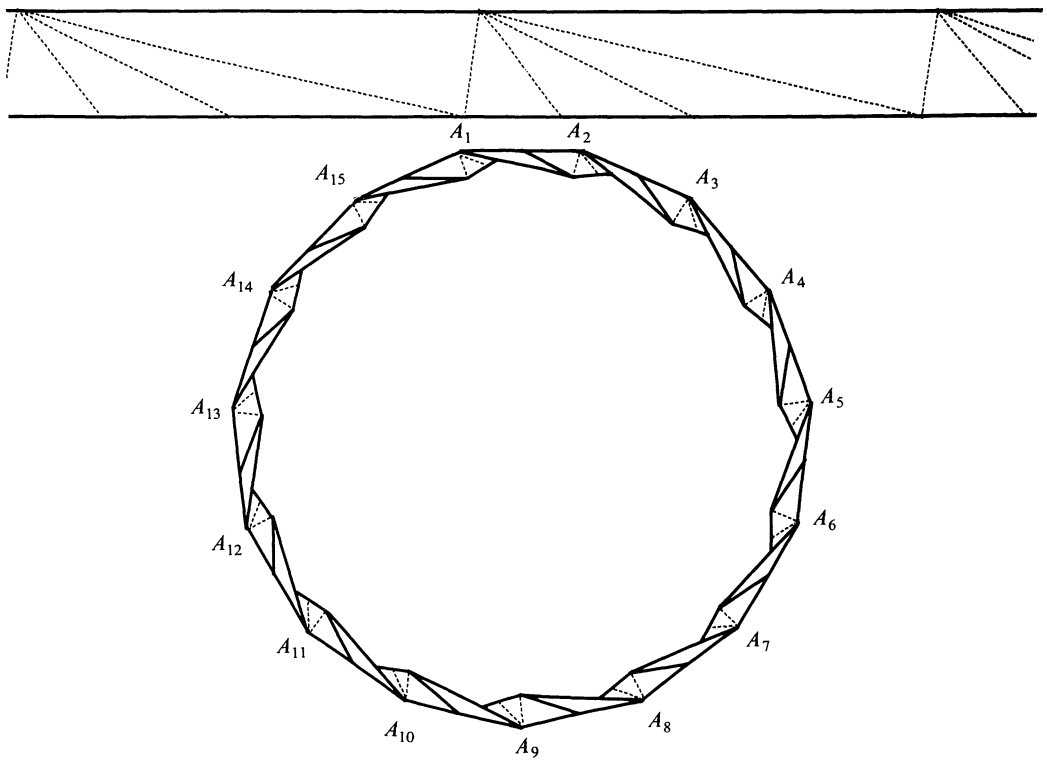


FIGURE 8

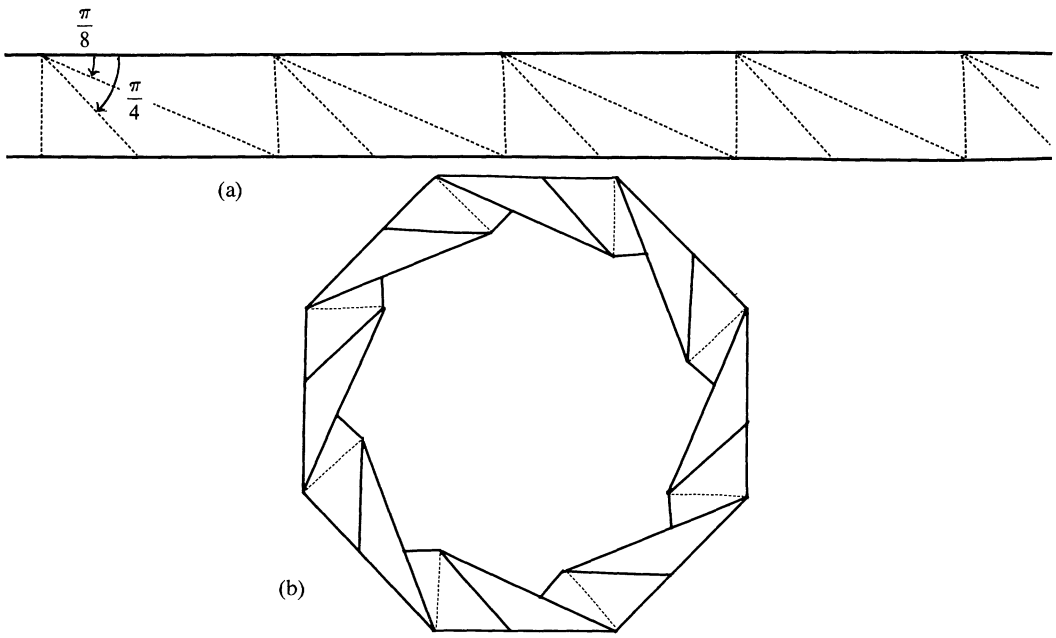


FIGURE 9

$\overline{A_1A_4}, \overline{A_4A_7}, \overline{A_7A_{10}}, \overline{A_{10}A_{13}}, \overline{A_{13}A_1}$) produces the desired 5-gon. This is not, as we have seen from the preceding example, the most efficient way to construct a 5-gon "from scratch," but it serves as an illustration of the type (iii) process.

Case (iv) constructions are achieved by folding an angle of $\pi/2$ at some point on the tape and then repeatedly bisecting the angle at the top of the tape to form angles of $\pi/2^2, \pi/2^3, \ldots, \pi/2^k$. We then follow the last transversal to the point where it intersects the bottom edge of the tape and repeat the folding process, making the $\pi/2$ fold through that point. Folding this tape on the transversals that make angles of $\pi/2^k$ and $\pi/2^{k-1}$ with the top edge of the tape will produce the desired 2^k-gon. FIGURE 9 illustrates this process for $N = 8$, so that $k = 3$.

We now make some observations about some interesting geometry that occurs on the $\{d^m u\}$-folded strip of paper (or tape). These concern what happens on the folded tape when we make secondary folds to produce tape suitable for approximating $2(2^{m+1} - 1)$-gons. FIGURE 10 shows a typical section (if we ignore the dot-dashed line - - - -) of the $\{d^m u\}$-tape where, according to the calculation given earlier, we know that, in the limit,

$$\angle DEC = \angle DBC = \frac{\pi}{2^{m+1} - 1}$$

and

$$\angle BAC = \angle ABC = \angle BDC = \angle BCD = \angle EDC = \angle ECD = \frac{(2^m - 1)\pi}{2^{m+1} - 1}.$$

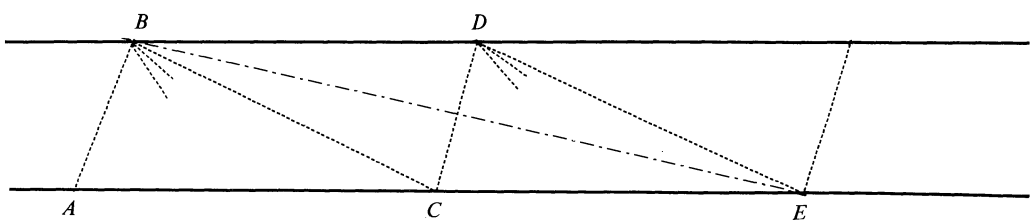


FIGURE 10

We can produce a $2(2^{m+1}-1)$-gon from this tape if we bisect every other small angle at the top of the tape. But, since $BDEC$ is a rhombus, the bisector of $\angle DBC$ will always pass through the point E. This is a useful fact to know when folding the paper, because it helps us to be more accurate with this secondary fold.

A second observation about this particular tape is that

$$\angle ABE = \angle ABC + \frac{1}{2}\angle CBD = \frac{(2^m-1)\pi}{2^{m+1}-1} + \frac{1}{2}\left(\frac{\pi}{2^{m+1}-1}\right) = \frac{\pi}{2},$$

always! This has a very nice practical consequence. It means that whenever we have tape of this type (i.e., that produces regular $2(2^{m+1}-1)$-gons) we can fold it along the shortest and longest fold lines and it will assume the shape of a rectangle. This makes the tape for this kind of polygon particularly easy to store.

We close by indicating how one may go through an actual procedure for constructing an approximation to a regular N-gon by paper-folding. First, factor N as $2^k a$ with a odd. If $a = 1$, use construction (iv). If $a > 1$, test to see if $a \in \mathfrak{F}$ by writing a in base 2. If $a \in \mathfrak{F}$ then it will assume the form (3) (next section) in base 2 so that one writes down the coordinates $x = p(a)$, $y = e(a)$ and then uses construction (i) or (ii), according as $k = 0$ or $k \geq 1$, with the $\{d^{x(y-1)}u^x\}$-procedure. If $a \notin \mathfrak{F}$ then Theorem 8 implies that $a|(2^{\mu(a)}-1)$, where one calculates $\mu(a)$ from formula (8)—of course, $k_0 = 0$ in our case. Suppose then that $2^{\mu(a)} - 1 = qa$. Then $2^k s_N = qN$ with $s_N = 2^{\mu(a)} - 1$. Thus the coordinates of s_N are $(1, \mu(a))$ and we are ready to apply construction (iii).

As emphasized previously, we are concerned with *efficient* folding procedures. Otherwise we could confine attention to $\{d^m u\}$-procedures. It is thus natural to ask: *What is the optimal primary folding procedure $\{d^m u^n\}$ to produce a regular N-gon for a given N, in the sense that $m+n$ is minimal*? Here we may assume N odd. If $N \in \mathfrak{F}$ then this is certainly the procedure given by construction (i). However, if $N \notin \mathfrak{F}$, then the procedure given by the strategy of the previous paragraph, with $m + n = \mu(N)$, is very likely not to be optimal. For example, if $N = 43$, then $\mu(N) = 42$; however, we could exploit the fact that $43|129$ and $129 \in \mathfrak{F}$ with coordinates $(7,2)$, to find a primary folding procedure $\{d^7 u^7\}$ with $m + n = 14$. Perhaps the reader will be motivated to find an answer to this open question.

Another open question is: *Does there exist a general procedure for folding regular N-gons that have all of their interior covered with paper*? (We described such a folding procedure for a pentagon, FIGURE 2(b).)

Properties of $\mathfrak{F}$, the set of folding numbers

We have defined the set $\mathfrak{F}$ of **folding numbers** to be the set of *integers* s, expressible in the form

$$s = \frac{2^{m+n}-1}{2^n-1}, \tag{1}$$

with $1 \leq n \leq m$. Some of the properties of $\mathfrak{F}$ were necessary to prove our main theorem; these and other properties of $\mathfrak{F}$ are explored in this section.

From a purely number-theoretic point of view, it is natural to generalize the definition of $\mathfrak{F}$ by defining, for any integer $t \geq 2$, the set $\mathfrak{F}_t$ to consist of all integers expressible as $(t^{m+n}-1)/(t^n-1)$, with $1 \leq n \leq m$; plainly $\mathfrak{F} = \mathfrak{F}_2$. The generalization is achieved gratis, since $\mathfrak{F}_t$ is as easy to study as $\mathfrak{F}_2$; but we thereby obtain results unavailable if we confine our attention to $\mathfrak{F}_2$.

We first determine under what circumstances $(t^{m+n}-1)/(t^n-1)$ is, in fact, an integer. This requires two elementary facts on greatest common divisors.

PROPOSITION 1. *Let a, b, k be positive integers, and $c = \gcd(a, b)$. Then*

(i) $\gcd(ka, kb) = kc$, *and*
(ii) *For all $t \geq 2$,* $\gcd(t^a - 1, t^b - 1) = t^c - 1$.

Proof. (i) Obvious.

(ii) Since $c|a$ and $c|b$, it is plain that $(t^c-1)|(t^a-1)$ and $(t^c-1)|(t^b-1)$. Also, there are integers r, s such that $c=ra+sb$, so $t^c-1=t^{ra+sb}-1=t^{sb}(t^{ra}-1)+(t^{sb}-1)$. This shows that t^c-1 is a linear combination of t^a-1 and t^b-1, so that $t^c-1=\gcd(t^a-1, t^b-1)$.

COROLLARY. *$(t^b-1)|(t^a-1)$ if and only if $b|a$.*

From the corollary, it follows that *$(t^{m+n}-1)/(t^n-1)$ is an integer precisely when $n|m$.* Moreover, we see that $(t^{m+n}-1)/(t^n-1)$ can only be an integer if $n \leqslant m$, so that it is unnecessary to refer to this restriction in discussing $\mathfrak{F}_t$.

THEOREM 1. *There is a one-one correspondence between pairs of positive integers m, n with $n|m$ and the elements of $\mathfrak{F}_t$, given by*

$$(m, n) \to \frac{t^{m+n}-1}{t^n-1}. \tag{2}$$

Proof. All that remains to be proved is that the given mapping is one-one. We do this by writing elements of $\mathfrak{F}_t$ in base t. We know that the elements of $\mathfrak{F}_t$ are integers s of the form $(t^{nq}-1)/(t^n-1)$ with $q \geqslant 2$, and

$$\frac{t^{nq}-1}{t^n-1} = 1 + t^n + t^{2n} + \cdots + t^{(q-1)n}.$$

If we write this in base t we get a representation of s as a string of 0's and 1's:

$$\underbrace{10\ldots0}\,\underbrace{10\ldots0}\,\underbrace{10\ldots0}\ldots\underbrace{10\ldots0}\,\underbrace{10\ldots0}\,\underbrace{10\ldots0}1, \tag{3}$$

where the repeating block $10\ldots0$ is of length n and consists of 1 followed by $(n-1)$ 0's; and where there are q 1's in the entire string. We call n the **period** and q the **extent** of this representation. Since the expression of any integer in base t is unique, it follows that we may attach to each element s of $\mathfrak{F}_t$ its period $p(s)$ and its extent $e(s)$. Moreover

$$p\left(\frac{t^{m+n}-1}{t^n-1}\right)=n,\ e\left(\frac{t^{m+n}-1}{t^n-1}\right)=\frac{m}{n}+1,$$

so that the correspondence given by (2) is certainly one-one.

Note that in the course of the proof we have characterized the elements of $\mathfrak{F}_t$ by their base t numeral representation. We will call $(n, q)_t$ the ***t*-coordinates** of $s=(t^{nq}-1)/(t^n-1)$. If $t=2$, we simply write (n, q) for $(n, q)_2$. We may also write $s=(n, q)_t$ if no confusion can result.

COROLLARY. *For each $s \in \mathfrak{F}$ there is precisely one folding procedure leading to the angular measure π/s, and such a folding procedure $\{d^m u^n\}$ must satisfy the condition $n|m$. In fact $n=p(s)$, $m=p(s)(e(s)-1)$, where $p(s)$ is the period of s, and $e(s)$ is the extent of s.*

TABLE 1 lists all the elements s of $\mathfrak{F}$ with $s<10^8$. Note that our definition of $\mathfrak{F}$ excludes the number 1, as it should if we are talking of folding numbers. However, it is sometimes convenient to include 1; we would then write $\overline{\mathfrak{F}}_t$ for $\mathfrak{F}_t \cup \{1\}$. Of course we cannot associate a unique period with the number 1, but we can say that it is of unit extent.

We next consider the following question, motivated by our study of folding procedures. Suppose we do not insist that $n|m$, and suppose that we express $(t^n-1)/(t^{m+n}-1)$ as a reduced fraction. What denominators can we get? In fact, we will show that the possible denominators are precisely the elements of $\mathfrak{F}_t$. To this end we introduce the idea of a **section** of an element of $\mathfrak{F}_t$. Thus a section of the given representation (3) of s is the number obtained by starting at any 1 and taking that 1 and all the digits to the *right*. It follows that a section of an element of $\mathfrak{F}_t$ always belongs to $\overline{\mathfrak{F}}_t$. Indeed if s_0 is a section of s, then

$$p(s_0)=p(s),\quad e(s_0) \leqslant e(s). \tag{4}$$

(We may assign to 1 *any* period.)

y	2^{y-1}													
26	67708863													
25	33554431													
24	16777215													
23	8389607													
22	4194303													
21	2097151													
20	1048575													
19	524287													
18	262143													
17	131071													
16	65535													
15	32767													
14	16383													
13	8191	22369621												
12	4095	5592405												
11	2047	1398101												
10	1023	349525												
9	511	87381	19173961											
8	255	21845	2396745											
7	127	5461	299593	17895697										
6	63	1365	37449	1118481	34636833									
5	31	341	4681	69905	1082401	17043521								
4	15	85	585	4369	33825	266305	2113665	16843009						
3	7	21	73	273	1057	4161	16513	65793	262657	1049601	4196353	16781313	67117057	
2	3	5	9	17	33	65	129	257	513	1025	2049	4097	8193	2^x+1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
y / x	1	2	3	4	5	6	7	8	9	10	11	12	13	$14 \leqslant x \leqslant 26$

TABLE 1. **The folding numbers, $s = \dfrac{2^{xy}-1}{2^x-1}$, up to 10^8. (An s-gon may be approximated by a $\{d^{x(y-1)}u^x\}$-folding.)**

Note: (i) for the folding number s, $p(s)=x$, $e(s)=y$; s is written as (x, y).
(ii) the folding number $(k,2)$ is just $(1, k)+2$, so that the entries $(x,2)$, $14 \leqslant x \leqslant 26$ can be read off the part of the table given.

We say that s_0 is a **proper section** if $e(s_0) < e(s)$. Note that (4) characterizes those elements s_0 of $\overline{\mathfrak{F}}_t$ which are sections of s.

THEOREM 2. *Let s_0 be a section of s. If $s_1 = \gcd(s, s_0)$, then s_1 is a section of s and hence in $\overline{\mathfrak{F}}_t$; moreover,*

$$p(s_1) = p(s),\ e(s_1) = \gcd(e(s), e(s_0)).$$

Thus $s_0 | s$ if and only if $e(s_0) | e(s)$.

Proof. Set $p(s) = p(s_0) = n$, $e(s) = q$, $e(s_0) = q_0$. Then

$$s = \frac{t^{nq} - 1}{t^n - 1},\ s_0 = \frac{t^{nq_0} - 1}{t^n - 1}.$$

Applying Proposition 1, we have that

$$\begin{aligned} s_1 = \gcd(s, s_0) &= \gcd\left(\frac{t^{nq} - 1}{t^n - 1}, \frac{t^{nq_0} - 1}{t^n - 1}\right) \\ &= \frac{\gcd(t^{nq} - 1, t^{nq_0} - 1)}{t^n - 1} \\ &= \frac{t^{nq_1} - 1}{t^n - 1} \end{aligned}$$

where $q_1 = \gcd(q, q_0) = \gcd(e(s), e(s_0))$. Thus $s_1 \in \overline{\mathfrak{F}}_t$ and

$$p(s_1) = n = p(s),\ e(s_1) = q_1 = \gcd(e(s), e(s_0)).$$

The following example illustrates Theorem 2.

EXAMPLE. Let $s = 1365$, $s_0 = 85$. Then, in base 2,

$$s = 10101010101, s_0 = 1010101.$$

Thus $s \in \mathfrak{F}$ and s_0 is a section of s. Moreover, $p(s) = p(s_0) = 2$, $e(s) = 6$, $e(s_0) = 4$. Theorem 2 asserts that if $s_1 = \gcd(s, s_0)$, then $s_1 \in \overline{\mathfrak{F}}$, $p(s_1) = 2$, $e(s_1) = \gcd(6, 4) = 2$. In other words $s_1 = 101$ in base 2, or $s_1 = 5$. It is easy to confirm that $\gcd(1365, 85) = 5$. The beauty of Theorem 2, exemplified here, is that the assertion

$$\gcd(10101010101, 1010101) = 101$$

is, in fact, true in *any* base (even base ten!).

Theorem 2 also enables us to write s/s_1 in base t. Indeed we may do this whenever s_1 is a section of s and $s_1 | s$. We find that

$$p\left(\frac{s}{s_1}\right) = p(s)e(s_1),\ e\left(\frac{s}{s_1}\right) = \frac{e(s)}{e(s_1)}.$$

Thus, in our Example, $p(1365/5) = p(273) = 2 \times 2 = 4$, $e(273) = \frac{6}{2} = 3$, or

$$273 = 100010001$$

in base 2, as may be easily verified.

We have the following corollary to Theorem 2.

COROLLARY. *Let s_0 be a section of s. Then the fraction $\frac{s_0}{s}$ is reduced if and only if $\frac{e(s_0)}{e(s)}$ is reduced.*

From this we deduce the answer to the question we posed.

THEOREM 3. *The reduced fraction forms of the rational numbers $(t^n - 1)/(t^k - 1)$, $n < k$, are precisely the fractions s_0/s where $s \in \overline{\mathfrak{F}}_t$ and s_0 is a proper section of s with $e(s_0)$ prime to $e(s)$. Moreover $\frac{e(s_0)}{e(s)} = \frac{n}{k}$.*

Proof. As we saw in the proof of Theorem 2,

$$\frac{s_0}{s} = \frac{t^{p(s)e(s_0)} - 1}{t^{p(s)e(s)} - 1}$$

and, by the Corollary, if $e(s_0)$ is prime to $e(s)$, the fraction s_0/s is reduced. Conversely, if $h = \gcd(k, n)$ then it follows by Proposition 1 that $(t^n - 1)/(t^k - 1)$ reduces completely to

$$\frac{(t^n - 1)/(t^h - 1)}{(t^k - 1)/(t^h - 1)}. \tag{5}$$

If s is the denominator in (5), we have $p(s) = h$, $e(s) = k/h$; and if s_0 is the numerator in (5) we have $p(s_0) = h$, $e(s_0) = n/h$. Thus s_0 is a proper section of s and $e(s_0)$ is prime to $e(s)$.

We infer that there is no advantage to us in considering the general $\{d^m u^n\}$ procedure, nor in considering the bigger angle rather than the smaller. For the angular measures we obtain by these procedures are all of the form $(s_0/s)\pi$, where $s \in \mathfrak{F}$, and s_0 is prime to s. Thus it is sensible to restrict attention to $\{d^m u^n\}$-folding procedures in which $n|m$.

We now study further properties of the elements of $\overline{\mathfrak{F}}_t$, and their representations, both in base t and as rational numbers $(t^{nq} - 1)/(t^n - 1)$. We prove a division algorithm, a factorization theorem and a counting theorem.

We again take $s \in \overline{\mathfrak{F}}_t$ and s_0 a proper section of s. We carry out the Euclidean algorithm with respect to s, s_0 and with respect to $e(s)$, $e(s_0)$, obtaining

$$s = Qs_0 + R, 0 \leqslant R < s_0 \text{ and } e(s) = qe(s_0) + r, 0 \leqslant r < e(s_0). \tag{6}$$

From Theorem 2 we know that $R = 0$ if and only if $r = 0$ and that, in that case, $Q \in \overline{\mathfrak{F}}_t$ and

$$p(Q) = p(s)e(s_0), \qquad e(Q) = q.$$

We now consider the generalization of these formulae; for convenience of statement, however, we now assume that $R \neq 0$, $r \neq 0$.

THEOREM 4. *Let $s \in \overline{\mathfrak{F}}_t$ and let s_0 be a proper section of s. Then, given the relationships in equations* (6) *with $R \neq 0$, $r \neq 0$, we have*

$$Q = t^{p(s)r}Q', Q' \in \overline{\mathfrak{F}}_t, p(Q') = p(s)e(s_0), e(Q') = q;$$

R is a section of s, $p(R) = p(s)$, and $e(R) = r$.

Proof. Let $x = p(s) = p(s_0)$, $y = e(s)$, $y_0 = e(s_0)$. Then

$$s = \frac{t^{xy} - 1}{t^x - 1}, s_0 = \frac{t^{xy_0} - 1}{t^x - 1}.$$

Now $y = qy_0 + r$, so $t^{xy} - 1 = t^{xr}(t^{qxy_0} - 1) + (t^{xr} - 1)$. Thus

$$s = Qs_0 + R, \text{ where } Q = \frac{t^{xr}(t^{qxy_0} - 1)}{t^{xy_0} - 1}, R = \frac{t^{xr} - 1}{t^x - 1}.$$

It follows that $Q = t^{xr}Q'$, $Q' \in \overline{\mathfrak{F}}_t$, $p(Q') = xy_0$, $e(Q') = q$, $R \in \overline{\mathfrak{F}}_t$, $p(R) = x$, $e(R) = r$; and the theorem is proved.

Let us give an illustration of this theorem, using the same example given earlier.

EXAMPLE. Let $s = 1365$, $s_0 = 85$. Thus, in base 2, $s = 10101010101$, $s_0 = 1010101$. We have $e(s) = 6$, $e(s_0) = 4$. $p(s) = p(s_0) = 2$. Now $6 = 1 \cdot 4 + 2$, so that $q = 1$, $r = 2$, whence $Q' = 1$, $Q = 2^4 = 16$, $R = 5 (= 101$ in base 2). This shows that the Euclidean algorithm, applied to s and s_0, produces $1365 = 16 \times 85 + 5$.

Notice that from q and r, together with the representations of s, s_0 in base t, we recover Q and R very easily using Theorem 4. For we know Q' in base t from $p(Q') = p(s)e(s_0)$, $e(Q') = q$, and then Q is obtained from Q' in base t by adjoining $p(s)r$ zeros on the right. Of course, R is

obtained simply in base t from the relations $p(R) = p(s)$, $e(R) = r$. Thus the Euclidean algorithm, for our special folding numbers s, s_0, is very easily executed—as shown in our example.

Let us now, as in TABLE 1, set $p(s) = x$, $e(s) = y$, so that

$$s = \frac{t^{xy} - 1}{t^x - 1}.$$

Using the notation $s = (x, y)_t$, we then have the following result.

THEOREM 5 (*Factorization Theorem*). *In the notation above,* $(x, y)_t(xy, z)_t = (x, yz)_t$.

The theorem merely asserts the obvious fact that

$$\frac{t^{xy} - 1}{t^x - 1} \cdot \frac{t^{xyz} - 1}{t^{xy} - 1} = \frac{t^{xyz} - 1}{t^x - 1}.$$

However, it does give us some very *unobvious* factorizations. Thus, if we consult TABLE 1 with $x = 3$, $y = 4$, $z = 2$, we deduce that $585 \times 4097 = 2396745$. The main interest in Theorem 5 lies in the fact that it is true *in any base*. For, consider the case $x = 1$, $y = 2$, $z = 3$, where we have the following instances of the rule $(1,2)_t(2,3)_t = (1,6)_t$:

$(t = 2)$	$3 \times 21 = 63$
$(t = 3)$	$4 \times 91 = 364$
$(t = 4)$	$5 \times 273 = 1365$
$(t = 5)$	$6 \times 651 = 3906$

It is not immediately obvious that these are all instances of the same rule! We next show how to count the elements of $\mathfrak{F}_t$.

THEOREM 6. *There are* $\psi(k)$ *elements of* $\mathfrak{F}_t$ *between* t^k *and* t^{k+1}, *where* $\psi(k)$ *is the number of divisors of* k. *Indeed to each factorization* $k = uv$ *of* k, *there corresponds the element* $(u, v + 1)_t$ *of* $\mathfrak{F}_t$ *and* $t^k < (u, v + 1)_t < t^{k+1}$.

Proof. We are considering numbers *strictly* between t^k and t^{k+1} since elements of $\mathfrak{F}_t$ are never multiples of t. Such numbers have $(k + 1)$ digits in base t. Since an element of $\mathfrak{F}_t$ terminates in 1, the period is a divisor of k, so we get exactly one such element for each integer dividing k. Indeed if $u|k$ we get such an element s with $p(s) = u$ and $e(s) = (k/u) + 1$.

It follows from Theorem 6 that the elements of $\mathfrak{F}_t$ sweep out homothetic rectangular hyperbolae

$$x(y - 1) = k.$$

Thus if we look at such a hyperbola in TABLE 1, we obtain precisely the $\psi(k)$ elements s of $\mathfrak{F}$ for which $2^k < s < 2^{k+1}$. For example, with $k = 6$, we have the four elements $(1,7)$, $(2,4)$, $(3,3)$, $(6,2)$, or 127, 85, 73, 65, satisfying $2^6 < s < 2^7$.

The set $\overline{\mathfrak{F}}_t$ has few standard algebraic properties (though, as we have demonstrated, it has certain unconventional ones). Neither $\overline{\mathfrak{F}}_t$ nor its complement is closed under multiplication, as we see from the examples $5 \times 5 = 25$, $11 \times 93 = 1023$. Indeed, Theorem 5 seems to express the single basic multiplicative property. However, $\mathfrak{F}$ does possess one property which is vital in the geometric argument of the preceding section.

THEOREM 7. *Every odd number is a factor of some element of* $\mathfrak{F}$.

Actually, a much stronger statement is available: *every odd number is a factor of some* (x, y) *with* $x = 1$, *that is, of some* $2^y - 1$. This is immediately obvious from Euler's Theorem, which implies that, for any odd number a,

$$2^{\phi(a)} \equiv 1 \bmod a,$$

where $\phi(a)$ is the number of positive integers less than and prime to a.

Of course, the elements of $\mathcal{F}$ are all prime to 2, so Theorem 7 asserts that the set of factors of elements of $\mathcal{F}$ gives us *all* numbers prime to 2. In this form the generalization to $\mathcal{F}_t$ is clear. The elements of $\mathcal{F}_t$ are all prime to t (they are, in fact, all congruent to 1 mod t), and Euler's Theorem will imply that the set of factors of elements of $\mathcal{F}_t$ gives us *all* numbers prime to t.

Although Theorem 7 implies that we could construct all regular polygons just by using $\{d^m u\}$ procedures, we stress that this result is theoretical and it would be highly uneconomical to confine oneself to such procedures. For example, as TABLE 1 shows, we can get an 85-gon with just 8 folds ($\{d^6u^2\}$); whereas, since $\phi(85) = 64$, exploiting Euler's function would require 64 folds ($\{d^{63}u\}$). Similarly we can get a 341-gon with just 10 folds ($\{d^8u^2\}$); whereas, since $\phi(341) = 300$, exploiting Euler's function would require 300 folds ($\{d^{299}u\}$)! Actually this last argument is somewhat unfair, since we may usually find a smaller exponent μ than $\phi(a)$ such that $2^\mu - 1$ is divisible by a (where a is odd). Indeed we have the following result, whose proof we omit.

Let the prime power factorization of the positive integer a be written as

$$a = 2^{k_0} p_1^{k_1} \cdots p_s^{k_s},\ k_0 \geqslant 0,\ k_i > 0,\ i = 1,2,\ldots,s. \tag{7}$$

Thus a is odd if and only if $k_0 = 0$. Set

$$\mu(a) = \operatorname{lcm}\left(\phi(2^{k_0}), \phi\left(p_1^{k_1}\right),\ldots,\phi\left(p_s^{k_s}\right)\right) \text{ if } k_0 \leqslant 2,$$

and (8)

$$\mu(a) = \operatorname{lcm}\left(\tfrac{1}{2}\phi(2^{k_0}), \phi\left(p_1^{k_1}\right),\ldots,\phi\left(p_s^{k_s}\right)\right) \text{ if } k_0 \geqslant 3.$$

THEOREM 8. *If q is prime to a, then $q^{\mu(a)} \equiv 1$ mod a.*

Notice, from (8), that $\mu(a)$ is always a factor of $\phi(a)$, but may be considerably smaller than $\phi(a)$. Let us give some examples (recall that $\phi(p^k) = p^{k-1}(p-1)$, $k \geqslant 1$):

$a = 24 = 2^3 \cdot 3.$	Then $\mu(a) = \operatorname{lcm}(2,2) = 2,$	while $\phi(a) = 8.$
$a = 63 = 7 \cdot 3^2.$	Then $\mu(a) = \operatorname{lcm}(6,6) = 6,$	while $\phi(a) = 36.$
$a = 85 = 5 \cdot 17.$	Then $\mu(a) = \operatorname{lcm}(4,16) = 16,$	while $\phi(a) = 64.$
$a = 341 = 11 \cdot 31.$	Then $\mu(a) = \operatorname{lcm}(10,30) = 30,$	while $\phi(a) = 300.$

The following result gives a general principle governing the reduction from $\phi(a)$ to $\mu(a)$.

THEOREM 9. *$\mu(a) < \phi(a)$ unless $a = 1, 2, 4, p^k, 2p^k$ (p an odd prime). If $\mu(a) < \phi(a)$, then*

$$\mu(a) = \tfrac{1}{2}\phi(a) \quad \text{if } a = 2^{k_0},\ k_0 \geqslant 3, \text{ or if } a = 4p^k;$$

$$\mu(a) \mid \tfrac{1}{4}\phi(a) \quad \text{if } a = 2^{k_0}p^k,\ k_0 \geqslant 3;$$

$$\mu(a) \mid \frac{1}{2^{s-1}}\phi(a) \quad \text{if } a \text{ is given by (7) with } k_0 \leqslant 1 \text{ and } s \geqslant 2;$$

$$\mu(a) \mid \frac{1}{2^s}\phi(a) \quad \text{if } a \text{ is given by (7) with } k_0 = 2 \text{ and } s \geqslant 2.$$

$$\mu(a) \mid \frac{1}{2^{s+1}}\phi(a) \quad \text{if } a \text{ is given by (7) with } k_0 \geqslant 3 \text{ and } s \geqslant 2.$$

Although the reduction from $\phi(a)$ to $\mu(a)$ may be considerable, it remains more economical to use the general $\{d^m u^n\}$-procedure, where possible, rather than depend on $\{d^m u\}$-procedures and the refinement of Euler's Theorem given by Theorem 8.

References

[1] James R. Newman, The World of Mathematics, Simon and Schuster, New York, 1956.

[2] Jean Pedersen, Combinatorics, Group Theory and Geometric Models, Cahiers de Topologie et Géométrie Différentielle, vol. XXII-4, Amiens (1981) 407–428.

NOTES

The Approximation of Factorial Fragments

SYLVAN BURGSTAHLER
University of Minnesota, Duluth
Duluth, MN 55812

Suppose one wished to estimate some enormous factorial, for instance 10000!. "Use Stirling's formula," everyone will suggest and, indeed,

$$N! \approx \sqrt{2\pi N}\left(\frac{N}{e}\right)^N\left[1+\frac{1}{12N}+\frac{1}{288N^2}+\cdots\right] \tag{1}$$

can be used but we are after something new here. Told that, almost everyone's next suggestion would be to break up the problem into a number of more tractable subproblems. One such approach is to consider 100 subproblems, the first to estimate 100!, the second to estimate $101 \cdot 102 \cdot 103 \cdots 200$, and so forth. But how should such individual "factorial fragments" be estimated? That, in essence, is the central question of this paper.

In searching for a fresh answer to that problem, I was momentarily chagrined to note that none of these fragments had a precise middle term since it would have seemed reasonable to estimate each fragment by the 100th power of its middle term. Then it occurred to me that perhaps the fact that these factorial fragments had an *even* number of factors might be turned to advantage. After all, even when there is a middle term, powers of that term overestimate the product—and this bias gets worse as the length of the string increases. Perhaps with an *even* number of factors, one could select one representative integer from the lower half of the string, another from the upper half, and then estimate the entire product by taking the 50th power of the product of these representatives. Furthermore, by choosing these representatives asymmetrically, one might also be able to correct for any systematic biases that might be present.

As we shall see, these thoughts do bear fruit. We will find that for factorial fragments of certain peculiar lengths (26 factors, for example, or 362 factors, or certain select longer lengths) unusually accurate two-factor approximations are possible. While an argument will be made that the formula for the 26-factor case might be of some use as an alternative to Stirling's formula in areas of application where factorial fragments occur (e.g., in combinatorics, probability, statistics, calculus of finite differences), for the most part our results possess charm rather than utility. For example, we will show that if n is "large" compared to 469061, then

$$\frac{(n+469061)!}{(n-469061)!} \approx [(n-282359)(n+282360)]^{469061}$$

with a relative error of n^{-4} as $n \to \infty$, a Ramanujan-like result not likely to be checked (directly) on even the most powerful modern computer!

Before turning to the derivation of these results, it might be mentioned that the fact that factorial fragments of length 100 do not lend themselves to two-factor approximation is, of course, bad news for purposes of pursuing our original illustrative example. Curiously, however, if we permit *three*-factor approximations to each subproblem, we will find that fragments involving strings of 99 consecutive integers can be estimated by an even more accurate formula. Propitiously, such strings work just fine when estimating 10000! (Does the reader see why?)

Derivation of the approximations

In the various branches of mathematics where products of consecutive integers occur, they are often given locally relevant names. In the calculus of finite differences, for example, they are sometimes called "rising factorials" and in combinatorics, the product of the k consecutive integers which follow n counts the number of sortings of $k+1$ objects into $n+1$ linearly ordered boxes [**1**]. For our purposes, however, we will generally think of this product as a polynomial of degree k in the variable n and hence we will designate it by $P_k(n)$, that is,

$$P_k(n) = (n+1)(n+2)(n+3)\cdots(n+k). \tag{2}$$

Either by using mathematical induction and recursive arguments or by summation arguments alone, one can show that the first few terms of $P_k(n)$ are as follows:

$$P_k(n) = n^k + \frac{k(k+1)}{2}n^{k-1} + \frac{(k-1)k(k+1)(3k+2)}{24}n^{k-2} + \frac{(k-2)(k-1)k^2(k+1)^2}{48}n^{k-3} + \cdots. \tag{3}$$

If desired, symbolic algebraic manipulations of the sort needed to expand $P_k(n)$ can now be done directly on microcomputers [**2**]. Specifically, to construct polynomial products out of factors as would be needed here, the relevant computer algorithms first compute the product of these factors at "enough" successive integers and then generate the product polynomial out of the resulting tables of finite differences.

As a first attempt at approximation, suppose we try to fit $P_k(n)$ by means of a single binomial expansion. Since

$$(n+a)^k = n^k + kan^{k-1} + \frac{k(k-1)}{2}a^2n^{k-2} + \frac{k(k-1)(k-2)}{6}a^3n^{k-3} + \cdots \tag{4}$$

it is clear that if the approximation of $P_k(n)$ by $(n+a)^k$ is to be effective for "large" n, then the constant a should be chosen so as to match as many terms as possible in the expansions in (3) and (4). The two first terms are already identical and the second coefficients will match if and only if $a = (k+1)/2$. With this choice of a, however, succeeding coefficients fail to match and hence the approximation

$$P_k(n) \approx \left[n + \frac{k+1}{2}\right]^2$$

isn't very good unless $n \gg k$. (This is the "middle-term" approximation mentioned earlier.)

The next (and more fruitful) attempt is to try to fit factorial fragments that have an even number of terms by using powers of two binomial terms, i.e., fit $P_{2k}(n)$ by an expression of the form

$$(n+a)^k(n+b)^k = n^{2k} + k(a+b)n^{2k-1} + \left[\frac{k(k-1)}{2}(a^2+b^2) + k^2ab\right]n^{2k-2} + \left[\frac{k(k-1)(k-2)}{6}(a^3+b^3) + \frac{k^2(k-1)}{2}(a^2b+ab^2)\right]n^{2k-3} + \cdots. \tag{5}$$

If (3) is rewritten with k replaced by $2k$, one finds without difficulty that $P_{2k}(n)$ and the expansion in (5) will match in their first three coefficients provided that

$$a+b = 2k+1 \text{ and}$$
$$3(k-1)(a^2+b^2) + 6kab = (2k-1)(2k+1)(3k+1). \tag{6}$$

By using the first of these equations to simplify the second, this system of equations can also be written as

$$a+b=2k+1$$
$$3ab=(2k+1)(k+1) \tag{7}$$

from which it follows that a and b can be thought of as the two solutions of the quadratic equation

$$3x^2-3(2k+1)x+(2k+1)(k+1)=0. \tag{8}$$

The solutions of (8) are:

$$x=\frac{2k+1}{2}\pm\frac{1}{6}\sqrt{3(2k-1)(2k+1)}\,. \tag{9}$$

If these solutions are inserted in (5), a perfectly horrendous approximation formula results. What is needed, clearly, is some further restriction on k so that a and b turn out to be integers. To accomplish this, note that since k is an integer, the first term in equation (9) is half of an odd integer. From this it follows that the solutions of (9) will be integers if and only if the argument of the radical is nine times the square of an odd integer, that is to say, if and only if there exists some other integer, I, such that

$$k^2=3I(I+1)+1. \tag{10}$$

The author found two solutions, $(k, I)=(13,7)$ and $(181,104)$, for this Diophantine equation by using a hand calculator and later found two further solutions, $(2521,1455)$ and $(35113,20272)$, on a computer. Still later, Gerald Bergum, South Dakota State University at Brookings, indicated how to obtain a complete solution to (10). He noted that (10) can be rewritten as

$$k^2=3\left[I+\tfrac{1}{2}\right]^2+\tfrac{1}{4},$$

which is equivalent to

$$4k^2-3(2I+1)^2=1. \tag{11}$$

Substituting $u=2k$ and $v=2I+1$ in (11) yields

$$u^2-3v^2=1 \tag{12}$$

which is a classical Fermat-Pellian equation. It is known that the positive integer solutions of (12) must satisfy $u+v\sqrt{3}=(c+d\sqrt{3})^m$ where c and d are the so-called "fundamental solutions" to (12) and where m is an arbitrary nonnegative integer. In this case, $c=2$ and $d=1$ and (since u must be even and v must be odd) there is the further restriction that m must be odd. Assembling all these facts, it follows that the positive integer solutions of (10) must satisfy

$$2k+(2I+1)\sqrt{3}=\left(2+\sqrt{3}\right)^{2m+1} \text{ for } m=0,1,2,\cdots. \tag{13}$$

If $m=0$, this equation tells us that $k=1$ which, by (9), means that $a=1$ and $b=2$. According to our theory, we would then expect that $P_2(n)$ ought to be "well approximated" by $(n+1)^1(n+2)^1$ which, of course, it is! Somewhat less trivially, if $m=1$ in (13), one finds that $k=13$ and hence, by (9), that $a=6$ and $b=21$. From this we are led to conclude that

$$P_{26}(n)\approx[(n+6)(n+21)]^{13} \text{ as } n\to\infty. \tag{14}$$

Replacing n by $(n-13)$ and simplifying, this can also be written as

$$\frac{(n+13)!}{(n-13)!}\approx[(n-7)(n+8)]^{13} \text{ provided } n\geqslant 13. \tag{15}$$

In a similar manner, the choice $m=2$ in (13) yields $k=181$ from which it follows that $a=77$ and $b=286$ and hence

$$P_{362}(n)\approx[(n+77)(n+286)]^{181}$$

or,

$$\frac{(n+181)!}{(n-181)!} \approx [(n-104)(n+105)]^{181}. \tag{16}$$

The next three formulas in this curious infinite family of approximations are:

$$\frac{(n+2521)!}{(n-2521)!} \approx [(n-1455)(n+1456)]^{2521}, n \geqslant 2521,$$

$$\frac{(n+35113)!}{(n-35113)!} \approx [(n-20272)(n+20273)]^{35113}, n \geqslant 35113,$$

$$\frac{(n+469061)!}{(n-469061)!} \approx [(n-282359)(n+282360)]^{469061}, n \geqslant 469061. \tag{17}$$

Needless to say, the author has not attempted a direct numerical verification of these last three results!

We note that by applying a similar technique—matching the first *four* coefficients of $P_{3k}(n)$ with the corresponding coefficients of the product of *three* binomial powers—we produced one nontrivial example of a three-factor approximation formula, namely

$$P_{99}(n) \approx [(n+15)(n+50)(n+85)]^{33} \text{ as } n \to \infty. \tag{18}$$

This is believed to be the lowest-order formula of this type involving integers after the formula for $P_3(n)$ which, of course, has only three factors.

The formulas in (14)–(18) are more accurate than the derivation given above might suggest. It turns out that, while each of the two-factor approximation formulas was derived by fitting just *three* coefficients in the expansions in (3) and (5), the matching is actually exact through the first *four* coefficients. (The verification of this fact is a straightforward exercise in the use of (7) to reduce the fourth term in (5) to the fourth term in $P_{2k}(n)$.)

From these results it follows that the *absolute* error made in approximating $P_{2k}(n)$ by a suitable two-factor expression for choices of k given by (13) is a polynomial in n of degree $2k-4$ and hence that the *relative* error is a rational function of n in which the degree of the denominator polynomial exceeds that of the numerator polynomial by four.

Yasuchi Mochizuki (as part of an undergraduate Senior Project) established similar results for the three-factor approximation formula given in (18). Specifically, he found that, although that formula was derived by matching just *four* coefficients of $P_{99}(n)$ to the corresponding coefficients of the products of the three relevant binomial factors, the fit is actually exact through the first *five* coefficients. From this, in turn, it follows that the relative error in that formula is of order n^{-5} as $n \to \infty$.

Examples and applications

If anyone really needed to estimate truly gigantic factorial fragments, the higher-order approximation formulas in this paper might be of some use; but they are presented more for their curiosity value than as practical aids to everyday calculations. Formulas (14) and (15), however, involving 26-factor fragments, are also "practical" in a limited sort of way: for doing factorial calculations of the sort commonly done on hand calculators. While scientific hand calculators often include a factorial key, its use is usually restricted to 69! since 70! is too large for most calculators to display. Stirling's formula (given in (1)) can be used on such calculators (with various stratagems to avoid register overflow) but it is relatively inconvenient for such purposes. For example, it takes some *two dozen* keyboard manipulations to enter particular cases of Stirling's formula into some calculators even if only the first term of the asymptotic series is used! Worse, the evaluation of $P_k(n)$ will generally require *two* such calculations! (To be fair, certain common factors can be ignored, which redresses the balance somewhat; or one can use a

programmable calculator.) However, formula (14) will less frequently produce register overflow than is the case with two applications of Stirling's formula and, if overflow is inescapable, it is easier to sidestep while using (14) than while using Stirling's formula.

To illustrate the increase in the accuracy of these approximations as n increases, let us first estimate $P_{26}(0)$. This is 26!, the smallest number one can estimate (directly) using the formulas of this paper. Its value is

$$403,291,461,126,605,635,584,000,000.$$

On a Texas Instruments SR-51-II calculator (that uses 13 digits internally but displays only ten of them), the calculator displays 4.0329146×10^{26}—the same value, incidentally, that is displayed when 26! is calculated using the factorial key. In this instance, formula (14) yields $(6 \times 21)^{13}$ which, by the calculator, is about 2.0175165×10^{27}, an overestimation by about 400%. The trouble, of course, is that we are using (14) with $n = 0$, but n was supposed to be "large" in that formula. If we were really insistent about using (14) to estimate 26!, a more intelligent approach would be something like the following:

$$26! = \frac{5!P_{26}(5)}{27 \cdot 28 \cdot 29 \cdot 30 \cdot 31} \approx \frac{120(11 \times 26)^{13}}{27 \cdot 28 \cdot 29 \cdot 30 \cdot 31} \approx 5.0412349 \times 10^{26}.$$

But, let's face it, Stirling's formula is simply superior in this instance. It yields 4.0200099×10^{26} on the calculator from just the first term of its asymptotic form!

Formula (2) is more accurate when approximating $70 \cdot 71 \cdot 72 \cdots 95$ which is $P_{26}(69)$. (Note: this product is the "other" factor one would need in order to compute 95! on a hand calculator that only gives factorials through 69!.) When the product $P_{26}(69)$ is multiplied out term by term on the calculator, the result is 6.0366003×10^{49}, a number in close agreement with the approximation $6.0387778 \times 10^{49} = (75 \cdot 90)^{13}$ from formula (14). In order to use Stirling's formula to approximate $P_{26}(69)$, we write it as 95!/69! and cancel out common factors from the separate approximations to these two factorials; but one still has to calculate

$$P_{26}(69) \approx \left(\frac{95}{69}\right)^{69.5}\left(\frac{95}{e}\right)^{26}. \tag{19}$$

The calculator value of the right side of (19) is 6.0385959×10^{49}, so Stirling's formula is a shade more accurate. However, the advantages of working with $(75 \cdot 90)^{13}$ rather than with the mess in (19) is obvious.

Formula (14) also lends itself to iteration. For example, to calculate 115! on a hand calculator, one could note that

$$115! = 89!P_{26}(89) = 63!P_{26}(63)\,P_{26}(89) \approx 63!(69 \cdot 84 \cdot 95 \cdot 110)^{13} \approx 2.92696546 \times 10^{188}.$$

The relative error introduced into this calculation because of the iterated use of the formula in (14) would be expected to be approximately of the order of $63^{-4} + 89^{-4} \approx 10^{-7}$ in view of our error expression and allowing for errors in the same direction, that is to say, the relative error should be only "a few" multiples of 10^{-7}.

A symbolic math system was used to find 115! exactly; it is enough for our purposes to note that it gave $115! \approx 2.92509369 \times 10^{188}$. The relative error of our approximation is .06399% and therefore "a few" turns out to be almost 6400 in this instance. Still, our approximation is satisfyingly accurate, in terms of relative error.

References

[1] M. Aigner, Combinatorial Theory, Springer-Verlag, New York, 1979, p. 76.

[2] Herbert S. Wilf, The disk with the college education, Amer. Math. Monthly, 89 (1982) 4–8.

Visualization of Matrix Singular Value Decomposition

CLIFF LONG
Bowling Green State University
Bowling Green, OH 43403

A real matrix is frequently used as a finite representation of a real function of two variables, especially as a tool for studying continuous functions in numerical analysis and computer graphics. It is also advantageous to use continuous functions to provide visualization for matrix techniques such as **singular value decomposition** (SVD). We will illustrate how this factorization technique can be thought of as providing least square best fit approximations to functions of two variables. The basic theory of SVD (sometimes called basic structure of a matrix) will be presented, one simple example given for clarification (similar to those found in [7]), and then a matrix representation of a sculptured head of Abe Lincoln will be used to illustrate the geometry involved. For ease in understanding, we'll restrict our attention to real matrices and refer the reader to [2], [4], [9], and [11] for the proofs.

Singular value decomposition of a matrix is a technique which represents any given matrix as a sum of rank 1 matrices, i.e., it yields a finite series expansion for a matrix. For example, the matrix

$$A = \begin{bmatrix} 3.01 & 0.01 & -2.99 \\ 2.99 & -0.01 & -3.01 \\ 2.00 & -4.00 & 2.00 \end{bmatrix} \tag{1}$$

can be written as the sum of three rank 1 matrices in a rather obvious decomposition:

$$A = \begin{bmatrix} 3 & 0 & -3 \\ 3 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & -4 & 2 \end{bmatrix} + \begin{bmatrix} .01 & .01 & .01 \\ -.01 & -.01 & -.01 \\ 0 & 0 & 0 \end{bmatrix}.$$

A less obvious decomposition (which results from the theorem stated below) is:

$$A = 6\begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} + 2\sqrt{6}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} + \frac{\sqrt{6}}{100}\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & 0 \end{bmatrix}. \tag{2}$$

In matrix singular value decomposition, the matrix need not be square nor real, and the rank 1 matrices are chosen, normalized and ordered for usefulness in solving problems. The theory of singular value decomposition is not new (according to [**10**, p. 78] it was established for real and square matrices in the 1870s by Beltrami and Jordan and later developments are referenced in [**8**]). However, its current importance and extensive use is due to the existence of an efficient and numerically stable algorithm developed by Golub in the 1960s ([**5**], [**6**]). The technique is regularly used in solving least square problems and computing pseudoinverses of matrices. It is certain to be used even more extensively now that good computer programs are readily available (e.g., Moler [**4**] and software packages such as EISPACK, LINPACK and IMSL). An application to digital image processing by Andrews and Patterson [**1**] inspired my interest in SVD, and comments such as [it is] "The most reliable method for computing the coefficients for general least square problems..." [**4**, p. 195] and "... it is not nearly as famous as it should be" [**11**, p. 142] have kept me going.

The key theorem for SVD of matrices is the following.

THEOREM. *Any real matrix A can be factored as $A = PSQ^T$, where P and Q are orthogonal and S is diagonal with diagonal elements $\sigma_i \geqslant 0$ (called the singular values of A)* [**9**, *p. 18*].

COROLLARY. *Any real $m \times n$ matrix A can be expressed as a finite sum of rank 1 matrices in normalized form, that is, $A = \sigma_1 R_1 + \sigma_2 R_2 + \cdots + \sigma_k R_k$, where $k = \min(m, n)$ and*

(1) $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0 = \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n$, *rank* $A = r \leqslant k$.
(2) $R_i = \bar{p}_i \bar{q}_i^T$ *where* $\bar{p}_i$ *is the* ith *column of* P *and a unit eigenvector of* AA^T, *and* $\bar{q}_i$ *is the* ith *column of* Q *and a unit eigenvector of* A^TA.
(3) *each* R_i *has the sum of the squares of its elements equal to* 1 (*this follows from* 2).

The proofs of these results depend on the fact that A^TA and AA^T are real, square and symmetric, and each has nonnegative eigenvalues and a complete set of orthogonal eigenvectors. (In fact A^TA and AA^T have precisely the same nonzero eigenvalues and the square roots of these are singular values σ_i, $1 \leqslant i \leqslant k$ of both A and A^T.)

To illustrate how the decomposition stated in the Corollary proceeds from the Theorem, consider our previous example. The matrix A is first factored as in the Theorem:

$$A = PSQ^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2\sqrt{6} & 0 \\ 0 & 0 & \frac{\sqrt{6}}{100} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}. \tag{3}$$

Next, this factorization can be written as the sum $\Sigma \sigma_i \bar{p}_i q_i^T$, where the σ_i are the diagonal entries of S, $\bar{p}_i$ the ith column of P and q_i the ith row of Q.

$$A = 6 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} + 2\sqrt{6} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

$$+ \frac{\sqrt{6}}{100} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}. \tag{4}$$

Multiplication of the $\bar{p}_i \bar{q}_i^T$ yields the desired decomposition $A = \sigma_1 R_1 + \sigma_2 R_2 + \sigma_3 R_3$, which we have noted in (2). As an illustration of some of the matrix ideas used in obtaining the Corollary from the Theorem, we now establish part 2 for a square matrix A and for $\bar{q}_1$, the first column of Q, i.e., we'll show that $A^TA\bar{q}_1 = \sigma_1^2 \bar{q}_1$. Since P is orthogonal and S diagonal, we have $A^TA = (PSQ^T)^T(PSQ^T) = (QS^TP^T)(PSQ^T) = QS^TP^{-1}PSQ^{-1} = QS^TSQ^{-1} = QS^2Q^{-1}$. But then, if $\bar{e}_1$ is the column vector $[1 \quad 0 \quad \cdots \quad 0]^T$, we have

$$A^TA\bar{q}_1 = QS^2Q^{-1}\bar{q}_1 = QS^2\bar{e}_1 = Q\sigma_1^2\bar{e}_1 = \sigma_1^2\bar{q}_1.$$

It is equally easy to show that the matrix $R_i = \bar{p}_i \bar{q}_i^T$ has the sum of the squares of its elements equal to 1 (i.e., it has Frobenius norm $\|R\|_F$ equal to 1). This property allows the SVD of a matrix A of rank r to be used to find an $m \times n$ matrix B of rank $l < r$ that minimizes $\|B - A\|_F$. This

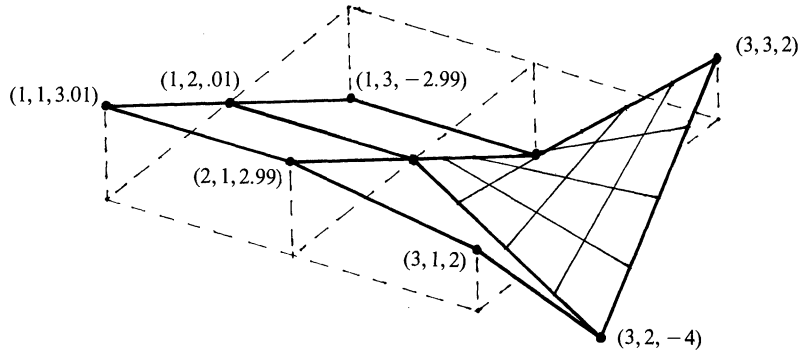


FIGURE 1

problem was posed and solved by Eckart and Young [3] who show that the first l terms of the SVD of A sum up to just such a matrix B [9, p. 26]. Thus the sum of the first l terms of the singular value decomposition is a best rank l approximation to the matrix A in the sense of the Frobenius norm (i.e., in the least squares sense). In fact,

$$\|A - \sigma_1 R_1 - \sigma_2 R_2 - \cdots - \sigma_l R_l\|_F = \sigma_{l+1}^2 + \cdots + \sigma_r^2.$$

While the Corollary suggests a way to determine the factorization $A = PSQ^T$, this method is not numerically stable for nearly singular matrices and should be replaced by an algorithmic approach such as that of Golub.

Now for some geometry! To a real $m \times n$ matrix A, we can associate a surface containing the points (x_i, y_j, a_{ij}) where (x_i, y_j) are lattice points on a rectangular grid. For example the matrix A given by (1) can be associated to the surface shown in FIGURE 1. This surface can be thought of as a piecewise hyperbolic function $z = ax + bxy + cy + d$ on the 3×3 grid points (x_i, y_j, a_{ij}).

Alternatively, given a continuous real function f defined over a rectangular grid, we may associate a real matrix A with entries the function values $f(x_i, y_j) = a_{ij}$ and treat the matrix A as a finite approximation to the surface $z = f(x, y)$. The singular value decomposition of this matrix A then gives a further approximation to the surface. Conversely, the related surface can be used to "visualize" the singular value decomposition. (Similar visualization techniques have been used for one-variable Taylor series and Fourier series expansions and should be utilized more often in the two-variable setting now that 3D computer graphics programs are more readily available.)

For illustrative purposes, we obtained a finite approximation to a bust of Abe Lincoln (using a crude homemade scanning device which allowed for a 49×36 matrix). The original sculpture and finite approximation (called ABE) are shown in FIGURE 2. The related matrix A was then factored to produce a finite expansion of ABE using rank 1 matrices from a SVD. The surfaces shown in FIGURE 3 represent surface approximations to ABE by keeping only a specified number of terms from the finite series expansion. The surface marked A_1 represents the approximation $A \doteq \sigma_1 R_1$, the surface A_2 represents the approximation of A by the two-term decomposition $A \doteq \sigma_1 R_1 + \sigma_2 R_2$, and so on. It is somewhat surprising that the rank 5 approximation to ABE (of a possible 36) is so recognizable. This means that the tail-end terms of the series $A = \Sigma \sigma_i R_i$ are not all that important, and suggests that the matrix might be somewhat ill-conditioned (actually $\sigma_1/\sigma_6 \doteq 33$ and the ratio σ_1/σ_{36}, called the condition number, is approximately 2000). Note that two different approximation techniques are used on the original sculpture. The first is the grid size which determines the matrix size $m \times n$. The second is related to the relative sizes of the singular values σ_i of the matrix. Thus our first step reduced ABE to 49×36 real numbers and our second step for surface A_5 reduced him to just $5 \times (49 + 36) + 5 = 430$ real numbers.

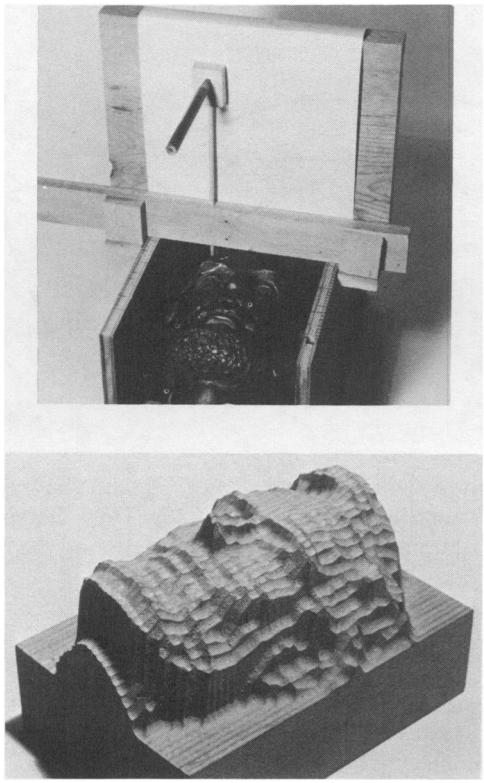


FIGURE 2

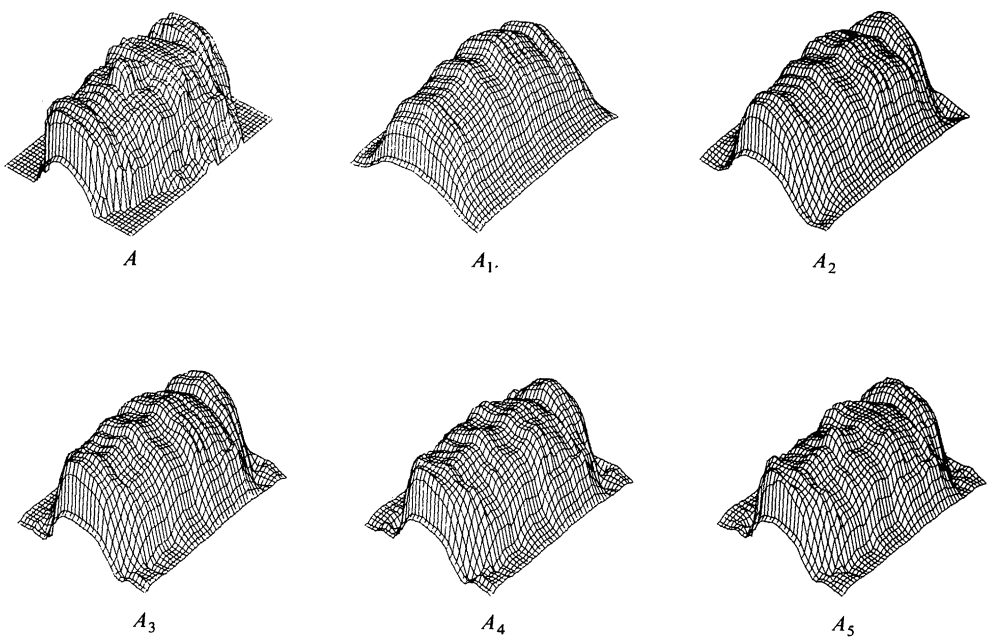


FIGURE 3. **Singular value decomposition of "*ABE*."**

These procedures for reducing a continuous image to a finite set of real numbers are of particular importance in image processing techniques [1]. Specifically, dropping the tail-end terms of the singular value decomposition can be associated with eliminating the "snowy" feature of a TV picture (i.e., noise elimination from a picture transmission). The essential information of the picture is carried by the earlier terms of the decomposition and associated with the larger singular values, while the more random noise (unless of significant size) is associated with the smaller singular values and discarded.

The omission of small singular values is also significant for handling problems involving inverses of ill-conditioned matrices. (Many least square approximation problems fall into this category.) If a matrix A is decomposed as in the Theorem, $A = PSQ^T$, and if A^{-1} exists, then since P is orthogonal and S is diagonal, it follows that $A^{-1} = (PSQ^T)^{-1} = QS^{-1}P^T$ where S^{-1} is a diagonal matrix with ith diagonal entry σ_i^{-1}. This is a factorization of A^{-1} as in the Theorem, so the Corollary applies. Thus if we know the singular value decomposition of a nonsingular matrix, then we also have a decomposition of A^{-1}. For example, the matrix A in (1) is shown in factored form in (3), and its SVD derived in (4). From the above discussion, we have

$$
\begin{aligned}
A^{-1} &= \begin{bmatrix} 3.01 & .01 & -2.99 \\ 2.99 & -.01 & -3.01 \\ 2 & -4 & 2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \frac{1}{6} & 0 & 0 \\ 0 & \frac{\sqrt{6}}{12} & 0 \\ 0 & 0 & \frac{100}{6}\sqrt{6} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \\
&= \frac{1}{6}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + \frac{\sqrt{6}}{12}\begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \\
&\quad + \frac{100}{6}\sqrt{6}\begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \\
&= \frac{1}{6}R_1^T + \frac{\sqrt{6}}{12}R_2^T + \frac{100\sqrt{6}}{6}R_3^T \\
&= \sigma_1^{-1}R_1^T + \sigma_2^{-1}R_2^T + \sigma_3^{-1}R_3^T.
\end{aligned}
$$

This shows that the least significant rank 1 matrix in the SVD of A (i.e., R_3 which has smallest

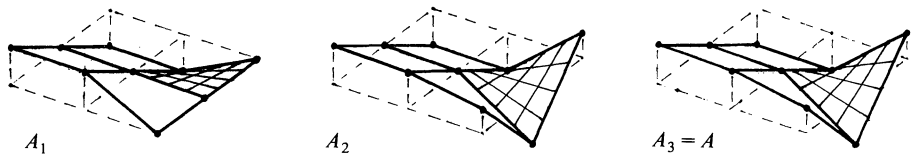


FIGURE 4(a). **Rank r approximations, A_r, to A.**

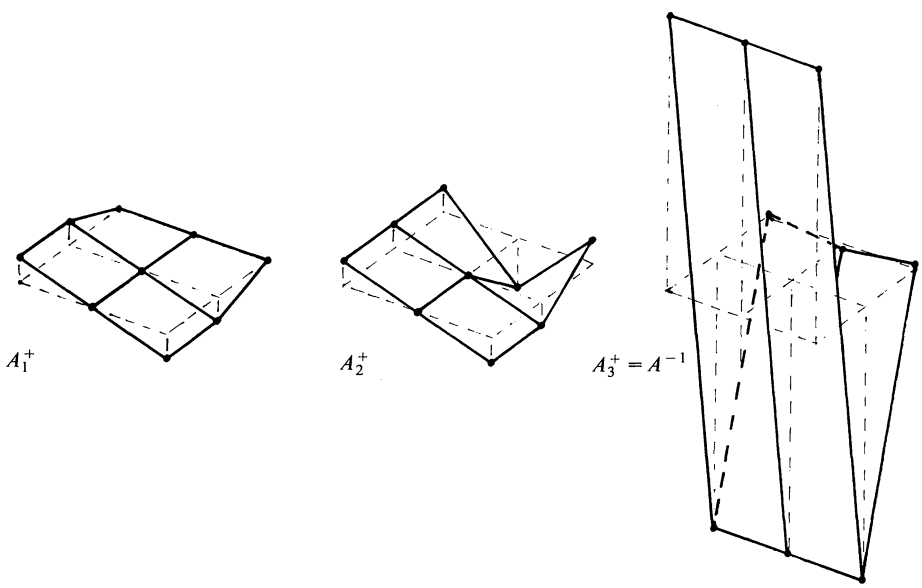


FIGURE 4(b). **Corresponding pseudoinverses, A_r^+ (not to scale).**

coefficient σ_3) becomes the most significant rank 1 matrix in the SVD of A^{-1}. Thus if the condition number of a matrix is large ($\sigma_n \ll \sigma_1$), the inverse is dominated by an insignificant part of the original matrix. This suggests that small changes in A (such as round-off errors or other noise) can seriously affect A^{-1} and such matrices are then called ill-conditioned. Rather than allow this noise to dominate the inverse, it seems more appropriate to ignore it and replace the corresponding diagonal terms of S^{-1} by 0. When this is done, the matrix A^{-1} is essentially replaced by an effective pseudoinverse. The decision of which values σ_i^{-1} to replace by 0 depends not only on the ratio σ_1/σ_i but also on the order of computer machine precision and the application involved.

For our 3×3 matrix A, the surfaces of FIGURE 4 show how A^{-1} is dominated by the smallest term of the singular value decomposition, and suggest that while A_2^+ might be a good replacement for A^{-1} in certain applications, this decision should not be taken lightly. The shape of the surface is being emphasized in FIGURE 4 with the scales chosen for viewing convenience. The maximum surface height for A_3^+ is actually about 200 times that of A_2^+.

When a matrix A is either square singular or nonsquare, then A^{-1} fails to exist and a pseudoinverse of A is given by $A^+ = QS^+P^T$ where S^+ is diagonal with $d_i = \sigma_i^{-1}$ if $\sigma_i \neq 0$ and $d_i = 0$ if $\sigma_i = 0$. Thus if A has rank r and singular value decomposition $A = \sum_{i=1}^{n} \sigma_i R_i$ then

$$A^+ = \sigma_1^{-1} R_1^T + \sigma_2^{-1} R_2^T + \cdots + \sigma_{r-1}^{-1} R_{r-1}^T + \sigma_r^{-1} R_r^T.$$

We show in FIGURE 5 the pseudoinverse A^+ of ABE, and A_1^+ through A_5^+, the pseudoinverses

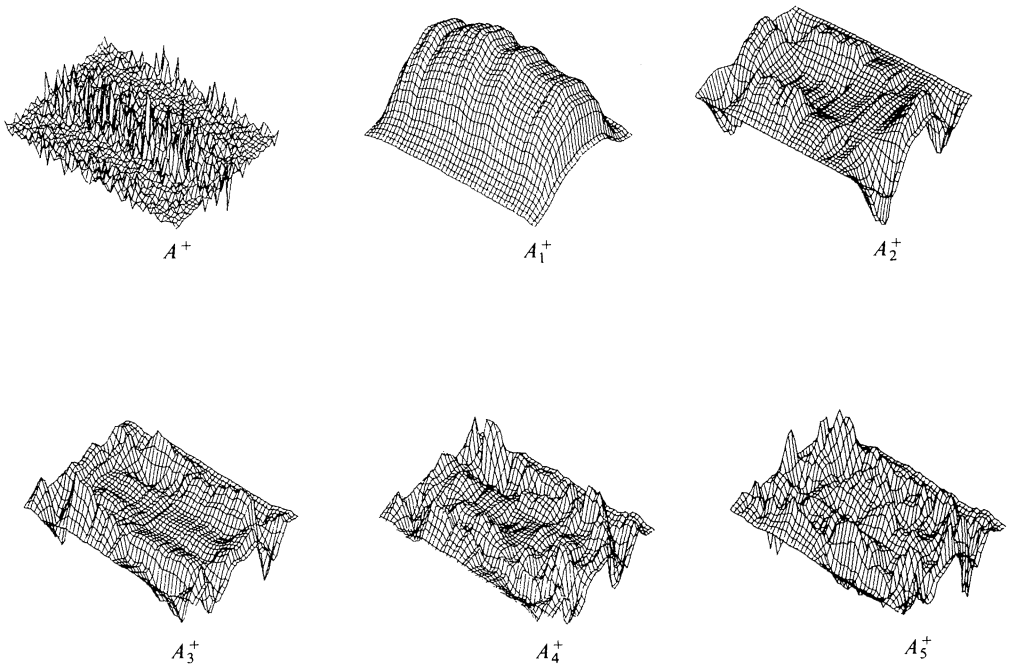


FIGURE 5. **Pseudoinverses of "ABE" and A_r.**

of our rank 1 through rank 5 approximations of ABE. The heights are again adjusted for viewing convenience (in fact A^+ has terms of much larger magnitude than the others). It is the "shape" of the matrices which is emphasized. It should be noted that while A_5 is a good approximation to A, A_5^+ is not a good approximation to A^+. Listing the singular values for a given problem frequently aids the user in deciding on an effective rank e for a matrix, and then A_e^+ is used in place of A^+. These substitutions provide the reliability for the SVD method in solving least square problems, since small changes in the original matrices are not allowed to dominate the pseudoinverse.

References

[1] H. C. Andrews and C. L. Patterson, Outer product expansions and their uses in digital image processing, Amer. Math. Monthly, 82 (1975) 1–12.

[2] P. J. Davis, Circulant Matrices, John Wiley and Sons, New York, 1979.

[3] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, Psychometrika, 1 (1936) 211–218.

[4] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, Computer Methods for Mathematical Computations, Prentice-Hall, Englewood Cliffs, NJ, 1977.

[5] G. H. Golub and W. Kahan, Calculating the singular values and pseudoinverse of a matrix, SIAM J. Numer. Anal., 2 (1965) 205–224.

[6] G. H. Golub and C. Reinsch, Singular value decomposition and least squares solutions, Numer. Math., 14 (1970) 403–420.

[7] P. E. Green, Mathematical Tools for Applied Multivariate Analysis, Academic Press, New York, 1976.

[8] V. C. Kema and A. J. Laub, The singular value decomposition: its computation and some applications, IEEE Transactions on Automatic Control, 25 (1980) 164–176.

[9] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[10] C. C. MacDuffee, The Theory of Matrices, Springer, Berlin, 1933.

[11] G. Strang, Linear Algebra and Its Applications, 2nd ed., Academic Press, New York, 1980.

An Algorithm for the Characteristic Polynomial

William A. McWorter, Jr.
Ohio State University
Columbus, Ohio 43210

In his text on linear algebra [3] Evar D. Nering wrote

> We introduce some topics from the theory of determinants solely for the purpose of finding the eigenvalues of a linear transformation. Were it not for this use of determinants we would not discuss them in this book.

This curious remark sparked my search for and discovery of a determinant-free algorithm for the characteristic polynomial, an algorithm which, happily, provides fast classroom procedures for finding eigenvalues and eigenvectors.

Computing the characteristic polynomial $C_A(x)$ of a square matrix A as the determinant of the matrix $xI - A$ is like programming in LISP, Lots of Irritating Single Parentheses. Linear algebra texts avoid the pain in the brace by discussing only small matrices or matrices of an embarrassingly simple form such as diagonal or triangular. For if a matrix is upper triangular, say

$$A = \begin{bmatrix} a_1 & & & \\ & \ddots & * & \\ & & \ddots & \\ 0 & & & a_n \end{bmatrix},$$

then determinant properties say that its characteristic polynomial is the simple product $C_A(x) = (x - a_1) \cdots (x - a_n)$. Less embarrassing is the **companion matrix** for a monic polynomial $P(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$. Here A has the form

$$A = \begin{bmatrix} 0 & & & & -a_0 \\ 1 & \ddots & & 0 & -a_1 \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \vdots \\ 0 & & & 0 & \vdots \\ & & & 1 & -a_{n-1} \end{bmatrix},$$

and its characteristic polynomial is simply $P(x)$. Combining triangular and companion matrix forms into block triangular form

$$A = \begin{bmatrix} A_1 & & & \\ & \ddots & * & \\ & & \ddots & \\ 0 & & & A_k \end{bmatrix}, \tag{1}$$

where the A_i are companion matrices of monic polynomials $P_i(x)$, we have what is called a **Frobenius matrix**; its characteristic polynomial is the product $C_A(x) = P_1(x)P_2(x) \cdots P_k(x)$.

Frobenius matrices are the key to our algorithm. *Every* square matrix is similar to a Frobenius matrix; i.e.,

(F) *For every square matrix A there exists an invertible matrix D such that $F = D^{-1}AD$ is a Frobenius matrix.*

Since similar matrices have the same characteristic polynomial and the characteristic polynomial of a Frobenius matrix involves no computation, finding $C_A(x)$ is no more difficult than finding a D such that $D^{-1}AD$ is a Frobenius matrix. The algorithm proposed here finds such a D

(for other characteristic polynomial algorithms see [1]). Note that by finding such a D, our algorithm also provides a constructive proof of the statement (F).

The columns of the matrix D have, almost, a very nice property. Let us illustrate with an example. Suppose A and D are 5×5 matrices such that $D^{-1}AD = F$, where F is the Frobenius matrix

$$F = \left[\begin{array}{ccccc} 0 & -1 & 1 & 2 & 3 \\ 1 & -2 & 4 & 5 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right].$$

Then, reading coefficients from the companion matrices in F, we have

$$C_A(x) = C_F(x) = (x^2 + 2x + 1)(x^3 - 2x^2 - 3x).$$

For our purposes it is convenient to express the similarity of A and F by the matrix equation $AD = DF$. Let $v_1, \ldots, v_5$ be the columns of D. Then the matrix equation $A[v_1 \ldots v_5] = [v_1 \ldots v_5]F$ is equivalent to the five vector equations

$$\begin{aligned} Av_1 &= v_2 \\ Av_2 &= -v_1 - 2v_2 \\ Av_3 &= v_1 + 4v_2 + v_4 \\ Av_4 &= 2v_1 + 5v_2 + v_5 \\ Av_5 &= 3v_1 + 6v_2 + 3v_4 + 2v_5. \end{aligned}$$

With two exceptions, the vectors $v_1, \ldots, v_5$ obey (for $n = 5$) the condition:

(*) *The vectors $v_1, \ldots, v_n$ are linearly independent, and for each i, the vector Av_i is either v_{i+1} or a linear combination of $v_1, \ldots, v_i$.*

The vectors v_3 and v_4 violate (*). However, a simple change of basis remedies the situation. Replace v_4 with $v_4' = Av_3$ and v_5 with $v_5' = Av_4'$. Using the basis $v_1, v_2, v_3, v_4', v_5'$ the vector equations become

$$\begin{aligned} Av_1 &= v_2 \\ Av_2 &= -v_1 - 2v_2 \\ Av_3 &= v_4' \\ Av_4' &= v_5' \\ Av_5' &= 6v_1 + 3v_4' + 2v_5' \end{aligned}$$

which are equivalent to the matrix equation

$$A[v_1 v_2 v_3 v_4' v_5'] = [v_1 v_2 v_3 v_4' v_5'] \left[\begin{array}{ccccc} 0 & -1 & 0 & 0 & 6 \\ 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right].$$

Hence A is similar to a Frobenius matrix via a matrix $D' = [v_1 v_2 v_3 v_4' v_5']$ whose columns satisfy condition (*).

Given any $n \times n$ matrix A it is, conceptually at least, easy to find a basis satisfying (*). To see this, let R^n be the set of $n \times 1$ matrices (or vectors) with real entries. Choose any nonzero vector v_1 in R^n. If Av_1 is independent of v_1, then set $v_2 = Av_1$; otherwise (if $n > 1$), choose any vector v_2 independent of v_1. Assuming we have constructed in this way k independent vectors $v_1, \ldots, v_k$,

form the vector Av_k. If Av_k is independent of $v_1,\ldots,v_k$, then set $v_{k+1}=Av_k$; otherwise (if $n>k$), set v_{k+1} equal to any vector independent of $v_1,\ldots,v_k$.

The polynomial $C_A(x)$ is immediately found from a basis satisfying (*). Let us construct such a basis for the 4×4 matrix

$$A=\begin{bmatrix} 2 & -4 & 8 & 8 \\ 1 & 0 & 1 & -2 \\ -1 & 1 & -4 & -2 \\ 1 & -1 & 2 & 0 \end{bmatrix}.$$

Choose any nonzero vector v_1, say $v_1=(1,0,0,0)^T$. Since $Av_1=(2,1,-1,1)^T$ is independent of v_1, set $v_2=Av_1$. Form $Av_2=(0,-1,1,-1)^T$. This vector depends on v_1 and v_2:

$$Av_2=2v_1-v_2. \tag{2}$$

Now choose any vector v_3 independent of v_1 and v_2, say $v_3=(0,1,0,0)^T$ (a standard basis vector will do). The vector $Av_3=(-4,0,1,-1)^T$ depends on v_1, v_2, and v_3:

$$Av_3=-2v_1-v_2+v_3. \tag{3}$$

Choose any vector v_4 independent of v_1, v_2, and v_3, say $v_4=(0,0,1,0)^T$. Form $Av_4=(8,1,-4,2)^T$. The vector Av_4 must depend on v_1, v_2, v_3, and v_4:

$$Av_4=4v_1+2v_2-v_3-2v_4. \tag{4}$$

At this point $\{v_1,v_2,v_3,v_4\}$ is a basis of R^4 and so the algorithm stops. Setting $D=[v_1v_2v_3v_4]$, the equations (2), (3), and (4) give

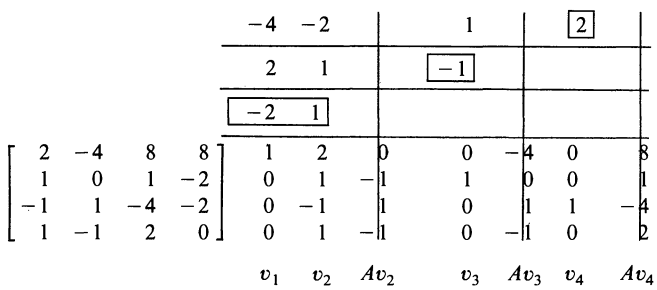
$$D^{-1}AD=\left[\begin{array}{c|c|c|c} 0 & 2 & -2 & 4 \\ \hline 1 & -1 & -1 & 2 \\ \hline 0 & 0 & 1 & -1 \\ \hline 0 & 0 & 0 & -2 \end{array}\right].$$

Hence

$$C_A(x)=(x^2+x-2)(x-1)(x+2)=(x-1)^2(x+2)^2.$$

Note that the coefficients of the monic polynomial $P_1(x)=x^2+x-2$ are the negatives of the coefficients in (2), those of $P_2(x)=x-1$ are the negatives of the coefficients in (3) of those terms beyond v_2, and those of $P_3(x)=x+2$ are the negatives of the coefficients in (4) of terms beyond v_3.

For hand calculation with this algorithm it is more efficient to use a human computer data structure, the scratch paper tableau in FIGURE 1. In this figure, the matrix A is in brackets at the left. The column vector v_1 is placed immediately to the right of A to facilitate the calculation of $Av_1=v_2$, and v_2 is placed to the right of v_1. The vector Av_2 is placed following v_2 and the dependence of Av_1 on v_1 and v_2 is visually apparent. The coefficients of v_1 and v_2 in the



$$C_A(x)=(-2+x+x^2)(-1+x)(2+x)$$

FIGURE 1

dependence equation $-2v_1 + v_2 + Av_2 = 0$ are placed above their corresponding vectors (the box containing -2 and 1). A line is placed through Av_2 to indicate its dependence and signal that a vector independent of v_1 and v_2 must be chosen. The standard basis vector $v_3 = e_2$ is clearly independent of v_1 and v_2 and is placed following the cancelled column Av_2. The algorithm continues by placing Av_3 after v_3. Dependence of Av_3 is easily seen; $2v_1 + v_2 - v_3 + Av_3 = 0$. The dependence coefficients are placed on a second level above their corresponding vectors. The vector $v_4 = e_3$ is independent of v_1, v_2, and v_3. Finally, the vector Av_4 is placed as the last column of the tableau. The dependence equation for Av_4 is easy and its coefficients are placed on a third level above the basis v_1, v_2, v_3, v_4. The values in the boxes are the coefficients of the monic polynomials $P_i(x)$, which are placed below the corresponding boxes, under the tableau, expressing the characteristic polynomial $C_A(x)$ as their product.

The dependence checks required to produce vectors satisfying condition (*) in this example were particularly easy and it is not difficult to construct such examples. However, classroom matrices are not much more difficult. If the matrix is small, dependence checks are eased by recalling that the second highest coefficient of $C_A(x)$ is the negative of the trace of A. Look at the tableau in FIGURE 2.

			813	−197	5	
−7	13	4	1	−7	229	−3337
12	−1	−5	0	12	−126	2994
6	0	3	0	6	−24	1302
			v_1	v_2	v_3	Av_3

$$C_A(x) = 813 - 197x + 5x^2 + x^3$$

FIGURE 2

When this algorithm is applied to a random $n \times n$ matrix, the first n vectors are likely to be independent as in FIGURE 2. Independence of v_1 and v_2 is obvious. That v_3 is independent of v_1 and v_2 follows from the fact that the boxed 2×2 block clearly has rank 2. The dependence equation for Av_3 is easily found by first noting that the coefficient of v_3 must be the negative of the trace of A.

If the matrix A is an unnoticed Frobenius matrix, then the algorithm proceeds rapidly because all dependence checks are obvious. If A has a low degree minimum polynomial, the algorithm often goes smoothly. Witness calculation with a combinatorial matrix whose tableau is in FIGURE 3.

For less cooperative matrices the dependence checks can be automated by the usual procedure of placing the column vectors $v_1, \ldots, v_k$, and Av_k in a matrix $[v_1 \ \ldots \ v_k \ Av_k]$ and reducing it to row echelon form. If the rank is $k+1$, then we have independence; otherwise, the last column provides the dependence coefficients. The product E of elementary row operations that reduce $[v_1 \ \ldots \ v_k \ Av_k]$ to row echelon form is useful for the next dependence check of the vectors $v_1, \ldots, v_{k+1}, Av_{k+1}$ since the matrix product $E[v_1 \ \ldots \ v_{k+1} \ Av_{k+1}]$ is in row echelon form except for the last column. Hence, if E is maintained, all but the last column of $E[v_1 \ \ldots \ v_{k+1} \ Av_{k+1}]$

	$a-b$	-1		0	$b-a$	
	$a-b$	-1		$b-a$		
	$a^2+2ab-3b^2$	$-2a-2b$				
$a\ b\ b\ b$	1	a	a^2+3b^2	0	0	b
$b\ a\ b\ b$	0	b	$2ab+2b^2$	1	0	b
$b\ b\ a\ b$	0	b	$2ab+2b^2$	0	1	a
$b\ b\ b\ a$	0	b	$2ab+2b^2$	0	0	b

$$C_A(x) = (a^2 + 2ab - 3b^2 - (2a+2b)x + x^2)(b-a+x)(b-a+x)$$

FIGURE 3

				0	−.5	−.05		0	
				0	−20	−16			
1	2	3	4	1	1	30	500	0	2
2	3	4	5	0	2	40	680	1	3
3	4	5	6	0	3	50	860	0	4
4	5	6	7	0	4	60	1040	0	5
				v_1	v_2	v_3	Av_3	v_4	Av_4

$$C_A(x) = (0 - 20x - 16x^2 + x^2)(0 + x)$$

FIGURE 4

can be ignored. When dependence occurs and an independent vector must be chosen, a column of E will signal which standard basis vector to choose. Here is an algorithm with dependence checks automated.

(1) (Initialize) Set $k = 1$, $E = I_n$, and $C_A(x) = 1$.
(2) (Choose an independent vector) Let s be the least positive integer such that the sth column of E has a nonzero entry in row k. Set $v_k = e_s$ (the sth standard basis vector) and $b = Ee_s$. Perform a product E' of elementary row operations on b so that $E'b = e_k$. Set E to $E'E$ and $t = k$.
(3) (Iteration loop) Form Av_k. Set $b = EAv_k$. If $k < n$ and the jth entry of b is nonzero for some $j > k$, perform a product E' of elementary row operations on b so that $E'b = e_{k+1}$. Set E to $E'E$, $v_{k+1} = Av_k$, k to $k + 1$, and repeat (3).
(4) (Form next factor of $C_A(x)$) We have $Av_k = \sum_{i=1}^{k} b_i v_i$, where b_i is the ith entry of b. Set

$$C_A(x) \text{ to } C_A(x)\left(x^{k-t+1} - \sum_{i=t}^{k} b_i x^{i-t}\right).$$

(5) (Test for end) If $k < n$, then set k to $k + 1$ and go to (2).
(6) (End) Output $C_A(x)$ (if needed, output $D = [v_1 \ \ldots \ v_n]$ and $E = D^{-1}$).

To show the algorithm in action, let's apply it to the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix}.$$

whose tableau is in FIGURE 4.

To clarify what happens to the transformed v_i's as the algorithm progresses, we'll maintain them. The algorithm begins with E equal to the 4×4 identity matrix:

$t=1$	1	0	0	0	1		$t=1$	1	0	0	0	1	1						
	0	1	0	0	0			0	1	0	0	0	2						
	0	0	1	0	0	reduce →		0	0	1	0	0	3	reduce →					
	0	0	0	1	0			0	0	0	1	0	4						
			E		Ev_1				E			Ev_1	Ev_2						

$t=1$	1	−.5	0	0	1	0	10		$t=1$	1	−2	1	0	1	0	0	0				
	0	.5	0	0	0	1	20			0	−2.5	2	0	0	1	0	20				
reduce →	0	−1.5	1	0	0	0	−10		reduce →	0	.15	−.1	0	0	0	1	16				
	0	−2	0	1	0	0	−20			0	1	−2	1	0	0	0	0				
		E			Ev_1	Ev_2	Ev_3				E			Ev_1	Ev_2	Ev_3	EAv_3				

At this point dependence occurs. $C_A(x)$ is updated to $x^3 - 16x^2 - 20x$ and we must choose a

vector independent of v_1, v_2, and v_3. The second column of the current E is the first column containing a nonzero entry in row 4 (since E is invertible, no row of E can be all zeros). Hence $v_4 = e_2$ is our next vector. The algorithm continues:

	1	−2	1	0	1	0	0	−2		1	0	−3	2	1	0	0	0	0
	0	−2.5	2	0	0	1	0	−2.5		0	0	−3	2.5	0	1	0	0	.5
	0	.15	−.1	0	0	0	1	.15	reduce →	0	0	.2	−.15	0	0	1	0	.05
$t=4$	0	1	−2	1	0	0	0	1	$t=4$	0	1	−2	1	0	0	0	1	0
		E			Ev_1	Ev_2	Ev_3	Ev_4				E		Ev_1	Ev_2	Ev_3	Ev_4	EAv_4

Dependence occurs again and $C_A(x)$ is updated to $(x^3 - 16x^2 - 20x)x$. We have a basis now so the algorithm stops. The final E is the inverse of $D = [v_1 \ \dots \ v_4]$ because it reduces D to the identity matrix $[Ev_1 \ \dots \ Ev_4]$. Note that the vectors Ev_i remain equal to e_i during later changes in E and so need not be saved.

The vectors Av_i constructed by the algorithm are useful for rapid calculation of eigenvectors. The algorithm executes step (2) k times. For each $i = 1, \dots, k$ a vector w_i is chosen at step (2) and then step (3) is executed repeatedly until dependence occurs. Next, step (4) is executed and constructs a polynomial $P_i(x)$. By the theorem below, the vector $P_1(A) \ \dots \ P_i(A)w_i$ is 0 for each $i = 1, \dots, k$. Now suppose that a is an eigenvalue of A. Then for each i for which $P_i(x)$ has the root a, the polynomial $P_1(x) \dots P_i(x)$ has the form $(x-a)^{t_i}Q_i(x)$, with $Q_i(x)$ relatively prime to $x - a$ and $t_i \geqslant 1$. Let s be the largest integer such that the vector $v = (A - aI)^s Q_i(A)w_i$ is not 0 (such a vector exists). Then v must be an eigenvector for a. We get in this way as many eigenvectors for a as there are $P_i(x)$ with a as a root. It can be shown that this recipe computes a basis of eigenvectors if and only if A is diagonalizable. The argument, which is omitted, uses facts about minimum polynomials of vectors and matrices.

As an example we apply the recipe to the matrix A whose tableau is in FIGURE 1. Here the polynomials are $P_1(x) = x^2 + x - 2$, $P_2(x) = x - 1$, and $P_3(x) = x + 2$, with vectors $w_1 = v_1$, $w_2 = v_3$, and $w_3 = v_4$. The eigenvalues of A are -2 and 1. Working first with the eigenvalue -2, the polynomial $P_1(x) = x^2 + x - 2$ has the root -2. Then

$$0 = (A^2 + A - 2)w_1 = (A + 2I)(A - I)w_1,$$

but $v = (A - I)w_1$ is not 0. Thus v is an eigenvector for -2, and, using the vectors in the tableau, the computation is easy:

$$(A - I)w_1 = (A - I)v_1 = Av_1 - v_1 = (1, 1, -1, 1)^T.$$

The polynomial $P_3(x) = x + 2$ also has the root -2. We have the vector equation

$$0 = P_1(A)P_2(A)P_3(A)w_3 = (A + 2I)^2(A - I)^2 v_4.$$

The vector $(A + 2I)(A - I)^2 v_4$ is 0 but $(A - I)^2 v_4$ is not:

$$(A - I)^2 v_4 = A^2 v_4 - 2Av_4 + v_4 = (-20, -2, 14, -5)^T.$$

This eigenvector for -2 is independent of the first. Next, working with the eigenvalue 1, the polynomial $P_1(x)$ has the root 1. We have

$$0 = (A^2 + A - 2I)w_1 = (A - I)(A + 2I)v_1,$$

so $(A + 2I)v_1 = (4, 1, -1, 1)^T$ is an eigenvector for 1. The eigenvalue 1 is also a root of $P_2(x)$. We have

$$0 = P_1(A)P_2(A)w_2 = (A - I)^2(A + 2I)v_3,$$

obtaining

$$(A - I)(A + 2I)v_3 = A^2 v_3 + Av_3 - 2v_3 = (-12, -3, 3, -3)^T,$$

a second eigenvector for 1. But this vector is a multiple of the eigenvector $(4, 1, -1, 1)^T$ and so A is

not diagonalizable. The recipe has provided us, however, with a basis for each eigenspace.

The recipe won't always produce a basis for each eigenspace because the Frobenius matrix

$$A = \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 1 & 0 & 1 \\ \hline 0 & 0 & 0 \end{array}\right]$$

is a counterexample. For this matrix the recipe produces only $(0,1,0)^T$ as an eigenvector for the eigenvalue 0 whose eigenspace is clearly of dimension 2. Note that the recipe is feasible only when the polynomials $P_i(x)$ are completely factored.

Our characteristic polynomial algorithm is similar to one due to A. M. Danilevskii [**1**]. Both compute a Frobenius matrix F similar to a given $n \times n$ matrix A. Danilevskii's method constructs a sequence $A^{(k)}$ of matrices similar to A by simple similarity operations. They terminate in F only in the case that F is actually a companion matrix, a common situation in practice. If his algorithm cannot continue after $A^{(s)}$, say, then the form of $A^{(s)}$ allows restarting his algorithm with a square submatrix of $A^{(s)}$. In this way Danilevskii's algorithm can handle any $n \times n$ matrix. It is interesting to note that when F is a companion matrix, the computations in step (3) of our algorithm are virtually identical to Danilevskii's computations.

We have now simplified two tedious computations of linear algebra. We have a determinant-free algorithm for the characteristic polynomial of a square matrix and a recipe for eigenvectors that avoids the odious computation of several null spaces. To further weaken the determinant's grip on eigenvalues we close with a few general remarks on the data the algorithm generates and a proof of the Cayley-Hamilton theorem, sans determinants.

Let A be an $n \times n$ real matrix. Assume that the algorithm has been applied to A and produced a basis $v_1, \ldots, v_n$ of R^n and polynomials $P_1(x), \ldots, P_k(x)$. Let w_i be that standard basis vector chosen at step (2) of the algorithm to begin the calculation of $P_i(x)$. Finally, let W_0 be the zero subspace, W_k the whole space R^n, and, for $i = 1, \ldots, k-1$, let W_i be the subspace spanned by all those v_j that precede w_{i+1} in the ordered basis $v_1, \ldots, v_n$. Then from steps (3) and (4) of the algorithm we have

(a) *For each* i $(i = 1, \ldots, k)$ *the vector* $P_i(A)w_i$ *is in* W_{i-1}.

Also, since every vector in R^n is a unique linear combination of the basis $v_1, \ldots, v_n$ which satisfies condition (*), we have

(b) *Every vector* v *in* R^n *has a unique representation in the form*

$$v = Q_1(A)w_1 + \cdots + Q_k(A)w_k,$$

with each $Q_i(x)$ *either zero or of degree less than that of* $P_i(x)$.

We prove

THEOREM. *For each* i $(i = 1, \ldots, k)$ *and every vector* w *in* W_i, *the vector* $v = P_1(A) \ldots P_i(A)w$ *is zero.*

Proof. We use induction on i. For $i = 1$ let w be any vector in W_i. By (b), w has the form $w = Q(A)w_1$ for some polynomial $Q(x)$. Therefore, using (a),

$$P_1(A)w = P_1(A)Q(A)w_1 = Q(A)P_1(A)w_1 = Q(A)0 = 0.$$

Now assume that the theorem is true for all $j \leqslant i$ and examine the vector $w^* = P_1(A) \ldots P_{i+1}(A)w$, with w in W_{i+1}. Since w is in W_{i+1} and (b) holds, w has the form

$$w = Q_1(A)w_1 + \ldots + Q_{i+1}(A)w_{i+1}.$$

Hence by induction, w^* reduces to

$$w^* = P_1(A) \ldots P_{i+1}(A)Q_{i+1}(A)w_{i+1}.$$

By (a), $w' = P_{i+1}(A)w_{i+1}$ is in W_i. Thus by induction again,

$$w^* = Q_{i+1}(A)P_1(A)\dots P_i(A)w' = Q_{i+1}(A)0 = 0.$$

COROLLARY (Cayley-Hamilton Theorem). $C_A(A) = 0$.

Proof. It suffices to prove $C_A(A)v$ is the zero vector for every v in R^n. But this is just the Theorem for $i = k$, since v is in $W_k = R^n$ and $C_A(x) = P_1(x)\dots P_k(x)$.

References

[1] D. K. Fadeeva and V. N. Fadeeva, Computational Methods of Linear Algebra, translated by Robert C. Williams, Freeman, San Francisco, 1963. (Originally published 1960, in Russian.)

[2] Hoffman and Kunze, Linear Algebra, Prentice-Hall, Englewood Cliffs, N.J., 1961.

[3] Evar D. Nering, Linear Algebra and Matrix Theory, 2nd ed., Wiley, New York, 1970.

[4] Rudolf Zurmuhl, Matrizen, 4th ed., Springer-Verlag, 1964.

Names of Functions: The Problems of Trying for Precision

R. P. BOAS

Northwestern University

Evanston, IL 60201

To students

Are you being confused by a textbook or a teacher who insists that $f(x)$ is the value of a function at the point x, whereas f is the name of the function? I wouldn't be surprised if you were, especially if your text goes on (most of them do) to present formulas like $\frac{d}{dx}\sin x = \cos x$ or $\frac{d}{dx}x^2 = 2x$, apparently without reflecting that, if $\sin x$ is the value of the sine function at x, then $\frac{d}{dx}\sin x$ is, strictly speaking, a meaningless formula. (We differentiate functions, not numbers.) You could get around this particular problem by writing $\frac{d}{dx}\sin = \cos$; but what *are* you to do about the function that has the value x^2 at x? You can't very well write ^{2} as the name of the function. What most people—you (probably), your teacher (quite likely), and your pocket calculator (almost certainly)—want to call the function is x^2, but this would violate the teacher's principles. I want to propose a simple way out of this bind.

First let's recall that there is one kind of function that has a well-established and unambiguous notation already: a sequence. The sequence $\{n^2\}_1^\infty$ is the function whose value at n is n^2, the domain being understood to be the set of positive integers. If you need a different domain you can write symbols like $\{n^2\}_3^\infty$ or $\{n^2\}_{10}^{100}$ or $\{n^2\}_{\text{odd } n}$. The identity sequence is $\{n\}_1^\infty$. In other words, we know a sequence when we meet one because it comes with its domain and its value at each domain point—which is just what every definition of a function is supposed to provide.

Why then don't we denote the function that has the value x^2 at the real number x by $\{x^2\}_R$ or $\{x^2\}_1^\infty$, and so on? The identity is $\{x\}$, and $\frac{d}{dx}\{x\} = \{1\}$, as it should be. This would be unambiguous, compact, and would require no special symbols to be learned.

If you use any notation a great deal you tend to abbreviate it, just as the word "radix" (meaning "root") got cut down to the modern square root sign. If your usual domain for functions is all real numbers, you will probably just write $\{\sin x\}$ instead of $\{\sin x\}_{-\infty<x<\infty}$. After doing that for a while, you will find yourself dropping the braces too—and there you will be with functions named x^2, $\sin x$, and so on, just as the keys on the calculator, or the tables in the textbook say. The difference is that you now ought to be able, on demand, to explain the

$$w^* = Q_{i+1}(A)P_1(A)\dots P_i(A)w' = Q_{i+1}(A)0 = 0.$$

COROLLARY (Cayley-Hamilton Theorem). $C_A(A) = 0$.

Proof. It suffices to prove $C_A(A)v$ is the zero vector for every v in R^n. But this is just the Theorem for $i = k$, since v is in $W_k = R^n$ and $C_A(x) = P_1(x)\dots P_k(x)$.

References

[1] D. K. Fadeeva and V. N. Fadeeva, Computational Methods of Linear Algebra, translated by Robert C. Williams, Freeman, San Francisco, 1963. (Originally published 1960, in Russian.)

[2] Hoffman and Kunze, Linear Algebra, Prentice-Hall, Englewood Cliffs, N.J., 1961.

[3] Evar D. Nering, Linear Algebra and Matrix Theory, 2nd ed., Wiley, New York, 1970.

[4] Rudolf Zurmuhl, Matrizen, 4th ed., Springer-Verlag, 1964.

Names of Functions: The Problems of Trying for Precision

R. P. BOAS

Northwestern University

Evanston, IL 60201

To students

Are you being confused by a textbook or a teacher who insists that $f(x)$ is the value of a function at the point x, whereas f is the name of the function? I wouldn't be surprised if you were, especially if your text goes on (most of them do) to present formulas like $\frac{d}{dx}\sin x = \cos x$ or $\frac{d}{dx}x^2 = 2x$, apparently without reflecting that, if $\sin x$ is the value of the sine function at x, then $\frac{d}{dx}\sin x$ is, strictly speaking, a meaningless formula. (We differentiate functions, not numbers.) You could get around this particular problem by writing $\frac{d}{dx}\sin = \cos$; but what *are* you to do about the function that has the value x^2 at x? You can't very well write ^{2} as the name of the function. What most people—you (probably), your teacher (quite likely), and your pocket calculator (almost certainly)—want to call the function is x^2, but this would violate the teacher's principles. I want to propose a simple way out of this bind.

First let's recall that there is one kind of function that has a well-established and unambiguous notation already: a sequence. The sequence $\{n^2\}_1^\infty$ is the function whose value at n is n^2, the domain being understood to be the set of positive integers. If you need a different domain you can write symbols like $\{n^2\}_3^\infty$ or $\{n^2\}_{10}^{100}$ or $\{n^2\}_{\text{odd } n}$. The identity sequence is $\{n\}_1^\infty$. In other words, we know a sequence when we meet one because it comes with its domain and its value at each domain point—which is just what every definition of a function is supposed to provide.

Why then don't we denote the function that has the value x^2 at the real number x by $\{x^2\}_R$ or $\{x^2\}_1^\infty$, and so on? The identity is $\{x\}$, and $\frac{d}{dx}\{x\} = \{1\}$, as it should be. This would be unambiguous, compact, and would require no special symbols to be learned.

If you use any notation a great deal you tend to abbreviate it, just as the word "radix" (meaning "root") got cut down to the modern square root sign. If your usual domain for functions is all real numbers, you will probably just write $\{\sin x\}$ instead of $\{\sin x\}_{-\infty<x<\infty}$. After doing that for a while, you will find yourself dropping the braces too—and there you will be with functions named x^2, $\sin x$, and so on, just as the keys on the calculator, or the tables in the textbook say. The difference is that you now ought to be able, on demand, to explain the

difference between $\sin x$, meaning a number, and $\sin x$, meaning a function—which is probably what your teacher was hoping for all along.

To teachers

Are you quite, quite sure that when you make students learn that f is a function and $f(x)$ is a value of a function, they are really learning what functions and values are? Or are they just parroting words? I've seen plenty of students who could give you a letter-perfect definition of a derivative but were helpless if you asked them what $\lim_{x\to\pi}\frac{\sin x}{x-\pi}$ is. How many of them can tell you what is the sine of the angle whose sine is x?

Maybe you deplore the calculator people's putting "x^2" on the squaring key. Your most impassioned arguments aren't going to stop them. "If you can't lick 'em, join 'em."

Abbreviations are an important mathematical tool. If we weren't allowed to use them, we'd still be writing $\lim_{y\to x}\frac{f(y)-f(x)}{y-x}$ instead of $f'(x)$. Bourbaki calls this sort of thing "abus de langage;" ordinary people call it shorthand. Admittedly it has disadvantages. The notation $\sin^2 x$ (Gauss is on record as detesting it) is shorthand for $(\sin x)^2$. The notation $\log^2 x$ can mean either $(\log x)^2$ or $\log(\log(x))$. However, people seem to prefer ambiguous notations to cumbersome ones.

In integration theory, a Lebesgue integrable function is not a function at all; it is an equivalence class of functions. Do we indicate this by our notation? Not that I've ever noticed.

I have heard a linguist claim that the eccentric orthography of English serves a useful purpose besides making it possible to have spelling bees: it helps us pick out the correct meanings of words that we see. Perhaps it's just as well that we *don't* use strictly consistent notations.

Paradoxes

Alas, poor Zeno!

Achilles and the tortoise
stomping 'round his bed
and that confounded arrow
whizzing past his head.

"All Cretans are liars."

Epimenides proposed it
to vex his friends (and you!):
If it's true it's clearly false
while if it's false it could be true.

The village barber

Shaves those and only those whose
razors stay on the shelf.
This makes a hairy problem:
does the barber shave himself?

Warning

The pitfalls of semantics
and logic's fickle weather—
are you prepared to cope with
"is not" and "is" together?

—KATHARINE O'BRIEN

difference between $\sin x$, meaning a number, and $\sin x$, meaning a function—which is probably what your teacher was hoping for all along.

To teachers

Are you quite, quite sure that when you make students learn that f is a function and $f(x)$ is a value of a function, they are really learning what functions and values are? Or are they just parroting words? I've seen plenty of students who could give you a letter-perfect definition of a derivative but were helpless if you asked them what $\lim_{x\to\pi}\frac{\sin x}{x-\pi}$ is. How many of them can tell you what is the sine of the angle whose sine is x?

Maybe you deplore the calculator people's putting "x^2" on the squaring key. Your most impassioned arguments aren't going to stop them. "If you can't lick 'em, join 'em."

Abbreviations are an important mathematical tool. If we weren't allowed to use them, we'd still be writing $\lim_{y\to x}\frac{f(y)-f(x)}{y-x}$ instead of $f'(x)$. Bourbaki calls this sort of thing "abus de langage;" ordinary people call it shorthand. Admittedly it has disadvantages. The notation $\sin^2 x$ (Gauss is on record as detesting it) is shorthand for $(\sin x)^2$. The notation $\log^2 x$ can mean either $(\log x)^2$ or $\log(\log(x))$. However, people seem to prefer ambiguous notations to cumbersome ones.

In integration theory, a Lebesgue integrable function is not a function at all; it is an equivalence class of functions. Do we indicate this by our notation? Not that I've ever noticed.

I have heard a linguist claim that the eccentric orthography of English serves a useful purpose besides making it possible to have spelling bees: it helps us pick out the correct meanings of words that we see. Perhaps it's just as well that we *don't* use strictly consistent notations.

Paradoxes

Alas, poor Zeno!

Achilles and the tortoise
stomping 'round his bed
and that confounded arrow
whizzing past his head.

"All Cretans are liars."

Epimenides proposed it
to vex his friends (and you!):
If it's true it's clearly false
while if it's false it could be true.

The village barber

Shaves those and only those whose
razors stay on the shelf.
This makes a hairy problem:
does the barber shave himself?

Warning

The pitfalls of semantics
and logic's fickle weather—
are you prepared to cope with
"is not" and "is" together?

—KATHARINE O'BRIEN

PROBLEMS

LEROY F. MEYERS, Editor
G. A. EDGAR, Associate Editor
The Ohio State University

Proposals

To be considered for publication, solutions should be mailed before October 1, 1983.

1170. In triangle ABC, the bisectors of the angles A, B, and C are the segments AP, BQ, and CR, respectively. If $AB = 4$, $AC = 5$, and $BC = 6$, find the size of $\angle QPR$. [*John P. Hoyt, Lancaster, Pennsylvania.*]

1171. Find all functions f such that $f''(x) + (f'(x))^2 = b^2$, where b is a given constant. Which of these solutions are polynomials? [*Mark Kantrowitz, student, Brookline, Massachusetts.*]

1172. Determine the number of real solutions x $(0 \leqslant x \leqslant 1)$ of the equation

$$(x^{m+1} - a^{m+1})(1-a)^m = \{(1-a)^{m+1} - (1-x)^{m+1}\}a^m,$$

where $0 \leqslant a \leqslant 1$ and m is a positive integer. [*M. S. Klamkin, University of Alberta, Canada.*]

1173. (a) Two positive integers are chosen. The sum is revealed to logician A, and the sum of squares is revealed to logician B. Both A and B are given this information and the information contained in this sentence. The conversation between A and B goes as follows: B starts.

B: "I can't tell what the two numbers are."
A: "I can't tell what the two numbers are."
B: "I can't tell what the two numbers are."
A: "I can't tell what the two numbers are."
B: "I can't tell what the two numbers are."
A: "I can't tell what the two numbers are."
B: "Now I can tell what the two numbers are."

What are the two numbers?

(b) When B first says he cannot tell what the two numbers are, A receives a large amount of information. But when A first says that he cannot tell what the two numbers are, B already knows that A cannot tell what the two numbers are. What good does it do B to listen to A? [*Thomas S. Ferguson, University of California, Los Angeles.*]

ASSISTANT EDITORS: DANIEL B. SHAPIRO and WILLIAM A. MCWORTER, JR., *the Ohio State University.*

We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) will be placed next to a problem number to indicate that the proposer did not supply 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. It is not necessary to submit duplicate copies.

Send all communications to the problems department to Leroy F. Meyers, Mathematics Department, The Ohio State University, 231 W. 18 Ave., Columbus, Ohio 43210.

1174. The sixth morning problem of the 1953 Putnam Competition was to determine the limit of the sequence

$$\left(\sqrt{k}\,,\sqrt{k-\sqrt{k}}\,,\sqrt{k-\sqrt{k+\sqrt{k}}}\,,\sqrt{k-\sqrt{k+\sqrt{k-\sqrt{k}}}}\,,\ldots\right)$$

in the specific instance when $k = 7$. (The limit is 2.) For which real numbers A does there exist a k such that the sequence converges to A? For such A, write k explicitly in terms of A. [*Thomas P. Dence, California State University, Los Angeles.*]

Quickies

Solutions to Quickies appear at the conclusion of the Problems section.

Q685. If two altitudes of a plane triangle are congruent, then the triangle must be isosceles. Does the same result hold for a convex spherical triangle? [*M. S. Klamkin, University of Alberta, Canada.*]

Solutions

A New Quadratic Factorization

May 1982

1145. Let $f(x) = ax^2 + bx + c$ be a quadratic polynomial with integral coefficients, where $a \neq 0$. Show that:

(a) if $f(x)$ is factorable into linear factors with integral coefficients, then there are integers d and e such that $d + e = b$ and $de = ac$; and

(b) if the integers d and e satisfy $d + e = b$ and $de = ac$, then

$$f(x) = \frac{ax + d}{(a,d)} \cdot \frac{ax + e}{a/(a,d)},$$

where each of the linear factors has integral coefficients. [*Kenneth A. Brown, Jr., Nova High School, Fort Lauderdale, Florida.*]

Solution: (a) If $f(x)$ is factorable, then $f(x) = (px + q)(mx + n)$ for some integers p, q, m, n, by Gauss's theorem. Then

$$f(x) = pmx^2 + (qm + pn)x + qn.$$

Let $d = qm$ and $e = pn$. Then

$$de = qm \cdot pn = pm \cdot qn = ac \quad \text{and} \quad d + e = qm + pn = b.$$

(b) We are given that $f(x) = ax^2 + bx + c$, where a, b, and c are integers and $a \neq 0$, and that d and e are integers which satisfy $d + e = b$ and $de = ac$. Let $m = (a, d)$. Then $a = pm$ and $d = qm$ for some relatively prime integers p and q.

Since $de = ac$, we have

$$c = de/a = qme/(pm) = qe/p \quad \text{and} \quad e = ac/d = pmc/(qm) = pc/q.$$

Then $(p, q) = 1$ implies that $p|e$ and $q|c$, i.e., that $e = pn$ and $c = qy$ for some integers n and y. Using $de = ac$ again, we find $qm \cdot pn = pm \cdot qy$, or $n = y$, so that $c = qn$. Now

$$f(x) = ax^2 + bx + c = ax^2 + (d + e)x + c = pmx^2 + (qm + pn)x + qn$$

and $pm = a \neq 0$. Hence

$$f(x) = \frac{p^2m^2x^2 + (pm \cdot qm + pm \cdot pn)x + pm \cdot qn}{pm} = \frac{(pmx + qm)(pmx + pn)}{pm}$$

$$= \frac{pmx + qm}{m} \cdot \frac{pmx + pn}{p} = \frac{ax + d}{(a, d)} \cdot \frac{ax + e}{a/(a, d)},$$

where the coefficients of each of the linear factors are integers since

$$(a, d) = m|pm = a, \quad (a, d) = m|qm = d, \quad a/(a, d) = p|pm = a,$$

and

$$a/(a, d) = p|pn = e.$$

RUTH KOELLE
Somerset County College

Also solved by Walter Blumberg, James Bolte, Alberto Facchini (Italy), M. A. Fitting, Bern-Kirchenfeld Literargymnasium Problem Solving Group (Switzerland), Chico Problem Group, Paul M. Harms, George C. Harrison, Joel Levy, Henry S. Lieberman, Gary Ling, Joe Marchione, Vania D. Mascioni (student, Switzerland), Bob Prielipp, John Putz, Stanley Rabinowitz, J. M. Stark, Mikel Thompson, Leon Warman (student), Bella Wiener, Michael Woltermann, and the proposer. Solutions to part (a) by Duane M. Broline, W. C. Igips, Mark Kantrowitz (student), L. Kuipers (Switzerland), Herby McKaig (student, Canada), Dan Rawsthorne, Sahib Singh, and John Samoylo. Many purported solutions to part (b) contained no proof that the coefficients were integers.

Interlaced Repeated Exponentials — May 1982

1146. Let $f(x) = 2^x$ and $g(x) = 3^x$ for all real x, and indicate iteration by superscripts. It is easy to check that $f(1) < g(1) < f^2(1) < f^3(1) < g^2(1) < f^4(1) < g^3(1) < f^5(1) < g^4(1)$. Is it true that $f^n(1) < g^{n-1}(1) < f^{n+1}(1)$ for all $n \geq 3$? [*James Propp, undergraduate, Harvard College.*]

Solution: Yes, it is true.

The first inequality follows easily by induction, since

$$f^{n+1}(1) = 2^{f^n(1)} < 2^{g^{n-1}(1)} < 3^{g^{n-1}(1)} = g^n(1)$$

if $f^n(1) < g^{n-1}(1)$.

To prove the second inequality, it suffices to prove that

$$g^{n-1}(1)/f^{n+1}(1) < \frac{1}{2}$$

for all $n \geq 3$. This can also be proved by induction, starting from

$$g^2(1)/f^4(1) = 27/65536 < \frac{1}{2}$$

and observing that from $g^{n-1}(1)/f^{n+1}(1) < \frac{1}{2}$ follows

$$\frac{g^n(1)}{f^{n+2}(1)} = \frac{3^{g^{n-1}(1)}}{2^{f^{n+1}(1)}} < \frac{3^{\frac{1}{2}f^{n+1}(1)}}{2^{f^{n+1}(1)}} = \left(\frac{\sqrt{3}}{2}\right)^{f^{n+1}(1)} < \left(\frac{\sqrt{3}}{2}\right)^{f^4(1)} < \frac{1}{2}.$$

VANIA D. MASCIONI, student
Swiss Federal Institute of Technology

Also solved by Duane M. Broline, Stephen D. Bronn, Alan Edelman (student), John L. Leonard, Joel Levy, Jan Söderkvist (student), J. M. Stark, Michael Woltermann, and the proposer. Partial solution by Chauncey Wayne Lovell (student). One person misinterpreted the problem.

Re-exploring a Rectangle Problem — May 1982

1147. Marion Walter, in Exploring a Rectangle Problem, this MAGAZINE, 54 (1981) 131–134, discussed variations on the problem *If O is in the interior of the rectangle $ABCD$ and $|OA| = a$, $|OB| = b$, and $|OC| = c$, what is $|OD|$?* and suggested that it leads to other questions. Here is one such question. For a given triple (a, b, c), what is the maximum area of the rectangle $ABCD$? [*James S. Robertson, Rochester, Minnesota.*]

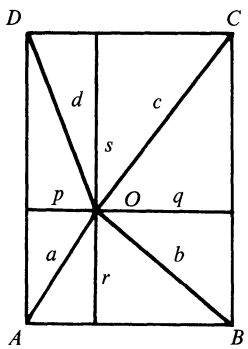


FIGURE 1

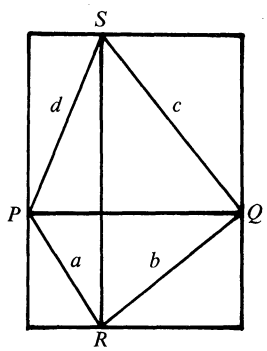


FIGURE 2

Solution: Figure 1 illustrates the situation described in the problem statement. Figure 2 is obtained by using the other diagonals of the small rectangles to form a new linkage $PRQS$ whose sides have lengths a, b, c, d, where, as in Walter's article, $d = \sqrt{a^2 + c^2 - b^2}$. (Note that a rectangle $ABCD$ exists if and only if $a^2 + c^2 > b^2$.) With the constraint $PQ \perp RS$, the new linkage generates the same family of rectangles $ABCD$ as the original linkage. The area of quadrilateral $PRQS$ is clearly $\frac{1}{2}$ that of $ABCD$, since each small rectangle is halved by its diagonals.

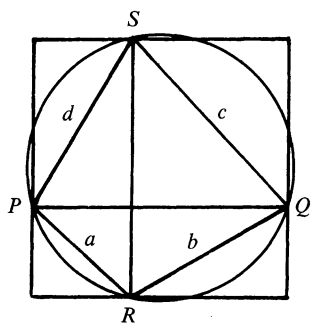


FIGURE 3

The maximum area of a quadrilateral $PRQS$ with given side lengths a, b, c, d (in order) occurs when $PRQS$ can be inscribed in a circle, as shown in Figure 3. (See Schultze and Sevenoak, *Plane Geometry*, Macmillan, New York, 1922, pp. 269,272,273; and Ivan Niven, *Maxima and minima without calculus*, Dolciani Mathematical Expositions no. 6, 1981, pp. 8–9,53, for a proof of this.) Then PQ and RS are chords of a circle, and so $pq = rs$. From this relation and the relation $a^2 + c^2 = b^2 + d^2$, the maximum area is readily calculated in terms of a, b, and c as follows.

Since

$$r^2 = a^2 - p^2, \quad q^2 = b^2 - r^2 = b^2 - a^2 + p^2, \quad \text{and} \quad s^2 = d^2 - p^2,$$

from $p^2q^2 = r^2s^2$ we deduce

$$p^2(b^2 - a^2 + p^2) = (a^2 - p^2)(d^2 - p^2),$$

and so $p = ad/(a^2 + c^2)^{1/2}$. Similarly,

$$q = bc/(a^2 + c^2)^{1/2}, r = ab/(a^2 + c^2)^{1/2}, \quad \text{and} \quad s = cd/(a^2 + c^2)^{1/2}.$$

Hence the maximal area for $ABCD$ is given by

$$(p+q)(r+s) = \frac{(ad+bc)(ab+cd)}{a^2+c^2}$$

$$= \frac{(a^2+c^2)bd + (b^2+d^2)ac}{a^2+c^2} = ac + bd,$$

a remarkably simple expression.

There is also a simple formula for the radius of the circle $PRSQ$. We have

$$R = \sqrt{\left(\frac{p+q}{2}\right)^2 + \left(\frac{r-s}{2}\right)^2}$$

(since the center of the circle must lie on the perpendicular bisector of the chords PQ and RS), from which, with $p^2 + q^2 + r^2 + s^2 = a^2 + c^2$ and $pq = rs$, we obtain $R = \frac{1}{2}\sqrt{a^2 + c^2}$. Thus R depends only on a and c.

JAMES S. ROBERTSON
Rochester, Minnesota

Also solved by Benny N. Cheng (student), George C. Harrison, Victor Hernandez (Spain), Geoffrey A Kandall, L. Kuipers (Switzerland), Vania D. Mascioni (student, Switzerland), J. M. Stark, and Harry Zaremba.

All but two of the solvers found the maximum by setting a derivative equal to 0. However, not everyone remembered that setting a derivative equal to 0 provides only a necessary, not a sufficient, condition for a maximum.

Swimming Date — May 1982

1148. Tom Trotter from Toronto was a guest of his friend Paul Porter of Peoria at a Fourth-of-July bicentennial celebration. The following conversation took place:

Paul: When will you be back here again?

Tom: Not this year, but I'll be back before the Fourth of July eight years from now, and I'll look you up on the fourth day of the month.

Paul: If you tell me the day of the week for that day, and the year, will I be able to figure out the month when you'll be here?

Tom: No, but if I tell you just the day of the week for the thirtieth of that same month, you'll be able to figure it all out, even if I don't tell you the year!

Paul: Right you are! I'll have my swimming pool open.

When will Tom be visiting Paul? [*John M. H. Olmsted, Southern Illinois University.*]

Solution: Let $x \in \{0, 1, \ldots, 6\}$ represent the day of the week (with, say, Mon $\leftrightarrow$ 1, Tues $\leftrightarrow$ 2,..., Sun $\leftrightarrow$ 0) on which Jan. 4 occurs in any given year. The table below shows the day of the week on which the fourth of each month occurs for the rest of the year. (Addition is performed modulo 7.) For example, since January has $31 \equiv 3 \pmod 7$ days, February 4 occurs 3 days later in the week than January 4.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Non-leap yrs.	x	~~$x+3$~~	$x+3$	$x+6$	~~$x+1$~~	~~$x+4$~~	$x+6$	~~$x+2$~~	$x+5$	x	$x+3$	$x+5$
Leap years	x	~~$x+3$~~	$x+4$	x	~~$x+2$~~	~~$x+5$~~	x	$x+3$	$x+6$	~~$x+1$~~	$x+4$	$x+6$

The table shows that no other fourth of the month occurs on the same day of the week as May 4, June 4 and August 4 in non-leap years and, similarly, as May 4, June 4 and October 4 in leap years. But since it is given that the arrival date cannot be determined from the day of the week and the year, these dates are eliminated. February 4 is eliminated, since February never has 30 days. Of those remaining, only August 4 in leap years occurs on a day of the week occupied by no other fourth. Since Tom will not visit again in 1976 or after July 4 in 1984, he will visit on August 4, 1980.

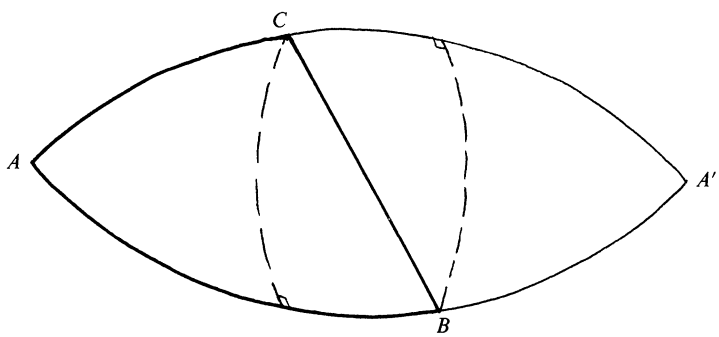
JOHN PUTZ
Alma College

Also solved by Clayton W. Dodge, Philip M. Dunson, Milton P. Eisner, Carrie Fujioka & Katsuhiko Nakano (students), Chico Problem Group, Bruce Hoffmeister, Kathy Jones, L. Kuipers (Switzerland), Kay P. Litchfield, Vania D. Mascioni (student, Switzerland), Dennis Wildfogel, and the proposer.

Answers

Solutions to the Quickies which appear near the beginning of the Problems section.

Q685. No. Consider a spherical triangle cut off from a lune AA' by an arc BC through its center. By symmetry, the altitudes from B and C are congruent.



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.

Frazier, Ian, *The killion*, New Yorker (6 September 1982) 32-33.

"Through a computer error Marcie's check was made out to an extremely high number....one killion dollars. The killion, as every mathematician knows, is a number so big that it kills you." *Sic perit* Marcie, in this farcical little piece. Enjoy it, but don't tell your students about it--no sense adding to math anxiety.

Stein, Kathleen, *The fractal cosmos*, Omni 5:5 (February 1983) 62-66, 71, 115.

Alan Norton's graphics team with Benoit Mandelbrot's fractals to produce marvelous new shapes which illustrate this popular-style explanation of what fractals are about.

Wrinkles in four dimensions, Scientific American (October 1982) 80-82.

Account of discovery by Michael Freedman (San Diego) of a "fake R^4," a crinkled Euclidean spacetime wildly different from the smoothly curved manifold R^4 physicists and mathematicians have grown used to.

Weiss, George H., *Random walks and their applications*, American Scientist 71:1 (January-February 1983) 65-71.

Summarizes mathematical results on random walks and notes details of their applications in describing configurational properties of polymers, energy transfer in amorphous solids, and motion of bacteria on surfaces.

Hethcote, Herbert W., *Measles and rubella in the United States*, American Journal of Epidemiology 117 (1983) 2-13.

Uses a mathematical model (integrodifferential equations) to examine the practical effects of different vaccination strategies on measles and rubella, diseases which may soon be eradicated in the U.S.

Harris, S., What's So Funny About Computers?, Kaufmann, 1983; $6.95 (P).

Zany cartoons! Sample: Couple viewing work in art gallery, man says to woman: "Not bad for a computer, but the chimpanzee's work had more feeling."

Reid, Constance, Neyman--from Life, Springer-Verlag, 1982, 298 pp, $19.80.

Constance Reid has already enriched our feeling for two famous mathematicians in *Hilbert* and *Courant in Göttingen and New York*. In this new book, she offers another sensitive masterpiece, describing the life of the premier statistician Jerzy Neyman (1894-1981).

Moore, Gregory H., Zermelo's Axiom of Choice: Its Origins, Development, and Influence, Springer-Verlag, 1982; xiv + 410 pp.

This is a remarkably thorough and learned history of the Axiom of Choice, offering insight into the spirit of the development of twentieth-century mathematics. Eleven tables at the end give deductive relations concerning "the Axiom" and its relatives.

Eves, Howard, An Introduction to the History of Mathematics, 5th ed., Saunders, 1983; xviii + 593 pp.

New edition of an old favorite, with extensive revision and augmentation of the second half, particularly in regard to biographical information.

Loweke, George P., The Lore of Prime Numbers, Vantage, 1982, viii + 259 pp, $17.95.

Intriguing and in-depth investigation of prime numbers, at a semi-popular level (i.e., the reader needs to be able to follow algebra). Occasional misstatements occur (e.g., "The Greeks had no methods of multiplication"), and there is no mention of the work of J. P. Jones, J. Robinson, and J. Matijaševič about prime-producing polynomials.

Ralston, Anthony, and Reilly, Edwin D., Jr., Encyclopedia of Computer Science and Engineering, 2nd ed., Van Nostrand Reinhold, 1983; xxix + 1664 pp, $87.50.

The first edition in 1976 had no mention of microcomputers. This edition has 90 new articles and modifications in most of the rest; it includes 29 articles on mathematics, from "Matrix Computations" (by J. H. Wilkinson) to "Roundoff Error." (Beware--in our review copy, print on some pages was cut off at the bottom.)

Arden, Bruce W., What Can be Automated? The Computer Science and Engineering Research Study (COSERS), MIT Pr., 1980, 934 pp.

Excellent and readable survey of research areas in computer science, at a non-technical level--in effect, an encyclopedia of computer science, valuable despite dated nature of some material (e.g., reference to a 1976 study of computer science Ph.D.'s).

Beltrami, Edward, The High Cost of Clean Water: Models for Water Quality Management, UMAP/Birkhauser, 1982; 53 pp.

Introduces the student to an important problem of increasing national concern and demonstrates the role of mathematical modelling in public policy. The details of the original problem setting (on Long Island) have been altered for the sake of making pedagogical points. First-order PDE's are used in the modelling of water quality, while the model for siting sewers results in a mixed integer programming problem with non-linear objective. Thirteen exercises (with solutions) are included.

Burghes, D. N., *et al.*, Applying Mathematics: A Course in Mathematical Modelling, Halsted, 1982; 194 pp.

Collection of 20 examples and 26 exercises which apply mathematics to real problems, together with some "theory" of modelling. Students should find these intriguing, and the necessary mathematics not too formidable.

NEWS & LETTERS

CORRECTION TO "BOOLE'S" ALGEBRA

Professor H. E. Heatherly of the University of Southwestern Louisiana has called attention to an erroneous statement in my paper "Boole's Algebra Isn't Boolean Algebra" (this *Magazine*, September 1981, p. 175); namely, that the axioms for Boolean algebra given there become axioms for a Boolean ring by replacing $a + a = a$ by $a + a = 0$. It is only true if the (redundant) axiom $a + bc = (a+b)(a+c)$ is deleted.

Theodore Hailperin
Lehigh University
Bethlehem, PA 18015

COMPUTATIONAL COMPLEXITY COURSE

The Northeast Section of the MAA and the University of Maine are sponsoring a short course on computational complexity to be held June 13-17, 1983 at the University of Maine, Orono. Principal lecturer is Herbert Wilf, U. of Pennsylvania, and cost (including course fee, room and board) is $175. For more information, contact Grattan Murphy, Dept. of Mathematics, U. of Maine at Orono, Orono, ME 04469.

MIAMI UNIVERSITY CONFERENCE

The 11th Annual Mathematics and Statistics Conference at Miami University, Oxford, Ohio, will be held September 23-24, 1983, on the theme "Operations Research and Mathematics as Applied in Business." William Lucas, Cornell, Harvey Wagner, U. of North Carolina, and Tom Shriber, U. of Michigan, are invited lecturers. Contributed papers relating to the general theme should be sent by June 1, 1983, to Stan Payne, Dept. of Mathematics and Statistics, Miami University, Oxford, OH 45056.

The Ohio Delta Chapter of Pi Mu Epsilon will also hold its annual student conference at the same time; students are invited to send abstracts of contributed papers to Milton Cox, at the above address.

43rd PUTNAM COMPETITION: WINNERS AND SOLUTIONS

Teams from 249 schools competed in the 1982 William Lowell Putnam mathematical competition. The top five winning teams, in descending rank, are:

Harvard University
Benji N. Fisher, Michael J. Larsen, Michael Raship.

University of Waterloo
David W. Ash, W. Ross Brown, Herbert J. Fichtner.

California Institute of Technology
Bradley W. Brock, Scott R. Johnson, Zinovy B. Reichstein.

Yale University
Alan S. Edelman, Paul N. Feldman, Nathaniel E. Glasser.

Princeton University
Gregg N. Patruno, David P. Roberts, Daniel J. Scales.

The five highest ranking individuals, named Putnam Fellows, are:

David W. Ash	Univ. of Waterloo
Eric D. Carlson	Michigan State Univ.
Noam D. Elkies	Columbia University
Brian R. Hunt	Univ. of Maryland
Edward A. Shpiz	Washington Univ.

Solutions to the 1982 Putnam problems were prepared for publication in this Magazine *by Bruce Hanson and Loren Larson, St. Olaf College.*

A-1. Let V be the region in the cartesian plane consisting of all points (x,y) satisfying the simultaneous conditions

$$|x| \le y \le |x| + 3 \text{ and } y \le 4.$$

Find the centroid $(\bar{x},\bar{y})$ of V.

Sol. Let A, B, C, D, E, F denote the points (-4,4), (0,0), (4,4), (-1,4), (0,3), (1,4) respectively. The centroid of ΔABC is (0,8/3), and the centroid of ΔDEF is (0,11/3). Using weighted averages,

$$\bar{y}\ (\text{Area } V) + \frac{11}{3}\ (\text{Area } \Delta DEF) = \frac{8}{3}\ (\text{Area } \Delta ABC)\ ,$$

and from this we find that

$$(\bar{x},\bar{y}) = (0,13/5)\ .$$

A-2. For positive real x, let $B_n(x) = 1^x + 2^x + 3^x + \ldots + n^x$. Prove or disprove the convergence of

$$\sum_{n=2}^{\infty} \frac{B_n(\log_n 2)}{(n \log_2 n)^2}$$

Sol. For $n \geq 2$ and $1 \leq m \leq n$, $1 \leq m^{\log_n 2} \leq n^{\log_n 2} = 2$. Therefore

$$0 < \frac{B_n(\log_n 2)}{(n \log_2 n)^2} \leq \frac{2n}{(n \log_2 n)^2} = \frac{2}{n(\log_2 n)^2}.$$

Thus, the series converges by the comparison test. (The series $\sum_{n=2}^{\infty} \frac{2}{n(\log_2 n)^2}$ converges by the integral test.)

A-3. Evaluate

$$\int_0^{\infty} \frac{\text{Arctan } (\pi x) - \text{Arctan } x}{x}\, dx\ .$$

Sol. We have

$$\int_0^{\infty} \frac{\text{Arctan } \pi x - \text{Arctan } x}{x}\, dx$$

$$= \lim_{c \to \infty} \left[\int_0^{c} \frac{\text{Arctan } \pi x}{x}\, dx - \int_0^{c} \frac{\text{Arctan } x}{x}\, dx \right]$$

$$= \lim_{c \to \infty} \int_c^{\pi c} \frac{\text{Arctan } x}{x}\, dx\ .$$

For each integer $n > 0$, there is a $c > 0$ such that for all $x > c$,

$$\frac{\pi}{2} - \frac{1}{n} < \text{Arctan } x < \frac{\pi}{2}\ .$$

Thus,

$$\int_c^{\pi c} \frac{\pi/2 - 1/n}{x}\, dx < \int_c^{\pi c} \frac{\text{Arctan } x}{x}\, dx < \int_c^{\pi c} \frac{\pi/2}{x}\, dx\ ,$$

$$\left(\frac{\pi}{2} - \frac{1}{n}\right) \log \pi < \int_c^{\pi c} \frac{\text{Arctan } x}{x}\, dx < \frac{\pi}{2} \log \pi\ .$$

The limit as $n \to \infty$ is $\frac{\pi}{2} \log \pi$.

A-4. Assume that the system of simultaneous differential equations

$$y' = -z^3,\ z' = y^3$$

with the initial conditions $y(0) = 1$, $z(0) = 0$ has a unique solution $y = f(x)$, $z = g(x)$ defined for all real x. Prove that there exists a positive constant L such that for all real x,

$$f(x + L) = f(x),\ g(x + L) = g(x).$$

Sol. The differential equations imply that

$$\frac{dy}{dz} = \frac{dy/dx}{dz/dx} = -z^3/y^3\ .$$

Now separate variables, integrate, and apply the initial conditions, and find that $y^4 + z^4 = 1$. If we think of x as a time variable, we see that the point governed by these equations will cycle around the curve $y^4 + z^4 = 1$ in a periodic manner, and we can let L denote the cycle time. (Note that the speed, given by $(\dot{y}^2 + \dot{z}^2)^{1/2} = (z^6 + y^6)^{1/2}$, along the orbit, is bounded away from zero.)

A-5. Let a, b, c, and d be positive integers and

$$r = 1 - \frac{a}{b} - \frac{c}{d}\ .$$

Given that $a + c \leq 1982$ and $r > 0$, prove that

$$r > \frac{1}{1983^3}\ .$$

Sol. Case 1. Suppose that $b > 1983$ and $d > 1983$. Then

$$\frac{a}{b} + \frac{c}{d} < \frac{a}{1983} + \frac{c}{1983} \leq \frac{1982}{1983},$$

and therefore $r > 1/1983$.

Case 2. Suppose that $b \leq 1983$. Then

$$\frac{a}{b} \leq \frac{b-1}{b} \leq \frac{1982}{1983}.$$

If $d \geq 1983^2$ then $\frac{c}{d} \leq \frac{1982}{1983^2}$, so that

$$\frac{a}{b} + \frac{c}{d} \leq \frac{(1982)(1983) + 1982}{1983^2} = \frac{(1982)(1984)}{1983^2} = \frac{1983^2 - 1}{1983^2},$$

and therefore $r \geq 1/1983^2$.

If $d < 1983^2$, then

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \leq \frac{bd - 1}{bd} < \frac{1983^3 - 1}{1983^3},$$

and therefore $r > 1/1983^3$.

A-6. Let σ be a bijection of the positive integers, that is, a one-to-one function from $\{1, 2, 3, \ldots\}$ onto itself. Let $x_1, x_2, x_3, \ldots$ be a sequence of real numbers with the following three properties:

(i) $|x_n|$ is a strictly decreasing function of n;

(ii) $|\sigma(n) - n| \cdot |x_n| \to 0$ as $n \to \infty$;

(iii) $\lim_{n \to \infty} \sum_{k=1}^{n} x_k = 1$.

Prove or disprove that these conditions imply that

$$\lim_{n \to \infty} \sum_{k=1}^{n} x_{\sigma(k)} = 1.$$

Sol. We will construct an example to show that the conditions do not imply the conclusion.

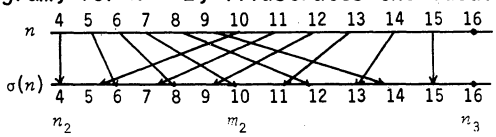
For $n \geq 2$, let $x_n = \frac{(-1)^n}{n \log n}$. The series $\sum_{n=2}^{\infty} x_n$ converges by the alternating series test. Define x_1 so that the series adds to 1. Thus, conditions (i) and (iii) are satisfied (except possibly $|x_1| \leq |x_2|$ in which case make obvious adjustments).

Define $\sigma(1) = 1$ and define σ as follows. Take $n_1 = 2$ and for each $k \geq 1$, set $n_{k+1} = n_k^2$. Let $m_k = (n_k + n_{k+1})/2$. Then define

$$\sigma(n) = n_k + 2(n - n_k) = 2n - n_k, \text{ for } n_k \leq n < m_k$$

$$\sigma(n) = n_k + 1 + 2(n - m_k) = 2n - n_{k+1} + 1, \text{ for } m_k \leq n < n_{k+1}$$

for $k = 1, 2, \ldots$. The following diagram, for $k = 2$, illustrates the idea.



It is straightforward to show that σ is a bijection of the positive integers.

If $n_k \leq n < m_k$, then $|\sigma(n) - n||x_n| = |n - n_k| \frac{1}{n \log n} \leq \frac{1}{\log n}$.

If $m_k \leq n < n_{k+1}$, $|\sigma(n) - n||x_n| \leq \left(\frac{n_{k+1} - n_k}{2}\right)|x_n| \leq m_k \frac{1}{n \log n} \leq \frac{1}{\log n}$.

These conditions show that condition (ii) is satisfied.

Finally, we will show that $\sum_{n=1}^{\infty} x_{\sigma(n)}$ diverges. It suffices to show that $\sum_{n=n_k}^{m_k - 1} x_{\sigma(n)}$ does not converge to 0 as $k \to \infty$. Since $\sigma(n)$ is even for $n_k \leq n < m_k$, all the terms in the sum are positive. Thus, for $k \geq 1$,

$$2 \sum_{n=n_k}^{m_k - 1} x_{\sigma(n)} \geq \sum_{n=n_k}^{n_{k+1} - 1} \frac{1}{n \log n} \geq \int_{n_k}^{n_k^2} \frac{dx}{x \log x}$$

$$= \log \log n_k^2 - \log \log n_k$$
$$= \log 2.$$

Hence, the series diverges.

B-1. Let M be the midpoint of side BC of a general ΔABC. Using the *smallest possible* n, describe a method for cutting ΔAMB into n triangles which can be reassembled to form a triangle congruent to ΔAMC.

Sol. Let N be the midpoint of side AB. Cut ΔAMB along MN. Turn ΔNBM over and bring side AN of ΔANM into juxtaposition with side BN of ΔNBM (A and B will coincide). It's easy to prove that the resulting figure is a triangle congruent to ΔAMC.

B-2. Let $A(x,y)$ denote the number of points (m,n) in the plane with integer coordinates m and n satisfying

$$m^2 + n^2 \leq x^2 + y^2 .$$

Let

$$g = \sum_{k=0}^{\infty} e^{-k^2} .$$

Express

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x,y)\, e^{-x^2-y^2}\, dx\, dy$$

as a polynomial in g.

Sol. Changing to polar coordinates (with $A(r) \equiv A(r,0)$) yields

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x,y)\, e^{-x^2-y^2}\, dx\, dy$$

$$= \int_0^{\infty} \int_0^{2\pi} A(r)\, e^{-r^2}\, r\, d\theta\, dr$$

$$= 2\pi \int_0^{\infty} A(r)\, d(-\tfrac{1}{2} e^{-r^2}) .$$

By the formula for integration by parts (for Stieltjes integrals--see Rudin, *Principles of Mathematical Analysis*, McGraw-Hill), we have

$$\int_0^{\infty} A(r)\, d(-\tfrac{1}{2} e^{-r^2})$$

$$= -\frac{A(r)\, e^{-r^2}}{2}\Big|_0^{\infty} + \frac{1}{2}\int_0^{\infty} e^{-r^2}\, dA(r)$$

$$= \frac{1}{2} + \frac{1}{2}\int_0^{\infty} e^{-r^2}\, dA(r) .$$

Let $0 < r_1 < r_2 < r_3 < \ldots$ be the points of discontinuity of the step function $A(r)$. Then

$$\int_0^{\infty} e^{-r^2}\, dA(r)$$

$$= \sum_{k=1}^{\infty} e^{-r_k^2}(A(r_k + 0) - A(r_k))$$

where $A(r_k + 0) - A(r_k)$ is the number of lattice points on the circle $x^2 + y^2 = r^2$. Thus we have

$$\sum_{k=1}^{\infty} e^{-r_k^2}\, (A(r_k + 0) - A(r_k))$$

$$= \sum_{\substack{m,n \text{ integers} \\ (m,n) \neq (0,0)}} e^{-m^2-n^2}$$

$$= \left(\sum_{m=-\infty}^{\infty} e^{-m^2}\right)\left(\sum_{n=-\infty}^{\infty} e^{-n^2}\right) - 1$$

$$= (2g - 1)^2 - 1 .$$

Putting the above results together we find that the original integral equals $\pi(2g - 1)^2$.

B-3. Let p_n be the probability that $c + d$ is a perfect square when the integers c and d are selected independently at random from the set $\{1,2,3,\ldots,n\}$. Show that

$$\lim_{n \to \infty} (p_n \sqrt{n})$$

exists and express this limit in the form $r(\sqrt{s} - t)$ where s and t are integers and r is a rational number.

Sol. For each n, let S_n denote the set of ordered pairs (c,d), with c and d in $\{1,2,\cdots,n\}$, such that $c+d$ is a perfect square. Let s_n denote the number of elements in S_n (i.e., $s_n = n^2 p_n$). Clearly, $S_n \subseteq S_{n+1}$. Let $D_{n+1} = S_{n+1} \setminus S_n$. If $(c,d) \in D_{n+1}$, either $c = n+1$ or $d = n+1$, and $c+d$ is is a square between $n+2$ and $2n+2$. There is at most one element (c,d) in

D_{n+1} such that $c+d = 2n+2$, and there are at most two elements in D_{n+1} such that $c+d = 2n+1$. For each square, q^2, between $n+2$ and $2n$, there are exactly two elements in D_{n+1} such that $c+d = q^2$. These observations imply that

$$(*) \quad s_n + 2k \le s_{n+1} \le s_n + 2k + 3 ,$$

where k is the number of squares between $n+2$ and $2n$.

Now, k = (number of integers between $\sqrt{n+2}$ and $\sqrt{2n}$) $\le \sqrt{2n} - \sqrt{n+2} + 1 \le \sqrt{2n} - \sqrt{n} + 1$. Also, $k \ge \sqrt{2n} - \sqrt{n+2} - 2 \ge \sqrt{2n} - \sqrt{n} - 4$. Substituting these inequalities for k into (*), we find that

$$2(\sqrt{2} - 1)\sqrt{n} - 8 \le s_{n+1} - s_n \le 2(\sqrt{2} - 1)\sqrt{n} + 5.$$

From this and the fact that $s_1 = 0$ it follows that

$$2(\sqrt{2} - 1) \sum_{k=1}^{n} \sqrt{k} - 8n \le s_{n+1} \le 2(\sqrt{2} - 1) \sum_{k=1}^{n} \sqrt{k} + 5n .$$

Now

$$\frac{2}{3} n^{3/2} = \int_0^n x^{1/2} dx \le \sum_{k=1}^{n} \sqrt{k} \le n + \int_0^n x^{1/2}\, dx = n + \frac{2}{3} n^{3/2} ,$$

and therefore

$$2(\sqrt{2} - 1) \frac{2}{3} n^{3/2} - 8n \le s_{n+1} \le 2(\sqrt{2} - 1)(n + \frac{2}{3} n^{3/2}) + 5n .$$

When we multiply these inequalities by $\sqrt{n+1}$ and divide by $(n+1)^2$, and take limits as $n \to \infty$, we find that

$$\lim_{n \to \infty} p_n \sqrt{n} = \lim_{n \to \infty} p_{n+1} \sqrt{n+1} = \lim_{n \to \infty} \frac{s_{n+1}}{(n+1)^2} \sqrt{n+1} = \frac{4}{3} (\sqrt{2} - 1).$$

B-4. Let $n_1, n_2, \ldots, n_s$ be distinct integers such that

$$(n_1 + k)(n_2 + k) \cdots (n_s + k)$$

is an integral multiple of

$$n_1 n_2 \cdots n_s$$

for every integer k. For each of the following assertions, give a proof or a counterexample:

(a) $|n_i| = 1$ for some i.

(b) If further all n_i are positive, then

$$\{n_1, n_2, \ldots, n_s\} = \{1, 2, \ldots, s\} .$$

Sol. a. Let $A = (n_1+1)(n_2+1)\ldots(n_s+1)$ and $B = (n_1-1)(n_2-1)\ldots(n_s-1)$. Then $0 \le AB = (n_1^2 - 1)(n_2^2 - 1)\ldots(n_s^2 - 1) < n_1^2 n_2^2 \cdots n_s^2$ (note from the assumptions that none of the n_i's are zero). Therefore, either $|A| < |n_1 n_2 \ldots n_s|$ or $|B| < |n_1 n_2 \ldots n_s|$. But $n_1 n_2 \ldots n_s$ divides both A and B, and therefore either A or B must be zero. It follows that

$|n_i| = 1$ for some i.

b. Suppose all the n_i are positive and suppose we have labeled them so that $n_1 < n_2 < \ldots < n_s$. From part a, $n_1 = 1$. Suppose there is an n_k such that $n_k > k$, and let k be the smallest such subscript (i.e., $n_i = i$ for $i = 1, 2, \ldots, k-1$). Our condition implies that $n_1 n_2 \ldots n_s$ divides $(n_1-k)\ldots(n_{k-1}-k)(n_k-k)\ldots(n_s-k)$ our choice of k,

$$\left|\frac{(n_1-k)(n_2-k)\ldots(n_{k-1}-k)}{n_1 n_2 \ldots n_{k-1}}\right| = \left|\frac{(-1)^{k-1}(k-1)!}{(k-1)!}\right| = 1 ,$$

and therefore $n_k n_{k+1} \ldots n_s$ must divide $(n_k-k)\ldots(n_s-k)$. But this is impossible because $0 < (n_k-k)\ldots(n_s-k) < n_k n_{k+1} \ldots n_s$. This contradiction implies that $n_i = i$ for $i = 1, 2, \ldots, s$ and the result is established.

B-5. For each $x > e^e$ define a sequence $S_x = u_0, u_1, u_2, \ldots$ recursively as follows: $u_0 = e$, while for $n \ge 0$, u_{n+1} is the logarithm of x to the base

u_n. Prove that S_x converges to a number $g(x)$ and that the function g defined in this way is continuous for $x > e^e$.

Sol. The hypothesis is that $e^e < x = u_0^{u_1} = e^{u_1}$, hence $e = u_0 < u_1$. This inequality plus the fact that $u_0^{u_1} = u_1^{u_2}$ (both equal x) means that $u_2 < u_1$. In general, if s and t are real numbers such that $e \le s < t$, then $s^t > t^s$. Therefore, $u_1^{u_2} = u_0^{u_1} > u_1^{u_0}$, so $u_2 > u_0$. Thus, we have $u_0 < u_2 < u_1$.

Assume that $e \le u_n < u_{n+1}$. Starting with the fact that $u_n^{u_{n+1}} = u_{n+1}^{u_{n+2}}$, the same argument as above shows that $u_n < u_{n+2} < u_{n+1}$.

The previous results imply that the subsequence $\{x_{2n}\}$ is increasing and bounded above by x_1. Therefore it converges, say to U. Also, the subsequence $\{x_{2n+1}\}$ is decreasing and bounded below by x_0. Therefore it converges, say to L.

From $u_{k+1} = \log_{u_k} x = \dfrac{\log x}{\log u_k}$, we find that $\dfrac{u_{2n+1}}{u_{2n}} = \dfrac{\log u_{2n-1}}{\log u_{2n}}$.

As $n \to \infty$, this yields $\dfrac{L}{U} = \dfrac{\log L}{\log U}$, or equivalently, $\dfrac{L}{\log L} = \dfrac{U}{\log U}$. Since $\dfrac{t}{\log t}$ is a strictly increasing function when $x \ge e$, it must be the case that $U = L$. It follows that S_x converges, say to $g(x)$.

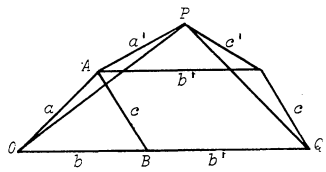
Now, $g(x) \log g(x) = \lim_{n \to \infty} u_{n+1} \log u_n = \lim_{n \to \infty} \log x = \log x$. From this it follows that $g(x)$ is continuous.

B-6. Let $K(x,y,z)$ denote the area of a triangle whose sides have lengths x,y, and z. For any two triangles with sides a,b,c and a',b',c', respectively, prove that

$$\sqrt{K(a,b,c)} + \sqrt{K(a',b',c')} \le \sqrt{K(a+a',\ b+b',\ c+c')}$$

and determine the cases of equality.

Sol. Let O denote the origin in the plane. Pick points B and B' on the positive x-axis and points A and A' in the upper half plane so that $|\overrightarrow{OA}| = a$, $|\overrightarrow{OB}| = b$, $|\overrightarrow{AB}| = c$, $|\overrightarrow{OA'}| = a'$, $|\overrightarrow{OB'}| = b'$, $|\overrightarrow{A'B'}| = c'$. Then, Area $\Delta AOB = K(a,b,c)$ and Area $\Delta A'OB' = K(a',b',c')$. Let P and Q be such that $\overrightarrow{OP} = \overrightarrow{OA} + \overrightarrow{OA'}$ and $\overrightarrow{OQ} = \overrightarrow{OB} + \overrightarrow{OB'}$ (see figure).



We claim

$$\sqrt{K(a,b,c)} + \sqrt{K(a',b',c')} \le \sqrt{\text{Area } POQ} \le \sqrt{K(a+a',b+b',c+c')}\ .$$

For the first inequality, let d and d' denote the lengths of the altitudes from A and A' respectively in ΔAOB and $\Delta A'OB'$. We must show that

$$\sqrt{bd/2} + \sqrt{b'd'/2} \le \sqrt{(b+b')(d+d')/2}\ ,$$

or equivalently, that

$$bd/2 + \sqrt{bdb'd'} + b'd'/2 \le (bd + bd' + b'd + b'd')/2,$$

or $\sqrt{bdb'd'} \le (bd' + b'd)/2$.

But this is just an instance of the arithmetic mean - geometric mean inequality.

For the second inequality, note that $|\overrightarrow{OQ}| = b+b'$, $|\overrightarrow{OP}| \le a+a'$ with equality if and only if $\overrightarrow{OA}$ and $\overrightarrow{OA'}$ are parallel, and $|\overrightarrow{PQ}| \le c+c'$ with equality if and only if AB and $A'B'$ are parallel. This implies that Area $\Delta POQ \le K(a+a', b+b', c+c')$ with equality if and only if the given triangles are similar. In the case of similar triangles, $b/d = b'/d'$, or equivalently, $bd' = b'd$, and in this case the leftmost inequality is also an equality. Thus, the desired inequality holds in all cases, and equality occurs if and only if the given triangles are similar.

THE CHAUVENET PAPERS

A Collection of Prize-Winning Expository Papers in Mathematics

James C. Abbott, editor

This two-volume collection of the twenty-four prize winning Chauvenet Papers contains the finest collection of expository articles in mathematics ever assembled!

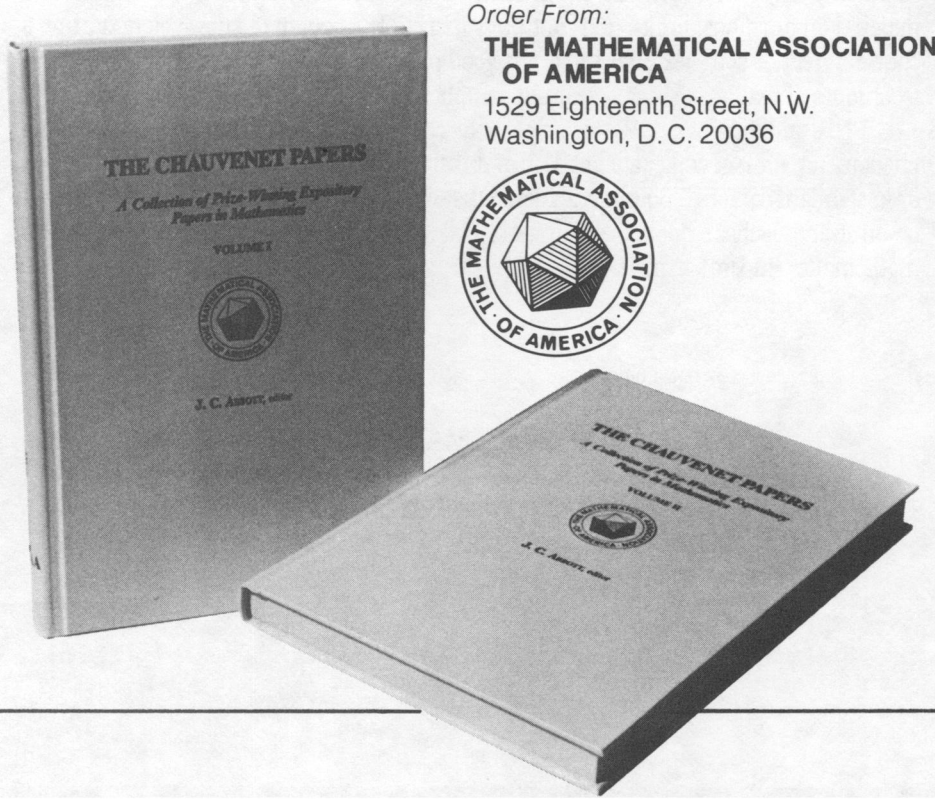
THE CHAUVENET PRIZE for special merit in mathematical exposition was first awarded by the Mathematical Association of America in 1925 to G. A. Bliss. Since that time, twenty-four Chauvenet Prizes have been awarded and the Prize has become the Association's most prestigious award for mathematical exposition. The list of authors is a veritable WHO's WHO of mathematical expositors, and contains some of the more prominent mathematicians of the past fifty years.

Clearly written, well organized, expository in nature, these papers are the jewels of mathematical writing during our times. They were selected by peer juries of experts judged for their presentation as well as their content. They are a sheer joy to read for those who delight in the beauty of mathematics alone.

Volume I—xviii + 312 pages. Hardbound. List: $21.00—MAA Member: $16.00

Volume II—vii + 282 pages. Hardbound. List: $21.00—MAA Member: $16.00

Special Package Price for Both Volumes: List $36.00—MAA Member: $27.00



Order From:

THE MATHEMATICAL ASSOCIATION OF AMERICA

1529 Eighteenth Street, N.W.
Washington, D. C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036
MATHEMATICS MAGAZINE VOL. 56, NO. 3, MAY 1983