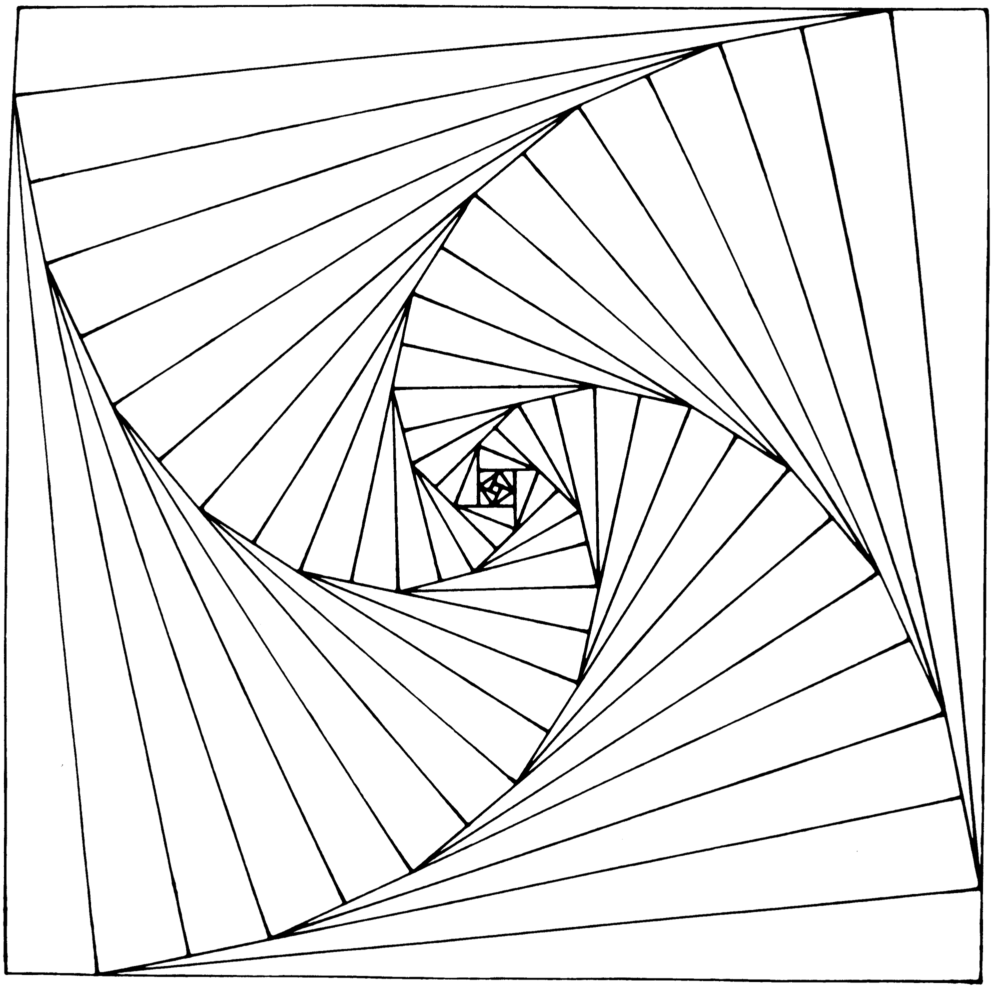
Vol. 62, No. 3 June 1989

MATHEMATICS MAGAZINE



An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

The aim of *Mathematics Magazine* is to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these are essential to the presentation of some good mathematics. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

The full statement of editorial policy appears in this *Magazine*, Vol. 54, pp. 44–45, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher.

Send new manuscripts to: G. L. Alexanderson, Editor, Mathematics Magazine, Santa Clara University, Santa Clara, CA 95053. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and one copy and keep one copy. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added.

AUTHORS

D.F. Bailey received his Ph.D. degree from Vanderbilt University in 1965. Although trained as a point set topologist, while teaching at Cornell College, a small liberal arts college in Iowa, staffing considerations forced him to teach numerical analysis. This occasioned his interest in the solution of equations by functional iteration. The present article actually had its origin several years ago when he was visiting at the University of Iowa. At that time he occupied an office near Howard Lambert. He was therefore amused to learn that a third order iterative method for computing nth roots was often referred to as Lambert's method and less often as Bailey's method. Since 1983 Dr. Bailey has been chairman of the department of mathematics at Trinity University.

Vol. 62, No. 3 June 1989

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $11 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $28 per year. Bulk subscriptions (5 or more copies) are available to colleges and universities for classroom distribution to undergraduate students at a 41% discount ($6.50 per copy—minimum order $32.50).
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.
Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, The Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

Second class postage paid at Washington, D.C. and additional mailing offices.
Postmaster: Send address changes to Mathematics Magazine Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20077-9564.
PRINTED IN THE UNITED STATES OF AMERICA

A Historical Survey of Solution by Functional Iteration

D. F. BAILEY
Trinity University
San Antonio, TX 78284

Introduction

A colleague once joked that the development of any idea could be traced back to the "primordial slime" if one were determined to do so. In this article we take the idea of solution of equations by iteration techniques and attempt to verify his claim. Alas, we do not succeed. We do, however, show that solution by iteration has a very long history indeed, and we attempt to outline that history. This article began when we read Bateman [1], Bruins [2], and Traub [3]. The majority of the references (especially the primary sources) were found in these works or by searching backward from them. We acknowledge our indebtedness to these authors. No attempt will be made to mention one or more of them explicitly each time a source was so discovered. It should also be pointed out that Goldstine [4] is relevant to the topic.

Even though the topic is an elementary one, it is still not universally a part of the mathematics curriculum. Therefore, to refresh the reader's memory and to define the particular type of iteration techniques to be considered, we begin with a quick review of the general technique as well as an example. Given an equation $f(x)=0$ the standard procedure is to introduce a new function Φ having the property that $\Phi(x)=x$ implies $f(x)=0$. A solution to the equation $\Phi(x)=x$ is then approximated by choosing x_1 in some manner and defining $x_{n+1}=\Phi(x_n)$. Under proper choice of Φ, or x_1, or both, one may be able to obtain convergence of the sequence $\{x_n\}$. If $\{x_n\}\to x_0$ it must be the case that $\Phi(x_0)=x_0$ if Φ is continuous at x_0.

To illustrate the foregoing we solve $f(x)=0$ when $f(x)=2x^3+4x^2-2x-5$. Note first that

$$2x^3+4x^2-2x-5=0 \tag{1}$$

is equivalent to

$$x^2=\frac{2x+5}{2x+4} \quad \text{if} \quad x\neq -2.$$

Thus if we are able to find a solution to

$$\Phi(x)=\sqrt{\frac{2x+5}{2x+4}}=x, \tag{2}$$

we likewise have a solution to equation (1). (It is easy to see that our choice of Φ is not unique. Indeed we might have used

$$\Phi_1(x) = x - \frac{2x^3+4x^2-2x-5}{6x^2+8x-2}$$

$$\Phi_2(x) = \sqrt[3]{\frac{2x+5-4x^2}{2}}$$

or

$$\Phi_3(x) = \frac{2x^3+4x^2-5}{2}.)$$

We now obtain a solution to equation (2) as follows. Choose $x_1 = 2$ and let $x_{n+1} = \Phi(x_n)$. Then

$$\begin{aligned} x_2 &= \Phi(x_1) = 1.0606602 \\ x_3 &= \Phi(x_2) = 1.0785933 \\ x_4 &= \Phi(x_3) = 1.078152 \\ x_5 &= \Phi(x_4) = 1.0781628 \\ x_6 &= \Phi(x_5) = 1.0781626 \\ x_7 &= \Phi(x_6) = 1.0781626 \\ &\text{etc.} \end{aligned}$$

Thus it appears that $\{x_n\} \to x_0 \doteq 1.0781626$. Indeed one can show that with Φ as in (2) and $x_1 \geqslant 0$, $\{x_n\}$ converges. (Other choices of Φ may not yield such desirable behavior. Attempts to locate a solution to $\Phi_3(x) = x$, for instance, produce disaster. Even with $x_1 = 1.07816$ one obtains a decreasing sequence of values which rapidly exceeds the capacity of any computer.)

Methods for Computing *N*th Roots

The earliest use of iteration techniques occurred prior to 1700 B.C.; Neugebauer and Sachs [5] report that a clay tablet of that period (#7289 of the Yale Babylonian Collection) shows a square figure 30 units on a side along whose diagonal the sexagesimal (base 60) numbers 1;24,51,10 and 42;25,35 appear. A conversion to decimal notation will show that 1;24,51,10 is nearly 1.4142129, which agrees with $\sqrt{2}$ through the first 6 digits. It is also easy to see that (operating base 60) 30 × 1;24,51,10 = 42;25,35. Neugebauer and Sachs deduce, therefore, that the square shown has edge length 30, diagonal 42;25,35 and that the Babylonian value for $\sqrt{2}$ is recorded along the diagonal as well. The point of interest, however, involves how the value $\sqrt{2} \doteq 1;24,51,10$ might have been obtained.

A well-known iteration technique for obtaining approximations to $\sqrt{N}$ uses the iteration formula

$$x_{n+1} = \frac{x_n + \dfrac{N}{x_n}}{2}. \tag{3}$$

If we use equation (3) to approximate $\sqrt{2}$, choose $x_1 = 1;30$ and operate base 60, we obtain $x_2 = 1;25$ and $x_3 = 1;24,51,10$. Neugebauer and Sachs conjecture that this iteration scheme was used by the Babylonians and they comment that "the same procedure is attested in another text in finding the approximate value of $\sqrt{28;20}$" [6].

There seems to be wide agreement that this reconstruction represents the manner in which the Babylonians computed square roots [7]. Boyer says the method has been attributed to Archytas (428–365 B.C.) [8]. Some writers feel that Archimedes (287–212 B.C.) used equation (3) to obtain square roots. This question is far from settled, however, and the literature on Archimedes' computation of square roots is large [9].

In any event, the method appears without question around A.D. 75 in the *Metrica* of Heron of Alexandria (c.75?) [10]. In Book I of this work Heron computes the area of a triangle with sides 7, 8, and 9 to be $\sqrt{720}$. He then says, "since 720 has no rational square root we shall obtain a very close approximation to $\sqrt{720}$ as follows: The nearest (perfect integral) square to 720 is 729. $\sqrt{729} = 27$, $\frac{720}{27} = 26\frac{2}{3}$, $27 + 26\frac{2}{3} = 53\frac{2}{3}$, $53\frac{2}{3}/2 = 26\frac{1}{2}\frac{1}{3}$, $\therefore$ $\sqrt{720} =$ approximately $26\frac{1}{2}\frac{1}{3}$. For $(26\frac{1}{2}\frac{1}{3})^2 = 720\frac{1}{36}$, the difference being only $\frac{1}{36}$ of a unit."

The text continues "Now if we desire that the difference be less than $\frac{1}{36}$ we commence with the number just obtained, $720\frac{1}{36}$ instead of 729, and by proceeding in the same way we shall find that the difference will be much less than $\frac{1}{36}$" [11]. That is, we may *iterate* the procedure to obtain better and better approximations.

Whether the "method of Heron" was transmitted or rediscovered is not clear but Smyly says it was "known to the Arabs, but... subsequently forgotten" [12]. It was known in the 14th century to one Nicolas Rhabdas [13]. It also appears in the works of the Indian writer Jñānarāja around 1503 [14] and in a letter of Isaac Newton (1642–1727) to John Collins (1625–1683) dated August 27, 1675 [15]. (Collins is known largely for his extensive correspondence with outstanding scientific figures; he was a mathematics teacher and at one time served as a seaman "in the Venetian service against ye Turke" [16].) In this same letter to Collins, Newton gives techniques for approximating $\sqrt[3]{N}$ and $\sqrt[4]{N}$ [17]. His scheme is familiar; namely,

$$x_{n+1} = \frac{p}{p+1}x_n + \frac{N}{(p+1)x_n^{p-1}}, \qquad p = 3,4. \tag{4}$$

Iterative techniques for computing nth roots become more sophisticated in the work of Thomas Fautet DeLagny (1660–1734). DeLagny spent over forty years in work on approximation of roots [18] "being almost altogether taken up in extracting the Roots of pure Powers (especially the Cubick)" [19]. DeLagny gave both rational and irrational bounds for $\sqrt[3]{a^3+b}$. If we let $z = \sqrt[3]{a^3+b}$ his bounds are [20]

$$\frac{a}{2} + \sqrt{\frac{a^2}{4} + \frac{b}{3a}} > z > \frac{a}{2} + \sqrt{\frac{a^2}{4} + \frac{b-1}{3a}}$$

and

$$a + \frac{ab+a}{3a^3+b+1} > z > a + \frac{ab}{3a^3+b}.$$

DeLagny published his results in 1691 in the *Journal des Scavans* [21].

Sometime between 1691 and 1694 Edmund Halley (1656–1742) learned from a friend that DeLagny had shown that $\sqrt[3]{a^3+b}$ was between $a + ab/(3a^3+b)$ and $a/2 + \sqrt{a^2/4 + b/3a}$ and that

$$\sqrt[5]{a^5+b} \doteq \sqrt{\sqrt{\frac{b}{5a} + \frac{a^4}{4}} - \frac{a^2}{4}} + \frac{a}{2}.$$

Halley says, "having by tryal found the goodness of them... I was willing to find out the Demonstration" [22].

As an example we give Halley's demonstration that

$$\sqrt[3]{a^3+b} \doteq a+\frac{ab}{3a^3+b}.$$

Note first that

$$(a+e)^3=a^3+b \quad \text{where} \quad b=3a^2e+3ae^2+e^3.$$

Neglecting e^3 on the assumption that e is small, we have $b \doteq 3a^2e+3ae^2$ or

$$\frac{b}{3a^2} \doteq e+\frac{e^2}{a} \doteq e.$$

Likewise from $b \doteq 3a^2e+3ae^2$ we have

$$\frac{b}{3a^2+3ae} \doteq e.$$

Now using $e \doteq b/3a^2$ in the foregoing we have [23]

$$e \doteq \frac{b}{3a^2+3a\dfrac{b}{3a^2}}=\frac{b}{3a^2+\dfrac{b}{a}}=\frac{ab}{3a^3+b}.$$

Thus

$$\sqrt[3]{a^3+b}=a+e \doteq a+\frac{ab}{3a^3+b}.$$

After satisfying himself of DeLagny's results, Halley gives

$$\sqrt[n]{a^n+b} \doteq a+\frac{ab}{na^n+\dfrac{n-1}{2}b} \quad \text{for} \quad n=2,3,4,5,6,7. \tag{5}$$

He further states that the rule holds "also of the other higher Powers" [24]. In the sources cited we can find no mention of iterating this procedure. Bateman, however, finds that "Halley clearly states that his approximations are to be used as a basis of a method of iteration" [25]. If in equation (5) we set $N=a^n+b$ and eliminate b, we obtain

$$\sqrt[n]{N} \doteq a\left(\frac{(n-1)a^n+(n+1)N}{(n+1)a^n+(n-1)N}\right), \tag{6}$$

which form can obviously be used for iteration.

In 1770 J. H. Lambert (1728–1777) published a version of this result in his *Beyträge zum Gebrauche der Mathematik und deren Anwendungen* [26]. Lambert's derivation is different from that of Halley, using as its start the binomial theorem. We outline his procedure below.

Let $x=(a+b)^n=a^n+na^{n-1}b+n((n-1)/2)a^{n-2}b^2+\cdots$. Thus

$$x\left(1+\frac{zb}{a}\right)=a^n+na^{n-1}b+n\frac{n-1}{2}a^{n-2}b^2+\cdots+zba^{n-1}+zb^2na^{n-2}+\cdots$$

$$=a^n+(n+z)ba^{n-1}+\left(z+\frac{n-1}{2}\right)nb^2a^{n-2}+\cdots.$$

Now in order to eliminate those terms involving b^2 we set $z = -(n-1)/2$ and obtain

$$x\left(1 - \frac{(n-1)b}{2a}\right) = a^n + \left(n - \frac{n-1}{2}\right)ba^{n-1} + b^3(\cdots).$$

If we ignore those terms involving b^3 and solve for x above we obtain

$$x \doteq \left(\frac{2a + (n+1)b}{2a - (n-1)b}\right)a^n.$$

In order to see that this result is indeed a form of equation (6) we first replace n by $1/n$ and recall that x is $(a+b)^{1/n}$. This yields

$$(a+b)^{1/n} \doteq \left(\frac{2an + (n+1)b}{2an + (n-1)b}\right)a^{1/n}.$$

Next replace a by a^n to obtain

$$(a^n + b)^{1/n} \doteq \left(\frac{2na^n + (n+1)b}{2na^n + (n-1)b}\right)a.$$

Finally now if we let $N = a^n + b$ and eliminate b as before we obtain equation (6).

Many texts in numerical analysis refer to the iteration scheme based on equation (6) as "Lambert's Method." The method has also been attributed to Hutton and to Bailey. Davies and Peck [27] give equation (6) with the comment that this "method of extracting the nth root of any number approximately, is due to Hutton." In a sense this may be true.

Charles Hutton (1737–1823) was professor of mathematics at the Royal Military Academy at Woolwich for 34 years [28]. Sometime prior to 1812 he published several rules for approximating roots. He was aware of the work of most of his predecessors, for he says "the persons who have best succeeded in their enquiries after such rules, have been successively Sir Isaac Newton, Mr. Raphson, M. deLagney, and Dr. Halley" [29]. He claims, however, that his results are an improvement with regard to "ease and universality." Hutton gives equation (6) explicitly in his *Tracts* and in a textbook as well. He may indeed be the first to do so. Also, in his text he states explicitly that "the operation may be repeated as often as we please, by using always the last found root for the assumed root" [30]. That is, we may use the iteration scheme

$$a_{p+1} = a_p\left(\frac{(n-1)a_p^n + (n+1)N}{(n+1)a_p^n + (n-1)N}\right) \tag{7}$$

to obtain a convergent sequence of approximations. In his *Tracts* Hutton even shows that the scheme given by equation (7) is of third order. That is, he shows

$$\left|\sqrt[n]{N} - a_{p+1}\right| \leqslant K\left|\sqrt[n]{N} - a_p\right|^3$$

for some constant K.

The iteration scheme of equation (7) appears many times. Dunkel attributes it to Peter Barlow in 1814 [31]. S. M. Jacob rediscovered the technique and treated it in detail in 1903 [32]. J. V. Uspensky published the result in 1927 although he "does not venture to say positively whether this method is new or not" [33]. V. A. Bailey on the other hand publishes the result in 1941 with the comment that the method "appears to be new" [34].

General Iteration Techniques

We turn now from the approximation of roots to the problem of solving a general equation of the form $f(x)=0$. A surprisingly sophisticated example occurs in the ninth century. Kennedy and Trasne [35] relate that Habash al-Hāsib al-Marwazi, an Arab writer whose name means "The Computer from Merv," found himself faced with the problem of solving the equation $\theta - m\sin\theta = t$ for θ when m and t are given. Al-Hāsib's method was as follows. Take $\theta_0 = t + m\sin t$, $\theta_1 = t + m\sin\theta_0$, $\theta_2 = t + m\sin\theta_1$, $\theta_3 = t + m\sin\theta_2$ and note that $\theta_3 - m\sin\theta_3 \doteq t$ so that θ_3 is an approximate solution.

In more familiar notation we see that one can solve $t=\theta - m\sin\theta$ if and only if one can solve $f(\theta)=\theta$ where $f(\theta) = t + m\sin\theta$. Thus the method of al-Hāsib is very familiar; take $\theta = t$ as an initial guess. Then $\theta_0 = f(t)$, $\theta_1 = f(\theta_0)$, etc. Of course, with the aid of modern computers, we might not be satisfied with θ_3 as an approximate solution; otherwise the procedure could have come from a modern day numerical analysis text.

The next examples of which we are aware come in 1674. Michael Dary (1613–1679), a self-taught mathematician, tobacco-cutter, gauger of wine casks, and sometime gunner in the Tower of London [36], described a method for solving $x^p = ax^q + n$ (where p and q are integers and $q<p$) in a letter to Newton dated August 15, 1674 [37]. His method in brief is to set $f(x) = (ax^q+n)^{1/p}$ and solve by iteration. Dary said "first guess at the root as nearly as you can, the nearer the better (not for necessity but for accommodation), and suppose that guess to be z." Then

$$b=(az^q+n)^{1/p}, \qquad c=(ab^q+n)^{1/p}, \qquad d=(ac^q+n)^{1/p}, \quad \text{etc.}$$

On April 2, 1674 James Gregory (1638–1675) writes to Collins on the problem of solving

$$c = a + \frac{a^2}{b} + \frac{a^3}{b^2} + \cdots + \frac{a^n}{b^{n-1}} \tag{8}$$

for a if c, b and n are given [38]. He first reduces equation (8) to $b^n c = ab^m f - a^p$, where $n-1=m$, $n+1=p$, and $c+b=f$. Now the above holds if and only if

$$a = \frac{bc}{f} + \frac{a^p}{b^m f}.$$

That is to say, we wish to solve $g(a)=a$ where

$$g(x) = \frac{bc}{f} + \frac{x^p}{b^m f}.$$

Gregory said one should, in our notation, take as our first guess $x_1 = 0$ and iterate via $x_{n+1} = g(x_n)$.

Of course functional iteration of this type reappears in the literature. As late as 1908 R. Ross gives a general discussion of solution by iteration [39]. As his first example he seeks a solution to $x^3-2x-5=0$. (This is a famous example of Newton which we shall see again.) Ross finds a solution by iterating $f(x)=\sqrt[3]{2x+5}$. Interestingly this is exactly Dary's method of 1674.

As one might expect, concern with the rigorous treatment of iteration techniques, as regards the convergence of the iterates, came quite late. We outline two very important results below, together with others which seem to us of interest, since they appear to complete a particular line of inquiry.

In a famous paper of 1922, Stefan Banach (1892–1945) gave a sufficient condition for convergence of the iterates of a function f [40]. Banach's condition, when specialized to the real line, is that $|f(x)-f(y)| \leqslant K|x-y|$ for all x and y, where K is some real number such that $0 \leqslant K < 1$. To illustrate we apply Banach's theorem to Ross's example above. Recall first that by the Mean Value Theorem one can establish Banach's condition if it can be shown that $|f'(x)| \leqslant K < 1$. Now suppose we have

$$f(x) = \sqrt[3]{2x+5}\,.$$

It then follows that

$$f'(x) = \frac{2}{3}\frac{1}{(2x+5)^{2/3}}$$

and thus, if we insist that $x \geqslant 0$, $0 < f'(x) < \frac{2}{3}$ and we take $K = \frac{2}{3}$.

Of course the question arises as to how one might weaken Banach's requirement. In 1959, Cheny and Goldstein embedded a small result in a much larger paper. This result, when specialized to the real line, gave the following. If $f\colon [a,b] \to [a,b]$ and $|f(x)-f(y)| < |x-y|$ then the sequence of iterates $\{f^n(x)\}$ converges to a unique fixed point of f [41].

The question which logically follows the result of Cheny and Goldstein is this: How does one treat functions in which $|f(x)-f(y)| \leqslant |x-y|$? Krasnoselski [42] published a result in 1955 which says if $f\colon [a,b] \to [a,b]$ and satisfies $|f(x)-f(y)| \leqslant |x-y|$, then the sequence defined by $x_{n+1} = \frac{1}{2}(x_n + f(x_n))$ converges to a fixed point of f.

A somewhat unexpected extension of Krasnoselski's result (on the real line) appeared in 1975. Hillam [43] showed that if $f\colon [a,b] \to [a,b]$ and $|f(x)-f(y)| \leqslant M|x-y|$ then the sequence defined by

$$x_{n+1} = \frac{1}{M+1}f(x_n) + \frac{M}{M+1}x_n$$

converges to a fixed point of f.

Perhaps the ultimate result on iteration of a real-valued real function is that of Franks and Marzec which appeared in 1971 [44]. The result to which we refer is the following. If f is any continuous function and $f\colon [a,b] \to [a,b]$, then the sequence defined by

$$x_{n+1} = \frac{nx_n + f(x_n)}{n+1}$$

converges to a fixed point of f.

Newton's Method and Variations

We turn our attention next to those general iteration techniques which explicitly involve the derivative. In 1669 Newton showed an unpublished tract entitled *De Analysi per aequations numero terminorum infinitas* to Isaac Barrow (1630–1677) and Collins [45]. In this tract he illustrates "the numerical resolution of affected equations." His example is $y^3 - 2y - 5 = 0$; his method is as follows. By substitution we see that our solution is near 2. Thus let $y = 2 + p$ and substitute in the equation above to obtain:

$$p^3+6p^2+12p+8-4-2p-5=0$$

or, equivalently,

$$p^3+6p^2+10p-1=0.$$

We now reason that p^3+6p^2 is small and thus $10p-1\doteq 0$ so that $p\doteq .1$. It follows that $y\doteq 2.1$. Setting $p=.1+q$ and substituting into $p^3+6p^2+10p-1=0$ we have

$$.001+.03q+.3q^2+q^3+.06+1.2q+6q^2+1+10q-1=0$$

or, equivalently,

$$q^3+6.3q^2+11.23q+.061=0.$$

Again we ignore $q^3+6.3q^2$ and obtain $11.23q+.061\doteq 0$. Thus $q\doteq -.0054$, $y\doteq 2.0946$ and so on [**46**].

From our point in time we are able to represent more concisely what is taking place. We have a function f (in Newton's example a polynomial) and we wish to solve $f(x)=0$. Given some first approximation y we set $f(y+p)=0$ and attempt to solve for p. But

$$f(y+p)=f(y)+f'(y)p+\frac{f''(y)p^2}{2}+\cdots;$$

thus if we, like Newton, ignore all terms involving powers of p higher than the first and recall that $f(y+p)$ was assumed to be 0, we have $f(y)+f'(y)p\doteq 0$ or equivalently $p\doteq -f(y)/f'(y)$. Hence $y-f(y)/f'(y)$ more nearly approximates a solution.

We should emphasize that the paragraph above is a modern reconstruction. Indeed we have used Taylor's Theorem for our exposition and Brook Taylor (1685–1731) published his theorem only in 1712. As an aside it is interesting to note that Taylor credits Halley's method of root extraction as the source of his original idea [**47**].

In 1690 Joseph Raphson (1648–1715) published a tract entitled *Analysis Aequationum Universalis seu Ad Aequationes Algebraicas Resolvendas Methodus Generalis, & Expedita, Ex nova Infinitarum Serierum Methodo.* In this tract he gives an application of Newton's procedure which is at once easier and more nearly the procedure we now call Newton's Method [**48**]. First he considers the equation $a^3=d$. Setting $a=g+x$ he then observes that $g^3+3g^2x+3gx^2+x^3=d$. Neglecting all powers of x above the first, he obtains

$$x\doteq\frac{d-g^3}{3g^2}$$

and it follows that

$$a\doteq g-\frac{g^3-d}{3g^2}.$$

Of course this is easily seen to be $a\doteq g-f(g)/f'(g)$ if we define $f(z)=z^3-d$.

Raphson likewise shows that if $a=g+x$ is a solution to $ba-a^3=c$, then

$$x\doteq\frac{c+g^3-bg}{b-3g^2}.$$

Again if $f(z) = bz - z^3 - c$ we have $a = g + x \doteq g - f(g)/f'(g)$. Later he considers the specific example $77284a - a^3 = 8083128$ [49]. Taking an initial estimate of $g_1 = 200$ he obtains successively $g_2 = 186$, $g_3 = 180.6$, $g_4 = 179.814$, $g_5 = 179.797652$.

Thomas Simpson (1710–1761) in 1740 in an essay *A New Method for the Solution of Equations in Numbers*, states the rule

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{9}$$

rhetorically. In 1798 in Lagrange's *Traité de la Résolution des Équations Numériques* the functional expression finally appears [50].

Many writers have given variations of Newton's Method (9) and it is interesting to note that in some sense most follow a hint given already by Newton himself. In *De Analysi*, following his treatment of $y^3 - 2y - 5 = 0$, Newton comments that he might have obtained p more accurately in his approximation $y = 2 + p$ (see above). He writes, "Moreover it is to be observed that in this Example, if I had any doubt whether $0.1 = p$ approached near enough to the Truth, instead of $10p - 1 = 0$, I had seigned $6p^2 + 10p - 1 = 0, \ldots$" [51].

That is, we might take y as an approximate solution to $f(x) = 0$ and set

$$0 = f(p + y) = f(y) + f'(y)p + \frac{f''(y)p^2}{2} + \cdots.$$

Then neglecting all powers of p higher than the second we may solve

$$0 = f(y) + f'(y)p + \frac{f''(y)}{2}p^2 \quad \text{for } p. \tag{10}$$

Proceeding thus we find

$$p = \frac{-f'(y) \pm \sqrt{f'(y)^2 - 2f(y)f''(y)}}{f''(y)}. \tag{11}$$

This yields two iteration formulas, namely:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} - \frac{\sqrt{f'(x_n)^2 - 2f(x_n)f''(x_n)}}{f''(x_n)} \tag{12}$$

and

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} + \frac{\sqrt{f'(x_n)^2 - 2f(x_n)f''(x_n)}}{f''(x_n)}. \tag{13}$$

Alternately we may set $y = x_n$ and $y + p = x_{n+1}$ in equation (10) and obtain

$$f(x_n) + f'(x_n)(x_{n+1} - x_n) + \frac{f''(x_n)(x_{n+1} - x_n)^2}{2} = 0. \tag{14}$$

From (14) it follows that

$$x_{n+1} - x_n = \frac{-f(x_n)}{f'(x_n) + \dfrac{f''(x_n)}{2}(x_{n+1} - x_n)}. \tag{15}$$

On the right-hand side of (15) one may now use

$$x_{n+1}-x_n \doteq -\frac{f(x_n)}{f'(x_n)}$$

from (9) and obtain, after some manipulation

$$x_{n+1}=x_n-\frac{2f(x_n)f'(x_n)}{2f'(x_n)^2-f(x_n)f''(x_n)}. \tag{16}$$

This last derivation follows Wall [52].

Equations (13) and (16) are found implicitly in Halley's work of 1694 [53]. Indeed, equation (16) is commonly called Halley's Method. Where the iterations first appear explicitly is not clear.

A different generalization of Newton's Method was discovered by Tschebyscheff. Returning to equation (10) we find

$$pf'(x_n) \doteq -f(x_n)-\frac{p^2}{2}f''(x_n)$$

and thus

$$p \doteq -\frac{f(x_n)}{f'(x_n)}-p^2\frac{f''(x_n)}{2f'(x_n)}. \tag{17}$$

If we now use

$$p \doteq -\frac{f(x_n)}{f'(x_n)}$$

in the right-hand side of (17) we obtain

$$p \doteq -\frac{f(x_n)}{f'(x_n)}-\left(\frac{f(x_n)}{f'(x_n)}\right)^2\frac{f''(x_n)}{2f'(x_n)}.$$

Finally, since $p=x_{n+1}-x_n$ we obtain Tschebyscheff's iteration scheme

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}-\left(\frac{f(x_n)}{f'(x_n)}\right)^2\frac{f''(x_n)}{2f'(x_n)}. \tag{18}$$

Tschebyscheff gave (18) in a paper written in 1838. This paper took a silver medal in a competition in 1840–41, but was not published until 1951 [54].

There is a great temptation at this point to include a recent variation of Newton's Method due to Burgstahler. Burgstahler's distinctly different and clever variation applies only to polynomials however. Since we have omitted mention of other polynomial rootfinders, we leave it to the reader to enjoy Burgstahler's paper [55] without commentary. Instead, we close by observing that we have come full cycle. That is, Newton's Method (9) applied to the polynomial $f(x)=x^n-N$ yields the Babylonian square-root algorithm (3) when $n=2$, and Newton's root-extraction formulae (4) when $n=3,4$. Moreover, if we apply the improved Newton method (16) to the same polynomial, we obtain (6), the nth-root algorithm of Halley; or do you prefer Lambert? Or how about Hutton?

REFERENCES

1. Harry Bateman, Halley's methods for solving equations, *Amer. Math. Monthly* 45 (1938), 11–17.
2. Evert M. Bruins, On the history of approximative computation, *Janus* 60 (1973), 199–206.
3. J. F. Traub, *Iterative Methods for the Solution of Equations*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.
4. Herman H. Goldstine, *A History of Numerical Analysis from the 16th through the 19th Century*, Springer-Verlag, New York, 1977.
5. O. Neugebauer and A. Sachs, *Mathematical Cuneiform Texts*, American Oriental Society, New Haven, Conn., 1945, pp. 42–43.
6. ______, p. 43.
7. C. B. Boyer, *A History of Mathematics*, John Wiley and Sons, Inc., New York, 1968, pp. 30–31; Bruins, 199–200; P. S. Jones, The square root of two in Babylonia and America, *Mathematics Teacher* 42 (1949), 307–308.
8. Boyer, pp. 30–31.
9. T. J. I'a. Bromwich, The methods used by Archimedes for approximating to square roots, *The Mathematical Gazette* 14 (1928), 253–257; Tobias Dantzig, *The Bequest of the Greeks*, Charles Scribner's Sons, New York, 1955, pp. 152–159; D. C. Gazis and Robert Herman, Square roots geometry and Archimedes, *Scripta Mathematica* 25 (1960), 229–241.
10. Howard Eves, *An Introduction to the History of Mathematics*, 3rd edition, Holt, Rinehart and Winston, Inc., 1969, pp. 157–158.
11. Morris R. Cohen and I. E. Doabkin, *A Source Book in Greek Science*, Harvard University Press, Cambridge, Mass., 1969, pp. 85–86.
12. O. C. Smyly, Square roots in Heron of Alexandria, *Hermathena* 63 (1944), 18–26.
13. Thomas Heath, *A History of Greek Mathematics*, Oxford University Press, Oxford, 1965, pp. 324-325.
14. B. Data, Nârâyana's method for finding approximate value of a surd, *Bull. Calcutta Math. Soc.* 23 #4 (1931), 187–194.
15. S. J. Rigaud, *Correspondence of Scientific Men II*, Georg Olms Hildesheim, 1965 (reprint of Oxford 1841), p. 372.
16. D. T. Whiteside, John Collins, *The Dictionary of Scientific Biography*, Vol. 3, edited by Charles C. Gillispie, Charles Scribner's Sons, New York, 1972, pp. 348–349.
17. It is interesting to note the persistence of interest in the iteration scheme (4). The scheme formed the basis of a paper in the *Annals* as late as 1909; C. L. Bouton, Discussion of a method for finding numerical square roots, *Annals of Mathematics*, ser. 2, Vol. 10 (1909), 167–172.
18. M. A. Nordgaard, *A Historical Survey of Algebraic Methods of Approximating the Roots of Numerical Higher Equations up to the Year 1819*, Teachers College, Columbia University, 1922, p. 38.
19. Edmund Halley, A new, exact and easier Method of finding Roots of Equations generally, and that without any previous Reduction, *Miscellanea Curiosa* 2, London (1708), 70–88.
20. Nordgaard seems to have a misprint. On page 39 he gives

$$a + \frac{ab + a}{3a^2 + b + 1} > z > a + \frac{ab}{3a^2 + b}.$$

    His example, however, applies the correct inequality and when DeLagny's result is given on page 41 it is correct.
21. Nordgaard, p. 39.
22. Halley, p. 71.
23. There is a misprint in Halley. There the quantity shown is $ab/(3a^2 + b)$.
24. Halley, p. 77.
25. Bateman (1938), p. 13.
26. The result appears in a section titled "Vorläufige Kenntnisse für die, so die Quadratur und Rectification des Circuls suchen." This appears in the second volume of the *Beyträge*, pp. 140–169. The above is reproduced in Lambert's *Opera Mathematica*, Vol. 1, pp. 194–212.
27. Charles Davies and W. G. Peck, *Mathematical Dictionary and Cyclopedia of Mathematical Science*, A. S. Barnes and Co., 1876, p. 242.
28. Margaret E. Baron, *The Dictionary of Scientific Biography*, Vol. 6, pp. 576–577.
29. *Hutton's Tracts on Mathematical and Philosophical Subjects*, Vol. 1, London, 1812, p. 210.
30. Charles Hutton, *A Course of Mathematics* (from the 5th and 6th London editions, revised and corrected by Robert Adrain), New York, 1812, p. 88.
31. Otto Dunkel, A note on the computation of arithmetic roots, *Amer. Math. Monthly* 34 (1927), 366.

32. S. M. Jacob, On sequences which determine the nth root of a rational number, *Proc. London Math. Soc.* ser. 2, Vol. 1 (1903), 166–174.

33. J. V. Uspensky, Note on the computation of roots, *Amer. Math. Monthly* 34 (1927), 130–134.

34. V. A. Bailey, Prodigious calculation, *Austral. J. Science* 3 (1941), 78–80.

35. E. S. Kennedy and W. R. Transne, A medieval iterative algorism, *Amer. Math. Monthly* 63 (1956), 80–83.

36. E. G. R. Taylor, *The Mathematical Practitioners of Tudor and Stuart England*, Cambridge, 1954, p. 217; see also the letter of Collins to Sir John Frederick, June 24, 1673. This letter appears in Rigaud I, 204; Dary is mentioned again in a letter from Collins to Gregory. See Rigaud II, 198.

37. Rigaud, II, p. 365.

38. Rigaud, II, pp. 255–256.

39. R. Ross, A method of solving algebraic equations, *Nature* 78 (1908), 663–665.

40. Stefan Banach, Sur les opérations dans les ensembles abstraits et leur applications aux équations intégrals, *Fund. Math.* 3 (1922), 160.

41. W. Cheny and A. A. Goldstein, Proximity maps for convex sets, *Proc. Amer. Math. Soc.* 10 (1959) 448–450; see also M. Edelstein, An extension of Banach's contraction principle, *Proc. Amer. Math. Soc.* 12 (1961), 7–10.

42. M. A. Krasnoselski, Two remarks on the method of successive approximations, *Uspekhi Mat. Nauk.* (n.s.) #1, Vol. 10 (1955), 123–127; For an English translation see F. F. Bonsall, *Lectures on some Fixed Point Theorems of Functional Analysis*, Tata Institute of Fundamental Research, Bombay, 1962, p. 30.

43. B. P. Hillam, A generalization of Krasnoselski's theorem on the real line, *Math. Mag.* 48 (1975), 167–168.

44. R. L. Franks and R. P. Marzec, A theorem on mean-value iterations, *Proc. Amer. Math. Soc.* 30 (1971), 324–326.

45. Nordgaard, p. 36.

46. Newton, Analysis by equations of an infinite number of terms, translated by John Stewart, London, 1745, *The Mathematical Works of Isaac Newton*, Vol. 1, edited by Derek T. Whiteside, Johnson Reprint Corporation, New York, 1964, pp. 10–11.

47. H. Bateman, The correspondence of Brook Taylor, *Bibliotheca Mathematica* ser. 3, Vol. 7 (1907), 367–371.

48. Nordgaard, pp. 37–38.

49. Nordgaard gives the equation as $77284a - a^3 = 8013128$ but since none of the following calculations agree with this statement we assume a typographical error.

50. Nordgaard, p. 45.

51. Newton, p. 11.

52. H. S. Wall, A modification of Newton's method, *Amer. Math. Monthly* 55 (1948), 90–94.

53. Bateman (1938), p. 12.

54. Bruins, p. 202.

55. Sylvan Burgstahler, An algorithm for solving polynomial equations, *Amer. Math. Monthly* 93 (1986), 421–430.

NOTES

Flipping a Coin over the Telephone

CHARLES VANDEN EYNDEN
Illinois State University
Normal, IL 61761

In the spring of 1983, when Ralph Sampson was about to graduate from the University of Virginia, Charlie Thomas, an owner of 20 automobile dealerships, and Frank Mariani, with interests in real estate and horse racing, had good reason to feel nervous. Thomas and Mariani were, respectively, chairman of the board of the Houston Rockets and president of the Indiana Pacers, two teams in the professional National Basketball Association. Sampson, as most people know, was a player of unusual skill and agility, sufficient, in fact, for him to play forward despite his height of 7 feet and 4 inches.

Sampson was of special interest to the Rockets and Pacers because these teams had finished last in the Western and Eastern divisions of the NBA, thus earning a possible first choice in the college draft. A flip of a coin was to determine which team would get to choose Sampson and thus, according to many observers, be assured of a successful franchise for years to come. *The conversation that follows is pure fiction.*

Thomas (in Houston): Frank, it looks like you and I will have to flip a coin to see who gets first choice in the draft this year. Why don't you fly down here and we'll get it over with?

Mariani (in Indianapolis): Here's a better idea—you fly up here. Texas starts to get so hot this time of year.

Thomas: I'm not sure I want to visit allergy country right now. Maybe we can avoid either of us traveling. Why don't I just flip a coin right now and you call it heads or tails?

Mariani: How will I know if I've won or not?

Thomas: I'll tell you!

Mariani (after a pause): That's a great idea, but I'll save you the trouble. I just happen to have a coin in my hand already, and I've flipped it. What's your call, heads or tails?

Thomas: Maybe we could meet in Arkansas.

The purpose of this note is to explain how the flip of a coin can be handled over the telephone so that both parties will be satisfied that everything is on the up and up. First I will lay out the mechanics of the transaction (usually called the "protocol" in the literature), and then explain the mathematics involved (only elementary number theory), including the actual computational methods used, and why both parties will be assured that they were treated fairly.

The protocol. Let us call the participants Mr. H and Mr. I. Mr. H starts by choosing two distinct large prime numbers p and q, each of the form $4k+3$. He calculates $n = pq$ and tells Mr. I the value of n.

Now the ball is in Mr. I's court. He chooses an integer x between 0 and n that is relatively prime to n, and calculates a number a such that $x^2 \equiv a \pmod{n}$ and $0 < a < n$. Of course a is just the remainder when x^2 is divided by n. (That x is relatively prime to n can be checked by computing the greatest common divisor of x and n.) This is where the actual flip of the coin occurs. Mr. I tells Mr. H the value of a.

Mr. H now finds all integers t such that

$$t^2 \equiv a \pmod{n}, \qquad 0 < t < n. \tag{1}$$

It turns out that there will be exactly four such values of t, two of which are x and $n - x$, and two of which are not. Mr. H chooses one of the four values, say y, and tells it to Mr. I. This choice corresponds to calling the coin heads or tails. If either y or $n - y$ equals x, then Mr. H wins; otherwise Mr. I wins. In the second case Mr. I proves that he has won by announcing the actual value of x, which is neither y nor $n - y$.

An example. We will illustrate the protocol with reasonably small numbers, although in practice much larger values would be used. Suppose Mr. H chooses the primes $p = 23$ and $q = 31$, and sends to Mr. I the value $n = 23 \cdot 31 = 713$. Mr. I responds by randomly choosing $x = 220$, and computes that $x^2 = 48400 \equiv 629 \pmod{713}$. He sends back the value $a = 629$. Notice that Mr. I knows two solutions to

$$t^2 \equiv 629 \pmod{713}, \qquad 0 < t < 713, \tag{2}$$

namely, 220 and $493 = 713 - 220$.

Now Mr. H computes all the solutions to (2), using his knowledge of the factorization $713 = 23 \cdot 31$. (We will explain his computational methods later.) He finds the solutions $t = 59, 220, 493$, and 654. Note that $493 \equiv -220$ and $654 \equiv -59 \pmod{713}$, so that the solutions can be split into two pairs. Mr. H does not know whether Mr. I started with the pair 220, 493 or the pair 59, 654, however. If Mr. H chooses 220 or 493, he wins the coin flip, but if he chooses 59 or 654, he loses.

Why Mr. I can't cheat. The obvious difficulty with flipping a coin over the telephone is that the person flipping the coin could cheat, telling the person who calls heads or tails that he is wrong no matter how the coin came up. Mr. I corresponds to the person flipping the coin in our protocol. Why couldn't he say that Mr. H was wrong no matter which value y he announces?

We said that Mr. I can prove Mr. H wrong by revealing his original value of x, which is neither y nor $n - y$. But why couldn't Mr. I also compute all four solutions to (1), then claim that his original x was one of the two other than y and $n - y$?

The answer is that Mr. I needs to know the factorization of n in order to compute the other two solutions. In fact knowing a solution other than y or $n - y$ is equivalent to being able to factor n, as we will show in the next paragraph. In practice p and q would be chosen to be very large, perhaps 100 decimal digits each, making n around 200 digits. No efficient method is known for factoring large numbers, even though mathematicians have been working on this problem for centuries. It can be estimated that to factor a 200-digit integer, using the most efficient methods known and fastest available computer, would take millions of years. Thus Mr. I cannot cheat because he cannot factor n, and thus only knows his original solution x to (1) and its twin $n - x$.

It is quite easy to see that knowing a third solution to the congruence (1) allows one to factor n. For suppose y is a solution to (1) and y is congruent to neither x nor $-x$ $(\bmod\ n)$. Now

$$y^2 \equiv a \equiv x^2 \pmod{n},$$

and so $n = pq$ divides $y^2 - x^2 = (y-x)(y+x)$. If n divides $y - x$ then $y \equiv x \pmod{n}$, contrary to our assumption. Likewise n cannot divide $y + x$. We conclude that p divides one of $y - x$ and $y + x$ and q divides the other, and so the greatest common divisor of $n = pq$ and $y - x$ must be p or q. Thus a factor of n can be found by applying the Euclidean algorithm to n and $y - x$. This algorithm works very quickly, even on numbers as large as 200 decimal digits, as we will see in the next section.

We can illustrate this with our example by supposing Mr. H sends $y = 59$. Then $y - x = 59 - 220 = -161$, and we easily compute the greatest common divisor of -161 and 713 to be 23, which was the factor p.

The efficiency of the Euclidean algorithm. The Euclidean algorithm allows us to compute the greatest common divisor of integers u and v, $v > 0$, by writing

$$\begin{aligned} u &= vq_1 + r_1, & 0 &< r_1 < v, \\ v &= r_1q_2 + r_2, & 0 &< r_2 < r_1, \\ r_1 &= r_2q_3 + r_3, & 0 &< r_3 < r_2, \\ r_{n-2} &= r_{n-1}q_n + r_n, & 0 &< r_n < r_{n-1}, \end{aligned} \tag{3}$$

where r_n is the last nonzero remainder. Then, as is well known, r_n is the greatest common divisor of u and v.

THEOREM 1. *If the Euclidean algorithm is applied to the integers u and v with $0 < v < u$, then the number of divisions is no more than $2\log_2 u$.*

Proof. It suffices to show that $r_1 < u/2$, so that the numbers at the left of the equations (3) decrease by at least a factor of 2 after each two divisions. If $v \leqslant u/2$ then this is easy, since $r_1 < v$. But if $v > u/2$ then $r_1 = u - vq_1 \leqslant u - v < u - u/2 = u/2$.

Although Lamé proved in 1844 that the number of divisions needed in the Euclidean algorithm is no more than 5 times the number of decimal digits in the smaller of u and v, the above theorem is sufficient for our purposes. For example, it shows that if the algorithm is applied to two numbers of no more than 200 digits, then the number of divisions necessary is no more than $2\log_2 10^{200} = 400\log_2 10 < 1400$. Such a computation would take a large computer less than a second.

In the next section we will use the fact that the Euclidean algorithm can also be used to find integers α and β such that

$$\alpha u + \beta v = d,$$

where $d = (u, v)$. This is best done on a computer by defining $\alpha_{-1} = \beta_0 = 1$, $\alpha_0 = \beta_{-1} = 0$, and, for $i \geqslant 1$,

$$\alpha_i = \alpha_{i-2} - q_i\alpha_{i-1}, \qquad \beta_i = \beta_{i-2} - q_i\beta_{i-1}.$$

Then it is easily proved by induction that $r_i = \alpha_i u + \beta_i v$ for $i > 0$, and, in particular, that $\alpha_n u + \beta_n v = r_n = d$. The α's and β's can be computed recursively at the same time as the Euclidean algorithm is applied. Since computing each α and β requires a multiplication and a subtraction, the total number of elementary arithmetic operations performed is five times the number needed for the Euclidean algorithm alone, and the computation is still quite reasonable.

We will illustrate this with $u = 31$ and $v = 23$. (The reason for choosing these numbers will become apparent later.) The Euclidean algorithm gives

$$\begin{aligned} 31 &= 1 \cdot 23 + 8, \\ 23 &= 2 \cdot 8 + 7, \\ 8 &= 1 \cdot 7 + 1, \end{aligned}$$

so $q_1 = 1$, $q_2 = 2$, and $q_3 = 1$. Then $\alpha_1 = \alpha_{-1} - q_1\alpha_0 = 1 - 1 \cdot 0 = 1$. Likewise $\alpha_2 = 0 - 2 \cdot 1 = -2$ and $\alpha_3 = 1 - 1(-2) = 3$. In a similar way $\beta_1 = -1$, $\beta_2 = 3$, and $\beta_3 = -4$. Thus

$$1 = (31, 23) = 3 \cdot 31 + (-4)23.$$

Can Mr. H do his part? Since the fairness of the protocol depends on the inability of Mr. I to compute another solution to (1) when n is a 200 digit number, the reader may wonder whether Mr. H will be able to compute all four solutions, even knowing the factorization of n. Of course, with such large numbers a computer will be necessary; but even the world's fastest computer is not enough to allow Mr. I to factor n. In this section we explain the methods Mr. H will use, all based on elementary number theory. We will have to show that our algorithms not only work, but do so in a reasonable time when applied to large numbers.

Mr. H's method will be to solve (1) via the pair of congruences

$$\gamma^2 \equiv a \pmod{p}, \qquad \delta^2 \equiv a \pmod{q}, \tag{4}$$

getting two incongruent solutions to each of these quadratic congruences. These solutions are then used with the Chinese Remainder Theorem to generate the four desired solutions to (1). The following theorem is an easy application of Fermat's Theorem, which says that if p is prime and p does not divide ξ, then $\xi^{p-1} \equiv 1 \pmod{p}$.

THEOREM 2. *If $p = 4k + 3$ is a prime not dividing a and if $\xi^2 \equiv a \pmod{p}$ has a solution, then a^{k+1} is a solution to*

$$\gamma^2 \equiv a \pmod{p}.$$

Recall that the primes p and q were assumed to be of the form $4k + 3$, so in theory Theorem 2 can be used to compute solutions to (4). Since $k + 1$ may be a 100-digit number, however, the computation of a^{k+1} modulo p needs more attention. It turns out that computing powers to a modulus can be done quite efficiently. We start by writing k + 1 in binary form, say,

$$k + 1 = b_m 2^m + b_{m-1} 2^{m-1} + \cdots + b_0,$$

where each b_i is 0 or 1. One way to do this is to start with $k + 1$ and successively divide by 2, throwing away the remainders, which are $b_0, b_1, \ldots, b_m$. Note that $m < \log_2 10^{100} < 400$, so no more than 400 divisions are needed. Now

$$a^{k+1} = a^{b_0}(a^2)^{b_1}(a^4)^{b_2} \cdots (a^{2^m})^{b_m}. \tag{5}$$

We compute $a^2, a^4, \ldots, a^{2^m}$, by successively squaring and reducing (mod n), which entails at most 800 more multiplications and divisions. Finally we compute the right side of (5) by successive multiplication and reduction (mod p), with at most 800 more operations. Notice that in these computations we never have to deal with integers exceeding p^2.

We will illustrate this method by solving $\gamma^2 \equiv 629 \pmod{23}$. Here $p = 4 \cdot 5 + 3$, so $k = 5$ and $k + 1 = 6$. Now $6 = 2 \cdot 3 + 0, 3 = 2 \cdot 1 + 1$, and $1 = 2 \cdot 0 + 1$, so $6 = 0 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2$. Then

$$\begin{aligned} a &= 629 \equiv 8 \pmod{23}, \\ a^2 &\equiv 64 \equiv 18 \pmod{23}, \quad \text{and} \\ a^4 &\equiv 18^2 = 324 \equiv 2 \pmod{23}. \end{aligned}$$

Then $a^{k+1} = a^6 = a^2a^4 \equiv 18 \cdot 2 = 36 \equiv 13 \pmod{23}$, giving the solution $\gamma = 13$. The reader may want to check that using the same method on the congruence $\delta^2 \equiv 629 \pmod{31}$ leads to the solution $\delta = 28$.

Mr. H uses this technique to find integers γ and δ satisfying (4). He also finds integers α and β such that

$$\alpha p + \beta q = 1, \tag{6}$$

using the Euclidean algorithm method explained previously. Now consider the four values of

$$t = \pm\delta\alpha p \pm \gamma\beta q, \tag{7}$$

where the plus-or-minus signs are taken independently. It is easy to check that no two of these values are congruent (mod n). Also

$$t^2 \equiv (\pm\gamma\beta q)^2 \equiv \gamma^2(\beta q)^2 \equiv a(1)^2 \equiv a \pmod{p},$$

where we have used (4) and (6). Likewise $t^2 \equiv a \pmod{q}$. Since the distinct primes p and q both divide $t^2 - a$, so does $n = pq$, which shows that t is a solution to (1). Thus (7) gives Mr. H his required four solutions to congruence (1).

In our example (6) becomes $23\alpha + 31\beta = 1$, for which we have already computed the solution $\alpha = -4$, $\beta = 3$. We have also found the solutions $\gamma = 13$, $\delta = 28$ to (4). Thus

$$\begin{aligned} \pm\delta\alpha p \pm \gamma\beta q &= \pm 28(-4)23 \pm 13 \cdot 3 \cdot 31 = \pm 3785 \text{ and } \pm 1367 \\ &\equiv 654, 59, 220, \text{ and } 493 \pmod{713}. \end{aligned}$$

History and reference guide. The idea presented in this paper is due to Manuel Blum [1], and short expositions of it can be found in [7] and [8]. For background on the Euclidean algorithm, congruences, Fermat's Theorem, and the Chinese Remainder Theorem see [5], [7], or [9]. The computational complexity of arithmetic and number theoretic algorithms is treated in [3], [4], and [7].

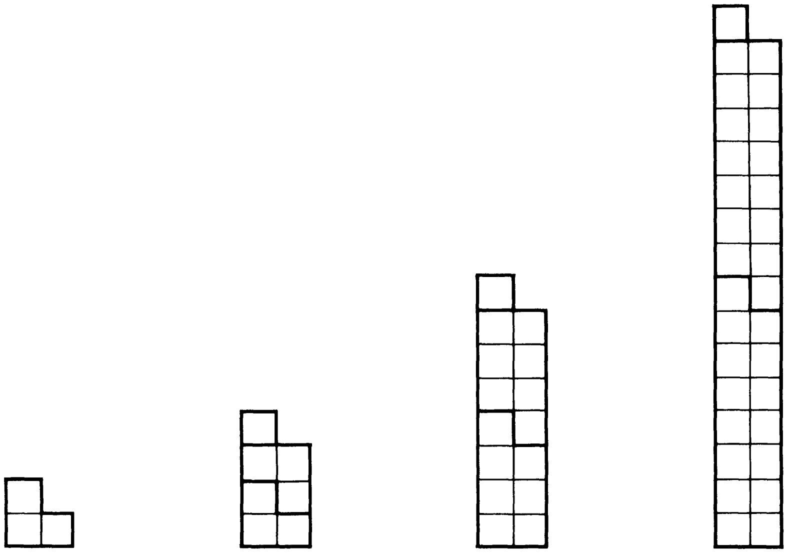
We have not considered how the 100-digit primes p and q can be found. It turns out that determining that a large number is prime can be done much faster than factoring a number of like size. See [2], [3], [4], or [6] for explanations of efficient primality tests, as well as the best present factorization methods.

Just before the 1983 NBA draft the Indianapolis Pacers were sold to Herb and Melvin Simon, local shopping center developers. Herb Simon and Charlie Thomas of the Houston Rockets were present on May 19 when NBA commissioner Larry O'Brien tossed a 100-year-old silver dollar to determine the first draft choice. Thomas called heads and won. By the 1985–86 season the Rockets, with the help of Sampson and another very tall player named Akeem Olajuwon, went to the NBA championship series, where they lost to the Boston Celtics in six games.

REFERENCES

1. Manuel Blum, Coin flipping by telephone—a protocol for solving impossible problems, *Proc. IEEE Spring Comp. Conf.* (1982), 133–137.
2. John Brillhart, D. H. Lehmer, J. L. Selfridge, Bryant Tuckerman, and S. S. Wagstaff Jr., *Factorizations of* $b^n \pm 1$, Amer. Math. Soc., 1983.
3. Donald E. Knuth, *The Art of Computer Programming*, vol. 2, *Seminumerical Algorithms*, 2nd ed., Addison-Wesley, 1981.
4. D. H. Lehmer, Computer technology applied to the theory of numbers, in *Studies in Number Theory*, W. J. LeVeque, ed., MAA, 1969.
5. Ivan Niven and H. S. Zuckerman, *The Theory of Numbers*, 4th ed., Wiley, 1980.
6. Carl Pomerance, *Lecture Notes on Primality Testing and Factoring*, MAA, 1984.
7. Kenneth H. Rosen, *Elementary Number Theory and Its Applications*, Addison-Wesley, 1984, 297–299.
8. M. R. Schroeder, *Number Theory in Science and Communication*, Springer-Verlag, 1984, 196–198.
9. Charles Vanden Eynden, *Elementary Number Theory*, Random House/Birkhäuser, 1987.

Proof without Words: Recursion



$A_2 = 3$ $\quad A_3 = 2A_2 + 1$ $\quad A_4 = 2A_3 + 1$ $\quad A_5 = 2A_4 + 1$

$$A_2 = 3 \,\&\, A_n = 2A_{n-1} + 1 \Leftrightarrow A_n = 2(2^{n-1}) - 1 = 2^n - 1$$

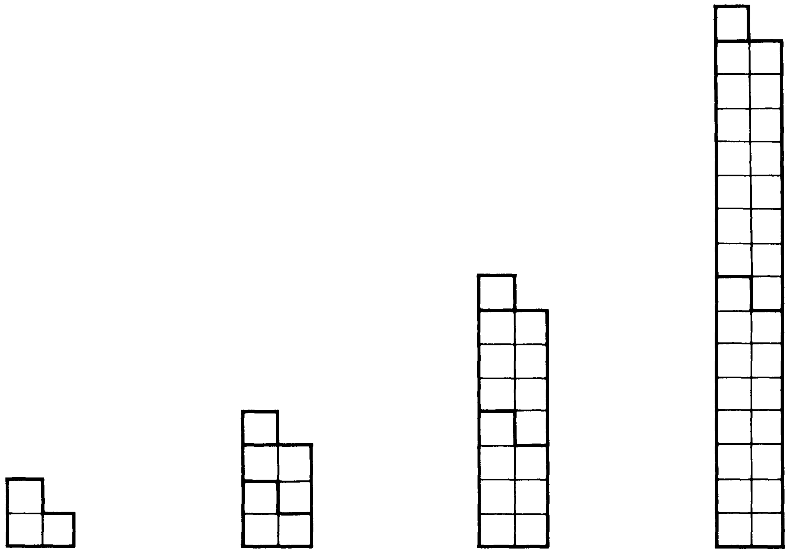
—Shirley Wakin
University of New Haven
New Haven, CT 06516

another very tall player named Akeem Olajuwon, went to the NBA championship series, where they lost to the Boston Celtics in six games.

REFERENCES

1. Manuel Blum, Coin flipping by telephone—a protocol for solving impossible problems, *Proc. IEEE Spring Comp. Conf.* (1982), 133–137.
2. John Brillhart, D. H. Lehmer, J. L. Selfridge, Bryant Tuckerman, and S. S. Wagstaff Jr., *Factorizations of* $b^n \pm 1$, Amer. Math. Soc., 1983.
3. Donald E. Knuth, *The Art of Computer Programming*, vol. 2, *Seminumerical Algorithms*, 2nd ed., Addison-Wesley, 1981.
4. D. H. Lehmer, Computer technology applied to the theory of numbers, in *Studies in Number Theory*, W. J. LeVeque, ed., MAA, 1969.
5. Ivan Niven and H. S. Zuckerman, *The Theory of Numbers*, 4th ed., Wiley, 1980.
6. Carl Pomerance, *Lecture Notes on Primality Testing and Factoring*, MAA, 1984.
7. Kenneth H. Rosen, *Elementary Number Theory and Its Applications*, Addison-Wesley, 1984, 297–299.
8. M. R. Schroeder, *Number Theory in Science and Communication*, Springer-Verlag, 1984, 196–198.
9. Charles Vanden Eynden, *Elementary Number Theory*, Random House/Birkhäuser, 1987.

Proof without Words: Recursion



$A_2 = 3$ $\quad$ $A_3 = 2A_2 + 1$ $\quad$ $A_4 = 2A_3 + 1$ $\quad$ $A_5 = 2A_4 + 1$

$$A_2 = 3 \,\&\, A_n = 2A_{n-1} + 1 \Leftrightarrow A_n = 2(2^{n-1}) - 1 = 2^n - 1$$

—Shirley Wakin
University of New Haven
New Haven, CT 06516

Multiplying Long Numbers

R. P. BOAS
Northwestern University
Evanston, IL 60201

Not everybody seems to realize how easy it is to multiply fairly long numbers with a pocket calculator. If your calculator (like mine) will produce an accurate 8-digit product of 4-digit integers, you break the factors into blocks of four (in effect, write them in base 10^4) and multiply by applying the scheme that you learned in grade school—except that you add the columns on the calculator. This process is quite fast for numbers of, say, 16 digits.

Incidentally, there is an early 19th-century book by Crelle which simply contains all products of pairs of 3-digit numbers; it was intended to simplify long multiplications in just this way. A few years ago somebody recomputed it and published it as *The Book Computer*, although it would have been simpler (and more attractive) to have photographed Crelle's book.

If, however, your numbers are much longer, the process just outlined (which is merely multiplication of polynomials by the distributive law, plus carries) can become tedious. If you have access to MACSYMA or a similar program, you can multiply numbers of any length. If the numbers are extremely long, it may save time to apply the Fast Fourier Transform; there is a succinct and lucid exposition of multiplication by this method in [1, p. 65]. Of course, you can write a program for long multiplication by using the distributive law; this might hardly seem worthwhile if you want to do only one or two multiplications.

I want to point out (more as a curiosity than as a serious contribution) that multiplication can be easier if you use a language like BASIC that has built-in matrix multiplication. If you can multiply two $k \times k$ matrices, and if you can multiply two n-digit integers to get a correct $2n$-digit product, then you can multiply two (kn)-digit integers rather easily, especially if you are willing to do the carries by hand. To illustrate, I take $k = 3$, $n = 6$; the principle is the same for any k.

Write the integers to be multiplied in base $B = 10^6$, and let $X = aB^2 + bB + c$, $Y = eB^2 + fB + g$. Let the computer evaluate the matrix product

$$\begin{bmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{bmatrix} \cdot \begin{bmatrix} g & 0 & 0 \\ f & g & 0 \\ e & f & g \end{bmatrix}.$$

Read the first column of the product matrix from the bottom up, then the rest of the first row from left to right. You get the product XY in base B, except that you still have to do the carries. Thus, if your computer can multiply $k \times k$ matrices, and can do $(2n)$-digit multiplication, you can multiply (nk)-digit integers.

At the time that I was interested in doing high-precision multiplication, there was supposedly an infinite-precision FORTRAN arithmetic package available on the central computer, but I didn't know FORTRAN and didn't want to learn it. The idea of using BASIC's multiplication occurred to me, although it is evidently inefficient because matrix multiplication evaluates more products than are actually used in multiplication by the distributive law. However, if you want only a few products, and don't already have a multiplication program, the matrix method saves the time that you might spend writing your own program. Later, I did write my own program to do multiplications directly.

One of my colleagues once found me in the process of evaluating a long product, and said, "That's a waste of time, because there's a canned program for it. Let me show you." I gave him the data, he ran the canned program—and got the wrong answer.

REFERENCE

1. P. Henrici, *Applied and Computational Complex Analysis*, vol. 3, Wiley, New York, 1986.

A Professor's Lot

Words: Hal Fredricksen

(Apologies to Gilbert & Sullivan)

When a student's not engaged in studying theories (studying theories),
Or solving little problems from his book (from his book),
He delights in forming complicated queries (cated queries)
Calculated to put teacher on the hook (on the hook).

Though these knotty questions often are beguiling (are beguiling),
They always put professors to their test (to their test).
Still with chalk in hand most teachers now are smiling (now are smiling),
For this is the time they find that they like best (they like best).

But, papers to be graded are no fun (are no fun),
A professor's lot is not a happy one.

A professor's job involves his writing papers (writing papers),
Of importance, even possibly profound (bly profound).
But his efforts then can vanish into vapors (into vapors),
If he fails to recognize some trivial bound (trivial bound).

But he perseveres considering each suggestion (each suggestion),
Trying theories bold and methods up to date (up to date).
Referees then subject his results to question (sults to question),
For even when new they can't be second rate (second rate).

Returning to those quizzes to be done (to be done),
A professor's lot is not a happy one.

And committee work's the bane of his existence (his existence),
Cause there's never any lack of it to do (it to do).
And he can't escape no matter his resistance (his resistance).
So he plods along no certain end in view (end in view).

And besides the time he labors with his lessons (with his lessons),
He keeps office hours for students of slow wit (of slow wit),
Answering endless questions and inane digressions (nane digressions)
Which go on until it's way past time to quit (time to quit).

Can he relax, no a seminar's begun (nar's begun).
A professor's lot is not a happy one.

you might spend writing your own program. Later, I did write my own program to do multiplications directly.

One of my colleagues once found me in the process of evaluating a long product, and said, "That's a waste of time, because there's a canned program for it. Let me show you." I gave him the data, he ran the canned program—and got the wrong answer.

REFERENCE

1. P. Henrici, *Applied and Computational Complex Analysis*, vol. 3, Wiley, New York, 1986.

A Professor's Lot

Words: Hal Fredricksen

(Apologies to Gilbert & Sullivan)

When a student's not engaged in studying theories (studying theories),
Or solving little problems from his book (from his book),
He delights in forming complicated queries (cated queries)
Calculated to put teacher on the hook (on the hook).

Though these knotty questions often are beguiling (are beguiling),
They always put professors to their test (to their test).
Still with chalk in hand most teachers now are smiling (now are smiling),
For this is the time they find that they like best (they like best).

But, papers to be graded are no fun (are no fun),
A professor's lot is not a happy one.

A professor's job involves his writing papers (writing papers),
Of importance, even possibly profound (bly profound).
But his efforts then can vanish into vapors (into vapors),
If he fails to recognize some trivial bound (trivial bound).

But he perseveres considering each suggestion (each suggestion),
Trying theories bold and methods up to date (up to date).
Referees then subject his results to question (sults to question),
For even when new they can't be second rate (second rate).

Returning to those quizzes to be done (to be done),
A professor's lot is not a happy one.

And committee work's the bane of his existence (his existence),
Cause there's never any lack of it to do (it to do).
And he can't escape no matter his resistance (his resistance).
So he plods along no certain end in view (end in view).

And besides the time he labors with his lessons (with his lessons),
He keeps office hours for students of slow wit (of slow wit),
Answering endless questions and inane digressions (nane digressions)
Which go on until it's way past time to quit (time to quit).

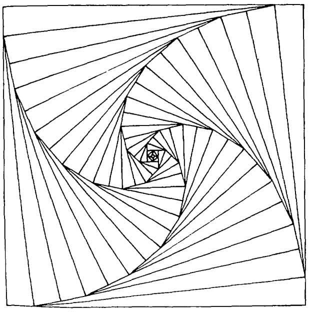
Can he relax, no a seminar's begun (nar's begun).
A professor's lot is not a happy one.

A Pythagorean Tiling of the Plane

ERNEST J. ECKERT
University of South Carolina
Aiken, SC 29801

HUGO HAAGENSEN
Aalborg, Denmark

Each of the four Pythagorean triangles surrounding the inner-most 1 × 1 square in the figure is a $(3,4,5)$ triangle. Reflecting each triangle in its hypotenuse we obtain a 7 × 7 square.



Each of the four Pythagorean triangles surrounding the 7 × 7 square is a $(5,12,13)$ triangle. Reflecting each triangle in its hypotenuse we obtain a 17 × 17 square.

This process may be continued indefinitely by appropriately choosing the sequence of Pythagorean triangles:

$$(a_1,b_1,c_1)=(3,4,5),(a_2,b_2,c_2)=(5,12,13),$$
$$(a_3,b_3,c_3)=(7,24,25),\dots,(a_{n-1},b_{n-1},c_{n-1}),(a_n,b_n,c_n),\dots.$$

The key to success is

$$b_n-a_n=b_{n-1}+a_{n-1}. \tag{1}$$

The sequence may be constructed by successively applying the matrix

$$A=\begin{pmatrix}1&-2&2\\2&-1&2\\2&-2&3\end{pmatrix}$$

to the 'vector' $(3,4,5)$. $A(3,4,5)=(5,12,13)$, $A(5,12,13)=(7,24,25),\dots$, $A(a_{n-1},b_{n-1},c_{n-1})=(a_n,b_n,c_n)$. It is a simple matter to check the validity of (1).

Figures similar to the one shown may be drawn by using other sequences of triangles satisfying (1). For example, had we started with $(a_1,b_1,c_1)=(8,15,17)$ the successive application of A would generate the sequence $(8,15,17),(12,35,37),(16,63,67),\dots$. The innermost square is now 7 × 7. We note that $c_n-b_n=2$, whereas in the first example $c_n-b_n=1$; c_n-b_n is, in fact, invariant under A.

Some Thoughts about Limits

RAY REDHEFFER
UCLA
Los Angeles, CA 90024

When I was a student at MIT one of my professors—a distinguished foreign scholar—expressed consternation at having received the following response on a written examination:

> Question: What is the meaning of $\lim_{x \to a} f(x) = L$?
> Answer: For every ε there is a δ.

This brings us to the purpose of this note, which is twofold. A first objective is to simplify the formulation and proof of theorems on limits. A second is to give precise definitions to, and so legitimize, certain modes of expression that are avoided by algebraists but embraced by analysts.

There is a sense in which theorems on limits are much easier than many of the standard homework problems in elementary calculus. For example, I would defend the view that it's easier to prove the theorem about the limit of a sum than to integrate $(\sec x)^5$. Yet many people have the opposite perception, as suggested by the anecdote above.

Why is this? Although various answers can be given, certainly a good deal of the trouble is caused by a piling up of logical terms: *For any ε there exists a δ such that for all x* of a certain kind something happens.... Quantifiers are the quicksand in which the path ends. Here we give a formulation in which quantifiers are kept separate, so that not too many of them pile up in one place. The approach has been classroom tested at UCLA, and the improvement in student performance is wholly out of proportion to the modest mathematical content of the innovation. The latter consists solely in the formulation and use of Theorem 1 (which follows).

Turning to the second objective, there seems to be a consensus that certain modes of expression common in analysis should be kept below stairs, in the scullery, and not admitted to polite company. As an extreme case, I recall reading somewhere that we musn't say a function is "increasing" because, after all, it's one and the same function all the time, so can't increase (or decrease or change in any other way). Still less, then, can one use phrases like "for x near a" or "for large x."

The view taken here is that such phrases have an honorable tradition in analysis, and express concepts of great importance, and should be used. If you want to say that $1/x$ is bounded near $x = 1$, or is less than 0.001 for large x, go ahead and say it. All that's missing is a precise definition. And not even that will be missing if you read far enough.

The main definitions. Although the methods suggested here apply to situations of great generality, it is assumed for simplicity that all functions and variables are real valued. We begin with the following:

DEFINITION 1. *A condition C holds for x near a if there exists a positive constant δ such that C holds when $x \neq a$ and $|x - a| < \delta$.*

For example, a function $f(x)$ is *bounded for x near a* if there exists a constant M such that $|f(x)| \leqslant M$ for x near a. Here the condition C is expressed by the inequality $|f(x)| \leqslant M$. Another example is given next:

Definition 2. *The statement* $\lim_{x \to a} f(x) = L$ *means the following: If* $\varepsilon > 0$ *then* $|f(x) - L| < \varepsilon$ *holds for x near a.*

Using lim as an abbreviation for $\lim_{x \to a}$ we shall develop the theory of limits by a series of examples. The first two of these depend only on the definitions.

Example 1 (boundedness). If $\lim f(x) = L$, then $f(x)$ is bounded for x near a. To see this, let $\varepsilon > 0$ be given. The condition $|f(x) - L| < \varepsilon$ holds for x near a and gives

$$|f(x)| = |f(x) - L + L| \leqslant |f(x) - L| + |L| < \varepsilon + |L|.$$

This shows that the bound M can be any number larger than $|L|$.

Example 2 (boundedness of reciprocal). If $L \neq 0$ in the above discussion then not only $f(x)$ but also $1/f(x)$ is bounded for x near a. This follows from

$$|f(x)| = |L - (L - f(x))| \geqslant |L| - |f(x) - L| > |L| - \varepsilon$$

with $0 < \varepsilon < |L|$. Taking reciprocals, we see that the bound for $|1/f(x)|$ could be any number larger than $|1/L|$.

Since Examples 1 and 2 both assert that a certain condition holds "for x near a" they involve Definition 1. The same applies to several of the examples given below. The fact that Definition 1 allows such a simple formulation is already worth the price of admission. However, the full scope of the definition is seen only in conjunction with the following theorem.

The main theorem. If C_1 and C_2 are two conditions, then $C_1 \wedge C_2$ holds when both C_1 and C_2 hold simultaneously. For example, if C_1 is the condition $f(x) > 2$, and C_2 is the condition $f(x) \leqslant 5$, then $C_1 \wedge C_2$ is the condition $2 < f(x) \leqslant 5$.

When two conditions C_1 and C_2 *each* hold for x near a, one would expect that C_1 and C_2 *both* hold for x near a. The following theorem shows that Definition 1 is consistent with this use of everyday language:

Theorem 1. *If* C_1 *and* C_2 *each hold for x near a, then* $C_1 \wedge C_2$ *holds for x near a.*

Proof. Let C_i hold for $x \neq a$ and $|x - a| < \delta_i$, where $i = 1$ or 2 and where each $\delta_i > 0$. Let $\delta = \min(\delta_1, \delta_2)$. Then $\delta > 0$, and $|x - a| < \delta$ implies $|x - a| < \delta_i$ for $i = 1, 2$. Hence $x \neq a$ and $|x - a| < \delta$ imply that C_1 and C_2 both hold. This completes the proof.

A corresponding result holds for any finite number of conditions $C_1, C_2, \ldots, C_n$. All we have to do is let i range over the indices $1, 2, \ldots, n$ instead of 1, 2. Alternatively, since $C_1 \wedge C_2 \wedge C_3 = (C_1 \wedge C_2) \wedge C_3$, the general result can be obtained by repeated use of Theorem 1 as it stands. In most applications, two or three conditions C_i suffice.

Further examples. Here we give some examples that require the main theorem. The first expresses the fact that $\lim f(x)$ depends only on the local behavior of f, that is, on the behavior near a. This is perhaps the most important of all the properties of limits:

Example 3 (localization). If $f(x) = g(x)$ for x near a, and if $\lim f(x) = L$, then $\lim g(x) = L$. For proof, let $\varepsilon > 0$ be given. Then $|f(x) - L| < \varepsilon$ holds for x near a by

Definition 2 and $f(x) = g(x)$ holds for x near a by hypothesis. Hence Theorem 1 tells us that *both* conditions hold for x near a. When both hold we have

$$|g(x) - L| = |f(x) - L| < \varepsilon.$$

Hence $|g(x) - L| < \varepsilon$ for x near a, and $\lim g(x) = L$ by Definition 2.

The localization theorem underlies familiar calculations like

$$\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = \lim_{x \to 1} (x + 1) = 2.$$

Here $f(x) = x + 1$ and $g(x)$ is the fraction on the left. It is not true that $f = g$ in the sense of functional equality, but we do have $f(x) = g(x)$ for x near 1 in the sense of Definition 1.

As another example, if $f(x)$ is the constant function $f(x) = c$ then it is a trivial consequence of the definition that $\lim f(x) = c$. But the localization theorem gives the same conclusion under the much weaker hypothesis that $f(x) = c$ for x near a.

The next result is an aid in proving the theorem about the limit of a product and it also gives results when the latter does not apply:

Example 4 (preservation of zero limit). If $f(x)$ is bounded near a and if $\lim g(x) = 0$, then $\lim f(x)g(x) = 0$. For proof choose $M > 0$ so that $|f(x)| \leqslant M$ for x near a and let $\varepsilon > 0$ be given. Then $\varepsilon/M > 0$ and hence $|g(x)| < \varepsilon/M$ for x near a. Thus, the conditions $|g(x)| < \varepsilon/M$ and $|f(x)| \leqslant M$ *each* hold for x near a. By Theorem 1 the conditions *both* hold for x near a. When both hold we have

$$|f(x)g(x)| < M\frac{\varepsilon}{M} = \varepsilon.$$

This gives $|f(x)g(x)| < \varepsilon$ for x near a and the result follows.

The next example requires the extension of Theorem 1 to three conditions C_i, as discussed above:

Example 5 (squeeze theorem). If $f(x) \leqslant g(x) \leqslant h(x)$ holds for x near a, and if $\lim f(x) = \lim h(x) = L$, then $\lim g(x) = L$. For proof, let $\varepsilon > 0$ be given. Then the conditions $|f(x) - L| < \varepsilon$ and $|h(x) - L| < \varepsilon$ each hold for x near a by Definition 2, and $f(x) \leqslant g(x) \leqslant h(x)$ holds for x near a by hypothesis. Hence *all three* conditions hold for x near a. When all three conditions hold we have

$$-\varepsilon < f(x) - L \leqslant g(x) - L \leqslant h(x) - L < \varepsilon.$$

This gives $|g(x) - L| < \varepsilon$ for x near a and completes the proof.

Example 6 (limit of a sum). If $\lim f_i(x) = L_i$ for $i = 1, 2$, then

$$\lim[f_1(x) + f_2(x)] = L_1 + L_2.$$

Proof. Let $\varepsilon > 0$ be given and note that also $\varepsilon/2 > 0$. The conditions $|f_i(x) - L_i| < \varepsilon/2$ *each* hold for x near a and $i = 1, 2$. Hence, by Theorem 1, *both* conditions hold for x near a. When both hold we have

$$\begin{aligned}|f_1(x) + f_2(x) - (L_1 + L_2)| &= |f_1(x) - L_1 + f_2(x) - L_2| \\ &\leqslant |f_1(x) - L_1| + |f_2(x) - L_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.\end{aligned}$$

Hence the left side is $< \varepsilon$ for x near a, and this gives the conclusion.

Example 7 (limit of a product). If $\lim f_i(x) = L_i$ for $i = 1, 2$, then $\lim f_1(x)f_2(x) = L_1L_2$. For proof, it follows from Definition 2 that $f_1(x)$ and $f_2(x)$ are *each* defined for x near a and hence by Theorem 1 they are *both* defined for x near a. When both are defined we have the identity

$$f_1(x)f_2(x) - L_1L_2 = f_1(x)[f_2(x) - L_2] + L_2[f_1(x) - L_1].$$

By Examples 1 and 4 each of the two terms on the right has the limit 0. Hence the sum also has limit 0 and the result follows. Here and in Example 8 below, we use the fact that $\lim f(x) = L$ is equivalent to $\lim[f(x) - L] = 0$. This is an immediate consequence of the definition.

Example 8 (limit of reciprocal). If $\lim f(x) = L \neq 0$ then $\lim(1/f(x)) = 1/L$. By Example 2, the function $1/f(x)$ is defined for x near a and we can write

$$\frac{1}{f(x)} - \frac{1}{L} = \frac{1}{Lf(x)}(L - f(x)).$$

The conclusion now follows from Examples 2 and 4. Writing $g/f = (1/f)g$ we see that a theorem regarding the limit of a quotient follows from Examples 7 and 8.

Example 9 (uniqueness). If $\lim f(x) = L_i$ for $i = 1$ and 2, then $L_1 = L_2$. For proof let $\varepsilon > 0$ and note that the two conditions $|f(x) - L_i| < \varepsilon$ for $i = 1$ and 2 *each* hold for x near a, hence *both* hold for x near a. If x is a value for which both hold we have

$$|L_1 - L_2| = |L_1 - f(x) + f(x) - L_2| \leqslant |L_1 - f(x)| + |f(x) - L_2| < 2\varepsilon.$$

Since ε is arbitrary, and the left side is independent of ε, the left side must be 0. This gives $L_1 = L_2$.

On logical grounds this theorem should probably have been presented first. However, its postponement is justified on the psychological ground that, of all theorems on limits, it is perhaps the least exciting. Two questions: Are the results of the previous examples meaningful before we know the uniqueness property asserted in Example 9? And can we get uniqueness by applying Example 3 (the localization theorem) to the two functions f and g, taking $f = g$? The answers are given at the end of this note.

In the following example the abbreviation lim for $\lim_{x \to a}$ is not used, because the limits involve two points, a and b.

Example 10 (composite function). Let $\lim_{t \to b} f(t) = L$, $\lim_{x \to a} g(x) = b$, and suppose further that $g(x) \neq b$ for x near a. Then $\lim_{x \to a} f[g(x)] = L$. For proof let $\varepsilon > 0$ be given and choose $\eta > 0$ so that the two conditions $|t - b| < \eta$ and $t \neq b$ together imply $|f(t) - L| < \varepsilon$. (Here we are using Definitions 1 and 2, with the roles of x, a, δ taken respectively by t, b, η). The two conditions $g(x) \neq b$ and $|g(x) - b| < \eta$ *each* hold for x near a, hence *both* do. When both hold the variable $t = g(x)$ satisfies the conditions required for t above and we get

$$|f[g(x)] - L| = |f(t) - L| < \varepsilon.$$

This shows that the left side is $< \varepsilon$ for x near a and completes the proof.

A remark on continuity. By definition, the function f is *continuous* at $x = a$ if $\lim f(x) = f(a)$. Hence, with one exception, theorems on continuous functions follow immediately from theorems on limits, as established above. The exception concerns the continuity of composite functions, the analog of Example 10. Namely, if $f(t)$ is continuous at $t=g(a)$, and if $g(x)$ is continuous at $x=a$, then $f[g(x)]$ is continuous at $x=a$. The trouble is that Example 10 requires the extraneous condition $g(x)\neq g(a)$ for x near a, which is not needed here.

To see what is going on let us interpret the equation $\lim f(x)=f(a)$ by Definition 2. Namely, if $\varepsilon>0$ the inequality $|f(x)-f(a)|<\varepsilon$ holds for x near a. In the definition of limit the point $x=a$ has to be excluded, but since $|f(a)-f(a)|=0$, the defining inequality for continuity holds at $x=a$ automatically and the exclusion is no longer necessary. In the case at hand, suppose $f(t)$ is continuous at $t=b$. Then we can allow $g(x)=b$ without harm and a repetition of the proof of Example 10 gives the correct theorem on continuity of composite functions.

One-sided conditions. If the inequality $x\neq a$ in the side condition in Definition 1 is replaced by $x<a$ or $x>a$, it is said that C holds for x near $a-$ or $a+$, respectively. Similar notation applies to concepts that depend on Definition 1. For example, $f(x)$ is bounded near $a+$ if there exists a constant M such that $|f(x)|\leqslant M$ for x near $a+$. If a is replaced by a $-$ or a $+$ in Definition 2 the result is a definition of the *one-sided limits*

$$\lim_{x\to a-} f(x) \quad \text{or} \quad \lim_{x\to a+} f(x),$$

respectively. Definitions and theorems involving $a-$ or $a+$ are termed "one-sided" to distinguish them from the two-sided conditions discussed above. There is no need to repeat the proofs, since one-sided results are obtained by the purely mechanical process of writing $a-$ or $a+$ for a wherever it occurs.

It is obvious that, if a condition C holds for x near a, then C must hold both for x near $a-$ and for x near $a+$. The following theorem states a converse:

THEOREM 2. *If a condition C holds for x near $a-$ and for x near $a+$, then C holds for x near a.*

Proof. Choose $\delta_1>0$ so that C holds for $x<a$ and $|x-a|<\delta_1$. Also choose $\delta_2>0$ so C holds for $x>a$ and $|x-a|<\delta_2$. Let $\delta=\min(\delta_1,\delta_2)$. Then $\delta>0$, and the condition $|x-a|<\delta$ implies $|x-a|<\delta_i$ for $i=1$ and 2. Hence C holds if $|x-a|<\delta$ and $x<a$, and also if $|x-a|<\delta$ and $x>a$. Thus C holds if $|x-a|<\delta$ and $x\neq a$, and this completes the proof.

We illustrate the theorem by two examples.

Example 11 (boundedness). If $f(x)$ is bounded for x near $a-$ and also for x near $a+$, then $f(x)$ is bounded for x near a. To see this, choose M^- so that $|f(x)|<M^-$ holds for x near $a-$, and choose M^+ so $|f(x)|<M^+$ holds for x near $a+$. Let $M=\max(M^-,M^+)$. Then $|f(x)|<M$ holds for x near $a-$ and for x near $a+$. By Theorem 2 the same inequality holds for x near a and the result follows.

Example 12 (limits). If $\lim f(x)=L$ as $x\to a-$ and also as $x\to a+$, then $\lim f(x)=L$ as $x\to a$. For proof let $\varepsilon>0$ be given. Then the condition $|f(x)-L|<\varepsilon$ holds for x near $a-$, and also for x near $a+$. By Theorem 2 we have $|f(x)-L|<\varepsilon$ for x near a and this completes the proof.

Since the converse of Example 12 is trivial, we can say that $\lim f(x)$ exists if, and only if, the left- and right-hand limits exist and have the same value. This gives the best procedure to establish nonexistence of limits in typical textbook problems. For example, $x/|x|$ has no limit as $x \to 0$ because the right- and left-hand limits are not equal. A direct proof of the nonexistence by going back to the (ε, δ) definition is harder.

The left- and right-hand limits as $x \to a$ are often denoted by $f(a-)$ and $f(a+)$, respectively. By definition, $f(x)$ is *continuous from the left* or *right* at $x = a$ if $f(a) = f(a-)$ or $f(a) = f(a+)$, respectively. It is a consequence of Example 12 that $f(x)$ is continuous at $x = a$ if, and only if, it is continuous both from the left and from the right. This gives a simple test for continuity, analogous to the test for existence of a limit noted above.

Associated with the left- and right-hand limits are the corresponding left- and right-hand derivatives,

$$D^-f(a) = \lim_{x \to a-} \frac{f(x) - f(a)}{x - a}, \qquad D^+f(a) = \lim_{x \to a+} \frac{f(x) - f(a)}{x - a}.$$

By Example 12 the derivative $Df(a) = f'(a)$ exists if, and only if, the left- and right-hand derivatives exist and have the same value. For example, this shows *by inspection* that $|x|$ is not differentiable at $x = 0$. As in the cases above, an (ε, δ) proof by reference to the definition of derivative is harder.

Large x. Following the pattern of Definition 1 we say that *a condition C holds for large x* if there exists a constant N such that C holds for $x > N$. The analog of Definition 2 is that $\lim_{x \to \infty} f(x) = L$ means the following: If $\varepsilon > 0$, then $|f(x) - L| < \epsilon$ holds for large x. The analog of Theorem 1 is that if two conditions C_1 and C_2 each hold for large x, then $C_1 \wedge C_2$ also holds for large x. Development of the theory parallels Examples 1–12 and is in some respects simpler, because we no longer have to worry about the side condition $x \neq a$. In the present case a corresponds to ∞ and we have $x \neq a$ automatically since x is a real number.

If x is confined to integer values n one generally writes $f(x) = f_n$ and the role of x is now taken by n. Thus, $\lim_{n \to \infty} f_n = L$ means the following: If $\varepsilon > 0$, the inequality $|f_n - L| < \varepsilon$ holds for large n. The fact that n is an integer is understood and need not be emphasized, just as in the former cases it was understood that x is real. The resulting development parallels that for limits as $x \to a$ and yields the theory of limits of sequences.

Uniformity. Although we have not emphasized it by the notation, the condition C in Definition 1 is understood to be a condition depending on x; thus, $C = C(x)$. Furthermore, the statement that C holds for $x \neq a$, $|x - a| < \delta$ means that C holds *for all* x satisfying these conditions. Hence C holds on any smaller set of the form $x \neq a$, $|x - a| < \delta$, and that is why we did not have to insist that δ be small. For example, a function bounded on a set is bounded on any subset, and the inequality $|f(x)| \leqslant M$ expressing boundedness is a condition of form $C(x)$. By contrast, the statement that f is *unbounded* is not of the form $C(x)$, is not preserved by passage to subsets, and would not be a suitable condition C in Definition 1.

Sometimes $C(x)$ involves another parameter t so that $C = C(x, t)$. The parameter t is assumed to range over a specified set E. The statement that $C(x, t)$ holds *for x near a* is defined just as before. Namely, it means that there exists $\delta > 0$ such that $C(x, t)$ holds for $x \neq a$, $|x - a| < \delta$. In general δ depends on t, so that $\delta = \delta(t)$. If δ can be chosen independently of t, for all $t \in E$, it is said that $C(x, t)$ holds *uniformly for x near a*. A similar definition is used for large x. That is, $C(x, t)$ holds *uniformly for* *large* x if $C(x,t)$ holds for $x>N$, where N can be chosen independently of t. When x is restricted to integral values we have $C(x,t)=C(n,t)$ and the condition for uniformity is that $C(n,t)$ holds for $n>N$ where N is independent of t. When Theorem 1 and its analogs are extended to this situation it is advisable to require that the various conditions $C_i(x,t)$ be associated with one and the same set E. Aside from this, there is no significant change.

As an illustration, if $C(x,t)=C(n,t)$ is a condition of the form

$$|f_n(t)-L(t)|<\epsilon, \qquad t\in E,$$

it will be found that the methods used here yield a substantial part of the theory of uniform convergence of sequences. We do not give details, since the details so closely parallel what has already been done.

The domain-dependent definition of limit. There is a certain awkwardness in defining continuity of a function $f(x)$ on a closed interval $a\leqslant x\leqslant b$ because the two-sided limits at a and b need not exist. Thus, we have to require $\lim_{t\to x}f(t)=f(x)$ for $a<x<b$ together with the one-sided conditions $f(a)=f(a+)$, $f(b)=f(b-)$. This problem becomes more acute in more complicated situations, for example, if we want to define continuity of a function $f(x,y)$ in a closed region of the (x,y) plane.

One way of dealing with problems of this kind is to introduce what we shall call the *domain-dependent definition* of limit. In this definition the inequality $|f(x)-L|<\varepsilon$ is required for x satisfying the three conditions $x\neq a$, $|x-a|<\delta$, $x\in D(f)$ where $D(f)$ is the domain of f. For example, if $D(f)$ is the closed interval $a\leqslant x\leqslant b$ then the limits at $x=a$ and $x=b$ are on the same footing as the limits at any other point of this interval and the condition for continuity is $\lim_{t\to x}f(t)=f(x)$, $a\leqslant x\leqslant b$. With the domain-dependent definition, the distinction between one-sided and two-sided limits at the endpoints of the domain disappears.

To study the domain-dependent definition by the methods of this note one can introduce a set F for x in Definition 1 analogous to the set E for t above, and require the side condition $x\in F$. It is said then that "C holds for x in F and near a." When applying this form of Definition 1 to the theory of limits via Definition 2, one takes $F=D(f)$.

If we prefer not to complicate Definition 1 as described above, another method is the following: Define a function f^* by $f^*(x)=f(x)$ when $x\in D(f)$, otherwise $f^*(x)=L$. Then $\lim f(x)=L$ holds with the domain-dependent definition if, and only if, $\lim f^*(x)=L$ holds in the sense of Definition 2. The foregoing results apply to the latter problem and hence to the former.

Further discussion of domain dependence. It is of some interest to contrast the domain-dependent definition with Definition 2, and this is done now. For brevity we introduce the abbreviations

DD = the domain-dependent definition of limit

SD = the simple definition of limit.

More specifically, SD stands for the definition given in Definition 2 or for its one-sided extensions involving $a-$ or $a+$. Use of the word "simple" in this connection suggests an implied judgment about DD vs. SD that I hope to justify in the following discussion.

Let us begin with the question of uniqueness. If we want to falsify the statement "$\lim f(x)=L$" using DD we must pick a suitable $\varepsilon>0$ and then, for each $\delta>0$, we must exhibit an $x \in D(f)$ satisfying $x \neq a$, $|x-a|<\delta$ for which the defining inequality $|f(x)-L|<\varepsilon$ fails. But if no point $x \neq a$ of $D(f)$ is in the interval $|x-a|<\delta$ then we can find no x of this kind, the statement $\lim f(x)=L$ cannot be falsified, and we must admit that L is a value of $\lim f(x)$. This happens whenever some set of the form $|x-a|<\delta$ contains no points of $D(f)$ except possibly a itself; in more technical language, it happens whenever a is not a limit point of $D(f)$. In that case $\lim f(x)$ is not a number but a set; in fact, it is the entire real axis.

Although in principle there is nothing wrong with set-valued expressions, their introduction in the present context can lead to strange results. For instance let $f(x)=0$ for x rational and let $f(x)$ be undefined otherwise. Let $g(x)=0$ for x irrational and let $g(x)$ be undefined otherwise. Then using DD we have $\lim f(x)=\lim g(x)=0$ at every value a, without exception. But the sum function $f+g$ is defined for no x and hence, at every value a, the expression $\lim[f(x)+g(x)]$ is the entire real axis! This is a spectacular failure of the expected theorem regarding limit of a sum.

Since most analysts would find situations such as the above intolerable, it is customary to agree that $\lim f(x)$ is undefined at any point a which is not a limit point of $D(f)$. Such an agreement solves the problem, but it is a complication.

Another complication arises when we try to give a simple statement of the localization theorem, Example 3 above. At first glance one might think that all would be well if we just require $D(f)=D(g)$ except for the point a, which might or might not belong to either set. But this is not a good idea, because $D(f)$ is a *global* concept and the whole purpose of the exercise is to show that $\lim f(x)$ is a *local* concept. Suppose, for example, that f and g are defined by the formulas

$$f(x)=x, \qquad g(x)=x\frac{x-1000}{x-1000}$$

with their natural domains. If we have ascertained that $\lim f(x)=0$ as $x \to 0$ we can't use the alleged localization theorem to conclude anything about $\lim g(x)$ as $x \to 0$ because, no matter whether we include the point 0 or not, the functions f and g do not have the same domain. Since $g(x)=f(x)$ for $|x|<1000$ this example describes a spectacular failure to come to grips with the problem of localization.

To deal effectively with such examples in the vocabulary of this paper, one can introduce the two conditions

$$C_1\colon\ g(x)=f(x) \text{ when } x \text{ is in } D(f)$$

$$C_2\colon\ g(x) \text{ is undefined when } x \text{ is not in } D(f)$$

and declare that "$f(x)=g(x)$ for x near a" means "$C_1 \wedge C_2$ holds for x near a." The latter phrase is defined unambiguously in Definition 1. For those familiar with the concepts of *deleted neighborhood* and *restriction* f_A of a function to a set A, an equivalent formulation is "$f_A=g_A$ in some deleted neighborhood A of a." The above methods lead to an adequate localization theorem for DD, but there is no denying that they involve extra complication.

Similar problems arise when we want to develop theorems about the limit of sums, products, or quotients. One method (which is actually used in textbooks) is to require $D(f)=D(g)$. To see why this is inappropriate let f and g be the functions defined by the formulas

$$f(x)=x, \qquad g(x)=\frac{x}{x-1000}$$

with their natural domains. If we have ascertained that $\lim f(x)=\lim g(x)=0$ as

$x \to 0$, we cannot use the alleged limit-of-sum theorem to deduce anything about the limit of $f(x)+g(x)$, because f and g have different domains. Of course the fact that $g(x)$ is undefined at $x=1000$ has nothing to do with the limit as $x \to 0$, and the correct hypothesis for theorems of this kind is that a is a limit point of $D(f) \cap D(g)$. But that too is a complication.

Let us see next what happens when DD is used to define the derivative, which is, after all, a principal motive for discussing limits in the first place. One of the things we lose is the familiar fact that, if $f(x)$ has a maximum or minimum at a point c where $f'(c)$ exists, then $f'(c)=0$. For example, let $f(x)=x$ for $0 \leqslant x \leqslant 1$, the latter interval being $D(f)$. If DD is used, the defining limit for $f'(x)$ exists for every x on $0 \leqslant x \leqslant 1$, including the end points, and has the value 1. But $x=0$ gives an absolute minimum of $f(x)$ and $x=1$ gives an absolute maximum.

A more striking example is the following. Let $f(x)$ be undefined for x rational and let $f(x)=1$ for x irrational. Then (if DD is used to define the derivative) $f'(x)$ exists at every point in $D(f)$ and in fact $f'(x)=0$, $x \in D(f)$. But if we had said that $f(x)=0$ for rational x, instead of being undefined, the resulting function would be discontinuous at every value in its domain and would not have a derivative anywhere. Yet the only "advantage" the former function has over the latter, as regards smoothness, is that the former is undefined at some points where the latter is defined.

It is perhaps worthwhile to reflect why it is that none of these problems are encountered with SD. The reason is that existence of $\lim f(x)$ with SD automatically implies that a is a limit point of $D(f)$, and it also implies that $f(x)$ is defined for x near a, or $a-$, or $a+$ as the case may be. When two or more functions are involved, Theorem 1 and its analogs ensure that all of them are defined at the relevant values of x and questions about the domain do not arise.

The above discussion lends color to the view that DD should not be used in introductory courses on calculus. If you begin by learning SD you will have no trouble in understanding DD when the need for it arises at a later time. But if you start with DD, chances are that you will never understand either DD or SD.

Answers to questions. If we do not know the uniqueness theorem, we must allow the possibility that $\lim f(x)$ is set-valued and interpret statements accordingly. Aside from this it can be said that the results of Examples 1–8 retain their validity. For a specific illustration let us consider Example 2, the theorem concerning boundedness of $1/f(x)$. In our presumed state of ignorance the theorem would say that if L is a value of $\lim f(x)$ and $L \neq 0$, then $1/f(x)$ is bounded for x near a, and the bound can be any number larger than $1/|L|$. This is still true, and the proof is identical to the proof given above.

The answer to the second question, whether uniqueness can be deduced from localization, is more subtle. At first glance it would seem that one could reason as follows: Suppose $\lim f(x)$ has two values L_1 and L_2. Then we can say $\lim f(x)=L_1$, and with $g(x)=f(x)$ we can also say $\lim g(x)=L_2$. By the localization theorem $\lim f(x)=\lim g(x)$, hence $L_1=L_2$.

Despite its plausibility, this line of thought is incorrect. So long as uniqueness is in doubt the localization theorem gives $\lim f(x)=\lim g(x)$ only in the sense of set equality; if a value L is in one of these sets, it is also in the other. (An examination of the argument will show that this is what is actually proved.) To clinch the matter, suppose we had defined "$\lim f(x)=L$" by use of the inequality $|f(x)-L|<1$ rather than $|f(x)-L|<\varepsilon$. Then localization would still hold in the sense of set equality. But uniqueness fails, as shown by the simplest examples.

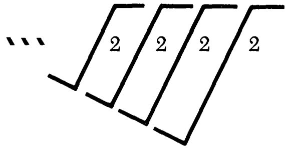
A Transcendental Sequence and a Problem on Commutativity of Exponentiation of Real Numbers

FOUAD NAKHLI
Michigan State University
East Lansing, MI 48824

The real-valued sequence a_n defined by

$$a_1 = m \in \mathbb{R}^+, \qquad a_{n+1} = m^{1/a_n}$$

came to me in an entertaining way. It arose from a doodle I drew. I draw doodles frequently, but for this one I used mathematical symbols. It was the following.



I decided quickly that this is 2 raised to the power which is the reciprocal of the "number" that stands behind the first $\sqrt{\ }$ sign. But this latter "number" is nothing but the whole thing altogether. (Pay attention to the three dots!) So our number (if it were a number at all) should satisfy the equation $x = 2^{1/x}$. I preferred to write it in the more explicit form

$$2^{\left(\frac{1}{2^{(1/\cdots)}}\right)},$$

where the three dots indicate that there is an infinite number of 2's extending to the inside in the same manner. My next step was to explore experimentally the subsequences u_{2n-1} and u_{2n} of the sequence $u_1 = 2$, $u_{n+1} = 2^{1/u_n}$. I found that these converge to the same limit which is the unique root of the equation $x^x = 2$ ($x \approx 1.5596104695$). So it turned out that my "number" was really a number! (I was lucky that I chose 2 for the doodle. Had I chosen a different number, it could have been a different matter, as we shall see later). At this stage, the general sequence a_n was ready for attack. In what follows we study its behavior as $m > 0$ varies and we find its limit when it converges. During the course of this study, a solution is found spontaneously to the following curious problem: Given a positive real number m, how many ordered pairs (x, y) of positive real numbers exist such that

$$x^y = y^x = m?$$

A couple of curious Numbers. Before we begin our study of a_n we state the answers of the previous problem because they are quite scattered in the body of the article (and because they are quite curious also).

1. For $0 < m < \mathrm{e}^{-e^{-1}}$, no pairs (x, y) exist.
2. For $m = e^{-e^{-1}}$, exactly one pair exists. This is $(x, y) = (e^{-1}, e^{-1})$.

3. For $e^{-e^{-1}} < m < 1$, exactly two pairs exist. These are (a, a) and (b, b) where a and b are the two distinct roots of the equation $x^x = m$.

4. For $1 \leqslant m \leqslant e^e$, exactly one pair exists. This is (c, c) where c is the unique root of the equation $x^x = m$.

5. For $m > e^e$, three pairs exactly exist. These are (d, f), (g, g), and (f, d) where g is the unique solution of the equation $x^x = m$, and (d, f) is the unique solution (x, y) of the system $x^y = y^x = m$, $x > y$.

The easy cases. The cases $0 < m < 1$ can be disposed of quite quickly. Here a_n is clearly positive. It is also decreasing. The facts $0 < m < 1$ imply that $1/m > 1$, and these together imply that $m^{1/m} < m$ which is $a_2 < a_1$. Moreover, if $a_{n+1} < a_n$ for some n, it will follow that $m^{1/a_{n+1}} < m^{1/a_n}$ or $a_{n+2} < a_{n+1}$. Hence a_n converges in these cases to a limit $a \geqslant 0$. This limit must satisfy $a^a = m$ except possibly for $a = 0$. But for $0 < m < e^{-e^{-1}}$, the equation $x^x = m$ where $x > 0$ is impossible. See FIGURE 1. Hence $a = 0$ for these values of m. For $e^{-e^{-1}} \leqslant m < 1$, a has three possibilities: zero or u or v where u and v $(u \leqslant v)$ are the two roots of $x^x = m$ (a double root for $m = e^{-e^{-1}}$). But here e^{-1} is a lower bound of a_n. Indeed, for $n = 1$, $a_1 = m \geqslant e^{-e^{-1}} > e^{-1}$, and $a_n \geqslant e^{-1}$ implies

$$a_{n+1} = m^{1/a_n} \geqslant m^e \geqslant \left(e^{-e^{-1}}\right)^e = e^{-1}.$$

So for $e^{-e^{-1}} \leqslant m < 1$, a is v since, unless $u = v$, we have $0 < u < e^{-1}$.

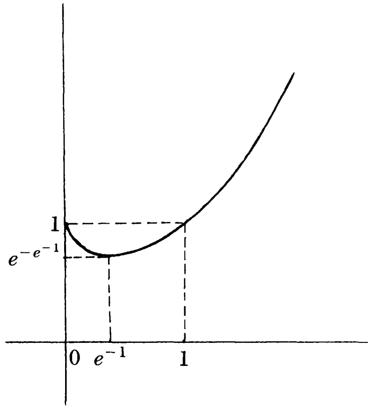


FIGURE 1.
The graph of the function $g(x) = x^x$, $x > 0$.

A second easy accomplishment is the proof of the convergence of the subsequences a_{2n-1} and a_{2n} for all $m > 1$. In fact if we consider here the sequence $b_n = a_{2n-1}$ $(n \geqslant 1)$, we find that it is defined by

$$b_1 = m,$$

$$b_{n+1} = m^{\left(\frac{1}{m^{1/b_n}}\right)}$$

because $b_1 = a_1 = m$ and

$$b_{n+1} = a_{2(n+1)-1} = a_{2n+1} = m^{1/a_{2n}} = m^{\left(\frac{1}{m^{1/a_{2n-1}}}\right)} = m^{\left(\frac{1}{m^{1/b_n}}\right)},$$

and this can be easily proved by induction to be decreasing. Since it is also bounded below by 1, it must be convergent. We can similarly see that for all $m > 1$, a_{2n} is increasing and bounded above by m and hence that it converges.

A nontrivial situation. What we know at this stage is that for $m > 1$, $a_{2n-1} \to c$ and $a_{2n} \to d$, where c and d are positive (in fact > 1). So the single problem that remains is to examine where the equality $c = d$ does hold if at all. Since c and d are positive, they must satisfy the equations $c^d = d^c = m$ by the definition of a_n. However we shall show that these equations are impossible for $1 < m \leqslant e^e$ unless $c = d$. The remaining case $m > e^e$ will be treated last. We will prove that there we have $d < c$.

Define the function $t(x)$ on the domain $(1, \infty) - \{e\}$ by $(\ln t(x))/t(x) = (\ln x)/x$ and $t(x) \neq x$ $\forall x \in (1, \infty) - \{e\}$. See FIGURE 2. We immediately see that $t(x)$ is strictly decreasing and bijective and that its range is equal to its domain. Also t is continuous on both of the intervals $(1, e)$ and (e, ∞), for its derivative

$$t'(x) = \frac{t^2}{x^2}\left(\frac{1 - \ln x}{1 - \ln t}\right)$$

is defined at any point there. Define next the function $m(x)$ on the same domain by $m(x) = x^{t(x)}$. We shall need to know some properties of this function. Here they are.

1. $m(x)$ is continuous on $(1, e)$ and on (e, ∞) since it is the composite function $e^{t(x)\ln x}$ where both e^x and $t(x)\ln x$ are continuous on each of these intervals.

2. As $x \to +\infty$, $m(x) \to +\infty$, since $t(x) \to 1$.

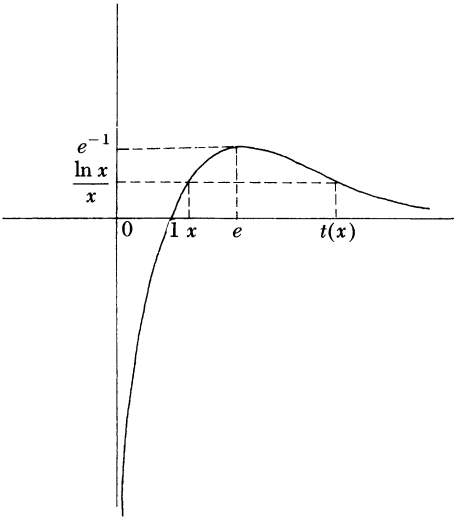


FIGURE 2.
The graph of the function $f(x) = (\ln x)/x$, $x > 0$.

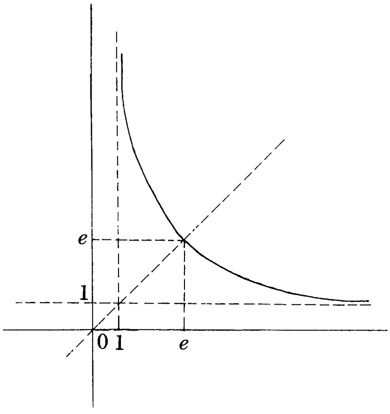


FIGURE 3.
The graph of the function $t(x)$, $x \in (1, \infty) - \{e\}$. The first bisector is an axis of symmetry.

3. As $x \to 1$, $m(x) \to +\infty$. Since $(\ln t)/t = (\ln x)/x$, $m(x) = [t(x)]^x$. Thus as $x \to 1$, $m(x) \to +\infty$ since $t(x) \to +\infty$ then.

4. For $1 < x < e$, $m(x)$ is strictly decreasing, and for $x > e$, it is strictly increasing. The derivative of $m(x)$ is

$$m'(x) = m(x)(t' \ln x + t/x) = \frac{m(x)t}{x}\left(\frac{1 - \ln x \ln t}{1 - \ln t}\right).$$

When $1 < x < e$, t is $> e$. Thus to show $m'(x) < 0$ for these values of x, we have to prove the inequality $\ln x \ln t < 1$. Let $a \in (1, e)$ and $b \in (e, \infty)$ be such that $(\ln b)/b = (\ln a)/a$. We prove that $\ln a \ln b < 1$. Let $z = \ln a$, thus $z \in (0,1)$. We have to show that $b < e^{1/z}$. Since $e^{1/z} > e$, it is sufficient to show that

$$\frac{\ln e^{1/z}}{e^{1/z}} < \frac{\ln b}{b}.$$

This is so because the function $f(x) = \ln x/x$ is strictly decreasing on (e, ∞). But $(\ln b)/b = (\ln a)/a = z/e^z$. Therefore, we have to prove that

$$\frac{1/z}{e^{1/z}} < \frac{z}{e^z}$$

for all z in $(0,1)$ or, equivalently, that

$$\frac{w}{e^w} < \frac{1/w}{e^{1/w}}$$

for all $w > 1$. The function

$$r(w) = \frac{e^{w-(1/w)}}{w^2}$$

is such that $r(1) = 1$ and its derivative

$$r'(w) = \frac{(w-1)^2}{w^4} e^{w-(1/w)} \geqslant 0$$

for all $w \geqslant 1$. So $r(w) > 1$ for all $w > 1$, that is,

$$\frac{w}{e^w} < \frac{1/w}{e^{(1/w)}},$$

for all $w > 1$ as desired. Now for $x > e$ the same inequality $\ln x \ln t < 1$ gives $m'(x) > 0$ since here $1 < t < e$.

5. $\lim_{x \to e} m(x) = e^e$. As $x \to e$, $t(x) \to e$ (FIGURE 2 again). Thus $m(x) \to e^e$.

From these properties we conclude that $m(x) > e^e$ for all x in $(1, \infty) - \{e\}$. See FIGURE 4. Now suppose that $c \neq d$. Then since $(\ln c)/c = (\ln d)/d$, we will have that $c = t(d)$ and $d^{t(d)} = d^c$ is in the range of $m(x)$. So we must have $d^c = m > e^e$. Thus for $1 < m \leqslant e^e$, $c = d$ follows as we asserted and hence here $a_n \to c$ where c is the unique root of the equation $x^x = m$. For the sake of making the answers to the problem in the introduction complete, we note that if c and d are such that $c^d = d^c = m$, then the result $c = d$ is true not only for $1 < m \leqslant e^e$ and c, $d > 1$, but also for $0 < m \leqslant e^e$ and c, $d > 0$.

Now we are ready for the case $m > e^e$. To start with we prove that e is an upper bound of a_{2n}. The function $k(x) = x^{1/x}$ $(x > 0)$ whose graph is shown in FIGURE 5 is

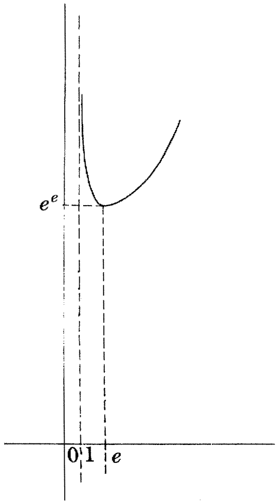


FIGURE 4.
The graph of the function $m(x)$, $x \in (1, \infty) - \{e\}$.

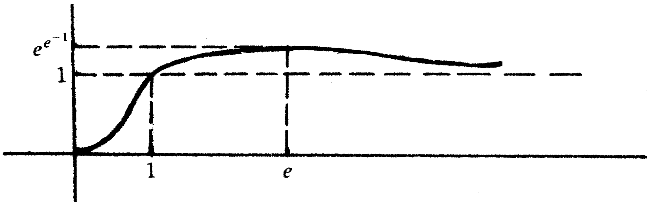


FIGURE 5.
The graph of the function $k(x) = x^{1/x}$, $x > 0$.

$< e$ for all $x > e^e$. Hence for $n = 1$, $a_2 = m^{1/m}$ is $< e$. Assume that for some $n \geqslant 1$ we have $a_{2n} < e$. Since

$$a_{2(n+1)} = m^{\left(\frac{1}{m^{1/a_{2n}}}\right)} \quad \text{and} \quad \ln a_{2(n+1)} = \frac{\ln m}{m^{1/a_{2n}}} > 0,$$

we find that

$$\ln\ln a_{2(n+1)} = \ln\ln m - \frac{\ln m}{a_{2n}} < \ln\ln m - \frac{\ln m}{e}$$

because we also have $a_{2n} < e$. Consider, thus, the function

$$q(x) = \ln\ln x - \frac{\ln x}{e}(x \geqslant e^e).$$

For $x > e^e$, the derivative $q'(x) = 1/x(1/\ln x - 1/e)$ is negative. So $q(x) < q(e^e) = 0$ and $\ln\ln m - (\ln m)/e < 0$. Consequently, $\ln\ln a_{2(n+1)} < 0$ or $a_{2(n+1)} < e$, and the proof is complete. A corollary to what preceded is that e is a lower bound of the subsequence a_{2n-1}. Since $a_{2n} < e$ for all $n \geqslant 1$, $1/a_{2n} > 1/e$. Thus

$$m^{1/a_{2n}} > m^{1/e}$$

or

$$a_{2n+1} > (e^e)^{1/e} = e$$

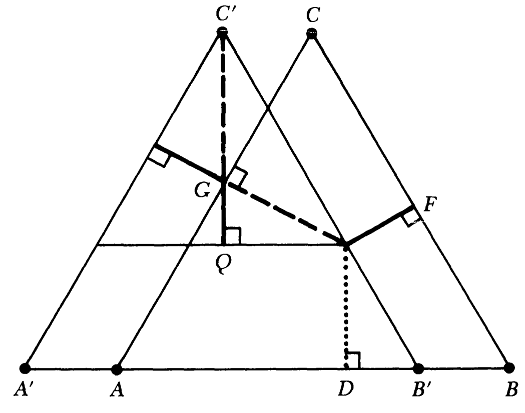
since $m > e^e$. Also $a_1 = m > e$ for $m > e^e$.

At last we prove that for $m > e^e$, we have $c > d$, so that a_n diverges in these cases. Since e is a lower bound of a_{2n-1} and an upper bound of a_{2n}, $c = d$ cannot hold unless $c = d = e$. But then the conditions $c^d = d^c = m$ become $m = c^c = e^e$. So for $m > e^e$, $c \neq d$, and from $d \leqslant e$ and $c \geqslant e$, $c > d$ follows. So c and d satisfy $c^d = d^c = m$ and $c > d$. But are they determined uniquely from these relations? The answer is yes, and this can be easily seen as follows. Suppose that the system $x^t = t^x = m$, $x > t$ has two solutions (x_1, t_1) and (x_2, t_2). Then $m(x_1) = m(x_2) = m(t_1) = m(t_2) = m$. But an equation of the form $m(x) = K > e^e$ has exactly 2 roots (FIGURE 5). Thus at most two of the four numbers x_1, t_1, x_2, and t_2 are different. But we have $1 < t_1$, $t_2 < e$ and $e < x_1$, x_2. Therefore, we must have $x_1 = x_2$ and $t_1 = t_2$.

A wide wild world. Finally we propose the following natural generalization of a_n. This is the sequence $a_1 = k$, $a_{n+1} = m^{[(a_n)^c]}$ where k, m, and c are complex numbers chosen such that all the terms of a_n make sense.

Proof without Words:

The perpendiculars to the sides from a point on the boundary or within an equilateral triangle add up to the height of the triangle.



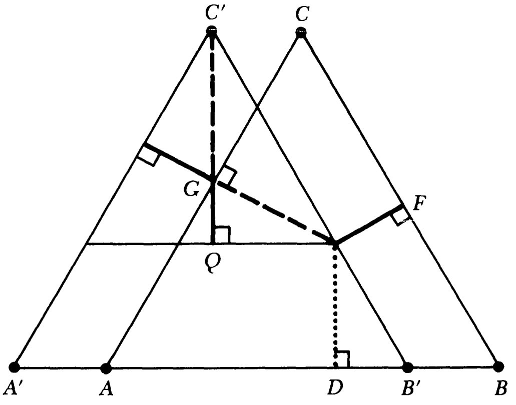
—SAMUEL WOLF
120 NORTH LONGCROSS ROAD
LINTHICUM HEIGHTS, MD 21090

At last we prove that for $m > e^e$, we have $c > d$, so that a_n diverges in these cases. Since e is a lower bound of a_{2n-1} and an upper bound of a_{2n}, $c = d$ cannot hold unless $c = d = e$. But then the conditions $c^d = d^c = m$ become $m = c^c = e^e$. So for $m > e^e$, $c \neq d$, and from $d \leqslant e$ and $c \geqslant e$, $c > d$ follows. So c and d satisfy $c^d = d^c = m$ and $c > d$. But are they determined uniquely from these relations? The answer is yes, and this can be easily seen as follows. Suppose that the system $x^t = t^x = m$, $x > t$ has two solutions (x_1, t_1) and (x_2, t_2). Then $m(x_1) = m(x_2) = m(t_1) = m(t_2) = m$. But an equation of the form $m(x) = K > e^e$ has exactly 2 roots (FIGURE 5). Thus at most two of the four numbers x_1, t_1, x_2, and t_2 are different. But we have $1 < t_1$, $t_2 < e$ and $e < x_1$, x_2. Therefore, we must have $x_1 = x_2$ and $t_1 = t_2$.

A wide wild world. Finally we propose the following natural generalization of a_n. This is the sequence $a_1 = k$, $a_{n+1} = m^{[(a_n)^c]}$ where k, m, and c are complex numbers chosen such that all the terms of a_n make sense.

Proof without Words:

The perpendiculars to the sides from a point on the boundary or within an equilateral triangle add up to the height of the triangle.



—SAMUEL WOLF
120 NORTH LONGCROSS ROAD
LINTHICUM HEIGHTS, MD 21090

Summations Involving Computer-Related Functions

LARRY HOEHN
JIM RIDENHOUR
Austin Peay State University
Clarksville, TN 37044

There are many interesting ways in which the disciplines of computer science and mathematics are interrelated. Often mathematical problems are solved only after insight is gained through the application of computers and the tremendous computational power they provide. On the other hand, many computer problems can be greatly simplified by the application of mathematics.

Our purpose is to present here a mathematical problem which arises in connection with some elementary functions encountered in computer science. We then relate the problem to traditional topics in mathematics and present mathematical formulas which greatly simplify the computations. Further mathematical analysis produces asymptotic formulas which give good approximate solutions to the problem at hand.

In most computer languages there are functions analogous to the DIV function found in PASCAL. This function is implemented for integers in a manner that is equivalent to division with truncation of the remainder. For example, 11 DIV 5 returns the value 2. There are a great many interesting results concerning the sum of the divisors of a positive integer n. These involve the sigma function which is defined as

$$\sigma(n) = \sum_{d|n} \left(\frac{n}{d}\right).$$

We wish to consider the analogous function $S(n)$ and a generalization $T(n)$, defined respectively by

$$S(n) = \sum_{d=1}^{n} (n \,\mathrm{DIV}\, d) = \sum_{d=1}^{n} \left[\frac{n}{d}\right],$$

and

$$T(n) = \sum_{d=1}^{n} \left[\sqrt[r]{\frac{n}{d}}\right],$$

where $[x]$ denotes the greatest integer function and r is a positive integer. First we consider some properties of $S(n)$.

The sum for $S(n)$ differs from $\sigma(n)$ in that if d is not a divisor of n, the term $[n/d]$ still makes a contribution to the sum since only the part truncated to the right of the decimal point is lost. TABLE 1 gives $S(n)$ and $\sigma(n)$ for $n \leqslant 16$.

TABLE 1

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\sigma(n)$	1	3	4	7	6	12	8	15	13	18	12	28	14	24	24	31
$S(n)$	1	3	5	8	10	14	16	20	23	27	29	35	37	41	45	50

While the sequence $\sigma(n)$ is not monotone, it appears from TABLE 1 that $S(n)$ increases monotonically. This is easily seen to be true since

$$S(n+1)=\sum_{d=1}^{n+1}\left[\frac{n+1}{d}\right]>\sum_{d=1}^{n}\left[\frac{n}{d}\right]=S(n).$$

TABLE 2 illustrates the nature of the individual terms of $S(n)$ when $n=24$.

TABLE 2

d	1	2	3	4	5	6	7	8	9	10	11	12	13	14	$\cdots$	24
$[n/d]$	24	12	8	6	4	4	3	3	2	2	2	2	1	1	$\cdots$	1

From TABLE 2 we note that the pattern for $[n/d]$ appears to change abruptly when d is near $\sqrt{n}$. In particular, we note that the numbers $[n/d]$ are distinct for $d=1,2,\ldots,[\sqrt{24}]$, but are not distinct for $d>[\sqrt{24}]$. We shall see in the lemma below that this is true in general.

For the remainder of the paper we let $s=[\sqrt{n}]$ and use the notation

$$q_1=\left[\frac{n}{1}\right],\quad q_2=\left[\frac{n}{2}\right],\ldots,q_s=\left[\frac{n}{s}\right].$$

We will also make frequent use of elementary properties of the greatest integer function; these properties can be found in many number theory texts, for example, see [**6**, pg. 78–79].

We begin with the following lemma:

LEMMA. *With notation defined above,*

a) $$q_1>q_2>\cdots>q_s,$$

b) $$q_{s+1}=\left[\frac{n}{s+1}\right]=\begin{cases}s-1 & \textit{if } s^2\leqslant n<s^2+s\\ s & \textit{if } s^2+s\leqslant n\leqslant s^2+2s,\end{cases}$$

c) $$q_s=\left[\frac{n}{s}\right]=\begin{cases}s & \textit{if } s^2\leqslant n<s^2+s\\ s+1 & \textit{if } s^2+s\leqslant n<s^2+2s\\ s+2 & \textit{if } n=s^2+2s,\end{cases}$$

d) $$q_d=\left[\frac{n}{d}\right]=k\quad \textit{if}\quad q_{k+1}<d\leqslant q_k,$$

e) $$\left[\sqrt[r]{\frac{n}{d}}\right]=k\quad \textit{if}\quad \left[\frac{n}{(k+1)^r}\right]<d\leqslant\left[\frac{n}{k^r}\right],$$

f) $$\sum_{d=q_s}^{n}\left[\frac{n}{d}\right]=(q_1+q_2+\cdots+q_s)+\left[\frac{n}{q_s}\right]-sq_s.$$

Proof. a) If $1<d<s$ then $d\leqslant s-1\leqslant\sqrt{n}-1$ which implies $d^2\leqslant n-2\sqrt{n}+1$ and hence $nd+d^2+d<n+nd$. Dividing by $d(d+1)$ we obtain $n/(d+1)+1<n/d$. Thus $[n/d]>[n/(d+1)]$ for $d=1,2,\ldots,s-1$.

Parts b) and c) follow in a straightforward manner so we will verify only the first assertion in part b). Assuming that $s^2\leqslant n<s^2+s$, we have $s^2/(s+1)\leqslant n/(s+1)<s$

which implies that

$$\left[s-1+\frac{1}{s+1}\right] \leqslant \left[\frac{n}{s+1}\right] < s,$$

or $s-1 \leqslant [n/(s+1)] < s$. Hence $[n/(s+1)] = s-1$.

Part d) is a special case of e) so we prove only e). If $d \leqslant [n/k^r]$, then $d \leqslant n/k^r$ and hence $k \leqslant \sqrt[r]{n/d}$. Thus $k \leqslant [\sqrt[r]{n/d}]$. On the other hand, if $[n/(k+1)^r] < d$, then $n/(k+1)^r - 1 < [n/(k+1)^r] \leqslant d-1$. Hence $n/(k+1)^r < d$ so that $\sqrt[r]{n/d} < k+1$ and $[\sqrt[r]{n/d} \leqslant k$. Therefore $k \leqslant [\sqrt[r]{n/d}] \leqslant k$ gives the desired result.

f) By part a) $q_s < q_{s-1} < \cdots < q_2 < q_1 = n$ so we can write

$$\begin{aligned}\sum_{d=q_s}^{n}\left[\frac{n}{d}\right] &= \left[\frac{n}{q_s}\right] + \left[\frac{n}{q_s+1}\right] + \cdots + \left[\frac{n}{n}\right]\\ &= \underbrace{\left[\frac{n}{q_s}\right] + \cdots + \left[\frac{n}{q_{s-1}}\right]} \underbrace{+ \cdots + \left[\frac{n}{q_{s-2}}\right]} + \cdots + \left[\frac{n}{q_2}\right] \underbrace{+ \cdots + \left[\frac{n}{q_1}\right]},\end{aligned}$$

where we have divided the terms into groups depending upon where the denominator d falls in relation to the numbers $q_s, q_{s-1}, \ldots, q_1$. Since the denominators are consecutive integers, the number of terms in the group where $q_{k+1} < d \leqslant q_k$ is $q_k - q_{k+1}$. Moreover, by part d) $[n/d] = k$ for each d in such a group. Hence

$$\begin{aligned}\sum_{d=q_s}^{n}\left[\frac{n}{d}\right] &= \left[\frac{n}{q_s}\right] + (q_{s-1}-q_s)(s-1) + (q_{s-2}-q_{s-1})(s-2)\\ &\quad + \cdots + (q_2-q_3)(2) + (q_1-q_2)(1)\\ &= \left[\frac{n}{q_s}\right] - q_s(s-1) + (q_{s-1}+q_{s-2}+\cdots+q_2) + q_1\\ &= \left[\frac{n}{q_s}\right] - sq_s + (q_1+q_2+\cdots+q_s).\end{aligned}$$

The following theorem gives one of our main results. The first part relates $S(n)$ to the sum of just the first s terms and the second part gives a result concerning the nature of $S(n)$ as $n \to \infty$. We note that Theorem 1 appears in some number theory texts (e.g. [1, p. 207–210] and [5, p. 168–169]). The proof given in these texts, however, is quite different from the one proposed here. Basically, these proofs involve counting the number of lattice points bounded by the x-axis, the y-axis, the line $x = n$, the line $y = n$, and the hyperbola $xy = n$. Our proof, on the other hand, depends on using various properties of the greatest integer function.

THEOREM 1. *With notation defined above,*

a) $$S(n) = 2\sum_{d=1}^{s}\left[\frac{n}{d}\right] - s^2$$

b) $$\lim_{n\to\infty}\left(\frac{S(n)}{n} - \ln n\right) = 2\gamma - 1 \textit{ where } \gamma \textit{ is Euler's constant.}$$

Proof. a) Since $s = [\sqrt{n}]$ it is easy to see that $s^2 \leqslant n \leqslant s^2 + 2s$. Because our proof requires that we be able to specify the values for q_s and q_{s+1}, we need to consider three cases and apply parts b) and c) of the Lemma. These cases are

$$\begin{aligned}&\text{Case 1:} && s^2 \leqslant n < s^2 + s,\\ &\text{Case 2:} && s^2 + s \leqslant n < s^2 + 2s,\\ &\text{Case 3:} && n = s^2 + 2s\end{aligned}$$

We first consider Case 1. Then $q_s = s$ so

$$\begin{aligned}S(n) &= \left[\frac{n}{1}\right] + \cdots + \left[\frac{n}{s}\right] + \left[\frac{n}{s+1}\right] + \cdots + \left[\frac{n}{n}\right]\\ &= q_1 + q_2 + \cdots + q_s - \left[\frac{n}{s}\right] + \left[\frac{n}{s}\right] + \left[\frac{n}{s+1}\right] + \cdots + \left[\frac{n}{n}\right].\end{aligned}$$

Using part f) of the Lemma, we get

$$\begin{aligned}S(n) &= (q_1 + q_2 + \cdots + q_s) - \left[\frac{n}{s}\right] + (q_1 + q_2 + \cdots + q_s) + \left[\frac{n}{q_s}\right] - sq_s\\ &= 2\sum_{d=1}^{s}\left[\frac{n}{d}\right] - \left[\frac{n}{s}\right] + \left[\frac{n}{q_s}\right] - s \cdot s\end{aligned}$$

from which the result follows. Case 2 is similar except that $q_s = s + 1$ and the last part of the sum for $S(n)$ is

$$\left[\frac{n}{s+1}\right] + \cdots + \left[\frac{n}{n}\right] = \left[\frac{n}{q_s}\right] + \cdots + \left[\frac{n}{n}\right],$$

and we can again apply part f) and simplify. In Case 3, $q_s = s + 2$ so

$$\left[\frac{n}{s+1}\right] + \left[\frac{n}{s+2}\right] + \cdots + \left[\frac{n}{n}\right] = s + \left[\frac{n}{q_s}\right] + \cdots + \left[\frac{n}{n}\right].$$

Again the result follows by applying part f) of the Lemma and simplifying.

b) It is well known that $\lim_{n\to\infty}(H_n - \ln n) = \gamma$ where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ is the nth partial sum of the harmonic series and $\gamma = .57721\ 56649\ldots$ is Euler's constant. Hence for any $\varepsilon > 0$, we have for large n that

$$\ln n + \gamma - \varepsilon < H_n < \ln n + \gamma + \varepsilon. \tag{1}$$

Since $s = [\sqrt{n}]$ we have $\sqrt{n} - 1 < s \leqslant \sqrt{n}$. Using this and part a) of Theorem 1, we see that

$$\begin{aligned}S(n) &= 2\sum_{d=1}^{s}\left[\frac{n}{d}\right] - s^2\\ &< 2\sum_{d=1}^{s}\frac{n}{d} - (\sqrt{n} - 1)^2\\ &= 2nH_s - n + 2\sqrt{n} - 1.\end{aligned}$$

Therefore, using (1), we have $S(n) < 2n(\ln s + \gamma + \varepsilon) - n + 2\sqrt{n} - 1$. But $2\ln s \leqslant 2\ln \sqrt{n} = \ln n$ so

$$\frac{S(n)}{n} - \ln n < 2\gamma + 2\varepsilon - 1 + \frac{2\sqrt{n} - 1}{n} < 2\gamma - 1 + 3\varepsilon \tag{2}$$

for large n. Similarly,

$$\begin{aligned} S(n) &= 2\sum_{d=1}^{s}\left[\frac{n}{d}\right] - s^2 \\ &> 2\sum_{d=1}^{s}\left(\frac{n}{d} - 1\right) - n \\ &= 2nH_s - 2s - n. \end{aligned}$$

Using (1), we have for large n, $S(n) > 2n(\ln s + \gamma - \varepsilon) - 2s - n$. But for large n,

$$\ln s = \ln\left(\frac{[\sqrt{n}\,]}{\sqrt{n}}\right) + \ln\sqrt{n} > \ln\left(\frac{\sqrt{n} - 1}{\sqrt{n}}\right) + \ln\sqrt{n} > \ln\sqrt{n} - \varepsilon.$$

Hence, for large n, $S(n) > 2n(\ln\sqrt{n} + \gamma - 2\varepsilon) - 2s - n$ and

$$\frac{S(n)}{n} - \ln n > 2\gamma - 1 - \frac{2s}{n} - 4\varepsilon > 2\gamma - 1 - 5\varepsilon.$$

The proof of part b) of Theorem 1 follows from (2) and (3).

In addition to whatever aesthetic value part a) of Theorem 1 may possess, it greatly simplifies the calculation of $S(n)$, and subsequently H_n. For example, in using the definition to calculate $S(n)$, when $n = 10{,}000{,}000$, it is necessary to calculate a sum involving 10,000,000 terms. However, when part a) of the theorem is used we need only find a sum with $s = 3162$ terms. This reduces the calculation time in the required loop by a factor of approximately 3000.

From part b) of Theorem 1, we have $S(n) \doteq n(\ln n + 2\gamma - 1)$. This provides a very good approximation for $S(n)$ as evidenced by TABLE 3.

TABLE 3

n	$S(n)$	$n(\ln n + 2\gamma - 1)$
10	27	25
100	482	476
1000	7,069	7,062
10,000	93,668	93,648
100,000	1,166,750	1,166,736
1,000,000	13,970,034	13,969,942
10,000,000	162,725,364	162,725,264
100,000,000	1,857,511,568	1,857,511,168

Similarly, from $S(n)/n - \ln n \doteq 2\gamma - 1$ and $H_n - \ln n \doteq \gamma$, we obtain $H_n \doteq S(n)/n + 1 - \gamma$. By using part a) of Theorem 1 to obtain $S(n)$, we can obtain very accurate approximations of H_n by summing only $[\sqrt{n}\,]$ terms as opposed to summing n terms. Furthermore, from TABLE 4 below we see that the accuracy of our approximation to H_n increases remarkably as n increases. This method of approximating partial sums of H_n provides a nice contrast to the Euler-Maclaurin asymptotic formula given in [2]. From [3, p. 237] we know that the summation of 10^{12} terms of a series is beyond the range of current day (i.e., 1977) computers. This formula, however, would make it possible (essentially) to sum 10^{12} terms of the harmonic series with only 10^6 summations.

Next we turn to a theorem similar to Theorem 1.

TABLE 4

n	$\frac{S(n)}{n}+1-\gamma$	H_n
10	3.12278	2.92897
100	5.24278	5.18738
1000	7.49178	7.48547
10,000	9.78959	9.78761
100,000	12.09028	12.09015
1,000,000	14.39282	14.39273
10,000,000	16.69532	16.69531

THEOREM 2. *With $T(n)$ defined above*

$$T(n)=\sum_{d=1}^{t}\left[\frac{n}{d^r}\right],\ \textit{where}\ t=\left[\sqrt[r]{n}\right].$$

Proof. By part e) of the Lemma and the technique used in the proof of part f) of the Lemma,

$$\begin{aligned}T(n)&=\left[\sqrt[r]{\frac{n}{1}}\right]+\left[\sqrt[r]{\frac{n}{2}}\right]+\cdots+\left[\sqrt[r]{\frac{n}{n}}\right]\\
&=t\left(\left[\frac{n}{t^r}\right]-\left[\frac{n}{(t+1)^r}\right]\right)+(t-1)\left(\left[\frac{n}{(t-1)^r}\right]-\left[\frac{n}{t^n}\right]\right)\\
&\quad+\cdots+2\left(\left[\frac{n}{2^r}\right]-\left[\frac{n}{3^r}\right]\right)+1\left(\left[\frac{n}{1^r}\right]-\left[\frac{n}{2^r}\right]\right)\\
&=(t-t+1)\left[\frac{n}{t^r}\right]+(t-1-t+2)\left[\frac{n}{(t-1)^r}\right]+\cdots+(2-1)\left[\frac{n}{2^r}\right]+(1)\left[\frac{n}{1^r}\right]\\
&=\sum_{d=1}^{t}\left[\frac{n}{d^r}\right].\end{aligned}$$

The formula in Theorem 2 dramatically reduces the number of terms needed to calculate $T(n)$, especially if r is large. For example, if $n=10{,}000{,}000$, calculating $T(n)$ from the definition requires 10,000,000 terms whereas the sum given by Theorem 2 involves only 10 terms.

REFERENCES

1. George E. Andrews, *Number Theory*, W. B. Saunders Company, Philadelphia, 1971.
2. R. P. Boas, Jr., and J. W. Wrench, Jr., Partial sums of the harmonic series, *Amer. Math. Monthly* 78 (1971), 864–870.
3. R. P. Boas, Jr., Partial sums of infinite series, and how they grow, *Amer. Math Monthly* 84 (1977), 237–258.
4. ______, Convergence, Divergence, and the Computer, *Mathematical Plums*, Ross Honsberger (Editor), MAA, 1979.
5. William J. LeVeque, *Fundamentals of Number Theory*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.
6. Ivan Niven and Herbert S. Zuckerman, *An Introduction to the Theory of Numbers*, John Wiley & Sons, Inc., New York, 1960.

The Riemann Conjecture

Mein lieber Herr Riemann,
All night I will dream on,
'Bout how you deserve a lecture.
But of course I allude
To your famous and shrewd
Outstanding and unsolved conjecture.

Oh, I owe you my life,
My 3 kids and my wife,
For proof of the Prime Number Theorem.
Your zeta function trick
Made the proof really slick,
And those primes—no more do I fear 'em.

But I just stop to think,
How I've taken (hic) to drink,
And evolved this hysterical laugh—
Because still I don't know
If ζ's roots all go
On the line Re $z = 1/2$!

So I don't sleep at night,
And I'm losing my sight
In search of this darn thing's solution.
As my mind starts to go
My calculations grow
In a flood of "complex" confusion.

I bought a computer;
Not any astuter,
It ran for nearly 10 years—no jive!
But *still* it doesn't know
If zeta's roots *all* go
On that line Re $z = .5$

Now I sit in my room—
I feel doomed in the gloom—
And entombed by mountains of paper.
Still, I pray that some night
My "ol' lightbulb" will light
With the clue that could wrap-up this caper!

—Jonathan P. Dowling
University of Colorado
Boulder, CO 80309

PROBLEMS

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

BRUCE HANSON, *associate editor*
St. Olaf College

Proposals

To be considered for publication, solutions should be received by November 1, 1989.

Murray Klamkin was awarded the MAA's Award for Distinguished Service to Mathematics in 1988, largely in recognition of his many contributions to problem solving. In this issue we further recognize this work: every problem posed in this issue is a proposal of Murray Klamkin's.

1322. An n-gon of consecutive sides $a_1, a_2, \ldots, a_n$ is circumscribed about a circle of unit radius. Determine the minimum value of the product of all its sides.

1323. A parallelepiped has the property that all cross sections which are parallel to any fixed face F have the same area as F. Are there any other polyhedra with this property?

1324. Determine the maximum value of

$$x_1 x_2 \cdots x_n\left(x_1^2 + x_2^2 + \cdots + x_n^2\right),$$

where $x_1 + x_2 + \cdots + x_n = 1$ and $x_1, x_2, \ldots, x_n \geqslant 0$.

1325. a. Determine the *minimum* value of

$$\max_{0 \leqslant x_i \leqslant 1} \left|F_1(x_1) + F_2(x_2) + \cdots + F_n(x_n) - x_1 x_2 \cdots x_n\right|$$

over all possible real-valued functions $F_i(t)$, $0 \leqslant t \leqslant 1$, $1 \leqslant i \leqslant n$.

b. Determine the *minimum* value of

$$\max_{0 \leqslant x_i \leqslant 1} \left|F_1(x_1) F_2(x_2) \cdots F_n(x_n) - (x_1 + x_2 + \cdots + x_n)\right|$$

over all possible real-valued functions $F_i(t)$, $0 \leqslant t \leqslant 1$, $1 \leqslant i \leqslant n$.

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

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

Solutions and new proposals should be mailed in duplicate to Loren C. Larson, Department of Mathematics, St. Olaf College, Northfield, MN 55057.

1326. A particle is projected vertically upwards in a uniform gravitational field and subject to a drag force mv^2/c. The particle in its ascent and descent has equal speeds at two points whose respective heights above the point of projection are x and y. It had been shown by Newton that if a denotes the maximum height of the particle, then x and y are related by

$$e^{2(a-x)/c}+e^{-2(a-y)/c}=2. \tag{1}$$

Consider the same problem except that the drag force is now $F(mv^2/c)$ where F is a smooth function. Show that if (1) still holds for all possible y values, then $F(u)=u$.

Quickies

Answers to the Quickies are on pages 204–205.

Q748. Determine the maximum value of

$$F=\frac{(x^{2n}-a^{2n})(b^{2n}-x^{2n})}{(x^{2n}+a^{2n})(x^{2n}+b^{2n})}$$

over all real x.

Q749. Prove that

$$\frac{x^{\lambda+1}}{y^{\lambda}}+\frac{y^{\lambda+1}}{z^{\lambda}}+\frac{z^{\lambda+1}}{x^{\lambda}}\geqslant x+y+z$$

where $x, y, z, \lambda>0$.

Q750. Ptolemy's inequality states that $ac+bd\geqslant ef$, where a, b, c, d are consecutive sides of a quadrilateral (it need not be planar), and e, f are its diagonals. There is equality if and only if the quadrilateral is cyclic (has a circumcircle). Determine a corresponding inequality for a spherical quadrilateral.

Q751. If $z_1, z_2, \dots, z_5$ are complex numbers such that

$$|z_{i+1}+z_{i+2}|=|z_{i+3}+z_{i+4}+z_{i+5}|$$

for $i=1,2,\dots,5$, and $z_{i+5}=z_i$, prove that $z_1+z_2+\cdots+z_5=0$.

Solutions

An odd recurrence **June 1988**

1297. *Proposed by Irl C. Bivens and Benjamin G. Klein, Davidson College, North Carolina.*

For k a positive integer, define A_n for $n=1,2\dots$ by

$$A_{n+1}=\frac{nA_n+2(n+1)^{2k}}{n+2}, \qquad A_1=1.$$

Prove that A_n is an integer for all $n\geqslant 1$, and A_n is odd if and only if n is congruent to 1 or 2 modulo 4.

Solution by Fred Dodd and Leon Mattics, University of South Alabama, Mobile.

Set $t = 2k+1$ and $S(n) = 1^t + 2^t + \cdots + n^t$. (Note that $t > 2$ is odd.) A simple induction argument shows that

$$A_n = \frac{2S(n)}{n(n+1)} \tag{1}$$

for all n. Since

$$2S(n) = \sum_{i=0}^{n} \left((n-i)^t + i^t\right) = \sum_{i=1}^{n} \left((n+1-i)^t + i^t\right)$$

then $2S(n)$ is divisible by both n and $n+1$. Consequently, as n and $n+1$ are relatively prime, it follows from (1) that A_n is an integer for all n.

Suppose that n is congruent to 1 or 2 modulo 4. Then $S(n)$ is odd since it has an odd number of odd terms. Also, $n(n+1) \equiv 2 \pmod 4$. Thus, by (1), A_n is odd.

Next suppose that $n \equiv 0 \pmod 4$. Then $(n/2)^t \equiv 0 \pmod n$ and so

$$S(n) = \left(\sum_{i=0}^{n/2} (n-i)^t + i^t\right) - (n/2)^t \equiv 0 \pmod n.$$

Thus, by (1), A_n is even.

Finally suppose that $n \equiv 3 \pmod 4$. Then $\left(\frac{n+1}{2}\right)^t \equiv 0 \pmod{n+1}$ and so

$$S(n) = \left(\sum_{i=1}^{(n+1)/2} (n+1-i)^t + i^t\right) - \left(\frac{n+1}{2}\right)^t \equiv 0 \pmod{n+1}.$$

Thus, by (1), A_n is even.

Also solved by J. M. Becker (France), Bilkent University Problem Solving Group (Turkey), Joshua Brandon (student), Duane M. Broline, Brown University Fly-Fishing Club, David Callan, David Doster, Lorraine L. Foster and Abe Achkinazi (student), Jayanthi Ganapathy, Francis M. Henderson, Farhood Pouryoussefi Kermany (student, Iran), Václav Konečný, Helen M. Marston, M. Riazi-Kermani, Adam Riese, John P. Robertson, Harry D. Ruderman, Harvey Schmidt, Jr., J. M. Stark, Western Maryland College Problems Group, A. Zulauf (New Zealand), and the proposers. There was one incomplete solution.

Circumscribable quadrangle June 1988

1298. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

A quadrilateral $ABCD$ is circumscribed about a circle, and P, Q, R, S are the points of tangency of sides AB, BC, CD, DA respectively. Let $a = |AB|$, $b = |BC|$, $c = |CD|$, $d = |DA|$, and $p = |QS|$, $q = |PR|$. Show that

$$\frac{ac}{p^2} = \frac{bd}{q^2}.$$

I. *Solution by J. M. Stark, Lamar University, Texas.*

Denote by r the radius of the circle tangent to the sides of $ABCD$, and let $\alpha, \beta, \gamma, \delta$ be the angles subtended at the center of the circle by the chords SP, PQ, QR, RS respectively.

We have $a = |AP| + |PB|$, $b = |BQ| + |QC|$, $c = |CR| + |RD|$, $d = |DS| + |SA|$ and right triangle geometry gives $|AP| = |SA| = r\tan(\alpha/2)$, $|BQ| = |PB| = r\tan(\beta/2)$, $|CR| = |QC| = r\tan(\gamma/2)$, $|RD| = |DS| = r\tan(\delta/2)$. It follows that

$$ac = r^2(\tan(\alpha/2) + \tan(\beta/2))(\tan(\gamma/2) + \tan(\delta/2)),$$

and

$$bd = r^2(\tan(\beta/2) + \tan(\gamma/2))(\tan(\delta/2) + \tan(\alpha/2)). \tag{1}$$

Application of the identity $\tan(x) + \tan(y) = (\sin(x + y))/(\cos(x)\cos(y))$ to (1) gives

$$\frac{ac}{bd} = \frac{\sin((\alpha + \beta)/2)\sin((\gamma + \delta)/2)}{\sin((\beta + \gamma)/2)\sin((\alpha + \delta)/2)}. \tag{2}$$

From $\alpha + \beta + \gamma + \delta = 2\pi$ we obtain $\sin((\gamma + \delta)/2) = \sin((\alpha + \beta)/2)$ and $\sin((\alpha + \delta)/2) = \sin((\beta + \gamma)/2)$, which, when combined with (2) yields

$$\frac{ac}{bd} = \frac{\sin^2((\alpha + \beta)/2)}{\sin^2((\beta + \gamma)/2)}. \tag{3}$$

Since $p^2 = (2r\sin((\alpha + \beta)/2))^2$ and $q^2 = (2r\sin((\beta + \delta)/2))^2$, it follows from (3) that $ac/bd = p^2/q^2$.

II. *Solution by O. Nouhaud, Faculté des Sciences de Limoges, France.*

Let A', B', C', D' be the inverses of A, B, C, D respectively under the inversion about the inscribed circle with center O and radius r. We know that

$$|A'B'| = r^2\frac{|AB|}{|OA||OB|}$$

(e.g., see *A Survey of Geometry*, Howard Eves, Allyn and Bacon, Boston, 1963, Theorem 3.4.20, p. 153). A circular permutation gives three similar relations. Moreover, $2|A'B'| = |SQ|$ because A' bisects SP and B' bisects PQ. Similarly, $2|C'D'| = |SQ|$ and $2|A'D'| = 2|B'C'| = |RP|$. The desired result follows from these relations.

Also solved by Mangho Ahuja, Wadie A. Bassali (Kuwait), J.-M. Becker (France), Bilkent University Problem Solving Group (Turkey), J. C. Binz (Switzerland), Duane M. Broline, Brown University Fly-Fishing Club, Onn Chan (student), Gang Chang (student), Chico Problem Group, Timothy Chow, Ragnar Dybvik (Norway), E. C. Friedman, Francis M. Henderson, J. Heuver (Canada), Geoffrey A. Kandall, Hans Kappus (Switzerland), Václav Konečný, L. Kuipers (Switzerland), Helen M. Marston, Richard E. Pfiefer, James S. Robertson, Harry D. Ruderman, Raul A. Simon (Chile), László Szücs, R. S. Tiberio, George Vakanas (student), and the proposer.

Balancing rational masses **June 1988**

1299. *Proposed by Isaac J. Schoenberg, Madison, Wisconsin.*

Eight positive point-weights of masses $m_i (i = 0, 1, \ldots, 7)$ are placed, respectively, on the vertices $A_i (i = 0, 1, \ldots, 7)$ of a regular octagon, and they are such that their centroid is at the center of the octagon. Assume all eight numbers m_i are *rational* numbers. Show that they balance out in diametrically opposite pairs; i.e., $m_0 = m_4$, $m_1 = m_5$, $m_2 = m_6$, $m_3 = m_7$.

I. *Solution by Mangho Ahuja, Southeast Missouri State University, Cape Girardeau.* Taking moments of the masses about the line A_0A_4,

$$m_2 - m_6 = (m_5 + m_7 - m_1 - m_3)(1/\sqrt{2}).$$

Here, both sides of the equation must equal zero, otherwise we have a rational number equal to an irrational number. Therefore, $m_2 = m_6$. Similarly, or by symmetry, $m_3 = m_1$, $m_4 = m_0$, $m_1 = m_5$.

II. *Solution by Daniel B. Shapiro, Ohio State University, Columbus.*

We consider more generally n point-weights of masses m_i placed at the vertices of a regular n-gon in the plane, having centroid at the center. Viewing this in the complex plane we let ω be the primitive nth root of unity $\exp(2\pi i/n)$, and consider the mass m_k placed at the point ω^k. The centroid condition becomes: $\Sigma_{k=0}^{n-1} m_k \omega^k = 0$.

The set of all such vectors $\mathbf{m} = (m_1, m_2, \dots, m_{n-1})$ forms a linear subspace M_n of $\mathbf{Q}^n$. Some elements of this space M_n arise from the fact that equal masses placed on any regular r-gon will balance. Then if $n = rs$, there are s different regular r-gons inscribed in the given n-gon. These correspond to the vertices $\{\omega^k, \omega^{k+s}, \omega^{k+2s}, \dots, \omega^{k+(r-1)s}\}$ for $k = 0, 1, \dots, s-1$. Placing unit masses at the vertices of such an r-gon corresponds to a vector $\mathbf{m}(r, k)$ in M_n, where the entry $m_i(r, k) = 1$ if $i \equiv k \pmod{s}$, and equals 0 otherwise.

THEOREM. *For any n the space M_n is generated by the vectors $\mathbf{m}(p, k)$, where p is a prime dividing n and $0 \leqslant k < n/p$.*

COROLLARY. *If $n = 2^t$ then the masses balance out in opposite pairs. That is, letting $s = n/2$ we have $m_0 = m_s, m_1 = m_{s+1}, m_2 = m_{s+2}, \dots$.*

This corollary follows since $\mathbf{m}(2, k)$ corresponds to a pair of diametrically opposite unit masses. This settles the original problem when $n = 8$. Similarly, the only way 5 rational masses can balance on a regular pentagon is when all masses are equal. The only way 9 rational masses can balance on a regular 9-gon is when the masses balance out in equilateral triangles.

Proof of the Theorem. Let $A = \mathbf{Q}[x]/(x^n - 1)$, a $\mathbf{Q}$-algebra of dimension n. We identify a vector $\mathbf{m} = (m_0, m_1, \dots, m_{n-1})$ in $\mathbf{Q}^n$ with the polynomial $\Sigma m_j x^j$. Let ε: $A \to \mathbf{Q}[\omega]$ be the evaluation map sending $f(x) \to f(\omega)$. Then by definition the space M_n is the kernel of ε, and this space is exactly the ideal in A generated by the cyclotomic polynomial $F_n(x)$. Since $\deg(F_n(x)) = \dim \mathbf{Q}[\omega] = \varphi(n)$, we know that $\dim M_n = n - \varphi(n)$. In fact a vector space basis of the space M_n is $\{F_n(x), xF_n(x), \dots, x^{\varphi(n)-1}F_n(x)\}$.

Using the identification of $\mathbf{Q}^n$ with A, the vector $\mathbf{m}(r, k)$ becomes $x^k g_r(x)$ where $g_r(x) = 1 + x^s + x^{2s} + \cdots + x^{(r-1)s}$. The theorem that the space M_n is generated by the vectors $\mathbf{m}(p, k)$ is equivalent to the statement that the ideal $M_n = \langle F_n(x) \rangle$ is generated by the elements $g_p(x)$ where p is a prime dividing n. Pulling back from A to the polynomial ring $\mathbf{Q}[x]$, we see that the theorem is equivalent to: $F_n(x)$ is the greatest common divisor of the polynomials $g_p(x)$. To prove it we factor $g_p(x)$ into irreducibles, noting that $g_p(x) = (x^n - 1)/(x^{n/p} - 1)$. The irreducible factors are those $F_d(x)$ where $d|n$ but $d \nmid (n/p)$. It easily follows that the only time $F_d(x)$ can be a factor of all the polynomials $g_p(x)$ for $p|n$ is when $d = n$. Hence the gcd equals $F_n(x)$ as claimed.

Note: If n is not a prime power there will be some dependence relations among the generating vectors $\mathbf{m}(p, k)$.

Also solved by S. F. Barger, Robert F. Barnes, J.-M. Becker (France), Bilkent University Problem Solving Group (Turkey), Ada Booth, Joshua Brandon, Duane M. Broline, Brown University Fly-Fishing Club, David Callan, Chico Problem Group (two solutions), Timothy Chow, Mary Lee Comer (student), Jim Delany, Brandon Dixon (student), Scott Ellett, E. C. Friedman, John F. Goehl, Jr., Thomas Jager, William J. Joel, Hans Kappus (Switzerland), Farhood Pouryoussefi Kermany (student, Iran), Benjamin G. Klein, Václav Konečný, Kee-Wai Lau (Hong Kong), Kathleen E. Lewis, Andreas Müller (Switzerland), Stephen Noltie, M. Riazi-Kermani, Adam Riese, H. Paul Smith (Canada), William P. Wardlaw, and the proposer.

Shapiro also investigated some special cases in 3 dimensions (viz., the Platonic solids, and other polyhedra) but did not find a unifying theorem.

A partition of $\{1,2,\ldots,n-2\}$ **June 1988**

1300. *Proposed by Edward Kitchen, Santa Monica, California.*

Let r, s, n be positive integers such that $r+s=n$. Prove that

$$\{1,2,\ldots,n-2\}=\left\{\left\lfloor\frac{n}{r}\right\rfloor,\left\lfloor\frac{2n}{r}\right\rfloor,\ldots,\left\lfloor\frac{(r-1)n}{r}\right\rfloor\right\}\cup\left\{\left\lfloor\frac{n}{s}\right\rfloor,\left\lfloor\frac{2n}{s}\right\rfloor,\ldots,\left\lfloor\frac{(s-1)n}{s}\right\rfloor\right\}$$

if and only if both r and s are relatively prime to n.

Solution by James Propp, University of California, Berkeley.

Note that the two sets

$$A=\left\{\left\lfloor\frac{n}{r}\right\rfloor,\left\lfloor\frac{2n}{r}\right\rfloor,\ldots,\left\lfloor\frac{(r-1)n}{r}\right\rfloor\right\}$$

and

$$B=\left\{\left\lfloor\frac{n}{s}\right\rfloor,\left\lfloor\frac{2n}{s}\right\rfloor,\ldots,\left\lfloor\frac{(s-1)n}{s}\right\rfloor\right\}$$

are subsets of $\{1,2,\ldots,n-2\}$ of cardinality $r-1$ and $s-1$, respectively. Since $\{1,2,\ldots,n-2\}$ itself has cardinality $n-2=(r-1)+(s-1)$, the stated condition $\{1,2,\ldots,n-2\}=A\cup B$ is equivalent to $A\cap B=\varnothing$.

If either of r,s has some nontrivial factor d in common with $n=r+s$, then d divides both r and s; setting $r'=r/d$ and $s'=s/d$ we get

$$\left\lfloor\frac{r'n}{r}\right\rfloor=\left\lfloor\frac{s'n}{s}\right\rfloor,$$

so that A and B are not disjoint.

Suppose now that r and s are relatively prime to n. For $1\leqslant k\leqslant n-1$, kr/n and ks/n are nonintegers summing to k, so that $\lfloor kr/n\rfloor+\lfloor ks/n\rfloor=k-1$. Note, however, that $\lfloor kr/n\rfloor=\lfloor k/(n/r)\rfloor$ is the number of positive multiples of n/r less than or equal to k, which in turn equals the cardinality of $A\cap\{1,2,\ldots,k\}$; similarly, $\lfloor ks/n\rfloor$ equals the cardinality of $B\cap\{1,2,\ldots,k\}$. We see that the two cardinalities sum to $k-1$. Considering this relation for successive values of k, we find that each time k increases by 1, either $A\cap\{1,2,\ldots,k\}$ or $B\cap\{1,2,\ldots,k\}$ (but not both) increases in size by 1. It follows that A and B are disjoint, as claimed.

Also solved by Bilkent University Problem Solving Group (Turkey), Joshua Brandon, Duane M. Broline, Brown University Fly-Fishing Club, Onn Chan (student), David Doster, Kazuhiro Ishikawa (Japan), Thomas Jager, L. Kuipers (Switzerland), David E. Manes, Helen M. Marston, Bruce Richter (Canada), Raul A. Simon (Chile), A. Zulauf (New Zealand), and the proposer.

This problem is a variation of Beatty's Theorem, which states that if α and β are positive irrational numbers satisfying

$$\frac{1}{\alpha}+\frac{1}{\beta}=1,$$

then $\{\lfloor\alpha\rfloor,\lfloor 2\alpha\rfloor,\ldots\}$ and $\{\lfloor\beta\rfloor,\lfloor 2\beta\rfloor,\ldots\}$ are disjoint sets with union $\{1,2,3,\ldots\}$. This theme has been studied in a much more general form by A. S. Fraenkel, The Bracket Function and Complementary Sets of Integers, *Canad. J. Math.* 21 (1969), 6–27.

Subrings and ideals **June 1988**

1301. *Proposed by Vidhyanath K. Rao, Ohio State University at Newark.*

Let T be the ring of symmetric polynomials in $\mathbf{Q}[x,y]$, where $\mathbf{Q}$ denotes the ring of rational numbers. Prove that the *subring* of T generated by $\{a(x^n+y^n)\colon a\in\mathbf{Q},\ n \text{ odd}\}$ is equal to the *ideal* of T generated by $x+y$.

Solution by David Callan, University of Bridgeport, Connecticut.

Let S and I denote the relevant subring and ideal, respectively. Clearly, $S \subseteq I$, since $x + y$ divides (in T) $x^n + y^n$ for odd n. We must show that $I \subseteq S$.

LEMMA. Suppose $0 \leqslant i, j$ and $i + j = 2n - 1$ is odd. Then $x^i y^j + x^j y^i \in S$.

Proof. By induction on n. The case $n = 1$ is obvious. If true for n, $x^{2n-1-i}y^i + x^i y^{2n-1-i} \in S$ for $i = 0, 1, \dots, n-1$. So the product of each by $(x+y)^2$ is in S. Also, by hypothesis, $x^{2n+1} + y^{2n+1}$ is in S. This says, in matrix form, that the product

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} x^{2n+1} + y^{2n+1} \\ x^{2n}y + xy^{2n} \\ \vdots \\ \vdots \\ x^{n+2}y^{n-1} + x^{n-1}y^{n+2} \\ x^{n+1}y^n + x^n y^{n+1} \end{pmatrix}$$

is a $(n+1) \times 1$ column matrix with elements in S. The coefficient matrix is invertible (it has determinant $= 2n + 1$), and multiplying by its inverse yields the result for $n + 1$.

Now $q \in I$ can clearly be expressed as a finite sum $q = \sum_{i,j} a_{i,j}(x^i y^j + x^j y^i)(x + y)$ with $a_{i,j} \in \mathbf{Q}$. Then, using the lemma, $(x^i y^j + x^j y^i)(x + y)$ is in S (i) if $i + j$ is odd, since S is closed under multiplication, (ii) if $i + j$ is even, since then the product is a sum of two odd-degree expressions as in the lemma. Thus $q \in S$, completing the proof.

Also solved by Duane M. Broline, Lorraine L. Foster, Thomas Jager, Andreas Müller (Switzerland), and the proposer.

Answers

Solutions to the Quickies on p. 199.

A748. On dividing,

$$F = -1 + \frac{2(a^{2n} + b^{2n})x^{2n}}{x^{4n} + (a^{2n} + b^{2n})x^{2n} + a^{2n}b^{2n}}.$$

Now by the arithmetic mean-geometric mean inequality,

$$x^{2n} + \frac{a^{2n}b^{2n}}{x^{2n}} + (a^{2n} + b^{2n})$$

takes on its minimum value when $x^2 = ab$ (we can assume that $a, b \geqslant 0$). Finally,

$$F_{\max} = \frac{(b^n - a^n)^2}{(b^n + a^n)^2}.$$

A749. Expanding out, we have to show that

$$z^\lambda x^\lambda (x^{\lambda+1} - y^{\lambda+1}) + x^\lambda y^\lambda (y^{\lambda+1} - z^{\lambda+1}) + y^\lambda z^\lambda (z^{\lambda+1} - x^{\lambda+1}) \geqslant 0.$$

Since the inequality is cyclic, we can assume without loss of generality that (i) $x \geqslant y \geqslant z$ or else (ii) $x \geqslant z \geqslant y$.

For (i), we can rewrite the inequality in the obvious form

$$z^{\lambda}(x^{\lambda}-y^{\lambda})(x^{\lambda+1}-y^{\lambda+1})+y^{\lambda}(x^{\lambda}-z^{\lambda})(y^{\lambda+1}-z^{\lambda+1})\geqslant 0.$$

For (ii), we rewrite the inequality in the form

$$z^{\lambda}(x^{\lambda}-y^{\lambda})(x^{\lambda+1}-z^{\lambda+1})+x^{\lambda}(z^{\lambda}-y^{\lambda})(z^{\lambda+1}-y^{\lambda+1})\geqslant 0.$$

More generally, the given inequality is the special case $a=y, b=z, c=x$ of the inequality

$$\frac{x^{\lambda+1}}{a^{\lambda}}+\frac{y^{\lambda+1}}{b^{\lambda}}+\frac{z^{\lambda+1}}{c^{\lambda}}\geqslant\frac{(x+y+z)^{\lambda+1}}{(a+b+c)^{\lambda}},$$

where $\lambda=(x+y+z)/(a+b+c)$, and $x, y, z, a, b, c>0$.

To prove this, let

$$F(\lambda)\equiv\left(\frac{a(x/a)^{\lambda+1}+b(y/b)^{\lambda+1}+c(z/c)^{\lambda+1}}{a+b+c}\right)^{1/(\lambda+1)}.$$

Then by the power mean inequality

$$F(\lambda)\geqslant F(0),$$

which gives the desired result. It is to be noted that λ may be 0 and the inequality can be extended to

$$\sum_i\frac{x_i^{\lambda+1}}{a_i^{\lambda}}\geqslant\frac{(\Sigma_i x_i)^{\lambda+1}}{(\Sigma_i a_i)^{\lambda}}.$$

Another proof of

$$\frac{x_1^{\lambda+1}}{x_2^{\lambda}}+\frac{x_2^{\lambda+1}}{x_3^{\lambda}}+\cdots+\frac{x_n^{\lambda+1}}{x_1^{\lambda}}\geqslant x_1+x_2+\cdots+x_n$$

follows immediately from the rearrangement inequality; i.e., if $a_1\geqslant a_2\geqslant\cdots\geqslant a_n\geqslant 0, b_1\geqslant b_2\geqslant\cdots\geqslant b_n\geqslant 0, c_i$'s are a permutation of the b_i's, then $a_1c_1+a_2c_2+\cdots+a_nc_n\geqslant a_1b_n+a_2b_{n-1}+\cdots+a_nb_1$.

A750. Let a, b, c, d and e, f denote the sides and diagonals of a spherical quadrilateral. Then the chord lengths of the spherical arcs of the sides and diagonals are given by $a'=2R\sin(a/2)$, $b'=2R\sin(b/2)$, etc., where R is the radius of the sphere. Then by Ptolemy's inequality above,

$$\sin(a/2)\cdot\sin(c/2)+\sin(b/2)\cdot\sin(d/2)\geqslant\sin(e/2)\cdot\sin(f/2).$$

Again there is equality if and only if the quadrilateral is cyclic.

A751. We will show more generally that if $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ are vectors in $\mathbf{E}^n$ such that

$$|\mathbf{A}_{i+1}+\mathbf{A}_{i+2}+\cdots+\mathbf{A}_{i+r}|=|\mathbf{A}_{i+r+1}+\mathbf{A}_{i+r+2}+\cdots+\mathbf{A}_{i+n}|$$

for $2r<n$, $i=0,1,\ldots,n-1$, and $\mathbf{A}_{i+n}=\mathbf{A}_i$, then

$$\mathbf{A}_1+\mathbf{A}_2+\cdots+\mathbf{A}_n=0.$$

Proof. Let $\mathbf{S}_i=\mathbf{A}_{i+1}+\mathbf{A}_{i+2}+\cdots+\mathbf{A}_{i+r}$ and $\mathbf{S}=\mathbf{A}_1+\mathbf{A}_2+\cdots+\mathbf{A}_n$. The given relations become $|\mathbf{S}_i|=|\mathbf{S}-\mathbf{S}_i|$ which on squaring becomes $\mathbf{S}^2=2\mathbf{S}\cdot\mathbf{S}_i$. On summing over i, we get $n\mathbf{S}^2=2r\mathbf{S}^2$. Hence, $\mathbf{S}=0$.

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.

Andrews, Edmund L., Equations patented; some see a danger, *New York Times* (15 February 1989) 25, 30.

In the last year the US Patent and Trademark Office has issued patents for a number of algorithms, including one to N. Karmarkar (Bell Labs) for his linear programming algorithm. A number of other data compression algorithms have also been awarded patents. This action represents a departure from past practice, including rulings by the Supreme Court that "scientific truth, or the mathematical expression of it, is not a patentable invention" (1939) and an "algorithm, or mathematical formula, is like a law of nature, which cannot be the subject of a patent" (1972). In 1980, however, the Court ruled that genetically-engineered bacteria could be patented. In 1981, in *Diamond vs. Diehr* (concerning computerized rubber-curing), the court held that an "application of a law of nature or mathematical formula" could be patentable. Conforming to the widening scope of such rulings, the Patent Office has been approving patents for inventions for which the algorithm is the main basis for the claim and applications are described only hypothetically. A key consideration in obtaining and being able to defend a patent is a sufficiently-detailed description of the invention, i.e., the algorithm. Karmarkar issued a public version of his algorithm several years ago; mathematicians who implemented it found that it did not perform as fast as Karmarkar's own private version, whose tuning heuristics he would not specify. Whether those heuristics are stated in the patent application remains to be seen. Meanwhile, despite claims of university officials and patent attorneys that "patents are necessary to provide incentives for future research," many mathematicians view the patenting of algorithms and equations with distaste and apprehension. If patents for algorithms had been awarded 20 years ago, most of the basic knowledge in computer science would be under patent, and its implementation in hardware or software would require licensing. The new development, though, seems characteristic of the Reagan-Bush era of privatization of public resources.

Goldberg, David E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, 1989; xiii + 412 pp. ISBN 0-201-15767-5

Introduction to the intriguing idea of search alorithms based on the metaphor and mechanics of natural selection and genetics. Genetic algorithms process populations of strings, via reproduction (in proportion to "fitness," i.e., the objective function to be optimized), crossover ("mating" of strings), and mutation. The book, which requires some general background in mathematics (sigma notation, normal probability density, combinatorics, occasional calculus) and computer programming, includes problems and computer assignments.

Paulos, John Allen, *Innumeracy: Mathematical Illiteracy and Its Consequences*, Hill and Wang, 1988; 135 pp, $16.95. ISBN 0-8090-7447-8

Breezy and readable, this entertaining book ranges over the foibles of innumeracy, touching upon large numbers, small probabilities, and pseudoscience. As a partial cure for the mass malady of innumeracy, Paulos prescribes the hiring of statistical ombudsmen by newspapers and TV networks and their promotion of the use of a logarithmic safety index for risks (like the Richter scale for earthquakes). Without being venomous, he criticizes those who are bent on "discovering meaning where there is only probability. . . [I]nnumerate people characteristically have a strong tendency to personalize--to be misled by their own experiences, or by the media's focus on individuals and drama." In that sense, to be innumerate is to be unobjective and unphilosophical; to be invincibly innumerate is incompatible with being educated; and flaunting innumeracy ("I never could do math") is gross anti-intellectualism.

Graham, Ronald L., Donald E. Knuth, and Oren Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, 1989; xiii + 625 pp. ISBN 0-201-14236-8

A delightful mathematical text in the spirit of Euler (to whom it is dedicated)! The title is that of a course taught at Stanford since 1970; originally intended as an antidote to "Abstract Mathematics," that title also reflects the book's content, "a blend of CONtinuous and disCRETE mathematics." Topics covered are sums, recurrences, number theory, binomial coefficients, generating functions, discrete probability, and asymptotic methods, with emphasis on "controlled manipulation" of mathematical formulas rather than existence theorems or combinatorial reasoning. The authors claim that the topics are "quite different" from those usually taught in undergraduate courses entitled "Discrete Mathematics"; they feel the course is appropriate for sophomores, as well as juniors, seniors, and graduate students; but they don't give any other guidance about prerequisites or intended audience. Mathematicians and theoretical computer scientists will find the book an utter pleasure and an indispensable reference. Apart from a section on hashing, though, there is no reference to motivations or applications from computing; so pragmatic computer science students are unlikely to be able to visualize "how what I've learned will ever help me." The text features margins full of such student graffiti (as well as quotes from famous mathematicians), and the inaugural use of a new mathematics typeface (AMS Euler); there is a $2.56 reward to the first finder of any error.

Cohen, Joel E., Big fish, little fish: The search for patterns in predator-prey relationships, *The Sciences* (January-February 1989) 37-42.

Mathematicians tend to exhibit the Lotka-Volterra predator-prey system as an application of differential equations; Cohen notes here that biologists reject it as a failed model. He goes on to describe in qualitative terms a new model, the cascade model, now being used to investigate the structure of food webs.

Resnikoff, Howard L., *The Illusion of Reality*, Springer-Verlag, 1989; x + 339 pp, $44. ISBN 0-387-96398-7

This introductory text for advanced undergraduates or graduates provides a "comprehensive view of information science as a fundamental constituent of other more established disciplines with a unity and coherence distinct from computer science, cognitive science, and library science." Chapters cover the mathematics of information measurement, physical measurements and information, principles of information-processing systems and signal detection, biological signal detection, and pattern structure and learning. There are no exercises.

Barnsley, Michael, *Fractals Everywhere*, Academic, 1988; xii + 396 + 31 color plates, $39.95. ISBN 0-12-079062-9

Splendid introduction to the mathematics behind fractals, based on a course at Georgia Tech whose prerequisite is "two years of calculus" (including linear algebra and mathematical induction). The book begins with metric spaces and their topology, then moves on through transformations and contraction mappings, to chaotic dynamics on fractals, fractal dimension, fractal interpolation, Julia sets, Mandelbrot sets, and measures on fractals. The author warns the reader, "You risk the loss of your childhood vision of clouds, forests, galaxies, leaves, feathers, flowers, rocks, mountains, torrents of water, carpets, bricks, and much else besides." The reader also risks learning a significant amount of advanced mathematics, with pleasant motivation.

Tanur, Judith M., et al. (eds.), *Statistics: A Guide to the Unknown*, 3rd ed., Wadsworth & Brooks/Cole, 1989; xxv + 284 pp, $16.95(P). ISBN 0-534-09492-9

A new and "leaner" edition of a useful supplement in statistics classes, featuring essays describing applications of statistics in different fields, with "study material." Only 17 essays of the 44 in the second edition (1978) have been carried over, and those have been updated; there are 12 new essays, on health insurance costs, employment discrimination, juror selection, employment statistics, economic indicators, earthquakes, and other topics.

van Laarhoven, P. J. M., and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, Reidel, 1987; xi + 186 pp, $49. ISBN 90-277-2513-6

Annealing is the process of heating a solid and then cooling it slowly so as to remove strain and crystal imperfections; the process involves minimization of the "free energy" of the solid. An analogy can be made for combinatorial optimization problems, with the cost function taking the role of energy and a control parameter corresponding to temperature. The consequent simulated annealing (also called *probabilistic hill climbing* or *stochastic relaxation*) provides a new (1982) general optimization technique, based on randomization and iterated improvement techniques. While iterated improvement approaches optimality by moving to points with monotonically lower "cost," simulated annealing paradoxically allows perturbations to points with higher cost function, in a limited way. Applications include computer-aided circuit design, neural-network machines, image-processing, and wheel balancing. The book itself is at an advanced level. Based on an anecdote with scandalous stereotyping, the authors recommend some chapters to mathematicians, others to physicists, still others to electrical engineers, "the complete book to combinatorial optimizers," and "another book" to biologists [perhaps Goldberg's *Genetic Algorithms*?].

Klamkin, Murray S. (comp.), *USA Mathematical Olympiads* 1972-1986, New Mathematical Library 33, MAA, 1988; xv + 127 pp, $13.50(P). ISBN 0-88385-634-4

Collection of the first 15 US Mathematical Olympiad competition essay-type problems, with solutions "more detailed than need be for the USAMO contestants." The problems are presented in chronological order but the solutions are grouped by subject matter.

Reid, Miles, *Undergraduate Algebraic Geometry*, Cambridge U Pr, 1988; viii + 129 pp, $12.95(P). ISBN 0-521-35662-8

Introduction to a subject "respected and feared much more than understood," at a level advanced undergraduates will understand and appreciate. Prerequisites include abstract algebra (including Galois theory) and some topology and geometry. There is a goodly number of exercises.

Case, Bettye Anne, et al. (eds.), *Response to the Challenge: Keys to Improved Instruction by Teaching Assistants and Part-Time Instructors*, MAA Notes 11, MAA, 1989; vi + 266 pp, $15.50(P). ISBN 0-88385-061-3

Contains data on use of TAs and part-time instructors in US institutions, models of programs at various institutions, suggestions on international teaching assistants (who should not be made "scapegoats for problems in U.S. education"), and surveys of student perceptions. Almost three-quarters of the book consists of an appendix of reproducible handouts: guides and manuals for TAs and temporary instructors, orientation materials, and evaluation forms, all of which the editors invite interested parties to reproduce and distribute.

Bern, Marshall W., and Ronald L. Graham, The shortest-network problem, *Scientific American* (January 1989) 84-89.

The problem of finding the shortest network of line segments interconnecting a set of points is called the *Steiner problem*. It is an NP-hard problem (no polynomial-time algorithm is known), and problems with 100 points are still well out of reach of exact solutions (though near-optimal solutions are attainable). A rectilinear version is useful in circuit design, and there is a potential application in phylogeny (only for small numbers of organisms).

Denning, Peter J., et al., Computing as a discipline, *Communications of the Association for Computing Machinery* 32:1 (January 1989) 9-23; *Computer* 22:2 (February 1989) 63-70. Condensations of the *Report of the ACM Task Force on the Core of Computer Science*, ACM order number 201880, $12.

What is computer science? How should it be introduced to students? Here is the "short definition" from this valuable report: "the systematic study of algorithmic processes that describe and transform information: their theory, analysis, design, efficiency, implementation, and application. The fundamental question underlying all computing is, 'What can be (efficiently) automated?'" (Both journals feature the body of the report, with curiously different editing. The *Communications* version also has the unabridged edition of an appendix that gives a full definition of computing as a discipline, while *Computer* abridges it slightly into a convenient table. Unfortunately, a second appendix, with detailed recommendations for the introductory course sequence, is omitted from both journal versions; but it is in the full report.)

Kenschaft, Patricia Clark, Confronting myths about math, *Journal of Career Planning & Employment* (Summer 1988) 41-44.

Important career information for math majors! Summarizes results from the author's survey of recent math graduates of her institution, which gives a picture of broad options available to math majors. In addition, eight myths are cited and countered (e.g., "Myth 7: Most math majors become teachers."). Not only should this article be posted outside your department office, it should also be in the hands of college admissions personnel, high-school counselors, and high-school math teachers.

NEWS AND LETTERS

SEVENTEENTH ANNUAL U.S.A. MATHEMATICAL OLYMPIAD–SOLUTIONS

1. By a *pure repeating decimal* (in base 10) we mean a decimal $0.\overline{a_1 \cdots a_k}$ which repeats in blocks of k digits beginning at the decimal point. An example is .243243243... $= \frac{9}{37}$. By a *mixed repeating decimal* we mean a decimal $0.b_1 \cdots b_m \overline{a_1 \cdots a_k}$ which eventually repeats, but which cannot be reduced to a pure repeating decimal. An example is .011363636··· $= \frac{1}{88}$.

Prove that if a mixed repeating decimal is written as a fraction $\frac{p}{q}$ in lowest terms, then the denominator q is divisible by 2 or 5 or both.

Sol. Let the decimal $0.b_1 \cdots b_m \overline{a_1 \cdots a_k}$ equal the fraction p/q. By elementary arithmetic

$$\frac{p}{q} = \frac{(10^k - 1)b_1 \cdots b_m + a_1 \cdots a_k}{10^m(10^k - 1)}$$

$$(*) \quad = \frac{10^k b_1 \cdots b_m + (a_1 \cdots a_k - b_1 \cdots b_m)}{10^m(10^k - 1)} .$$

We must consider what happens when the fraction $(*)$ is reduced to lowest terms. For this purpose we introduce the notion:

> A repeating decimal is *properly presented* if the repeating part has been moved as far to the left as possible.

Without loss of generality, we can assume that the decimal $0.b_1 \cdots b_m \overline{a_1 \cdots a_k}$ is properly presented. Since this is a mixed decimal, $m, k \geq 1$. A little thought shows that the above decimal is properly presented if and only if the digit $b_m \neq a_k$. Hence $(a_1 \cdots a_k - b_1 \cdots b_m)$ is not divisible by 10, and neither is

$$10^k b_1 \cdots b_m + (a_1 \cdots a_k - b_1 \cdots b_m).$$

Thus, in reducing the fraction $(*)$ to lowest terms, some (or all) of the 2's may cancel, or some (or all) of the 5's may cancel, but not both.

2. The cubic equation $x^3 + ax^2 + bx + c = 0$ has three real roots. Show that $a^2 - 3b \geq 0$, and that $\sqrt{a^2 - 3b}$ is less than or equal to the difference between the largest and smallest roots.

Sol. Let the roots of the cubic be $p \leq q \leq r$, so that

$$\begin{aligned} a^2 - 3b &= (-p - q - r)^2 - 3(pq + qr + rp) \\ &= p^2 + q^2 + r^2 - pq - qr - rp. \end{aligned}$$

We may prove $a^2 - 3b \geq 0$ by using the Cauchy-Schwarz inequality:

$$pq + qr + rp \leq p^2 + q^2 + r^2$$

since $(pq+qr+rp)^2 \leq (p^2+q^2+r^2)(q^2+r^2+p^2) = (p^2 + q^2 + r^2)^2$.

To obtain the second required result, note that the following four inequalities are equivalent:

$$a^2 - 3b \leq (r - p)^2,$$

$$p^2 + q^2 + r^2 - pq - qr - rp \leq r^2 - 2rp + p^2,$$

$$0 \leq pq + qr - rp - q^2,$$

$$0 \leq (r - q)(q - p),$$

and that the last of these follows immediately from the hypothesis $p \leq q \leq r$.

3. A function $f(S)$ assigns to each nine-element subset S of the set $\{1, 2, 3, ..., 20\}$ a whole number from 1 to 20. Prove that, regardless of how the function f is chosen, there will be a ten-element subset $T \subset \{1, 2, 3, ..., 20\}$ such that $f(T - \{k\}) \neq k$ for all $k \in T$.

Sol. The key observation is that there are at most $\binom{20}{9}$ "bad" (k, T) pairs with $k \in T$ and $f(T - \{k\}) = k$. This is true because each of the $\binom{20}{9}$ nine-element sets $S \subset \{1,\ldots,20\}$ can appear as $T - \{k\}$ in *at most one* of the bad relations $f(T - \{k\}) = k$. (For this to happen, the number k could only equal $f(S)$; set T could only equal $S \cup \{k\}$; and even then (k, T) would only be a bad pair if T had ten elements.)

But while there are at most $\binom{20}{9}$ of these bad (k, T) pairs, the set $\{1,\ldots 20\}$ has $\binom{20}{10}$ different ten-element subsets T, and of course $\binom{20}{10} > \binom{20}{9}$. Thus some ten-element subset T is involved in *none* of the bad pairs, and for that T we have $f(T - \{k\}) \neq k$ for all $k \in T$.

4. Let I be the incenter of triangle ABC, and let A', B', and C' be the circumcenters of triangles IBC, ICA, and IAB, respectively. Prove that the circumcircles of triangles ABC and $A'B'C'$ are concentric.

Sol. If we let O be the circumcenter of $\triangle ABC$, it will suffice to show $OA' = OB'$, and this can be established simply by proving $\angle A'B'O = \angle B'A'O$.

Segment IC is a common chord of the two given circles centered at A' and B', therefore $A'B' \perp IC$. Similarly, $B'O \perp CA$ and $A'O \perp CB$.

The acute angle between two lines is equal to the acute angle between their respective perpendiculars. Thus, if we define ***the reference angle of*** $\angle XYZ$ [ref($\angle XYZ$)] to equal the measure of $\angle XYZ$ or its supplement, whichever is less than or equal to 90°, then our perpendicularity results tell us that

$$\text{ref}(\angle A'B'O) = \text{ref}(\angle ICA)$$

and

$$\text{ref}(\angle B'A'O) = \text{ref}(\angle ICB).$$

We now recall that I is the incenter of the original triangle, therefore $\angle ICA = \angle ICB$ and hence ref($\angle A'B'O$) = ref($\angle B'A'O$).

But $\angle A'B'O$ and $\angle B'A'O$ are two angles of a triangle, and so they cannot be supplementary. Therefore the two angles must be equal.

5. A polynomial product of the form

$$(1\text{-}z)^{b_1}(1\text{-}z^2)^{b_2}(1\text{-}z^3)^{b_3}(1\text{-}z^4)^{b_4}(1\text{-}z^5)^{b_5}\cdots(1\text{-}z^{32})^{b_{32}},$$

where the b_k are positive integers, has the surprising property that if we multiply it out and discard all terms involving z to a power larger than 32, what is left is just $1 - 2z$. Determine, with proof, b_{32}. (The answer can be written as the difference of two powers of 2.)

Sol. Let

$$g(z) = (1 - z)^{b_1}(1 - z^2)^{b_2}(1 - z^3)^{b_3}(1 - z^4)^{b_4}\cdots$$
$$\cdots(1 - z^{31})^{b_{31}}(1 - z^{32})^{b_{32}}$$
$$\equiv 1 - 2z \pmod{z^{33}},$$

where the $\equiv$ means that some terms involving z to a power larger than 32 have been discarded.

Note that

$$g(-z) = (1 + z)^{b_1}(1 - z^2)^{b_2}(1 + z^3)^{b_3}(1 - z^4)^{b_4}\cdots$$
$$\cdots(1 + z^{31})^{b_{31}}(1 - z^{32})^{b_{32}}$$
$$\equiv 1 + 2z \pmod{z^{33}}.$$

Hence

$$g(z)g(-z) \equiv$$
$$(1 - z^2)^{b_1+2b_2}(1 - z^4)^{2b_4}(1 - z^6)^{b_3+2b_6}(1 - z^8)^{2b_8}\cdots$$
$$\cdots(1 - z^{30})^{b_{15}+2b_{30}}(1 - z^{32})^{2b_{32}}$$
$$\equiv 1 - 2^2z^2 \pmod{z^{33}}.$$

Let $q = z^2$ and $c_i = \begin{cases} b_i + 2b_{2i} & \text{if } i \text{ is odd} \\ 2b_{2i} & \text{if } i \text{ is even.} \end{cases}$

Then

$$g_1(q) = g(z)g(-z)$$
$$\equiv (1\text{-}q)^{c_1}(1\text{-}q^2)^{c_2}(1\text{-}q^3)^{c_3}$$
$$\ldots(1\text{-}q^{15})^{c_{15}}(1\text{-}q^{16})^{2b_{32}}$$
$$\equiv 1 - 2^2q \pmod{q^{17}},$$

where here, of course, the $\equiv$ refers to the discarding of terms involving q to a power larger than 16.

Similarly let $r = q^2$ and $g_2(r) = g_1(q)g_1(-q)$ to obtain

$$g_2(r)$$
$$\equiv (1\text{-}r)^{d_1}(1\text{-}r^2)^{d_2}(1\text{-}r^3)^{d_3}\ldots(1\text{-}r^7)^{d_7}(1\text{-}r^8)^{4b_{32}}$$
$$\equiv 1 - 2^4r \pmod{r^9},$$

use $s = r^2$ and $g_3(s) = g_2(r)g_2(-r)$ to obtain

$$g_3(s) \equiv (1\text{-}s)^{e_1}(1\text{-}s^2)^{e_2}(1\text{-}s^3)^{e_3}(1\text{-}s^4)^{8b_{32}}$$
$$\equiv 1 - 2^8s \pmod{s^5},$$

and use $t = s^2$ and $g_4(t) = g_3(s)g_3(-s)$ to obtain

$$g_4(t) \equiv (1\text{-}t)^{f_1}(1\text{-}t^2)^{16b_{32}} \equiv 1 - 2^{16}t \pmod{t^3}.$$

Equating coefficients of t^2 gives $\binom{f_1}{2} - 16b_{32} = 0$ and equating coefficients of t gives $-f_1 = -2^{16}$, therefore

$$b_{32} = \frac{1}{16}\binom{2^{16}}{2} = \frac{2^{16}(2^{16} - 1)}{16\cdot 2} = 2^{27} - 2^{11}.$$

Note. Our hypothetical polynomial actually does exist, and in general the b_k can be computed via the recursion $\sum_{k|n} kb_k = 2^n$.

29th INTERNATIONAL MATHEMATICAL OLYMPIAD—SOLUTIONS

1. Consider two coplanar circles of radii R and r ($R > r$) with the same center. Let P be a fixed point on the smaller circle and B a variable point on the larger circle. The line BP meets the larger circle again at C. The perpendicular l to BP at P meets the smaller circle again at A. (If l is tangent to the circle at P then $A = P$.)

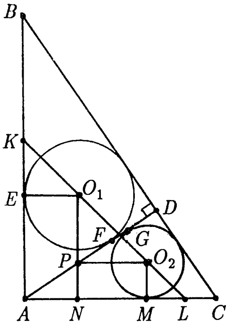
(i) Find the set of values of $BC^2 + CA^2 + AB^2$.

(ii) Find the locus of the midpoint of AB.

Sol. Let the center of the circles be the origin of a Cartesian coordinate system with the x-axis parallel to AP. We assign coordinates to A, B, C, and P as shown in the figure, and easily conclude that

$$BC^2 + CA^2 + AB^2$$
$$= (2y_1)^2 + [(2x)^2+(y_1+y_2)^2] + [(2x)^2+(y_1-y_2)^2]$$
$$= 6(x^2 + y_1^2) + 2(x^2 + y_2^2)$$
$$= 6R^2 + 2r^2,$$

which is constant.



Let the perpendicular to PB at B meet the larger circle again at D, so that $APBD$ is a rectangle. The midpoint M of diagonal AB is also the midpoint of diagonal PD, so $\vec{PM} = \frac{1}{2}\vec{PD}$. Since point P is fixed and point D traces the entire outer circle, the locus of M is a $\frac{1}{2}$ dilation of the outer circle with respect to the point P. More specifically, the locus of M is a circle of radius $\frac{R}{2}$ centered midway between P and the center of the given circles.

2. Let n be a positive integer and let A_1, A_2, ..., A_{2n+1} be subsets of a set B. Suppose that

(a) each A_i has exactly $2n$ elements,

(b) each $A_i \cap A_j$ $(1 \leq i < j \leq 2n + 1)$ contains exactly one element, and

(c) every element of B belongs to at least two of the A_i.

For which values of n can one assign to every element of B one of the numbers 0 and 1 in such a way that each A_i has 0 assigned to exactly n of its elements?

Sol. We prove that such an assignment is possible if and only if n is even.

I. To begin, we show that conditions (a)-(c) imply a strengthened version of (c), viz.

(c∗) Every element of B belongs to *exactly* two of the A_i.

First note that

$$(*) \qquad A_i = \bigcup_{\substack{j=1 \\ j \neq i}}^{2n+1} (A_i \cap A_j) \quad \text{for } i = 1,2,\ldots,2n+1.$$

The inclusion $\supseteq$ is trivial, and the reverse inclusion follows from (c).

Second, suppose, contrary to (c*), that for some labeling of the A_i the set $A_1 \cap A_2 \cap A_3$ is not empty. Then, by (b), the $2n$ - 1 sets $(A_1 \cap A_2) \cup (A_1 \cap A_3)$, $A_1 \cap A_4$, $A_1 \cap A_5$ ..., $A_1 \cap A_{2n+1}$ each contain exactly one element. Hence, by (∗), A_1 contains at most $2n$-1 elements, contradicting (a).

II. Next we show that if 0's and 1's can be assigned to the elements of B in the required manner, then n must be even. Let $B = \{b_1, b_2, \ldots b_m\}$, and consider all pairs (i,j) for which $b_i \in A_j$ and b_i is assigned the number 0. By assumption, there are n of these pairs for each value of j, and hence $n(2n + 1)$ such pairs altogether. On the other hand, by (c*) there are either 0 or 2 of these (i,j) pairs for each value of i, and hence an even number of such pairs altogether. Therefore $n(2n + 1)$ is even, and n is even.

III. Finally we show that if n is even then the required assignment of 0's and 1's is possible. Conditions (b) and (c*) show that the elements of B are in exact correspondence with the pairwise intersections $A_i \cap A_j$, $j \neq i$. Assign the number 0 to elements with

$$i - j \equiv \pm 2, \pm 4, \ldots \pm n \pmod{2n+1}$$

and the number 1 to elements with

$$i - j \equiv \pm 1, \pm 3, \ldots, \pm(n\text{-}1) \pmod{2n+1}.$$

For each value of i, the $2n$ differences i-j, $j \neq i$, generate this entire set of residues, so for each A_i the $2n$ elements will be assigned n 0's and n 1's.

3. A function f is defined on the positive integers by

$$f(1) = 1, \; f(3) = 3,$$
$$f(2n) = f(n),$$
$$f(4n + 1) = 2f(2n + 1) - f(n),$$
$$f(4n + 3) = 3f(2n + 1) - 2f(n),$$

for all positive integers n.

Determine the number of positive integers n, less than or equal to 1988, for which $f(n) = n$.

Sol. The definition of $f(n)$ suggests a connection with powers of two. If we tabulate n and $f(n)$ in binary notation, we note that $f(n)$ seems to be reversal of the binary digits of n. We prove that this is always the case by induction on n, taking for our base cases the equations

$$f(1_2) = 1_2,$$
$$f(10_2) = 1_2,$$
$$f(11_2) = 11_2.$$

For integers $n > 3$ we distinguish three cases: $n = A0_2$, $n = A01_2$, and $n = A11_2$, where A is some sequence of binary digits. We will use $\tilde{A}$ to denote the reversal of A, including all the 0's. For each case, we apply the appropriate recursion equation to complete the induction:

$$\begin{aligned} f(A0_2) &= f(A_2) = \tilde{A} \ , \\[1ex] f(A01_2) &= 2\cdot f(A1_2) - f(A_2) \\ &= 2\cdot[1\tilde{A}_2] - \tilde{A}_2 \\ &= [1\tilde{A}_2 + 1\tilde{A}_2] - \tilde{A}_2 \\ &= 10\tilde{A}_2, \\[1ex] f(A11_2) &= 3\cdot f(A1_2) - 2\cdot f(A_2) \\ &= 3\cdot[1\tilde{A}_2] - 2\cdot\tilde{A}_2 \\ &= [1\tilde{A}_2 + 1\tilde{A}0_2] - \tilde{A}0_2 \\ &= 11\tilde{A}_2. \end{aligned}$$

To finish the problem, we have to count the number of integers from 1 to 1988 ($= 11111000100_2$) which have palindromic binary expansions.

The number of $2m$-digit binary palindromes and the number of $(2m$-$1)$-digit palindromes are both equal to 2^{m-1}, since in each case the first digit must be 1, the next m - 1 digits can be 0 or 1 independently, and the remaining digits are uniquely determined by the previous ones.

There are

$$1 + 1 + 2 + 2 + 4 + 4 + 8 + 8 + 16 + 16 + 32 = 94$$

palindromes with eleven or fewer digits, and only two of these exceed 1988 (namely 11111011111_2 and 11111111111_2), so the required number is 92.

4. Show that the set of real numbers x which satisfy the inequality

$$\sum_{k=1}^{70} \frac{k}{x - k} \geq \frac{5}{4}$$

is a union of disjoint intervals, the sum of whose lengths is 1988.

Sol. Each term of the sum is piecewise continuous and piecewise decreasing in x. Therefore the sum itself is a piecewise continuous, piecewise decreasing function of x with asymptotes $x = k$, $k = 1, 2, \ldots 70$. This function decreases continuously from 0 to $-\infty$ on the interval $(-\infty,1)$, from $+\infty$ to $-\infty$ on each of the intervals $(1,2)$, $(2,3)$,...,$(69,70)$, and from $+\infty$ to 0 on the interval $(70,+\infty)$. It follows from the Intermediate Value Theorem that the set in question is a union of intervals of the form $(j,x_j]$, $j = 1,2,\ldots,70$, where $1 < x_1 < 2 < x_2 < 3 < \cdots < 70 < x_{70}$. Furthermore, the x_j's are the roots of the polynomial

$$0 = \left[1 - \frac{4}{5}\sum_{k=1}^{70} \frac{k}{x - k}\right]\cdot\prod_{j=1}^{70} (x - j)$$

$$= \prod_{j=1}^{70} (x - j) - \frac{4}{5}\sum_{k=1}^{70} k \prod_{\substack{j=1 \\ j\neq k}}^{70} (x - j)$$

$$= \left[x^{70} - \sum_{j=1}^{70} j\cdot x^{69} + \cdots\right] - \frac{4}{5}\left[\sum_{k=1}^{70} k\cdot x^{69} + \cdots\right]$$

$$= x^{70} - \frac{9}{5}\sum_{j=1}^{70} j\cdot x^{69} + \cdots .$$

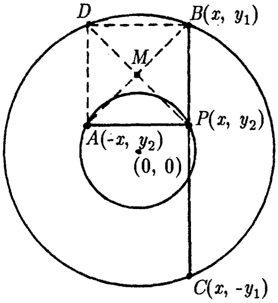
The sum of the interval lengths is

$$\begin{aligned} \sum_{j=1}^{70} (x_j - j) &= \sum_{j=1}^{70} x_j - \sum_{j=1}^{70} j \\ &= \frac{9}{5}\sum_{j=1}^{70} j - \sum_{j=1}^{70} j \\ &= \frac{4}{5}\left(\frac{70\cdot 71}{2}\right) = 1988, \end{aligned}$$

as required.

5. ABC is a triangle right-angled at A, and D is the foot of the altitude from A. The straight line joining the incenters of the triangles ABD, ACD intersects the sides AB, AC at the points K, L respectively. S and T denote the areas of the triangles ABC and AKL respectively. Show that $S \geq 2T$.

Sol. Let $AD = h$ and denote the circle inscribed in $\triangle ABD$ by C_1 and the circle inscribed in $\triangle ADC$ by C_2. Let O_1, O_2 and r_1, r_2 be the centers and radii of C_1 and C_2 respectively, let E and F be the points of contact of C_1 with AB and AD respectively, and let G and M be the points of contact of C_2 with AD and AC respectively. Let N be on AC with $NO_1 \perp AC$, and let P be on NO_1 with $PO_2 \perp NO_1$.



Then $PO_1 = NO_1 - NP = AE - MO_2 = AF - r_2 = h - r_1 - r_2$ and similarly $PO_2 = h - r_1 - r_2$. Therefore

$PO_1 = PO_2$, $\angle O_1O_2P = 45^\circ$, $\angle O_2LM = 45^\circ$,
$AL = AM + ML = AG + MO_2$
$= (h - r_2) + r_2 = h$
and similarly $AK = h$. Hence

$$\frac{S}{T} = \frac{\frac{1}{2}BC\cdot h}{\frac{1}{2}h^2} = \frac{BC}{h} = \frac{BD + DC}{\sqrt{BD\cdot DC}},$$

and by the arithmetic-geometric mean inequality the last expression is at least 2.

6. Let a and b be positive integers such that $ab + 1$ divides $a^2 + b^2$. Show that $\dfrac{a^2 + b^2}{ab + 1}$ is the square of an integer.

Sol. Suppose instead that $\dfrac{a^2 + b^2}{ab + 1} = k$, where k is an integer but not a perfect square, and where $\max\{a,b\}$ is as small as possible. Since $a = b$ would imply that $0 < k = \dfrac{2a^2}{a^2 + 1} < 2$ and hence that $k = 1$, a square, we may assume without loss of generality that $a < b$.

The equation $a^2 + b^2 - k(ab + 1) = 0$ is a quadratic in b for which the sum of the roots is ka and the product of the roots is $a^2 - k$, hence there is a second set of integers (a,b') satisfying this equation, with $b' = ka - b$ and $b'b = a^2 - k$.

Since a and k are positive integers, the hypothesis $b' < 0$ is inconsistent with the equation $a^2 + (b')^2 = k(ab' + 1)$. Since k is not a perfect square, the hypothesis $b' = 0$ is inconsistent with $bb' = a^2 - k$. Therefore $b' > 0$, and the pair (a,b') satisfies all the conditions we had set for (a,b) except possibly $a < b$. However, we now have $b' = \dfrac{a^2 - k}{b} < \dfrac{b^2 - k}{b} < b$, contradicting our requirement that $\max\{a,b\}$ be as small as possible. Therefore our original assumption is false, and the desired result is established.

[Solutions to the Olympiads are taken from a publication of the MAA Committee on the Mathematics Competitions, *Mathematical Olympiads* 1988, where the solutions appear with some additional detail. Copies can be ordered from Dr. Walter Mientka, Department of Mathematics, University of Nebraska, Lincoln, NE 68588-0322. Copies are $1.00 for each year, 1976-1988, with a minimum order of $5.00, payable in US funds.]

A CORRECTION

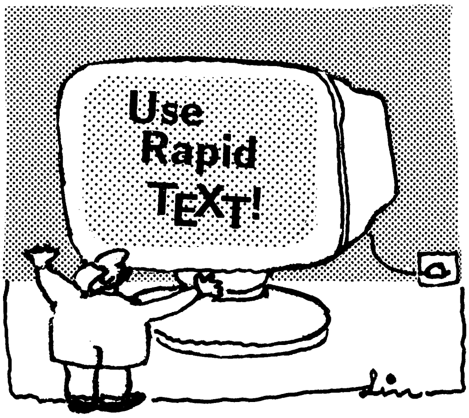
My choice of using a theorem of Miquel (this MAGAZINE, April 1988, pp. 253-259) to demonstrate the finding of a new theorem by means of point and circle duality was unfortunate, as has been pointed out to me by John P. Robertson (CIGNA Property/Casualty Group, Philadelphia). There are pitfalls in using point and circle duality. For example, while a circle can always be drawn through any three non-collinear points, it is certainly not true that three non-concentric circles always have a point in common. Figures 6 and 9 of my note are correct point and circle dual figures and hence they correctly represent 8_36_4 and 6_48_3 point and circle configurations. However, one word needs to be inserted in the final sentence of the statement of the so-called "New Theorem" (p. 259). This sentence should read, "Then the circles C_5, C_2, C_7, C_4 all can pass through a point 6." Actually, almost any point in the plane can be used as point 6 and thus the theorem reduces to a triviality! Can any reader suggest a better choice of theorem for the above mentioned purpose of demonstrating point and circle duality?

Harold L. Dorwart
Lower Cobble Road
Salisbury, CT 06068

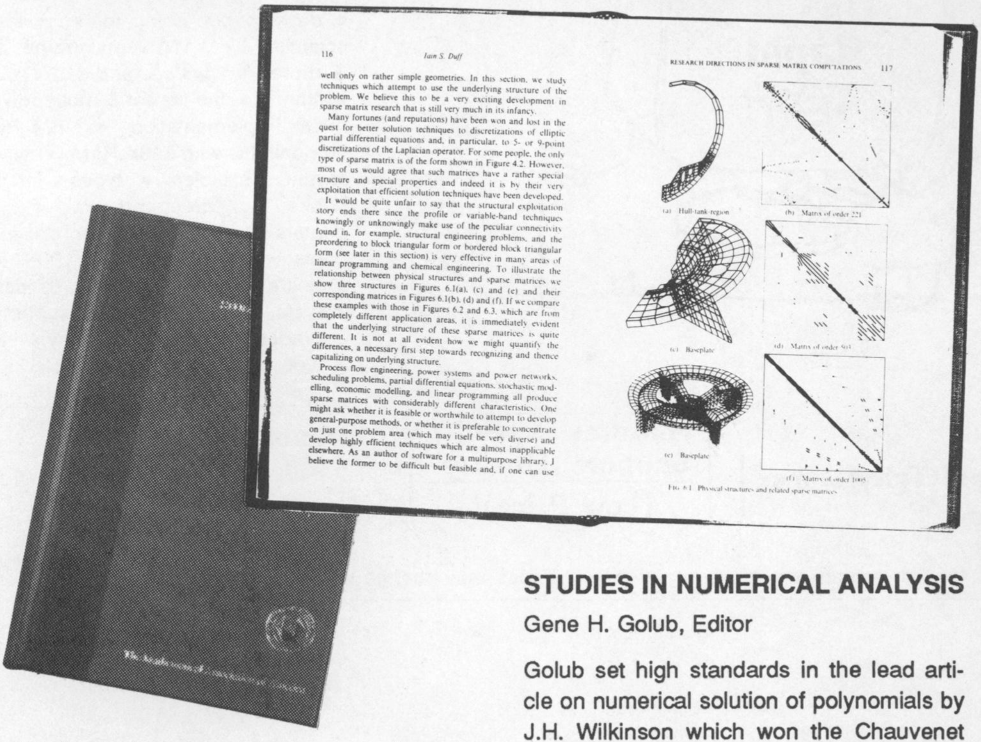
PASCAL'S TRIANGLE YET AGAIN

Peter Hilton and Jean Pedersen have written to point out an alternate method of extending Pascal's triangle to that used by Gloria Olive in "When does the symmetry property hold?", this MAGAZINE 61 (1988), 305-308. They have retained the condition $\binom{a+b}{a} = \binom{a+b}{b}$ while abandoning the condition $\binom{x}{-n} = 0$, n a positive integer. This has led to a number of interesting generalizations that appear in "Extending the binomial coefficients to preserve symmetry and pattern", *Computers Math. Applic.* 17 (1989), 89-102; "Binomial coefficients in the Pascal hexagon", *Koll. Math.-Didaktik Univ. Bayreuth* 14 (1988), 3-24, among others. Readers of the MAGAZINE may wish to look into this alternate definition and compare the advantages of the two generalizations.

—Editor


Use
Rapid
TEXT!

STUDIES IN NUMERICAL ANALYSIS



116 *Iain S. Duff*

well only on rather simple geometries. In this section, we study techniques which attempt to use the underlying structure of the problem. We believe this to be a very exciting development in sparse matrix research that is still very much in its infancy.

Many fortunes (and reputations) have been won and lost in the quest for better solution techniques to discretizations of elliptic partial differential equations and, in particular, to 5- or 9-point discretizations of the Laplacian operator. For some people, the only type of sparse matrix is of the form shown in Figure 4.2. However, most of us would agree that such matrices have a rather special structure and special properties and indeed it is by their very exploitation that efficient solution techniques have been developed.

It would be quite unfair to say that the structural exploitation story ends there since the profile or variable-band techniques knowingly or unknowingly make use of the peculiar connectivity found in, for example, structural engineering problems, and the preordering to block triangular form or bordered block triangular form (see later in this section) is very effective in many areas of linear programming and chemical engineering. To illustrate the relationship between physical structures and sparse matrices we show three structures in Figures 6.1(a), (c) and (e) and their corresponding matrices in Figures 6.1(b), (d) and (f). If we compare these examples with those in Figures 6.2 and 6.3, which are from completely different application areas, it is immediately evident that the underlying structure of these sparse matrices is quite different. It is not at all evident how we might quantify the differences, a necessary first step towards recognizing and thence capitalizing on underlying structure.

Process flow engineering, power systems and power networks, scheduling problems, partial differential equations, stochastic modelling, economic modelling, and linear programming all produce sparse matrices with considerably different characteristics. One might ask whether it is feasible or worthwhile to attempt to develop general-purpose methods, or whether it is preferable to concentrate on just one problem area (which may itself be very diverse) and develop highly efficient techniques which are almost inapplicable elsewhere. As an author of software for a multipurpose library, I believe the former to be difficult but feasible and, if one can use

RESEARCH DIRECTIONS IN SPARSE MATRIX COMPUTATIONS 117

FIG. 6.1 Physical structures and related sparse matrices

STUDIES IN NUMERICAL ANALYSIS

Gene H. Golub, Editor

Golub set high standards in the lead article on numerical solution of polynomials by J.H. Wilkinson which won the Chauvenet Prize as the best piece of mathematical exposition to appear in 1985. These papers describe the wide range of research in numerical analysis. A variety of problems and of computational tools is treated. Some of the computations require only a hand-held calculator, while others require a powerful computer.

422 pp., 1985, ISBN 0-88385-126-1

List: $46.50 MAA Member: $34.50

Catalog Number MAS-24

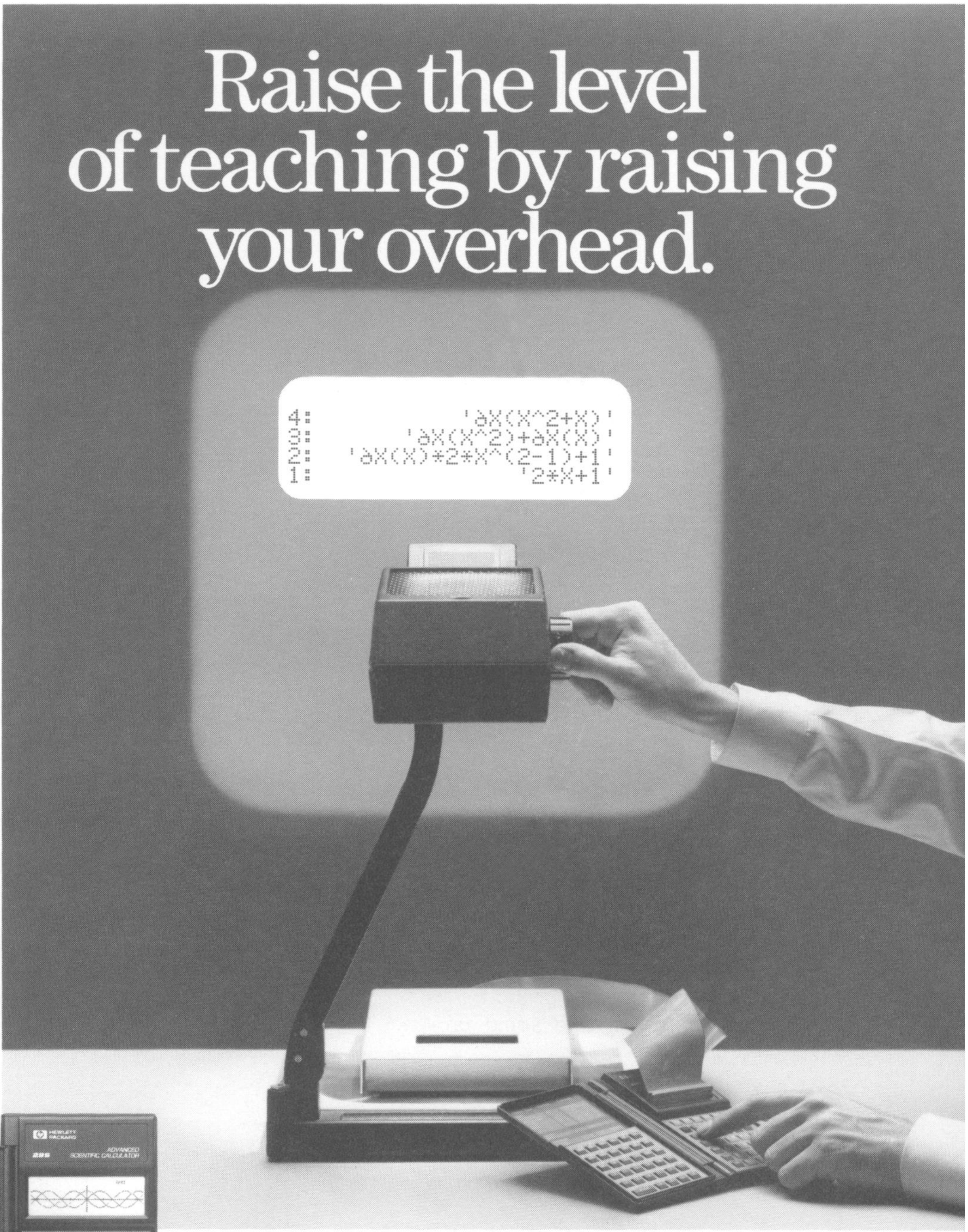
This book is not just for numerical analysts, it is accessible to anyone with some mathematical foundation who wants to taste selections from a numerical analysis menu. The papers do a good job in motivating and introducing their subjects, and some provide a nice historical background. I recommend the articles in this book for anyone interested in learning more about the current research interests in numerical analysis. They may excite their readers to investigate further this relevant field of mathematics.

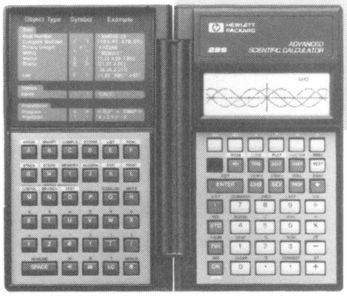
David C. Arney in *Mathematics and Computer Education*

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

Raise the level of teaching by raising your overhead.





Now there's a better way to teach algebra and calculus in the classroom. In fact, two ways.

First, introduce your students to the HP-28S. It's the only calculator that offers symbolic algebra and calculus.

Then, introduce yourself to the overhead display for the HP-28S. It allows you to project your calculations on an HP-28S for everyone in the classroom to see.

A scholastic offer for you.

If your department or students purchase a total of 30 HP-28S calculators, we'll give you a classroom overhead display for the HP-28S absolutely free. (A $500 retail value.) Plus, your very own HP-28S calculator free. (A $235 retail value.)

To learn more, and get free curriculum materials, call (503) 757-2004 between 8am and 3pm, PT. Offer ends October 31, 1989.

There is a better way.

HEWLETT PACKARD

Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington D.C. 20036