MATHEMATICS MAGAZINE

DIOPHANTI ALEXANDRINI ARITHMETICORVM LIBRI SEX,

ET DE NVMERIS MVLTANGVLIS LIBER VNVS.

CVM COMMENTARIIS C. G. BACHETI V. C. & obseruationibus D. P. de FERMAT *Senatoris Tolosani.*

Accessit Doctrinæ Analyticæ inuentum nouum, collectum ex varijs eiusdem D. de FERMAT Epistolis.

TOLOSÆ,
Excudebat BERNARDVS BOSC, è Regione Collegij Societatis Iesu.
M. DC LXX.

Vol. 59 No. 2
April 1986

WHEN MATHEMATICS SAYS NO • TIME SHARING
ABUNDANCY OF INTEGERS • INCIRCLES WITHIN

MATHEMATICS MAGAZINE

THE MATHEMATICAL ASSOCIATION OF AMERICA

Volume 59, Number 2, April 1986
(ISSN— 0025- 570X)

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

COVER: The title page of Fermat's edition of *Diophantus*, Toulouse, 1670.

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.

In addition to articles and notes the *Magazine* solicits proofs without words, mathematical verse, anecdotes, cartoons, and other such material consistent with the level and aims described above. Letters and comments are also welcome.

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, University of Santa Clara, 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

Philip J. Davis is professor of applied mathematics at Brown University, Providence, Rhode Island. He has written numerous papers and books on mathematical research. He is a Chauvenet Award Winner of the Mathematical Association of America. His book, *The Mathematical Experience*, coauthored with Reuben Hersh, won an American Book Award in 1983.

The present article is part of an anthology entitled: *No Way. Essays on the Nature of the Impossible,* edited by Philip J. Davis and David Park, and to be published by W. H. Freeman Co.

This project had its genesis some thirty years ago when Prof. Davis read E. T. Whittaker's book on the philosophy of science, *From Euclid to Eddington*, particularly the section entitled "Postulates of Impotence."

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, five times a year: January, March, May, September, and November.

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 non-member/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. Back issues may be purchased, when in print, from P. and H. Bliss Company, Middletown, CT 06457. 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 Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D. C. 20036.

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

When Mathematics Says No

PHILIP J. DAVIS
Brown University
Providence, RI 02912

"Experience is never limited and it is never complete"—Henry James

Statements of impossibility

When all the impermanencies of the world are considered, when one thinks of vast empires that have fallen, of religious beliefs and customs consigned to the ash-heaps of time, of facts and systems of science patched up as a result of body blows received from pummeling nature; when one sees day-to-day arrangements of life changing rapidly even as we live them, in what quarter are we to find a yearned for permanence? One answer has been—and it has been an answer for a very long time indeed—mathematics. It is asserted that the proven statements of mathematics are true and indubitably so; that they are universal, that their truth is independent of time and of national (or even intergalactic) origin. These are commonly held views; and since they are by no means self-evident, they have naturally been the subject of discussions for rather a long time. Such discussions have, over the years, constituted a good fraction of what is called the philosophy of mathematics. In the opinion of the writer (and of many observers of the mathematical scene) these views are naive and lead to a picture of mathematical activity that is inadequate.

In this article, I shall explore these views from a particular point of view, namely, that of the statements of impossibility that occur in mathematics. Of such statements there is an abundance:

"It is impossible for parallel lines to meet,"

"It is impossible to square the circle,"

"It is impossible that the sum of two even numbers be an odd number,"

"It is impossible to give a proof of the consistency of Zermelo's Axioms,"

and many, many others. We might say that since the phrase 'it is impossible that' is simply the negation of the phrase 'it is the case that,' any statement of mathematics that asserts that something is the case can be converted to an impossibility by denying its denial. Thus, 'two and two is four' converts to 'it is impossible that two and two not be four.' Nonetheless, some statements seem to fit the impossibility format more naturally than others. Examples: children say in subtraction "six from four you cannot take," and mathematicians say that "a transcendental number is one that cannot possibly satisfy a polynomial equation with integer coefficients." Furthermore, the psychological import of a statement that asserts impossibility is different from one that asserts actuality. ("You mean that such and such is impossible? You mean that no matter what I do, no matter how hard I try, I will never succeed in ...?") There seems to be a time element at work in such statements. Actuality is here, actuality is now, it is complete; an impossibility seems to bargain with an uncommitted future.

Consider the following three statements:

(1) It is impossible for two integers p and q to exist such that $p/q=\sqrt{2}$

(2) It is impossible to define the terms in an axiom.

(3) It is impossible to display $10^{1000000000}$ decimal digits of $\sqrt{2}$.

The first statement is an old theorem of mathematics. The second is a statement about mathematics and about the language in which mathematics is written. Since formalized mathematics proceeds by deduction from axioms to theorems, the former are arbitrary starting points and hence indefinable. In the first two statements, the action takes place within the world of the mathematical imagination. In the third statement, if we interpret the words "to display" in some physical sense, the exterior world and judgments about it now play a role. Isolated, self-contained, wholly formalized mathematics exists only as an idealization; common discourse, together with the facts of the real world, constantly intrudes to provide meaning and direction.

The object of this article, then, is to present typical statements of mathematical impossibility and to discuss their epistemological status; that is, to discuss what they are really saying and our reasons for believing that the statements are meaningful and true.

Squaring the circle

I shall begin with what is probably the most familiar of all the impossibility statements of mathematics: "It is impossible to square the circle."

This statement is very old—the playwright Aristophanes (400 B.C.) uses the term "circle squarer" as a term of derision. A circle squarer, metaphorically speaking, is one who persists in trying to do the impossible.

The term 'circle squarer,' much more narrowly construed, refers to one of a group of people, operating on the fringes of mathematical activity, who persist in believing they have discovered how to square the circle despite what mathematicians tell them. The impossible exerts a lure that cannot be matched by the possible.

The meaning of the statement is complex, and I shall begin with a deliberate oversimplification. First, the problem: given a circle, construct a square whose area equals that of the circle. There is a second version of the problem whose statement is simpler: given a circle, construct a square whose perimeter equals the length of the circumference of the circle. The two problems are intimately related. If one is impossible, so is the other and conversely. What follows relates to the second formulation. The impossibility statement, in its simplified form, is this: try as you will, you cannot succeed; it is impossible to square the circle.

If you are hearing this for the first time, you may say, "The impossibility statement is ridiculous. It goes against common sense. Suppose the circle is a beer barrel, a tire or the trunk of a tree. Just draw a rope around the circumference tightly, snip off the length of rope, measure it, divide the length in four equal parts and viola! you have the side of the square whose perimeter equals that of the circle."

Perhaps you are not experimentally inclined, but are arithmetically inclined and you recall some grade school geometry. You might refute the claim of impossibility in this way. "Let us suppose the radius of the circle is one foot. I know from my Euclid that the circumference of the circle is therefore 2π feet. I know from my little hand-held computer that the value of π is 3.1416, to four figures after the point. Therefore, the value $2\pi/4$, which is the side of the required square, is 1.5708 feet. Go build a square whose side is 1.5708 feet and you have squared the circle!

These arguments asserting possibility have brought in new elements: (1) The circle as a physical object, (2) Measurement as a physical act, (3) Construction as a physical act, (4) Approximation within mathematics itself. There is no doubt that we can 'square the circle' in either of the ways just mentioned and our solution would be a good practical solution. But these solutions are open to criticism; while they provide good approximations, they are not exact in the strict mathematical sense of the word.

If our task is to arrive at an ideal mathematical square, residing in the mathematical world, by a sequence of ideal mathematical operations of a certain specified kind, then we must pursue the task totally within the mathematical world. Our experiences with mathematics and physics have led us to two different places: "nature and artifice have collided." [3]

The Greek mathematicians of classical antiquity were rather less interested in going the way of physical measurement than in going the way of pure mathematical theory. It was they who located the problem totally and firmly within the mathematical world, and this placement becomes an essential part of the final assertion of impossibility. But we still have a job to do before the problem becomes well set and leads to an impossibility statement. We must clarify the means by which we are allowed to 'construct' line segments. Geometry is pursued in the real world by both physical and conceptual means. One draws real lines and real curves on a real piece of paper using certain drawing instruments. The physical environment becomes a laboratory in which reasoning, constructions, and discovery can all take place. The favorite drawing instruments of the Greeks were the ruler and the compass. With these instruments, the straight line and the circle can be produced. The Greeks also had instruments for drawing other curves, and discussions of their use occur in advanced material. In view of the simplicity of the ruler and compass, the notion of ruler and compass constructions took on a distinguished status. Such a construction leads to a figure all of whose parts are built up successively by the application only of these two instruments. Euclid gives many such constructions in *The Elements*.

We must now make a mathematical model of the physical act of construction by ruler and compass. Such an act is replaced by a formalized, mathematical surrogate in which is stated clearly what we are allowed to do. Once we have done this, the problem of squaring the circle is located in the realm of abstract mathematics; more precisely, with the aid of the algebraization of geometry initiated by Descartes and of the calculus initiated by Newton and Leibnitz, it is taken away from geometry itself and becomes a problem within the complex of ideas now known as algebra and analysis. To argue about this problem no longer requires that you own a set of drawing instruments, but that you have a profound knowledge of theorems of algebra and of advanced calculus. The problem can now be stated precisely, and the meaning of the impossibility of circle squaring is this: it is a *theorem* of mathematics that no finite number of algebraic operations of such and such a type can lead to the desired result. The truth status of the impossibility of squaring the circle is identical to that of any other theorem of pure mathematics, and its acceptance is on that basis.

As a matter of historical fact, in the days when Aristophanes was laughing at the circle squarers, it was not known whether or not such a construction was possible. Repeated failures led the mathematical community to conjecture that it was not. In the course of these failures many ingenious ruler-and-compass constructions were devised whose accuracy was very high (but not perfect). Proof of the impossibility was not reached until the 1800's. Over the years, a stronger assertion emerged; π is a transcendentental number in the sense just defined, and for a long time this was considered to be *the outstanding unsolved mathematical problem*. Finally, in 1882, the transcendentality of π was established by the German mathematician Ferdinand Lindemann.

Impossibilities within deductive mathematical structures

As we have just seen, to refine and make precise the notion of a mathematical impossibility requires that one confine oneself to a certain limited area of mathematics. Such an area will embrace its own mathematical objects, it will set forth definitions and axioms and, ultimately, will yield a set of true statements (theorems) relating those objects. This limitation is known as "working within a deductive mathematical structure." Historically, the details of the axiomatizations are often laid down *after* a considerable bulk of results has become clear on a more informal basis.

Let me mention a few more interesting impossibilities and the areas of mathematics in which they are now located.

(a) It is impossible for the sum of two even integers to be an odd integer. (Arithmetic of positive integers)

(b) Given the edge of a cube, it is impossible to construct by ruler and compass the edge of another cube whose volume is twice that of the first. (Galois Theory)

(c) It is impossible with ruler-and-compass operations to trisect a general angle. (Galois Theory)

Incidentally, (b), (c), and circle squaring together constitute the three classical impossibilities of Greek geometry.

(d) It is impossible to find a formula that involves only a finite number of arithmetic operations and a finite number of root extractions that solves the general quintic equation (or indeed any equation of degree higher than 4). (Galois Theory)

(e) It is impossible to solve the "Fifteen Puzzle." (Two-dimensional combinatorial geometry)

(This is a popular puzzle made of plastic. Little movable square pieces labelled $1, 2, 3, \ldots, 14, 15$ are placed in rows and columns within a four-by-four frame. One space is left blank. The order of the pieces can be altered by pushing the squares successively into the blank space. The pieces are placed initially in the inverted order $1, 2, 3, \ldots, 15, 14$, and the object of the puzzle is to rearrange them to the usual order $1, 2, 3, \ldots, 14, 15$. The initial order is very important.)

(f) It is impossible that there exist more than five types of regular polyhedra: the tetrahedron, the cube, the octahedron, the dodecahedron, and the icosahedron. (Three-dimensional Euclidean Geometry)

(g) It is impossible for $\sqrt{-1}$ to exist. (Arithmetic of positive and negative real numbers)

(h) It is impossible to set up a one-to-one correspondence between the set of integers and the set of real numbers. (Cantorian Set Theory)

Each of the impossibilities (a)–(h) is a mathematical theorem, and has been proved on the basis of a precise statement of conditions within the indicated mathematical structure. The truth status of each of these impossibilities is precisely the same as that of any proved theorem. It is different from that of the statements of, e.g., physics, and it will pay to look into just what it is.

No probable possible shadow of doubt*

Why are the theorems of mathematics true? There are probably more answers to this question than there are people who have thought deeply about it. Schools of mathematical philosophy have crystallized around the answers. To those of a skeptical mind, no answer may be convincing; taken as a whole, the answers may even be contradictory.

It has been said that mathematics is true because it is God-given. Mathematics is true because man has constructed it. Mathematics is true because it is nothing but logic, and what is logical must be true. Mathematics is true because it is tautological. It is true because it is proved. It is true because it is constructed; its fabric is knit from its axioms as a sweater is knit from a length of yarn. It is true in the way that the rules and subsequent moves of a game are true. It is true because it is beautiful, because it is coherent. Mathematics is true because it is useful. Mathematics is true because it has been elicited in such a way that it reflects accurately the phenomena of the real world.

Mathematics is true by agreement. It is true because we want it to be true, and whenever an offending instance is found, the mathematical community rises up, extirpates that instance and rearranges its thinking. Mathematics is true because, like all knowledge, it is based upon tacit understanding. Mathematics is true because it is an accurate expression of a primal, intuitive knowledge. Mathematics is true because there are numerous independent but supportive avenues to its kind of knowledge which are constantly being reconciled.

It has also been said that mathematics isn't true at all in a rock-bottom sense, it is true only in a probabilistic sense. Mathematics is true only in the sense that it is refutable and corrigible; its truths are eternally provisional. Mathematical truth is not a condition, it is a process. Truth is an idle notion, to mathematics as to all else. Walk away from it with Pilate.

Rattling off this list in rat-a-tat-tat fashion has very likely induced some vertigo in the minds of readers and a feeling that chaos must prevail in this most fundamental question of this most fundamental field. But the chaos is something that only philosophers of mathematics contend with. The majority of mathematicians hardly worry about it at all and often regard philosophical speculation with disdain or amusement. As they pursue their individual researches, inner peace and tranquillity are the rule and not the exception. The reasons for their feelings of security are

*W. S. Gilbert, "The Gondoliers."

hardly ever verbalized, and not the least of their reasons is that it is perfectly legitimate to do what they are doing because so many of their brilliant colleagues are doing it also.

Without in the least hoping to resolve the question of mathematical truth or of expressing adequately my own feelings about it here, I return to the impossibility of squaring the circle and ask: why should one believe Lindemann's Theorem, which implies this impossibility.

Imagine an encounter between a mathematician M and a person P who believes he has found a way to square the circle. Such encounters have occurred so often that this one is not entirely fictitious. (If the reader wants amusing descriptions of encounters with flesh-and-blood circle squarers, Augustus De Morgan's *A Budget of Paradoxes* is highly recommended.) I shall assume that mathematician M is an establishment fellow and that while P has a good opinion of his own work, he is not an utter fanatic.

P comes to M and says "See here, I have found a ruler and compass construction that squares the circle. What do you say to that?"

M's probable reply is "No. You must be mistaken. Your conclusion must be erroneous." M asserts this on the basis of the fact that the world mathematical community has accepted, since Lindemann, the proof of the impossibility. Lindemann's work has passed through the normal processes of validation. M, as an individual, may not have participated personally in this validation, but the record is open, and any time M wants, he may choose to sit in judgement of Lindemann. If M is living in the 1980's the acceptance of the impossibility is simply part of his mathematical inheritance.

Now P's claim contradicts what M and the whole mathematical community believe to be the case. If now M believes further that the portion of mathematics in which the discussion takes place is logically consistent—and he must make this assumption, else the whole enterprise falls to the ground* — then P's construction must be erroneous.

As a matter of psychology, P will probably not be satisfied by M's reply. After all, P has worked hard to arrive at his construction. His work has the same *prima facie* value as the established answer, except, and this is vital, that it has not been through the validation process. P claims that his work should have its moment in court, in the course of which the validity of the established answer should be reexamined. If M is a soft-hearted individual, he will agree.

M should now do several things. M (or the mathematical community) should reexamine Lindemann's work (or later versions of it) critically. This is not likely to occur except in the case of a severe crisis. M should therefore examine P's work critically. He must verify that P is dealing with the same problem as the established problem. (Very often, amateurs do not understand the precise statement of a problem.) Assuming that P understands the problem, M must verify that P's steps are correct in the sense that the logical inferences are correct, that appeals to other 'well established' results are made properly, and that these latter well established results are themselves correct. (It is often the case that amateurs assume what they want to prove, prove it afresh and say: Q.E.D.) Depending upon the complexity of P's work, this process may be exceedingly difficult and time consuming, and in any case, M will hardly want to reopen the validations of all the 'well established' results appealed to by P.

One may write a scenario in which M, after considerable labor, points his finger at such a page and such a line and says to P: "Look. You made an error right here and it is such and such." P, being a decent and reasonable sort, reexamines his own work, agrees with M, and the encounter is terminated. But if P's work is validated and the established work is discredited, then P's work, of course, prevails. If P's work is validated and the established work is revalidated, then we are in a crisis condition.

In point of fact, what M most likely says in a typical encounter is "Go away. My belief is that

*Allow one contradiction and any assertion becomes logically true. For example, suppose that $0 = 1$. Suppose, further, you would like to show that $4 = 11$. Multiply both sides of the first equation by 7. This produces $0 = 7$. Now add 4 to both sides of this. Although the world wouldn't collapse if such a mathematics were introduced, what role could it conceivably play?

(a) Lindemann's work is valid, and (b) mathematical analysis is consistent. Therefore, your work is invalid and I assert it without even examining your work." What is behind this brush-off?

As regards (a), M might reason this way: I personally have not examined Lindemann's work nor any modern formulations of it, but the mathematical community has. This community operates as an open forum in which material is constantly exposed to surveillance, criticism, revision, improvement, extension. This forum has intellectual links to the historical record and to the informal discussions of the past and present. It has links to other fora in allied areas such as in physics and philosophy, and all this constitutes the mathematical experience. The record of this experience is good, and while it does itself point out certain historical errors, ultimately corrected, it claims a very high batting average.

My belief in the validity of Lindemann's work is therefore not absolute, but very strong. It ultimately rests upon my faith in the collective judgement of the mathematical community. The margin of error that I allow to this judgement does not inhibit my personal work, nor does it create neurotic indecision. While my belief in this specific piece of mathematics may be somewhat less than, say, my belief in my own existence, my belief in the mathematical process as a whole is at the same level.

When it comes to (b), the consistency of mathematical analysis, M might reason this way: Experience has shown that it is consistent thus far, and that mathematical analysis has always possessed enough resilience to be able to handle any seeming inconsistencies that have popped up. I am aware of work such as that of Kurt Gödel (validated by the community) that if a mathematical discipline is rich enough (i.e., contains sufficient definitions and axioms) to be able to formulate the problem of its own consistency, then that consistency cannot be demonstrated within the discipline itself. Such a demonstration might be forthcoming within an enveloping-meta-discipline. My belief in the consistency of mathematical analysis rests less with the possibility of such a demonstration than in the fact that inconsistency would rob mathematics of its unique status among intellectual disciplines, a status to which I am committed. I have thus pursued the meaning of impossibility to the point where I have been compelled to formulate my acceptance as a personal commitment. The pursuit must now be terminated.

But there is more to the notion of impossibility than merely a conclusion derived within a deductive structure. For structures can give way and impossibilities can be converted to possibilities.

The theatre of the absurd: conversion of the impossible to the possible

It is an error—or at the very least it contributes to a misleading view of mathematics—if mathematics is seen as a set of static, formal, deductive structures, permanent in arrangement, and fixed for all time. A truer picture of the subject is obtained if these structures are viewed as historic but provisional, emerging as new thinking elicits new rules and delineates their scope, and as new creative pressures alter their individual relevances. The rules may change, hypotheses may change, the order of deduction may be turned upside down to allow hypotheses to become conclusions and vice versa, new interpretations may be found within larger milieus. Consider the following impossibilities, each of which has given rise to a mathematical crisis.

(a) It is impossible for $\sqrt{2}$ to exist. (The crisis of Pythagoras, discussed below.)
(b) It is impossible for $\sqrt{-1}$ to exist. (The crisis of Cardano.)
(c) Given a straight line L in the plane, and a point P not lying on it, it is impossible that there be other than precisely one straight line through P and parallel to L. (The crisis of synthetic a priori geometry, discussed below.)
(d) It is impossible for infinitesimal quantities to exist. (The crisis of Berkeley.)
(e) It is impossible to have a function which is zero for all values of x, except at $x=0$, and

whose integral (area) is positive. (The crisis of Heaviside-Dirac.)*

(f) It is impossible to put all the real numbers into a one-to-one correspondence with the positive integers. (The independence of the continuum hypothesis. The crisis of Galileo—considerably updated.)

In a word now: all of these impossibilities have been converted into active, useful, and logical possibilities.

Impossible, and yet possible. The $\sqrt{2}$ is the first instance of the "absurd" in mathematics. Impossible, absurd, within a rigid axiomatic frame of arithmetic to which mathematics was unable to confine itself and still remain creative. Possible upon relaxation or extension of the frame. The theatre of the absurd has been wildly successful. Though each of the impossibilities (a)–(f), and their subsequent dissolution form large and important chapters in the development of mathematics, I shall only talk about (a) and (c) at any length.** These two have particular interest because the crises they engendered spilled over to the larger nonscientific world.

It is amusing to observe that, in some instances, the way out of a crisis actually emerges far in advance of a proper statement of the impossibility itself. It is impossible working with integers only to divide the integer 1 by the integer 2 or the integer 3 by the integer 2, because the results are not integers. But there is hardly a number system in which, historically, $1/2$, and $3/2$ do not exist simultaneously with one, two, three. Impossibility is a status conferred by axioms.

Axiomatization has fluctuated in importance. At the moment, and since 1800, it is very strong and influential indeed. Arithmetic and analysis (calculus) were hardly axiomatized until last century. Classic Greek geometry was, of course, axiomatized . That is where it all began; that is where "were forged the chains with which human reason was bound for 2300 years." [**2**]

Ancient Indian and Oriental mathematics were innocent of axioms.

The impossibility associated with $\sqrt{2}$ is that of finding two integers p and q such that $p/q=\sqrt{2}$ (i.e., $(p/q)^2=2$). The discovery that this is impossible is attributed to Pythagoras (6th century B.C.). Since $\sqrt{2}$ exists as the palpable length of the diagonal of the unit square, and since it does not exist as the ratio of two integers (what other kind of numbers can there be?), then it both exists and doesn't exist. Here, indeed, is collision between nature and artifice.

A way out was hinted at by the Babylonians. In a tablet that has been dated approximately 1700 B.C. (see [**1**]), one finds an excellent sexigesimal approximation to $\sqrt{2}$. Greek reasoning led to a crisis in Greek mathematical thought. Ultimately, after many centuries, the dilemma was resolved when geometry was relocated within the real number system, i.e., the system of all "decimals" with infinitely many digits, a process already begun in the Babylonian tablet. The impossibility of $\sqrt{2}$ as the ratio of two integers becomes a dead issue in the reformulated geometry, but remains a very live issue, though nonparadoxical, within the theory of numbers. The crisis of Pythagoras had implications that went far deeper than mathematics itself. Coming at the time of, or just prior to, the development of axiomatic deduction, it is not inconceivable that this discovery was instrumental in the promulgation of deduction as a device for arriving at truth. Sir Karl Popper [**4**] speculates that the crisis and its residue went far deeper than mathematics: "(The Platonic Theory of forms and ideas) cannot be understood except in the context of the critical problem situation in Greek science (mainly in the theory of matter) which developed as a result of the discovery of the irrationality of the square root of two." One wonders what the subsequent history of Western thought would have been if Plato had not dreamed up ideal forms to keep us enthralled and in thrall.

Yesterday's shocks are often the bases of today's stabilities. Every undergraduate in mathematics and physics knows that there are three principal systems of continuous geometry. The most

---

*A crisis—of sorts—in the mathematical community that physicists in their naiveté hardly noticed. Physicists wisely reserve their energies for crises in physics.

**The transformation of (f) to a possibility is less familiar, even to a professional. It is embodied in the Lowenheim-Skolem Theorem of set theory. See, e.g., *Set Theory*, by Thomas Jech, p. 81.

famous is Euclidean geometry. Then, there are two kinds of non-Euclidean geometry: the hyperbolic geometry of Bolyai and Lobatchevsky and the elliptic geometry of Riemann. It is now hard to realize that in the early 19th century the possibility of a geometry alternate to that of Euclid was approached with fear and indignation. Undoubtedly the authority of the Ancients was at stake, as were ancient views of the relationship between the world and a mathematical description of the world. Euclid's geometry was held to provide an accurate description of the way the spatial world is. It had a priori truth, and this being the case, an alternate geometry was unthinkable, impossible.

The matter may now be viewed as a conflict between axiomatics and experience. Geometry, in the hands of Euclid, was built up from certain common notions and "self-evident" axioms. His Fifth Axiom, the famous parallel axiom, asserts that if a straight line is given in a plane, and if a point not on the line is given, then one and only one line may be drawn through the point parallel to the given line (proposition (c), above). Of course, this may be rephrased dramatically as an impossibility: it is impossible that there be other than one parallel through a given point.

To subsequent generations, this axiom seemed rather less transparent than the others, and there is a long history of attempts to deduce it from them. These attempts failed, and when something fails over a period of time, common sense, if you can call it that, would seem to say that what you're trying to do is impossible to do. At any rate, common sense of one sort was in conflict with that of another sort, and out of the conflict emerged the impossibility of the deduction.

How does one prove impossibility in this case? It is done by finding a set of mathematical objects, together with an interpretation that links the objects to the terms appearing in the axioms, that satisfy all the axioms of Euclid other than the fifth. The objects may be part of conventional Euclidean geometry, but the basic terms such as 'point,' 'line,' 'intersect,' 'parallel,' etc., are given special meanings. This process, which has since become commonplace in mathematical logic and set theory, is known as finding a "model" for the axioms. People have found a number of models —in fact, infinitely many—none of which satisfy Axiom 5. In this way, its independence is established. The possibility of more than one parallel is established by the geometry of Lobatchevsky and Bolyai and of no parallels by the geometry of Riemann.

Will now the real geometry please stand up? None can answer this request, for in the wake of the new discoveries, there is now no logically privileged geometry. Each non-Euclidean geometry is as logical and as consistent as Euclidean geometry. What remains is not the essential truth of geometry in some mystic sense, but questions of utility, simplicity and convenience. What geometry is useful? In the small arena of terrestrial and planetary measurements, Euclidean geometry is the shoe that fits the foot; over the vast reaches of intergalactic space-time, Riemannian geometry appears to be what is wanted.

When bells ring and lights flash

It is not uncommon for today's lecturer in mathematics to begin a lecture by saying something like: "Let X be a function space of such and such a kind and let f be a typical function in X." Later on, the lecturer might say, "Since f is *living in X* such and such is true." This current colloquialism underscores the formal limitations to which f has been subjected and because of which certain consequences flow. It also underscores the fact that f is prior to X, that X may not be the natural habitat for f; indeed, there may be no natural habitat, and residency in a particular spot may carry a price.

This is familiar from the larger world. We live in a room, the room is in the house, the house is on a lot, etc. What is impossible to do in a room may be quite possible in a house, what is impossible in the house may be quite possible in the backyard. Of course, the reverse may be equally true in that structures may evolve precisely to bring about desired impossibilities, e.g., that it be impossible for rain to fall on our heads or for thieves to lift our videorecorders.

Now one of the major contemporary tendencies in mathematics is to view it as the science of deductive structures. This view has reinforced analogous views in anthropology, psychology, education, literary theory. As mentioned above, a structure is a set of objects together with a set of rules that govern the combination or interaction of the objects. For the most part, the objects in their primitive conceptualization, predate the formation of the rigorous structure, and a given object may reside in many, many structures. Thus the mathematical object 2 may be an element of the set of integers, of a group, a ring, a field, etc. The operation of squaring a number (the function $v = x^2$) may be an element of many different sets of functions.

Structures are something like social clubs whose laws of admission are set up on the basis of exclusion principles. Thus, the infinite sequence of ones: $(1,1,\dots)$ is excluded from membership in the Hilbert Space called l^2 because the sum of its squared components is infinite. The elements that are allowed in the structure l^2 are "good," or at least in Dr. Johnson's terminology, eminently "clubbable." The purpose of club formation is to insure that members may interact properly. Within the club, some things are possible for members and others impossible. The structural emphasis of 20th century mathematics has taken the limelight off the individual members (numbers, vectors, functions, permutations, etc.) and placed it on the structure itself as an entity within which one looks for such features as closure, substructures, homomorphisms, etc.

Insofar as mathematical objects have an existence apart from the structure in which they are temporarily living,* questions may be asked of them in their wider, unstructured character. The answers may be various. Are there any members, x, of my structure for which $x^2 = 2$? No, if all the x are residing in the set of positive integers. Yes, if the x's comprise the real numbers. Does the equation $x^2 = -1$ have a solution? No, if x is a real number; yes, if x is a complex number.

Do you really want to solve the 15 puzzle? Then act like Alexander cutting the Gordian knot: lift the little squares out of their frame into the third dimension and rearrange them properly. This is what you would do if your life depended upon getting the desired arrangement.

Impossibilities are converted to possibilities by changing the structural background, by altering the context, by embedding the context in a wider context. It is very likely the case that all mathematical impossibilities may be altered in a nontrivial way so as to become possibilities. With regard to such conversion, a number of very important questions can be raised. (a) Why does one want to do it? What are the internal or external pressures for it? Surely one would require a better answer than merely that the stuff of mathematics is sufficiently pliable to allow it. (b) What are the larger consequences of the act? While there are many instances of conversion of which, perhaps, examples (a)–(f) of the last section are the most famous, to my knowledge there has been no general critique of the act.

When bells ring and lights flash, we are duly warned. Presumably, we are skirting on the edge of what is dangerous. When we shove a piece of plastic into the bank slot and ask for a hundred dollars at a time when our balance is seventy, mathematics says no and an error message is returned. When we take our scientific computer in hand, read in 1 and attempt to divide it by 0, those flashing lights remind us of our impossible, desperate, meaningless act.

Yet, tomorrow, the bank manager might have his machine programmed to allow everybody a line of credit of one thousand dollars. Washington does it all the time. It's called deficit financing, and giants of economic theory are associated with this structural change.

Do you want to divide by zero? Well, maybe not $1/0$, but $0/0$. This absurd ratio, when properly interpreted as the limit of legitimate ratios, is the very stuff out of which differential calculus is built. But why have we excluded $1/0$? Sorry, an oversight. There was no need to exclude it, for within projective geometry or modern matrix theory, answers are given routinely, usefully, and without shame. (*Every* rectangular matrix has a Moore-Penrose generalized inverse.)

When bells ring and lights flash, you may, indeed, want to pay attention. Then again, you may only want to look at the mechanism that has set them ringing.

*Up to a point, of course. The number 2, as a mere symbol, can serve only to distinguish, as with a hat check. Mathematics begins when 2 is related somehow to 1 and 3. This would be a very primitive structure.

A summing up

When placed within abstract deductive mathematical structures, impossibility statements are simply definitions, axioms, or theorems, as the case may be. If theorems, then assessments of their validity are arrived at by the standard processes and criteria of mathematical validation. These include an examination of the intuitive, deductive, and interactive bases of the theorems.

The history of mathematics displays a long and vitally important record of impossibilities being broken by instituting structural changes.

Meaning in mathematics derives not from naked symbols but from the relationship between these symbols and the exterior world. This relationship is established through the mediation of the mathematical community. Insofar as structures are added to primitive ideas to make them precise, flexibility is lost in the process.

In a number of ways, then, the closer one comes to an assertion of an absolute "no," the less is the meaning that can be assigned to this "no."

References

[1] Asger Aaboe, Episodes from the Early History of Mathematics, Random House, New York, 1964, pp. 26–27.
[2] Eric Temple Bell, The Search for Truth, Reynal, New York, 1934.
[3] Stuart Hampshire, Modern Writers and Other Essays, Knopf, New York, 1970.
[4] Karl Popper, Conjectures and Refutations, Routledge and Kegan Paul, London, 1963, p. 75.

This article will appear in an anthology entitled *No Way: Essays on the Nature of the Impossible*, edited by P. J. Davis and D. Park, to be published by W. H. Freeman and Co.

Fermat's Last Theorem

Mssr. Fermat—what have you done?
Your simple conjecture has everyone
Churning out proofs,
Which are nothing but goofs!
Could it be that your statement's an erudite spoof?
A marginal hoax
That you've played on us folks?
But then you're really not known for your practical jokes.
Or is it then true
That you knew what to do
When n was an integer greater than two?
Oh then why can't we find
That same proof... are we blind?
You must be reproved, for I'm losing my mind.

—Jonathan P. Dowling
1600 Hillside Road
Boulder, CO 80302

A summing up

When placed within abstract deductive mathematical structures, impossibility statements are simply definitions, axioms, or theorems, as the case may be. If theorems, then assessments of their validity are arrived at by the standard processes and criteria of mathematical validation. These include an examination of the intuitive, deductive, and interactive bases of the theorems.

The history of mathematics displays a long and vitally important record of impossibilities being broken by instituting structural changes.

Meaning in mathematics derives not from naked symbols but from the relationship between these symbols and the exterior world. This relationship is established through the mediation of the mathematical community. Insofar as structures are added to primitive ideas to make them precise, flexibility is lost in the process.

In a number of ways, then, the closer one comes to an assertion of an absolute "no," the less is the meaning that can be assigned to this "no."

References

[1] Asger Aaboe, Episodes from the Early History of Mathematics, Random House, New York, 1964, pp. 26–27.
[2] Eric Temple Bell, The Search for Truth, Reynal, New York, 1934.
[3] Stuart Hampshire, Modern Writers and Other Essays, Knopf, New York, 1970.
[4] Karl Popper, Conjectures and Refutations, Routledge and Kegan Paul, London, 1963, p. 75.

This article will appear in an anthology entitled *No Way: Essays on the Nature of the Impossible*, edited by P. J. Davis and D. Park, to be published by W. H. Freeman and Co.

Fermat's Last Theorem

Mssr. Fermat—what have you done?
Your simple conjecture has everyone
Churning out proofs,
Which are nothing but goofs!
Could it be that your statement's an erudite spoof?
A marginal hoax
That you've played on us folks?
But then you're really not known for your practical jokes.
Or is it then true
That you knew what to do
When n was an integer greater than two?
Oh then why can't we find
That same proof... are we blind?
You must be reproved, for I'm losing my mind.

—JONATHAN P. DOWLING
1600 Hillside Road
Boulder, CO 80302

Incircles Within

HÜSEYIN DEMIR
Middle East Technical University
Ankara, Turkey

The study of "remarkable" elements related to the triangle such as medians, altitudes, and angle bisectors has attracted people of diverse interests and intellectual stature, culminating in a human venture of extraordinary aesthetic content. Quite a number of results in this field are in the form of inequalities, of which there are collections [**1**] and systematic studies [**3**]. All these, however, seem to concern the individual triangle only. In this note we examine a triangle ABC and introduce some inequalities and related results about subtriangles of ABC and their incircles. Perhaps the reader will find here a new vein to explore.

Consider, as in FIGURE 1, a triangle ABC with P an interior or a boundary point, and let ρ_a, ρ_b, ρ_c be the inradii of the triangles PBC, PCA, PAB, respectively. If a, b, c and r are the sides and the inradius of ABC, we expect to have a relation between $\rho_a + \rho_b + \rho_c$ and a, b, c, r. Indeed, the following holds.

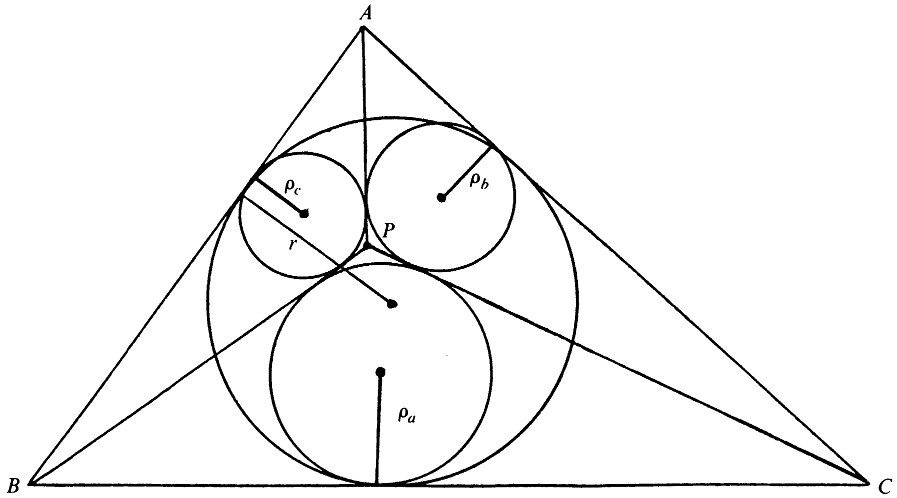


FIGURE 1

THEOREM 1. *For the elements defined above,*

$$r \leqslant \rho_a + \rho_b + \rho_c < \min(a, b, c). \tag{1}$$

For the proof of (1) we use the following result.

LEMMA. *Let P_a be a point on the side BC of triangle ABC and let r, r_2, r_3 be the inradii of ABC, P_aCA, and P_aBA, respectively. If h_a is the altitude of ABC from A, then*

$$r \leqslant r_2 + r_3 < h_a, \tag{2}$$

and equality is achieved on the left-hand side if P_a coincides with B or C.

Proof. Let I, I_2, I_3 be the incenters of triangles ABC, P_aCA, and P_aBA, respectively (see FIGURE 2). Let the lines through P_a drawn parallel to BI and to CI meet the segments CI and BI at I_2', I_3', respectively. Since P_aI_3 is the bisector of the external angle AP_aB of triangle P_aCA, and

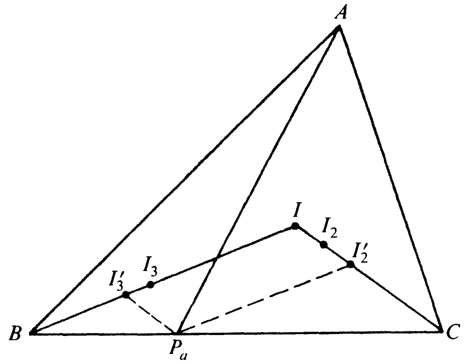


FIGURE 2

P_aI_3' is parallel to CI, the point I_3' lies between B and I_3. Similarly, I_2' lies between C and I_2. Denoting the distances from BC to I, I_i, I_i' by r, r_i, r_i' $(i=2,3)$, respectively, and considering the parallelogram $P_aI_2'II_3'$, we have

$$r = r_2' + r_3' \leqslant r_2 + r_3,$$

which proves the left-hand side of (2).

On the other hand, using the fact that any altitude of a triangle exceeds the diameter of its incircle, we obtain

$$r_2 + r_3 = \tfrac{1}{2}(2r_2 + 2r_3) < \tfrac{1}{2}(h_a + h_a) = h_a,$$

which implies the right-hand side of (2).

We can now prove Theorem 1. If P is an interior point of ABC, extend AP to intersect BC in P_a; otherwise, assume P is on BC and $P = P_a$ in what follows. Let J_2 and J_3 be the incenters of PCA and PBA (see FIGURE 3). Let the lines through P drawn parallel to P_aI_2, P_aI_3 meet the lines AI_2, AI_3 at J_2', J_3'. By the argument employed in the proof of the Lemma, J_i' is between A and J_i $(i=2,3)$. Denoting the distances of I_i and J_i' to AP_a by r_i and ρ_i' $(i=2,3)$ and setting $u = PA$, $v = PP_a$, we have

$$\frac{\rho_2'}{r_2} = \frac{AJ_2'}{AI_2} = \frac{AP}{AP_a} = \frac{u}{u+v} = \frac{\rho_3'}{r_3},$$

which yields

$$\rho_2' + \rho_3' = \frac{u}{u+v}(r_2 + r_3).$$

Since ρ_b, ρ_c are the distances of J_2, J_3 to PA, we have

$$\rho_b \geqslant \rho_2' \quad \text{and} \quad \rho_c \geqslant \rho_3',$$

hence

$$\rho_b + \rho_c \geqslant \rho_2' + \rho_3' = \frac{u}{u+v}(r_2 + r_3), \tag{3}$$

so from the Lemma we conclude

$$\rho_b + \rho_c \geqslant \frac{u}{u+v} r.$$

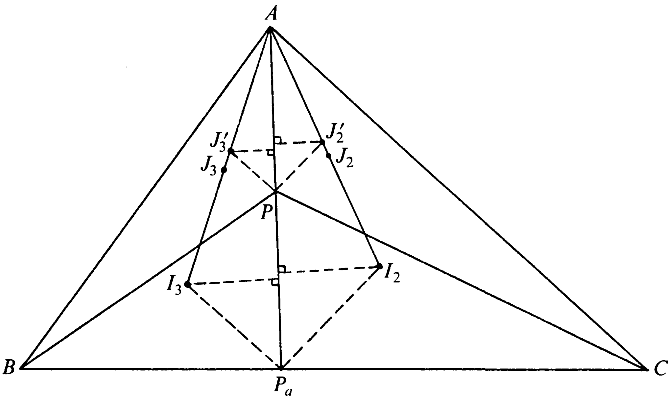


FIGURE 3

Now denote the areas and semiperimeters of the triangles ABC, PBC by Δ, Δ_a and s, s_a respectively. Then $\Delta = rs$ and $\Delta_a = \rho_a s_a$, so

$$\begin{aligned}\rho_a &= \frac{\Delta_a}{s_a} = \frac{\Delta_a}{\Delta} \cdot \frac{\Delta}{s_a} = \frac{\Delta_a}{\Delta} \cdot \frac{s}{s_a} r \\ &= \frac{v}{u+v} \cdot \frac{a+b+c}{a+PB+PC} r \\ &\geqslant \frac{v}{u+v} \cdot 1 \cdot r = \frac{v}{u+v} r,\end{aligned}$$

which, together with (3), implies

$$\rho_a + \rho_b + \rho_c \geqslant r.$$

As for the right-hand side in (1), let the inradii of triangles P_aCP, P_aBP be t_2, t_3 and let h_2', h_3' be the altitudes of P_aCP, P_aBP from C, B. Apply the Lemma to each of the triangles PBC, ACP_a, and ABP_a in turn to obtain the three inequalities

$$\begin{aligned}\rho_a &\leqslant t_2 + t_3, \\ t_2 + \rho_b &< h_2', \\ t_3 + \rho_c &< h_3'.\end{aligned}$$

These together imply

$$\rho_a + \rho_b + \rho_c < h_2' + h_3' \leqslant a.$$

Since the left-hand side of this inequality is symmetric in a, b, c, we obtain

$$\rho_a + \rho_b + \rho_c < \min(a, b, c),$$

which completes the proof of Theorem 1.

It would be interesting to obtain similar inequalities involving other symmetric functions of ρ_a, ρ_b, ρ_c.

Notice that equality can be achieved on the left-hand side of (1) should P coincide with one of the vertices of ABC. The right-hand side of (1) cannot be improved by multiplying $\min(a, b, c)$ by a constant. (Consider very thin triangles with very thin subtriangles.)

Although the Lemma gives an inequality which relates the inradii of a triangle and its two subtriangles to an altitude of the triangle, the following result relates these quantities in an equation. In addition, several nice constructions are made possible. (The reader unfamiliar with triangle identities used in the proof may consult [**2**, chap. 1].)

THEOREM 2. *Let P be a point on the side BC of a triangle ABC and let r, r_2, r_3 be the inradii of the triangles ABC, ACP, ABP, respectively. Then*

$$\frac{1}{r_2}+\frac{1}{r_3}-\frac{r}{r_2 r_3}=\frac{2}{h_a}, \tag{4}$$

where h_a is the altitude of ABC from A.

Proof. Let a, b, c be the sides, 2α, 2β, 2γ the angles, s the semiperimeter, R the circumradius, and Δ the area of triangle ABC (see FIGURE 4). Let the incircles of ACP and ABP

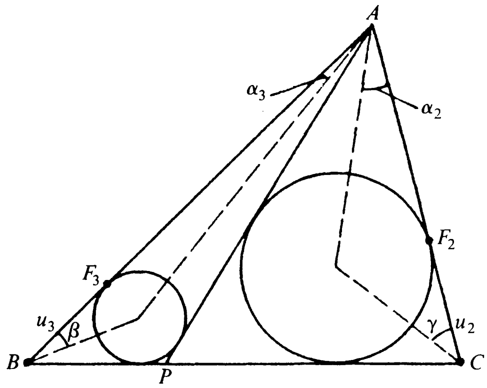


FIGURE 4

touch the sides AC and AB at F_2 and F_3, respectively, and set $u_2 = CF_2$, $u_3 = BF_3$. If the angles CAP and BAP are $2\alpha_2$ and $2\alpha_3$, we have

$$\cot\alpha=\cot(\alpha_2+\alpha_3)=\frac{\cot\alpha_2\cot\alpha_3-1}{\cot\alpha_2+\cot\alpha_3},$$

or, equivalently,

$$\cot\alpha(\cot\alpha_2+\cot\alpha_3)=\cot\alpha_2\cot\alpha_3-1. \tag{5}$$

Since

$$\cot\alpha_2=\frac{b-u_2}{r_2}=\frac{b}{r_2}-\cot\gamma,$$

$$\cot\alpha_3=\frac{c-u_3}{r_3}=\frac{c}{r_3}-\cot\beta,$$

substitution in (5) gives

$$\cot\alpha\left(\frac{b}{r_2}+\frac{c}{r_3}-\cot\gamma-\cot\beta\right)=\left(\frac{b}{r_2}-\cot\gamma\right)\left(\frac{c}{r_3}-\cot\beta\right)-1.$$

Regrouping, we obtain

$$\frac{b}{r_2}(\cot\alpha+\cot\beta)+\frac{c}{r_3}(\cot\alpha+\cot\gamma)-\frac{bc}{r_2r_3}=\cot\alpha\cot\beta+\cot\alpha\cot\gamma+\cot\gamma\cot\beta-1. \quad (6)$$

Using cyclic sum notation, we can write the right side of (6) as $\sum\cot\beta\cot\gamma-1$. Since $r\cot\alpha+r\cot\gamma=b$ and $r\cot\alpha+r\cot\beta=c$, we find

$$\begin{aligned}\cot\alpha+\cot\gamma&=\frac{b}{r}\\ \cot\alpha+\cot\beta&=\frac{c}{r}.\end{aligned} \quad (7)$$

Furthermore,

$$\begin{aligned}\sum\cot\beta\cot\gamma&=\sum\frac{(s-b)\cdot(s-c)}{r\cdot r}\\ &=\frac{1}{r^2}\sum\left[s^2-(b+c)s+bc\right]\\ &=\frac{1}{r^2}\left[-s^2+\sum bc\right].\end{aligned} \quad (8)$$

Now square each side of Heron's formula

$$sr=\sqrt{s(s-a)(s-b)(s-c)}$$

to get

$$\begin{aligned}sr^2&=(s-a)(s-b)(s-c)\\ &=s^3-2s\cdot s^2+s\cdot\sum bc-abc\end{aligned}$$

and solve for $\sum bc$:

$$\begin{aligned}\sum bc&=s^2+r^2+\frac{abc}{s}=s^2+r^2+\frac{2\Delta\cdot 2R}{s}\\ &=s^2+r^2+4rR.\end{aligned} \quad (9)$$

Substituting (9) into (8) gives

$$\sum\cot\beta\cot\gamma=1+\frac{4R}{r}. \quad (10)$$

Substituting (7) and (10) in (6), we obtain

$$\frac{bc}{r_2r}+\frac{bc}{r_3r}-\frac{bc}{r_2r_3}=\frac{4R}{r}. \quad (11)$$

Dividing (11) by bc/r, then using triangle relations gives the final result:

$$\begin{aligned}\frac{1}{r_2}+\frac{1}{r_3}-\frac{r}{r_2r_3}&=\frac{4R}{bc}=\frac{4aR}{abc}\\ &=\frac{4aR}{4\Delta R}\\ &=\frac{a}{\Delta}=\frac{2}{h_a}.\end{aligned}$$

Equation (4) provides us with a solution of the following construction problem.

PROBLEM. Construct, using only straightedge and compass, the cevian AP of a triangle ABC such that the triangles ABP and ACP have congruent incircles.

Setting $r_2 = r_3$ in (4), we obtain

$$2r_2^2 - 2h_a r_2 + rh_a = 0$$

the solution of which gives

$$r_2 = \tfrac{1}{2}\left(h_a - \sqrt{h_a(h_a - 2r)} \right).$$

The desired construction is as follows (see FIGURE 5).

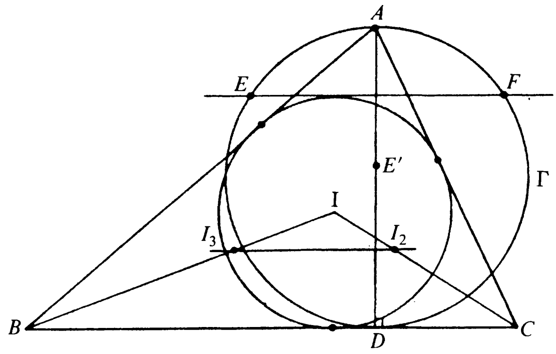


FIGURE 5

1. Draw the circle Γ, admitting the altitude AD as diameter.
2. Draw the tangent line to the incircle of ABC parallel to BC, meeting Γ at E and F.
3. Lay off on AD the segment AE' congruent to AE. (Then $DE' = 2r_2 = 2r_3$.)
4. The perpendicular bisector of DE' intersects the bisectors of the angles C, B at the centers I_2, I_3 of the congruent incircles.
5. Draw the common tangent line through A to these circles; this line intersects BC in the desired point P.

An unexpected result that we can derive from (4) is one that may be called the Five-Circle Theorem:

THEOREM 3. *Let P and Q be two points on the side BC of a triangle in the order B, P, Q, C. If the triangles ABP, APQ, AQC have congruent incircles, then the triangles ABQ, APC have congruent incircles.*

Proof. Let ρ be the common radius of the three congruent incircles and ρ_2, ρ_3 the inradii of the triangles ABQ, APC (see FIGURE 6). Applying (4) to ABQ and APC we have

$$\frac{2}{\rho} - \frac{\rho_1}{\rho^2} = \frac{2}{h_a}, \qquad \frac{2}{\rho} - \frac{\rho_2}{\rho^2} = \frac{2}{h_a},$$

which implies $\rho_1 = \rho_2$.

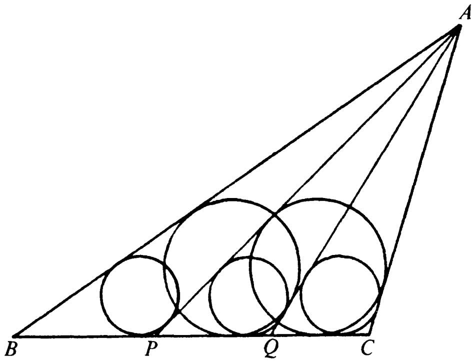


FIGURE 6

Readers are invited to obtain a geometric proof of the Five-Circle Theorem.

References

[1] O. Bottema, et al., Geometric Inequalities, Wolters-Noordhof Publishing, Groningen, 1969.
[2] H. S. M. Coxeter, Introduction to Geometry, 2nd ed., John Wiley & Sons, 1969.
[3] N. D. Kazarinoff, Geometric Inequalities, MAA New Mathematical Library, no. 4, 1961.

Measuring the Abundancy of Integers

RICHARD LAATSCH
Miami University
Oxford, OH 45056

The origins of the study of perfect numbers are lost in antiquity, but the concept was clearly recognized well over 2000 years ago and involves such contributors as Euclid, Fermat, Descartes, Mersenne, Legendre, and Euler. The classification of positive integers as perfect, abundant, or deficient is also an ancient idea, one which dates from before A.D. 100. It is the goal of this note to make explicit an idea used by researchers since the seventeenth century (and perhaps before) and reproduce, comment on, and extend their ideas. For the newcomer to these concepts, the basic definitions are as follows: A **perfect number** is a positive integer which is equal to the sum of its proper divisors ($1 + 2 + 3 = 6$; $1 + 2 + 4 + 7 + 14 = 28$), an **abundant number**, is one for which the sum of the proper divisors is greater than the number ($1 + 2 + 3 + 4 + 6 > 12$), and a **deficient number** has the sum of its proper divisors less than the number ($1 + 2 + 5 < 10$).

As the theory of this classification of numbers developed, investigators realized that it was computationally easier to make use of $\sigma(N)$, the sum of *all* the divisors of an integer N, including N itself. Thus a number N is abundant if $\sigma(N) > 2N$, perfect if $\sigma(N) = 2N$, and deficient if $\sigma(N) < 2N$.

Certainly the prime 41 is very deficient since the only divisors of 41 are 1 and 41, and $1 + 41$ is much less than $2 \cdot 41$, whereas 8 is, relatively speaking, just barely deficient since $1 + 2 + 4 + 8$ is close to $2 \cdot 8$. On the other hand, 360 is very abundant since the sum of its divisors, 1170, is greater than $3 \cdot 360$. By such examples we are led to a natural measure of the abundancy or deficiency of numbers.

DEFINITION. The **abundancy index** of a positive integer N is the number

$$I(N) = \frac{\sigma(N)}{N}.$$

For example, $I(41) = 42/41 = 1.024^+$, $I(8) = 15/8 = 1.875$, $I(360) = 1170/360 = 3.25$, and $I(6) = 12/6 = 2$. Thus N is perfect if $I(N) = 2$, deficient if $I(N) < 2$, and abundant if $I(N) > 2$. When $I(N)$ is an integer $r > 2$, N is said to be **multiperfect of index** r. For example, 120 is multiperfect of index 3 since $I(120) = 3$ (the smallest such example); also $I(30240) = 4$. The study of multiperfects was pursued by Descartes and others and blossomed around 1900 [7]; an excellent bibliography can be found in [**12**]. We shall return to the topic later, but first we establish a couple of easy theorems about the abundancy index.

THEOREM 1. *If $k > 1$, then $I(kN) > I(N)$.*

Proof. If $1, a_1, a_2, \ldots, a_t, N$ are the divisors of N, then $1, k, ka_1, ka_2, \ldots, ka_t, kN$ is a (perhaps incomplete) list of divisors of kN. Thus,

$$\begin{aligned} I(kN) &\geqslant (1 + k + ka_1 + ka_2 + \cdots + kN)/kN \\ &= 1/kN + (1 + a_1 + a_2 + \cdots + N)/N \\ &= 1/kN + I(N) > I(N). \end{aligned}$$

This theorem captures the 13th century assertions that every multiple of a perfect number is abundant and every divisor of a perfect number is deficient. It also assures us that there are infinitely many abundant numbers (all multiples of 6 are abundant) just as there are infinitely many deficient numbers (including all of the primes). What the theorem doesn't tell us is which multiples of a given N will have the larger abundancy indexes. When $N = 8$,

$$I(5 \cdot 8) > I(11 \cdot 8) > I(2 \cdot 8) > I(8).$$

We will soon make several observations about the way in which $I(N)$ grows.

The existence of arbitrarily large primes guarantees that the abundancy index has 1 as its lower bound, since

$$I(P) = \frac{P+1}{P} = 1 + \frac{1}{P}$$

for primes P. A little less obvious is the fact that the abundancy index takes on very large values.

THEOREM 2. *The function I is not bounded above.*

Proof. Let N be an integer divisible by all of the integers $1, 2, \ldots, k$. (One such choice is $N = k!$, but smaller choices are possible if $k \geqslant 4$.) Then

$$\sigma(N) \geqslant N + N/2 + N/3 + \cdots + N/k$$

and so

$$I(N) = \sigma(N)/N \geqslant 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{k}.$$

Since this is a partial sum of the divergent harmonic series, $I(N)$ can be made as large as desired.

This proof, although constructive, leads quickly to very, very large values of N. For example, $\sum_{i=1}^{n} 1/i$ first exceeds 4 when $n = 31$ and the smallest integer divisible by all positive integers less than or equal to 31 is about $7 \cdot 10^{13}$. However, the least integer N with $I(N) \geqslant 4$ is 27720.

We have now established that the numbers $I(N)$, all of which are rational numbers, are scattered throughout the interval $(1, \infty)$ of the real line. Before discussing how they are distributed, let us write an explicit formula for $I(N)$.

It is noted in every elementary book on number theory (see, e.g., [**8**]) that, if P is prime, then the sum of the divisors of P^n is the sum of a geometric progression

$$\sigma(P^n) = 1 + P + P^2 + \cdots + P^n = \frac{P^{n+1} - 1}{P - 1}, \tag{1}$$

and that, if $N = P^r \cdot Q^s \cdot \cdots \cdot V^t$ with $P, Q, \ldots, V$ distinct primes, then

$$\sigma(N) = \sigma(P^r) \cdot \sigma(Q^s) \cdot \cdots \cdot \sigma(V^t). \tag{2}$$

Combining (1) and (2), we obtain formula (3) below for the abundancy index $I(N)$.

THEOREM 3. *If $N = P^r \cdot Q^s \cdot \cdots \cdot V^t$ is the prime power decomposition of N, then*

$$\begin{aligned} I(N) &= \frac{1 + P + \cdots + P^r}{P^r} \cdot \frac{1 + Q + \cdots + Q^s}{Q^s} \cdot \cdots \cdot \frac{1 + V + \cdots + V^t}{V^t} \\ &= \frac{P^{r+1} - 1}{P^r(P-1)} \cdot \frac{Q^{s+1} - 1}{Q^s(Q-1)} \cdot \cdots \cdot \frac{V^{t+1} - 1}{V^t(V-1)}. \end{aligned} \tag{3}$$

We will explore here some elementary consequences of this formula; there are many additional properties of $\sigma(N)$ and $I(N)$ which involve inequalities and limits. One particularly intriguing result is that the limit of the "average" value of $\sigma(N)/N$ as N increases exists and equals $\pi^2/6$ [**13**, p. 226]. Other results of this type can be found in [**9**] and [**20**].

Prime factors and index growth

We begin our analysis of the behavior of the abundancy index by observing that, for a prime P and positive integer n,

$$I(P^n) = \frac{P^{n+1} - 1}{P^n(P-1)} = \frac{P - 1/P^n}{P - 1} \tag{4}$$

is a decreasing function of P when n is fixed, but it is an increasing function of n when P is fixed. In particular, for a fixed P, the sequence

$$\{I(P^n)\}_{n=1}^{\infty}$$

has for its first term $I(P)=(P+1)/P$ and increases to the limit $P/(P-1)$ as n becomes infinite. Thus for $P=5$, $6/5 \leqslant I(5^n) < 5/4$ for all n; and for $P=71$, $72/71 \leqslant I(71^n) < 71/70$, a very narrow range of values (about .0002 in width)—all slightly larger than 1. In other words, the contribution of a prime the size of 71 to the abundancy index of a number N having 71 as a factor is limited and does not depend strongly on the number of times the prime occurs as a factor of N. The contribution of a still larger prime is proportionally less. In fact, for primes P larger than 41, the fourth term of the sequence $I(P^4)$ agrees with the limit $P/(P-1)$ of the sequence to 8 significant digits. Note also that the ranges of these sequences (and their intervals) are disjoint for any two distinct primes P.

Given a number N, if the goal is to multiply N by a prime which is not one of its factors in order to increase the size of the abundancy index, it is clear that to get a maximum increase, the best choice is the smallest such prime. Sometimes, however, a bigger increase in the index will occur if N is multiplied by one of its prime factors. For example, given $M=360360=2^3\cdot 3^2\cdot 5\cdot 7\cdot 11\cdot 13$, you could multiply M by 2, 3, 5, or 17. It turns out that 17 is the best choice since $I(17M)=3.62+$, $I(2M)=I(5M)=3.50+$, and $I(3M)=3.47+$. In contrast, given $L=90=2\cdot 3^2\cdot 5$, the choice that maximizes the index is 2 since $I(2L)=91/30$ and $I(7L)=104/35$.

In assessing the maximal contribution of the primes in the prime factorization of N to the abundancy index $I(N)$, we will make extensive use of the following theorem, which follows immediately from Theorem 3 and equation (4).

THEOREM 4. *If $N=P^r\cdot Q^s\cdot\cdots\cdot V^t$ with $P, Q,\ldots, V$ distinct primes, then the least upper bound of the sequence $I(N^n)$ is*

$$\frac{P}{P-1}\cdot\frac{Q}{Q-1}\cdot\cdots\cdot\frac{V}{V-1}. \tag{5}$$

Theorem 4 yields several interesting small results. For example, *if N is an even integer but not a power of 2, then N^n is abundant for some n.* To see this, let $N=2PT$, where P is any odd prime factor of N. Then

$$\lim_{n\to\infty} I((2P)^n)=\frac{2}{2-1}\cdot\frac{P}{P-1}=\frac{2P}{P-1}>2,$$

so that some exponent r exists for which $I((2P)^r)>2$. But then, invoking Theorem 1,

$$I(N^r)=I((2P)^rT^r)\geqslant I((2P)^r)>2.$$

Similarly, *if N is not a power of 2, then 2^nN is abundant for some exponent n.*

For odd integers, we can show that *if P and Q are distinct odd primes, then $(PQ)^n$ is deficient for all n.* Since I is a decreasing function of P and 3 and 5 are the smallest odd primes, we only need to note that

$$\lim_{n\to\infty} I((3\cdot 5)^n)=(3/2)\cdot(5/4)=15/8<2.$$

From this it follows that not only is the product of any two distinct odd primes deficient, but its abundancy index is smaller than $15/8=1.875$. On the other hand,

$$\lim_{n\to\infty} I((3\cdot 5\cdot 7)^n)=35/16>2,$$

so there are abundant odd integers with three distinct prime factors. Thus we have a 1913 theorem of L. E. Dickson [**6**]: *Every abundant odd integer has at least three distinct prime factors.* The smallest odd abundant number turns out to be $945=3^3\cdot 5\cdot 7$. We note the dominant role played by the smaller primes in this kind of result: if an odd number is not divisible by either 3 or 5, then in order to be abundant, that number must have at least fifteen distinct prime factors—a fact which can be verified in a couple of minutes using Theorem 4 and a calculator.

It is an interesting historical anomaly that the first four perfect numbers (6, 28, 496, 8128) were all known by A.D. 100, and probably much earlier, but that the 13th century found scholars still unaware that odd abundant numbers such as 945 existed. The difference is partly explained by the fact that Euclid published in his *Elements* a formula for even perfect numbers.

Using formula (4), it is a simple matter to extend Dickson-type results—especially with the aid of a computer. We have prepared TABLE A, which gives, in tabular form, pairs of numbers J and K for the following statement: *Every number N with $I(N) \geq J$ must have at least K distinct prime factors.*

N Even		N Odd	
J	K	J	K
2	2	2	3
3	3	3	8
4	4	4	21
5	6	5	54
6	9	6	141
7	14	7	372
8	22	8	995
9	35	9	2697
10	55	10	7397
11	89	11	20502
12	142	12	57347
13	230		
14	373		
15	609		
16	996		
17	1637		
18	2698		
19	4461		
20	7398		
21	12301		
22	20503		
23	34253		
24	57348		

TABLE A. **Every number N with $I(N) \geq J$ must have at least K distinct prime factors.**

Given J, the number K is computed as follows:

$$\text{for } n \text{ even:} \quad K = \min\left\{ n : \prod_{i=1}^{n} \frac{P_i}{P_i - 1} > J \right\},$$
$$\text{for } n \text{ odd:} \quad K = \min\left\{ n : \prod_{i=2}^{n} \frac{P_i}{P_i - 1} > J \right\} \tag{6}$$

where P_i is the ith prime in the natural ordering of the primes ($P_1 = 2,\ P_2 = 3,\ P_3 = 5$, etc.).

The first few entries in TABLE A were known to R. D. Carmichael [4] in 1907; his paper explicitly states formula (5). Also Paul Poulet [19] in 1929 gave the first seven entries in the "even" table. Recently, computers have been called upon to generate this and similar tables (see, for example, [18]).

While compiling TABLE A for N even, it became apparent that there was a pattern in the entries in the K-column. The sequence of first differences

$$\{1,1,2,3,5,8,13,20,34,53,88,143,236,387,\dots\}$$

bears a resemblance, which is too obvious to be ignored, to the Fibonacci sequence

$$\{1,1,2,3,5,8,13,21,34,55,89,144,233,377,\dots\}$$

Thus, with K as defined in (6), it is natural to conjecture: If $K(J)$ is the least integer such that

$$\prod_{i=1}^{K(J)} \frac{P_i}{P_i - 1} > J$$

and f_J is the Jth Fibonacci number ($f_1 = f_2 = 1$), then $K(J+2) - K(J+1) \sim f_J$ in the sense that

$$\lim_{J \to \infty} [K(J+2) - K(J+1)]/f_J = 1.$$

This is an attractive conjecture because it relates a product of primes to the additively defined Fibonacci sequence.

Although the conjecture seems plausible for small values of J (from 2 to 16), in the range of values of J from 17 to 24 the ratio $[K(J+2) - K(J+1)]/f_J$ increases. (It is 1.304 for the last value, $J = 22$, which TABLE A allows us to compute.) The conjecture is probably false, but it may be possible to resuscitate a version of it by taking equal steps of some number $J < 1$ rather than integer steps, or by making a similar adjustment. Alternatively, perhaps $[K(J+2) - K(J+1)]/f_J$ approaches *some* limit—which, if true, would also be an interesting result.

It is reasonable to question the value of the numbers in the latter part of TABLE A. It is unlikely that anyone will ever want to exhibit a number N with abundancy index greater than 20, so the knowledge that N must contain at least 7398 distinct prime factors (all of the primes 2 through 75037, or worse!) is unlikely to be translated into a blueprint for finding N. Even for smaller values of the index there is no assurance that the *smallest* integer S with index $I(S) \geqslant J$ is obtained using the value of K given by TABLE A. In TABLE B, we give a few "smallest" integers S for given integer values of J and compare L, the number of distinct primes which actually occur in the factorization of S, with the number K in TABLE A.

J	S even	K	L
2	$6 = 2 \cdot 3$	2	2
3	$120 = 2^3 \cdot 3 \cdot 5$	3	3
4	$27720 = 2^3 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11$	4	5
5	$122522400 = 2^5 \cdot 3^2 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17$	6	7
	S odd		
2	$945 = 3^3 \cdot 5 \cdot 7$	3	3
3	$1018976683725 = 3^3 \cdot 5^2 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29$	8	9

TABLE B. **The smallest integer S with $I(S) \geqslant J$.**

A much more extensive table appears in [**1**], where Alaoglu and Erdös extended some work initiated by Ramanujan and, as a part of their efforts, tabulated the first 74 **superabundant numbers**—integers whose abundancy indexes are greater than those of all smaller integers.

One final note on the growth of the abundancy index: After a certain level has been reached by the index, those index values do not occur with increasing frequency. For example, the index first reaches 4 at 27720, is equal to, but not greater than, 4 twice between there and 50000, and the number of integers with index greater than 4 in the successive intervals of length 50000 between 50000 and 500000 are, respectively, 10, 10, 11, 15, 12, 10, 12, 12, and 11. Some quantitative results on the density $A(x)$, which is the limit as $N \to \infty$ of the relative frequency of occurrence of integers N with $I(N) \geqslant x$, can be found in [**21**]. A lively and entertaining discussion of these statistics is coupled with comments on the history of perfect and abundant numbers in an article by David G. Kendall [**15**] in which philosophical attitudes and applications—ancient, medieval, and his own—are considered. One of the interesting aspects of Kendall's article is the evidence he gives to support the conjectures that 1) the density $A(2)$ of the nondeficient numbers is $1/4$, or 2) the density of the nondeficient even numbers is $1/4$. We can use TABLE B to show that these conjectures cannot both be true, for the density of odd nondeficients is at least $1/1890$ because every odd multiple of 945 is odd and abundant.

The set of abundancy indexes

Near the end of the first section of this paper, we deferred the question of the nature of the set of abundancy indexes

$$D = \{ I(N) : N \geqslant 2 \}.$$

Here we resume that discussion.

Our immediate goal is to show that this set is dense in the interval $(1, \infty)$. We first examine the behavior of the product $\prod (P_i + 1)/P_i$, where P_i is the ith prime. The behavior is quickly established, for

$$\lim_{n \to \infty} \prod_{i=1}^{n} \frac{P_i + 1}{P_i} = \infty \Leftrightarrow \lim_{n \to \infty} \prod_{i=1}^{n} 1 + \frac{1}{P_i} = \infty \Leftrightarrow \lim_{n \to \infty} \sum_{i=1}^{n} \frac{1}{P_i} = \infty .$$

This second equivalence is a special case of the basic theorem on infinite products, that $\prod(1 + a_i) = \infty$ if, and only if, $\sum a_i = \infty$ (see [**14**] or [**16**]). However, $\sum 1/P_i$ is a known divergent series ([**17**, or **13**, p. 17). Thus $\prod (P_i + 1)/P_i$ is a divergent product whose partial products $\prod_{i=1}^{n} (P_i + 1)/P_i$ form a strictly increasing sequence and whose component factors $(P_i + 1)/P_i$ form a strictly decreasing sequence converging to 1. These observations yield an alternate proof of Theorem 2 and make possible the following theorem.

THEOREM 5. *The set* $D = \{ I(N) : N \geqslant 2 \}$ *is dense in the interval* $(1, \infty)$.

Proof. The method of proving this theorem is comparable to approximating a positive number by a subseries of $\sum 1/n$. We generate an increasing sequence $\{ I(N_k) \}$ converging to an arbitrary number $x > 1$, as follows: let $N_0 = P_1 P_2 \cdots P_t$, where t is the largest integer such that

$$I(N_0) = \prod_{i=1}^{t} (P_i + 1)/P_i \leqslant x.$$

(Since $I(P_1) = I(2) = 3/2$, this construction works only if $x \geqslant 3/2$, so we arbitrarily set $I(N_0) = 1$ if $x < 3/2$.) Let $d_1 = x/I(N_0)$; let Q_1 be the smallest prime such that $Q_1 > P_t$ and $(Q_1 + 1)/Q_1 \leqslant d_1$; and let $N_1 = N_0 Q_1$. Now $I(N_0) < I(N_1) \leqslant x$. Next let $d_2 = x/I(N_1)$, and let Q_2 be the smallest prime such that $Q_2 > Q_1$ and $(Q_2 + 1)/Q_2 \leqslant d_2$. Let $N_2 = N_1 Q_2$. With the proviso that we will stop if $I(N_j) = x$ at any stage (including the 0th stage), we continue this process to generate a sequence $\{ N_j \}$ of integers such that $\{ I(N_j) \}$ increases to x. The existence of the needed Q_j's are guaranteed by the convergence of $\{ (P_i + 1)/P_i \}$ to 1, and the fact that the gap from $I(N_j)$ to x will always be bridged follows from the divergence of $\prod (P_i + 1)/P_i$ to ∞.

The numbers N_j generated in the proof are all products of distinct primes. There are infinitely many integers which would not be used to generate the sequences $\{ I(N_j) \}$. Also, there are many duplications in the set of abundancy indexes; one such duplication occurs when

$$I(332640) = I(360360) = 48/11.$$

Further, some of these duplications are integers, as is evidenced by the existence of more than one perfect number. (Erdős [**9**] has done a sophisticated analysis of the frequency of occurrence of duplications.) However, the number 2 has an abundancy index which is uniquely its own. To see that $N = 2$ is the only solution of the equation $I(N) = 3/2$, note first that if N is even and $N > 2$, then $I(N) > 3/2$ (Theorem 1); second, if N is odd, then the denominator in formula (3) is odd and cannot have 2 as a factor.

Open questions

Since every $I(N)$ is a rational number and since $D = \{I(N): N \geqslant 2\}$ is dense in $(1, \infty)$, we are led to the (perhaps unanswerable) question: Is every rational number $q > 1$ the abundancy index of some integer? If we note that the construction in the proof of Theorem 5 can be carried out with only minor modifications after deleting the number 2 from the list of primes, we get another interesting question: Is every rational number $q > 1$ the abundancy index of some *odd* integer? A positive answer to the first question would be of some interest, and a positive answer to the second one would be a major advance since it would imply that there exist odd perfect numbers and odd multiperfect numbers, no examples of which are known. But the suggested modification of the proof of Theorem 5 does assert that there are odd integers whose "imperfections" are as small as we choose since we can, for any $\varepsilon > 0$, find an odd integer N with abundancy index $I(N)$ in the interval $(2 - \varepsilon, 2]$. In this context we might regard ε, or its reciprocal, as a "coefficient of frustration."

The search for multiperfect numbers, those numbers with integral abundancy indexes, was vigorously undertaken by Descartes, Fermat, and others in the 17th century. Those two were the most successful searchers of that era, and the competition between them apparently was not without rancor [**7**, pp. 33–38]. Many other multiperfects were discovered in the early 1900's by Carmichael [**4**], Mason [**5**], and Poulet [**19**] and, more recently, by Brown [**3**] and Franqui and Garcia [**10**; **11**]. These last four lists, all over 30 years old, contain seven multiperfects of index 8, each of which has from 41 to 43 distinct prime factors and is of the order of 10^{200}. Index 8 is still the greatest index achieved in the search for multiperfect numbers. Franqui and Garcia's second list also includes a multiperfect of index 7 with a 27-digit prime factor.

The ideas of the abundancy index and formula (3) have been the principal tools in the investigations of multiperfects from the beginning. Substitutions discovered and used by Descartes [**7**, p. 35], Poulet, and Carmichael and Mason have the effect of preserving the abundancy index or changing it by an integer amount to construct new multiperfect numbers from known ones. One example of such a substitution which preserves index is the replacement of $3^5 \cdot 7^3 \cdot 13$ by $3^7 \cdot 7^2 \cdot 19 \cdot 41$. A comparable substitution, $2^5 \cdot 3^3$ for $2^3 \cdot 3^2 \cdot 13$, was implicit in an earlier observation that $I(332640) = I(360360)$, or $I(2^5 \cdot 3^3 \cdot 5 \cdot 7 \cdot 11) = I(2^3 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 13)$. It is entirely possible that many substitutions were discovered in just this way—by comparing the prime factorizations of numbers (mostly multiperfects) known to have the same index. An excellent survey of current knowledge on multiperfect numbers and an extensive list of references can be found in [**12**].

One of the oldest and most famous (notorious) of the unanswered questions of mathematics is: Is there an odd perfect number? The theory of *even* perfect numbers was thoroughly settled by Euclid and Euler (see, for example, [**8**, Chapter 8].) Apparently, none of that theory is helpful in searching for *odd* perfect numbers.

To begin, we wish to satisfy equation (3)

$$I(N) = \frac{1 + P + P^2 + \cdots + P^r}{P^r} \cdot \frac{1 + Q + \cdots + Q^s}{Q^s} \cdot \cdots \cdot \frac{1 + V + \cdots + V^t}{V^t} = 2,$$

where the distinct prime factors $P, Q, \ldots, V$ of N are all odd. The denominator of this fraction is an odd number, so the numerator contains only a single factor of 2; but if r is an odd integer, then $1 + P + P^2 + \cdots + P^r$ is even. Thus exactly one of the exponents $r, s, \ldots, t$ is odd, and all of the other primes occur to even powers only. (The reader is invited at this point to walk in the footsteps of Euler and show that, for this prime with odd exponent, both the prime and the exponent are congruent to 1 modulo 4.) If we wish to have some 3's in the prime factorization of

N, we will need an even number of them, but the penalty for having none is that the minimum number of distinct prime factors of N increases from eight to eleven, **[25]**, **[27]**, **[29]**.

As a variation on the approach using formula (3), we might try the approach in the proof of Theorem 5 and approximate the number $x = 2$ from below by a sequence of abundancy indexes—keeping in mind the restrictions of even exponents and the like discussed above—and hope to be lucky enough to find at some stage that $2/I(N_j)$ is equal to some realizable $(P_i + 1)/P_i$ or $(P_i^2 + P_i + 1)/P_i^2$.

In light of this discussion, it is not surprising that no odd perfect numbers have yet been found and that those adventurers who look for them can become quickly discouraged. On the other hand, the possibility that an odd perfect or multiperfect number may exist has spurred several modern investigators to develop and improve criteria and conditions such numbers must satisfy. A few of these are noted in our references; further criteria and references are available in **[2]** and **[12]**.

The value of investigation

This note grew out of a "computer literacy" project for classes of humanities majors and prospective elementary teachers in which I introduced the abundancy index as a concept whose computational aspects were readily assigned to a computer. I asked them to compute a few values of $I(N)$ and search for an odd abundant number. Later findings of $\sigma(N)/N$ in older literature **[4]**, **[6]** led me to an enjoyable investigation of properties of the index. The two referees of an earlier version of the paper further stimulated my interest with a number of additional references (as well as being helpful in all the ways referees can be).

The abundancy index as a hierarchical classification of numbers is an interesting concept in its own right—at least in part for its recreational value when used to investigate the general topic of abundant and deficient numbers. In addition, it has growth and density properties to intrigue both the serious and recreational students of number theory. Its analysis provides a vehicle for unifying several parts of the theory; in so doing, it suggests new unsolved problems and illuminates old ones. It is hard to ask anything more of a completely intelligible one-line definition!

References

[1] L. Alaoglu and P. Erdös, On highly composite and similar numbers, Trans. Amer. Math. Soc., 56 (1944) 448–469.

[2] A. H. Beiter, Recreations in the Theory of Numbers, 2nd ed., Dover, New York, 1966, Chapter 3.

[3] A. L. Brown, Multiperfect numbers, Scripta Math., 20 (1954) 103–106.

[4] R. D. Carmichael, A table of multiply perfect numbers, Bull. Amer. Math. Soc., 13 (1907) 383–386.

[5] R. D. Carmichael and T. E. Mason, Note on multiply perfect numbers including a table of 204 new ones and the 47 others previously published, Proc. Indiana Acad. of Sci., (1911) 257–270.

[6] L. E. Dickson, Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors, Amer. J. Math., 35 (1913) 413–422.

[7] ______, History of the Theory of Numbers, Carnegie Institute, 1919 (reprinted by Chelsea, 1966), Vol. 1, Chapter 1.

[8] U. Dudley, Elementary Number Theory, 2nd ed., Freeman, 1978.

[9] P. Erdös, Remarks on number theory. II. Some problems of the σ function, Acta Arith., 5 (1959) 171–177.

[10] B. Franqui and M. Garcia, Some new multiply perfect numbers, Amer. Math. Monthly, 60 (1953) 459–462.

[11] ______, 57 new multiply perfect numbers, Scripta Math., 20 (1954) 169–171.

[12] R. K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981, pp. 25–30.

[13] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford, 1960.

[14] E. Hille, Analytic Function Theory, Vol. 1, Ginn, 1959, p. 223.

[15] D. G. Kendall, The scale of perfection, J. Appl. Probability, 19A (1982) 125–138.

[16] K. Knopp, Theory and Application of Infinite Series, Hafner, 1947, p. 219.

[17] I. Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, 4th ed., Wiley, 1980, p. 221.

[18] K. K. Norton, Remarks on the number of factors of an odd perfect number, Acta Arith., 6 (1961) 365–374.

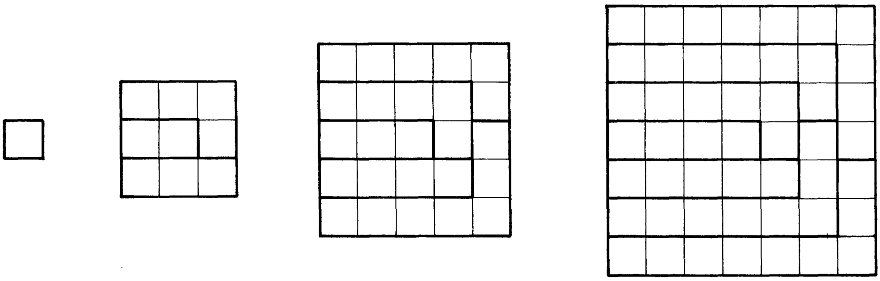
[19] P. Poulet, La Chasse aux Nombres, Fascicule I, Bruxelles, 1929, 9–27.

[20] M. Satyanarayana, Bounds of $\sigma(N)$, Math. Student, 28 (1960) 79–81.
[21] C. R. Wall, P. L. Crews, and D. B. Johnson, Density bounds for the sum of divisors function, Math. Comp., 26 (1972) 773–777.

Supplementary references on odd perfect numbers

[22] G. L. Cohen, On odd perfect numbers, Fibonacci Quart., 16 (1978) 523–527. (Improved upper bounds on $\sum_{i=1}^{t} 1/p_i$ when $N = p_1^{\alpha_1} \cdots p_t^{\alpha_t}$ is an odd perfect number.)
[23] G. L. Cohen and M. D. Hendy, On odd multiperfect numbers, Math. Chronicle, 10 (1981) 57–61. (If $\sigma(N) = KN$, K an integer $\geqslant 2$ and N odd with t distinct prime factors, then $t \geqslant (K^5 + 387)/70$.)
[24] G. G. Dandapat, J. L. Hunsucker, and C. Pomerance, Some new results on odd perfect numbers, Pacific J. Math., 57 (1975) 359–364. (A negative answer to a 1970 conjecture.)
[25] P. Hagis, Jr., Outline of a proof that every odd perfect number has at least eight prime factors, Math. Comp., 35 (1980) 1027–1032.
[26] P. Hagis, Jr., On the second largest prime divisor of an odd perfect number, Lecture Notes in Mathematics 899, Springer-Verlag, New York, 1971, pp. 254–263. (The prime of the title is > 1000.)
[27] ______, Sketch of a proof that an odd perfect number relatively prime to 3 has at least eleven prime factors, Math. Comp., 40 (1983) 399–404.
[28] P. Hagis, Jr., and W. L. McDaniel, On the largest prime divisor of an odd perfect number, Math. Comp., 29 (1975) 922–924. (The prime of the title is $\geqslant 100129$.)
[29] M. Kishore, Odd perfect numbers not divisible by 3, Math. Comp., 40 (1983) 405–411. (An alternative proof of the result of [**27**].)

Proof without words: Arithmetic progressions with sum equal to the square of the number of terms



$n = 4$

$4 + 5 + 6 + 7 + 8 + 9 + 10 = 7^2$

$$\sum_{k=n}^{3n-2} k = (2n-1)^2; \qquad n = 1, 2, 3, \ldots$$

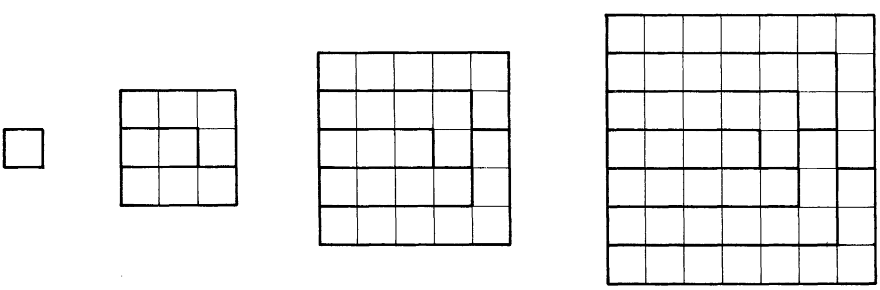
—JAMES O. CHILAKA
Long Island University
C.W. Post Campus

[20] M. Satyanarayana, Bounds of $\sigma(N)$, Math. Student, 28 (1960) 79–81.
[21] C. R. Wall, P. L. Crews, and D. B. Johnson, Density bounds for the sum of divisors function, Math. Comp., 26 (1972) 773–777.

Supplementary references on odd perfect numbers

[22] G. L. Cohen, On odd perfect numbers, Fibonacci Quart., 16 (1978) 523–527. (Improved upper bounds on $\sum_{i=1}^{t} 1/p_i$ when $N = p_1^{\alpha_1} \cdots p_t^{\alpha_t}$ is an odd perfect number.)
[23] G. L. Cohen and M. D. Hendy, On odd multiperfect numbers, Math. Chronicle, 10 (1981) 57–61. (If $\sigma(N) = KN$, K an integer $\geqslant 2$ and N odd with t distinct prime factors, then $t \geqslant (K^5 + 387)/70$.)
[24] G. G. Dandapat, J. L. Hunsucker, and C. Pomerance, Some new results on odd perfect numbers, Pacific J. Math., 57 (1975) 359–364. (A negative answer to a 1970 conjecture.)
[25] P. Hagis, Jr., Outline of a proof that every odd perfect number has at least eight prime factors, Math. Comp., 35 (1980) 1027–1032.
[26] P. Hagis, Jr., On the second largest prime divisor of an odd perfect number, Lecture Notes in Mathematics 899, Springer-Verlag, New York, 1971, pp. 254–263. (The prime of the title is > 1000.)
[27] ______, Sketch of a proof that an odd perfect number relatively prime to 3 has at least eleven prime factors, Math. Comp., 40 (1983) 399–404.
[28] P. Hagis, Jr., and W. L. McDaniel, On the largest prime divisor of an odd perfect number, Math. Comp., 29 (1975) 922–924. (The prime of the title is $\geqslant 100129$.)
[29] M. Kishore, Odd perfect numbers not divisible by 3, Math. Comp., 40 (1983) 405–411. (An alternative proof of the result of [**27**].)

Proof without words:
Arithmetic progressions with sum equal to the square of the number of terms



$n = 4$

$4 + 5 + 6 + 7 + 8 + 9 + 10 = 7^2$

$$\sum_{k=n}^{3n-2} k = (2n-1)^2; \qquad n = 1, 2, 3, \ldots$$

—James O. Chilaka
Long Island University
C.W. Post Campus

Time-Sharing and the DeMoivre-Laplace Theorem

STEPHEN H. FRIEDBERG
Illinois State University
Normal, IL 61761

One of the most important theorems of statistics is the central limit theorem. Roughly, it states that a sum of (stochastically) independent and identically distributed random variables has an approximately normal distribution, with the approximation improving as the number of random variables increases. A precursor of this theorem is the DeMoivre-Laplace theorem, which states that a sum of independent Bernoulli random variables (with common mean) is approximately normal. The latter theorem was first proved in 1718 by DeMoivre and then generalized by Laplace in 1812. The more general result, known as the central limit theorem, is attributed to Lindeberg (1922). For additional information regarding these theorems and the terms used in this article, the reader may consult any text in mathematical statistics (see, for example, Hoel, Port, and Stone [**1**]).

A Bernoulli random variable is one that takes on only two values, say, 1 with probability p, and 0 with probability $q = 1 - p$. The mean of such a Bernoulli random variable is p, and its variance is given by pq. These variables arise in many settings. For example, the values taken on may be male-female, yes-no, off-on, candidate A-candidate B, and so on. The DeMoivre-Laplace theorem is formally stated below.

DEMOIVRE-LAPLACE THEOREM. *Let $X_1, \dots, X_n$ be a random sample from a Bernoulli population with mean p and let $S_n = X_1 + X_2 + \cdots + X_n$. Then for any real numbers a and b, we have*

$$\lim_{n\to\infty} P\left(a < (S_n - np)/\sqrt{npq} < b\right) = \int_a^b (1/\sqrt{2\pi})\exp(-z^2/2)\,dz.$$

Notice that the random variable $(S_n - np)/\sqrt{npq}$ within the left side of the equation above is formed by *standardizing* the (binomial) random variable S_n, that is, the mean np is subtracted and the resulting expression is divided by the standard deviation $\sqrt{npq}$. On the other hand, the integrand of the right side is the probability density function of a standard normal random variable. Intuitively, the theorem states that a standardized binomial random variable may be approximated by a standard normal random variable.

One of the most common applications of this theorem involves the derivation of approximate confidence intervals to estimate the population mean p. In this note, we will be concerned with a problem which arose in the discussion of time-sharing. Specifically, a large corporation had allotted (rented) 10 computer lines to 40 smaller companies. The corporation later decided to extend this service to encompass 80 customers. Executives of the corporation felt that since the number of customers had doubled, it would be reasonable to double the number of allotted lines to 20. However, computer simulations seemed to indicate that *less* than 20 (actually, 15) lines would suffice. Why? It should be noted that any reduction in the number of lines would save the company an enormous amount of money. We will attempt to model the problem here and show that the DeMoivre-Laplace theorem provides a plausible explanation.

We begin by defining the following random variables:

$$X_i(x) = \begin{cases} 1 & \text{if company } i \text{ is on line at time } x \\ 0 & \text{otherwise.} \end{cases}$$

By definition, each X_i is a Bernoulli random variable. Let p denote the probability that X_i takes on the value 1. For this application, it turns out that each company uses the service about one hour each day. So, we take $p = 1/24$. We let $X = X_1 + X_2 + \cdots + X_{40}$. Notice that we may interpret X as the number of companies that are on line at a particular time. In fact, because the corporation is allotting 10 lines, the probability that a line is open is $\beta = P(X < 10)$. Roughly, β is a measure of customer satisfaction in the sense that the closer β is to 1, the more likely it is that a customer can go on line without waiting. The corporation's goal is to increase the number of lines to accommodate the additional customers without decreasing β.

If we let Z denote the standard normal random variable, the DeMoivre-Laplace theorem tells us that

$$(X - 40p)/\sqrt{40pq} \sim Z \tag{1}$$

or,

$$X \sim \sqrt{40pq}\, Z + 40p,$$

where $\sim$ means "has approximately the same distribution as." We should note that we have made the assumption that the X_i's are independent.

Now, let $W = X_1 + X_2 + \cdots + X_{80}$. If k denotes the number of lines that the company should use for 80 customers and still maintain satisfaction β, then we have the condition

$$P(W < k) = \beta = P(X < 10). \tag{2}$$

Applying the DeMoivre-Laplace theorem to W as we did earlier to X, we obtain

$$W \sim \sqrt{80pq}\, Z + 80p. \tag{3}$$

Combining (1) and (3), we have

$$W \sim \sqrt{80pq}\left[(X - 40p)/\sqrt{40pq}\right] + 80p.$$

Or,

$$W \sim \sqrt{2}\, X + (80 - 40\sqrt{2})\, p. \tag{4}$$

Equations (2) and (4) yield

$$\begin{aligned} P(X < 10) &= P(W < k) \\ &= P\left(\sqrt{2}\, X + (80 - 40\sqrt{2})\, p < k\right) \\ &= P\left(X < \left[k + (40\sqrt{2} - 80)\, p\right]/\sqrt{2}\right). \end{aligned}$$

From these equations, we set

$$10 = \left[k + (40\sqrt{2} - 80)\, p\right]/\sqrt{2}.$$

Solving for k, we obtain

$$k = 10\sqrt{2} + (80 - 40\sqrt{2})\, p.$$

For $p = 1/24 = .04$, we have $k = 15.08$, that is, about 15 lines should provide the same level of satisfaction for 80 customers as 10 lines provide for 40 customers. To see how k varies for selected values of p, we construct the following table.

Usage p	Number of lines k
.01	14.38
.04	15.08
.08	16.02
.13	17.19
.17	18.13
.21	19.06
.25	20.00

We see that as the *usage p* for a fixed number of customers increases, the number of lines necessary to maintain customer satisfaction also increases.

References

[1] P. Hoel, S. Port, and C. Stone, Introduction to Probability Theory, Houghton Mifflin Company, 1971, p. 184.

Ladder Approximations of Irrational Numbers

J. R. Ridenhour
Austin Peay State University
Clarksville, TN 37044

The approximation of irrational numbers by rationals is a topic which has interested mathematicians since the time of the Greeks. In fact, the Pythagoreans, who are credited with the discovery of $\sqrt{2}$ and its irrationality, had a clever device, known as the system of "side and diameter" numbers, for obtaining successively better rational approximations to $\sqrt{2}$. While the method of side and diameter numbers is thought to have originated before 450 B.C., the earliest account of which we now have record is in the writings of Theon of Smyrna (c. A.D. 140) and may be found in the English translation of Ivor Thomas [**6**, pp. 133–137].

The method of side and diameter numbers is best described algebraically in terms of "Ladder Arithmetic." The ladder consists of a succession of rungs where each rung contains a pair of integers s_k and d_k. The s_k's are associated with the sides and the d_k's with the diameters in the geometric description. For the initial rung, both s_0 and d_0 are 1 and the successive rungs are generated by using the recurrence relations $s_{k+1} = s_k + d_k$ and $d_{k+1} = 2s_k + d_k$ (Table 1). Then the quotients d_k/s_k approach the desired limit. The sequence of approximations for $\sqrt{2}$ so

s_k	d_k
1	1
2	3
5	7
12	17
$\vdots$	

Table 1

obtained is $1, 3/2, 7/5, \ldots$. The algebraic verification rests on the identity $d_k^2 = 2s_k^2 + (-1)^{k+1}$, which can easily be verified by induction using the recurrence relations given above. Once we have this identity, all that is necessary is to divide through by s_k^2 and take the limit, since the s_k's clearly approach ∞. An interesting observation is that this ladder is closely related to the

We see that as the *usage p* for a fixed number of customers increases, the number of lines necessary to maintain customer satisfaction also increases.

References

[1] P. Hoel, S. Port, and C. Stone, Introduction to Probability Theory, Houghton Mifflin Company, 1971, p. 184.

Ladder Approximations of Irrational Numbers

J. R. RIDENHOUR
Austin Peay State University
Clarksville, TN 37044

The approximation of irrational numbers by rationals is a topic which has interested mathematicians since the time of the Greeks. In fact, the Pythagoreans, who are credited with the discovery of $\sqrt{2}$ and its irrationality, had a clever device, known as the system of "side and diameter" numbers, for obtaining successively better rational approximations to $\sqrt{2}$. While the method of side and diameter numbers is thought to have originated before 450 B.C., the earliest account of which we now have record is in the writings of Theon of Smyrna (c. A.D. 140) and may be found in the English translation of Ivor Thomas [**6**, pp. 133–137].

The method of side and diameter numbers is best described algebraically in terms of "Ladder Arithmetic." The ladder consists of a succession of rungs where each rung contains a pair of integers s_k and d_k. The s_k's are associated with the sides and the d_k's with the diameters in the geometric description. For the initial rung, both s_0 and d_0 are 1 and the successive rungs are generated by using the recurrence relations $s_{k+1} = s_k + d_k$ and $d_{k+1} = 2s_k + d_k$ (TABLE 1). Then the quotients d_k/s_k approach the desired limit. The sequence of approximations for $\sqrt{2}$ so

s_k	d_k
1	1
2	3
5	7
12	17
⋮	

TABLE 1

obtained is $1, 3/2, 7/5, \ldots$. The algebraic verification rests on the identity $d_k^2 = 2s_k^2 + (-1)^{k+1}$, which can easily be verified by induction using the recurrence relations given above. Once we have this identity, all that is necessary is to divide through by s_k^2 and take the limit, since the s_k's clearly approach ∞. An interesting observation is that this ladder is closely related to the continued fraction expansion for $\sqrt{2}$:

$$\sqrt{2}=1+\cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{2+\cdots}}}.$$

The convergents are $1, 1+\frac{1}{2}, 1+1/(2+\frac{1}{2}), \ldots,$ and these are precisely the ladder approximations.

While Theon refers to "sides" and "diameters" (or diagonals) of squares, he does not give a precise description of the geometric construction. He merely states that each new side is obtained by adding the previous side and diameter while each new diameter is the sum of the previous diameter and twice the previous side. He then asserts that the squares of the diameters are alternately greater than or less than twice the squares of the sides by one unit. That is,

$$d_k^2-2s_k^2=(-1)^{k+1}.$$

Obviously, s_k and d_k cannot be the side and diagonal of the same square for then the ratio d_k/s_k would be exactly $\sqrt{2}$. One possible and relatively simple construction is given in FIGURE 1. The

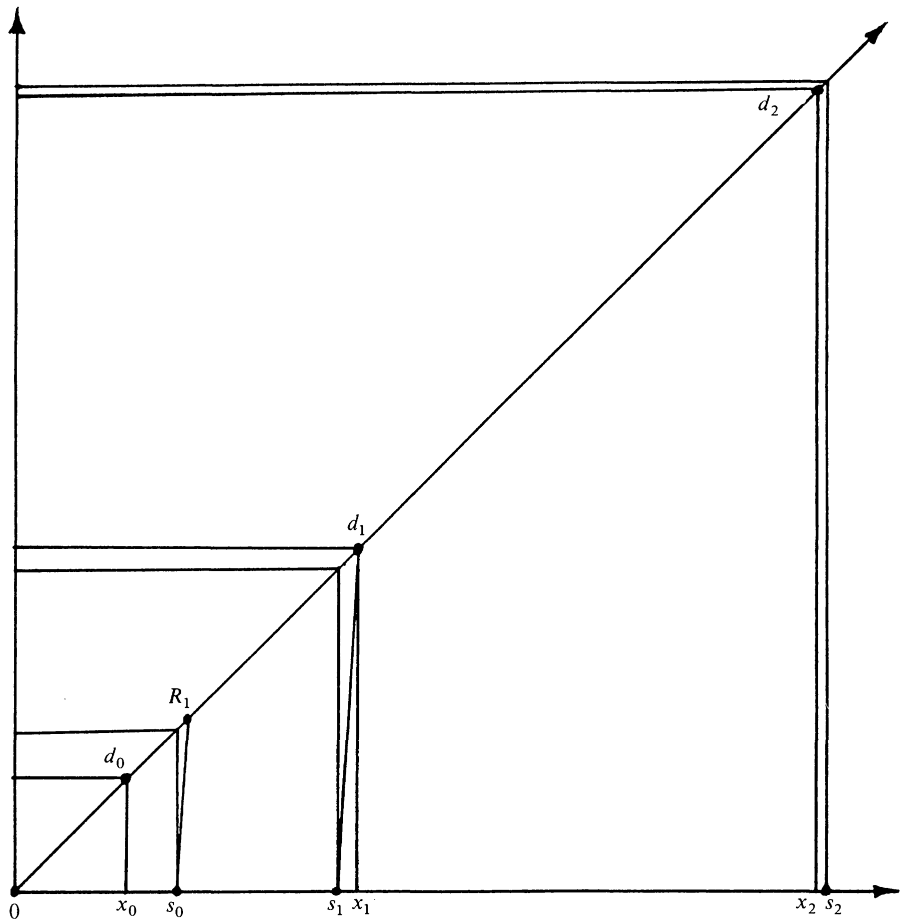


FIGURE 1

segment from 0 to d_k represents the diameter d_k, the segment from 0 to s_k represents the side s_k, and the segment from 0 to x_k represents a side of the square with diameter d_k. First construct perpendicular rays and bisect to get the 3 rays emanating from 0. Then construct the two smaller squares, one having diameter $d_0 = 1$ and the other having side $s_0 = 1$. Then construct the next two smaller squares, one having side $s_1 = s_0 + d_0$ and the other with diameter $d_1 = d_0 + 2s_0$. This process can be continued indefinitely where each pair of squares is constructed from the previous pair using the relations $s_{k+1} = s_k + d_k$ and $d_{k+1} = d_k + 2s_k$. Since $d_k^2 = 2x_k^2$, d_k^2 is larger than or less than $2s_k^2$, depending upon whether x_k is to the right or left of s_k. From the figure, it appears that d_k^2 will be alternately less than and greater than $2s_k^2$, with x_k very close to s_k after a few iterates. (If the reader extends the figure to include one more pair of squares, he will find that x_3 is very close to s_3 and the two squares are very nearly congruent.) When enough iterates are done that the two squares are nearly congruent, then d_k/s_k is nearly the ratio of the diameter and side of the same square, which is $\sqrt{2}$. It is interesting to note that for each pair of squares the area of the strip between them is $1/2$, since

$$d_k^2 = 2x_k^2 = 2s_k^2 + (-1)^{k+1}$$

implies

$$x_k^2 = s_k^2 + (-1)^{k+1}/2.$$

Another nice feature of FIGURE 1 is that the geometric construction of d_k/s_k is neatly built into the figure. All that is necessary is to construct the line through s_0 that is parallel to the segment joining s_k and d_k and let R_k be the intersection of that line with the bisecting ray. (This has been drawn in for s_1 and d_1 to produce R_1.) Then, since $s_0 = 1$, we have by similar triangles that the segment from 0 to R_k is the ratio of d_k to s_k, which is the rational approximation to $\sqrt{2}$.

The relationship between d_k^2 and $2s_k^2$ plays an important role, since it assures us that the strip between the squares has constant area $1/2$, which forces x_k to become close to s_k as the sides become longer. Theon merely states this result without proof. Thomas [**6**, p. 133] feels that this is because he considered it to be well known. How might the Pythagoreans have deduced this result? Heath suggests the following line of reasoning in [**3**]. From Euclid's Book II, Proposition 10, they

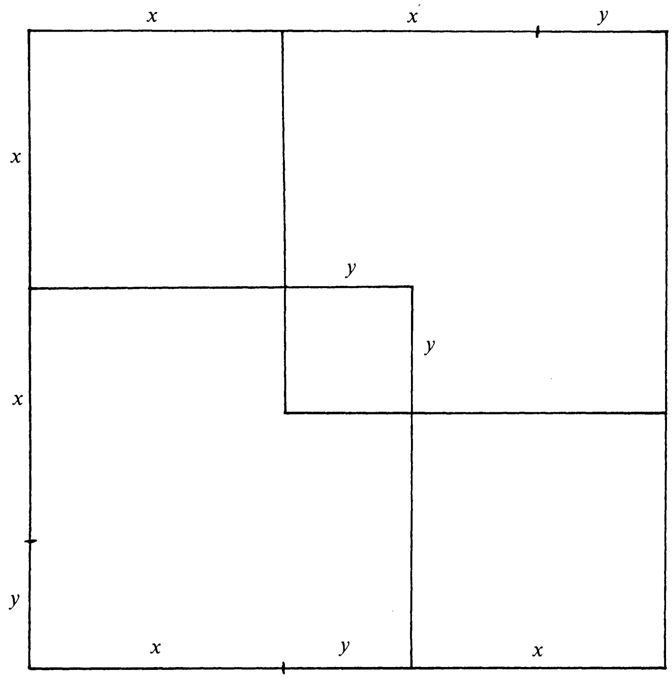


FIGURE 2

were aware of the identity

$$(2x+y)^2 = 2x^2 + 2(x+y)^2 - y^2,$$

which is apparent from FIGURE 2. Using s and d and rearranging, we see that

$$(2s+d)^2 - 2(s+d)^2 = -(d^2 - 2s^2).$$

Since, for a given diameter and side, d and s, the next diameter and side are $2s+d$ and $s+d$, this implies that, if $d^2 - 2s^2$ is ± 1, then, for the next pair, the difference will be ∓ 1.

While the author is not aware of any geometric interpretations like that given by FIGURE 1, there are other geometric constructions which produce sequences of sides and diameters satisfying the basic recurrence relations. Some of these are given by Vedora in [7]. Vedora also gives some generalizations of the methods involved to approximate $\sqrt{n}$, but his techniques are incapable of producing rational approximations for $n^{1/r}$ where $r > 2$. A relatively recent discussion of side and diagonal numbers is contained in [**8**].

There has been much speculation on the scope of such ladder approximations known to the Greeks. Archimedes in his treatise *Measurement of a Circle*, states, with no explanation, that

$$\frac{1351}{780} > \sqrt{3} > \frac{265}{153}.$$

There have been numerous ingenious attempts to explain how Archimedes arrived at such an excellent approximation for $\sqrt{3}$. A thorough discussion of some of the more prominent explanations may be found in Heath's book on the works of Archimedes [**3**, pp. lxxxi–xcix]. Among these explanations is a method of Heilermann which is a generalization of the Pythagorean ladder approximation. Again a ladder is formed containing successive rungs $(s_0, d_0), (s_1, d_1), (s_2, d_2), \ldots,$ but this time the recurrence relations used for generating the successive pairs of numbers are

$$s_{k+1} = s_k + d_k \quad \text{and} \quad d_{k+1} = ns_k + d_k,$$

where n represents an arbitrary positive number. It is then possible to show that when $s_0 = d_0 = 1$

$$\frac{d_k^2}{s_k^2} = n + \frac{(1-n)^{k+1}}{s_k^2},$$

and that the second term on the right-hand side of the equation tends to 0 as $k \to \infty$. Then it follows that $d_k/s_k \to \sqrt{n}$ as $k \to \infty$. However, using $s_0 = 1$, $d_0 = 1$ with $n = 3$ we get the ladder given in TABLE 2. A problem arises immediately. If this is indeed how Archimedes arrived at his

s_k	d_k	d_k/s_k
1	1	1
2	4	2
6	10	5/3
⋮	⋮	⋮
2448	4240	265/153
6688	11584	362/209
18272	31684	989/571
49920	86464	1351/780

TABLE 2

approximation, why would he not state instead the better approximation

$$\frac{989}{571} < \sqrt{3} < \frac{1351}{780}?$$

Heilermann explains a modification of the use of the above method which results in Archimedes' approximation as successive approximations, and the interested reader is referred again to Heath [**3**, p. xcvii].

Our purpose is to consider some ladder approximations which are extensions of those mentioned above. These extensions prove to be rather interesting in that while the results are straightforward generalizations of the classical ones, the techniques involved in their proof require results from several areas of mathematics. We will need some combinatorial identities as well as some elementary results from the theory of difference equations and the distribution of zeros of polynomials.

We shall assume throughout the remainder of this note that n is the real number, larger than 1 but not necessarily an integer, whose roots we wish to approximate. Each row of our ladder will contain r numbers rather than the usual two. The first row contains only 1's, just as the Pythagorean ladder. To obtain the first entry on a given row, we merely add all the entries on the previous row. To find the entry just to the right of a given entry we add the given entry to the product of $n-1$ and the entry just above the given entry. The entries will be denoted by a_{ij}, where the subscripts are associated with the row and column of the ladder in which the entry lies. What results is a rectangular array of the following type

$$\begin{array}{ccccc} a_{00} & a_{01} & a_{02} & \cdots & a_{0,r-1} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1,r-1} \\ a_{20} & a_{21} & a_{22} & \cdots & a_{2,r-1} \\ \vdots & \vdots & \vdots & & \vdots \end{array}$$

where the fundamental properties of the array are

$$a_{0j} = 1; \qquad j = 0,1,\ldots,r-1, \tag{1}$$

$$a_{i+1,0} = \sum_{j=0}^{r-1} a_{ij}; \qquad i \geqslant 0, \tag{2}$$

and

$$a_{i,j+1} = a_{ij} + (n-1)a_{i-1,j}; \qquad i \geqslant 1; \qquad j = 0,1,\ldots,r-2. \tag{3}$$

The main result is that, as we proceed down the ladder, the quotients of the entries in the various columns and those in the first column approach powers of the rth root of n.

THEOREM 1. *If a_{ij} are defined by* (1), (2), (3), *then*

$$\lim_{i\to\infty} \frac{a_{ij}}{a_{i0}} = n^{j/r}; \qquad j = 1,2,\ldots,r-1.$$

The proof of the theorem will employ tools from several branches of mathematics. We first note, however, that as a result of this theorem we are able to obtain successive approximations to rational powers of positive numbers. Moreover, if n is a rational number then the approximants are also rational. TABLE 3 gives the first six lines of the ladder $3^{1/4}$, $3^{2/4}$, and $3^{3/4}$ (these numbers rounded to 4 decimal places are 1.3161, 1.7321, and 2.2795).

Proof of the Theorem. We begin our mathematical analysis by demonstrating that the following combinatorial identities follow from (1), (2), (3):

$$a_{ij} = \sum_{k=0}^{j} \binom{j}{k} (n-1)^k a_{i-k,0}; \qquad i \geqslant j \tag{4}$$

$$a_{i0}=\sum_{k=1}^{r}\binom{r}{k}(n-1)^{k-1}a_{i-k,0};\qquad i\geqslant r. \tag{5}$$

Ladder				Quotients		
a_{i0}	a_{i1}	a_{i2}	a_{i3}	a_{i1}/a_{i0}	a_{i2}/a_{i0}	a_{i3}/a_{i0}
1	1	1	1	1	1	1
4	6	8	10	1.5	2	2.5
28	36	48	64	1.2857	1.7243	2.2857
176	232	304	400	1.3182	1.7273	2.2727
1112	1464	1928	2536	1.3165	1.7338	2.2806
7040	9264	12192	16048	1.3159	1.7318	2.2795

TABLE 3

Here (4) expresses entries on or below the main diagonal of the ladder as a linear combination of entries in the first column while (5) relates entries in the first column of the array to the r entries immediately above the given entry.

To prove (4) we shall use induction on j. Let S be the set

$$S=\{\,j\,|\,\text{equation (4) holds for each } i\geqslant j\,\}.$$

We wish to show that $S=\{0,1,\dots,r-1\}$. Obviously, $0\in S$. Suppose $j\in S$ where $j\leqslant r-2$. Then if $i\geqslant j+1$ we have

$$a_{i,j+1}=a_{ij}+(n-1)\,a_{i-1,j}.$$

But since $j\in S$ and $i\geqslant j+1$, we may use (4) to expand both terms on the right side and

$$\begin{aligned}a_{i,j+1}&=\sum_{k=0}^{j}\binom{j}{k}(n-1)^{k}a_{i-k,0}+\sum_{k=0}^{j}\binom{j}{k}(n-1)^{k+1}a_{i-1-k,0}\\&=\sum_{k=0}^{j+1}\binom{j+1}{k}(n-1)^{k}a_{i-k,0}\end{aligned}$$

where we have changed the index of summation on the second sum before combining with the corresponding terms of the first sum. Hence $j+1\in S$ and this completes the proof of (4).

To prove (5) we first write a_{i0} as a sum using (2) and then use (4) to express each term of the sum in terms of entries in the first column. That is, if $i\geqslant r$,

$$\begin{aligned}a_{i0}&=\sum_{k=0}^{r-1}a_{i-1,k}\\&=\sum_{k=0}^{r-1}\sum_{m=0}^{k}\binom{k}{m}(n-1)^{m}a_{i-1-m,0}.\end{aligned}$$

Interchanging the order of summation and using the identity

$$\sum_{k=m}^{r-1}\binom{k}{m}=\binom{r}{m+1}$$

(see [**1**, p. 49]), we may rewrite the above as

$$\begin{aligned} a_{i0} &= \sum_{m=0}^{r-1}\left[(n-1)^m a_{i-1-m,0}\sum_{k=m}^{r-1}\binom{k}{m}\right] \\ &= \sum_{m=0}^{r-1}\binom{r}{m+1}(n-1)^m a_{i-1-m,0} \\ &= \sum_{k=1}^{r}\binom{r}{k}(n-1)^{k-1}a_{i-k,0}. \end{aligned}$$

This proves (5).

To prove Theorem 1 we need some elementary concepts from the theory of difference equations. A treatment of the necessary results may be found in chapter 5 of [**4**]. We shall consider rth order linear homogeneous difference equations with constant coefficients where the difference equation is of the form

$$x_{i+r} = \alpha_{r-1}x_{i+r-1} + \alpha_{r-2}x_{i+r-2} + \cdots + \alpha_0 x_i. \tag{6}$$

This difference equation may be solved in a manner somewhat analogous to the solution of an rth order linear homogeneous differential equation and we shall briefly outline the facts we shall need later. The characteristic polynomial is

$$P(\lambda) = \lambda^r - \alpha_{r-1}\lambda^{r-1} - \alpha_{r-2}\lambda^{r-2} - \cdots - \alpha_0. \tag{7}$$

If the characteristic polynomial has distinct zeros $\lambda_1, \lambda_2, \dots, \lambda_r$, then the general solution of (6) is

$$x_i = c_1\lambda_1^i + c_2\lambda_2^i + c_3\lambda_3^i + \cdots + c_r\lambda_r^i. \tag{8}$$

That is, every sequence x_i which satisfies (6) must be of the form (8) where the c_i's are suitably chosen.

If the first r members of the sequence are given, then we can substitute these values into (8) and obtain the system of equations:

$$\begin{array}{lllll} c_1 & +c_2 & +\cdots & +c_r & = x_0 \\ c_1\lambda_1 & +c_2\lambda_2 & +\cdots & +c_r\lambda_r & = x_1 \\ \vdots & & & & \\ c_1\lambda_1^{r-1} & +c_2\lambda_2^{r-1} & +\cdots & +c_r\lambda_r^{r-1} & = x_{r-1}. \end{array} \tag{9}$$

We can then solve this system to determine the values for $c_1, c_2, \dots, c_r$. The determinant of the matrix of coefficients is

$$V(\lambda_1, \lambda_2, \dots, \lambda_r) = \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_r \\ \vdots & & & & \\ \lambda_1^{r-1} & \lambda_2^{r-1} & \lambda_3^{r-1} & \cdots & \lambda_r^{r-1} \end{vmatrix}.$$

This is the well-known Vandermonde determinant whose value is

$$V(\lambda_1, \lambda_2, \dots, \lambda_r) = \prod_{i>j}(\lambda_i - \lambda_j),$$

which is nonzero if the λ_i's are distinct.

Before proceeding further we present a theorem concerning the zeros of the polynomial (7). Although this theorem is only a slight modification of a classic theorem of Cauchy [**5**, p. 122], we present the simple proof for completeness.

THEOREM 2. *Consider the polynomial* (7) *where* $\alpha_0, \alpha_1, \ldots, \alpha_{r-1}$ *are all positive real numbers and let* λ_1 *be the positive real zero of this polynomial. Then all the other zeros are of smaller absolute value than* λ_1.

Proof. The theorem is obviously true if $r=1$, so assume that $r>1$. If z is a complex number which is not a positive real number, then z may be written in polar form as $z=\rho e^{i\theta}$, where $0<\theta<2\pi$. Then

$$|P(z)| = \left|\rho^r e^{ir\theta} - \left(\alpha_{r-1}\rho^{r-1}e^{i(r-1)\theta} + \cdots + \alpha_1\rho e^{i\theta} + \alpha_0\right)\right|$$
$$< \left|\rho^r e^{ir\theta}\right| - \left|\alpha_{r-1}\rho^{r-1}e^{i(r-1)\theta} + \cdots + \alpha_1\rho e^{i\theta} + \alpha_0\right|.$$

But $|\rho^r e^{ir\theta}| = \rho^r$ and so

$$\left|\alpha_{r-1}\rho^{r-1}e^{i(r-1)\theta} + \cdots + \alpha_1\rho e^{i\theta} + \alpha_0\right| \leqslant \alpha_{r-1}\rho^{r-1} + \cdots + \alpha_2\rho^2 + \left|\alpha_1\rho e^{i\theta} + \alpha_0\right|.$$

But if $0<\theta<2\pi$, $|\alpha_1\rho e^{i\theta}+\alpha_0| < \alpha_1\rho + \alpha_0$. Consequently,

$$|P(z)| > \rho^r - \left(\alpha_{r-1}\rho^{r-1} + \alpha_{r-2}\rho^{r-2} + \cdots + \alpha_1\rho + \alpha_0\right)$$
$$= P(|z|).$$

Since $P(\lambda)$ has only one positive real zero, λ_1, and $P(0) = -\alpha_0 < 0$, it follows that $P(\lambda)$ is negative on $[0, \lambda_1)$ and positive on $(\lambda_1, +\infty)$. If z is a zero of $P(\lambda)$ which is not a positive real number, we have that $|P(z)| > P(|z|)$ so $P(|z|) < 0$. But since $|z|$ is a positive real number it follows that $|z| < \lambda_1$.

Theorem 1 is a statement about the nature of the limit of a sequence of numbers. For this reason we are interested in the limiting behavior of sequences of numbers satisfying difference equations such as (6). We need the following theorem.

THEOREM 3. *Consider the difference equation* (6) *where* $\alpha_0, \alpha_1, \ldots, \alpha_{r-1}$ *are positive real numbers and the zeros of* $P(\lambda)$ *are distinct. If* $\{x_i\}$ *is any sequence satisfying* (6) *such that* $x_0, x_1, x_2, \ldots, x_{r-1}$ *are all positive numbers, then*

$$\lim_{i\to\infty} \frac{x_{i+1}}{x_i} = \lambda_1, \tag{10}$$

where λ_1 *is the positive zero of* $P(\lambda)$.

Proof. Let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be the zeros of $P(\lambda)$. Then

$$x_i = c_1\lambda_1^i + c_2\lambda_2^i + \cdots + c_r\lambda_r^i$$

where we wish to show that $c_1 \neq 0$. By Cramer's rule, $c_1 = D_1/D$ where D is the Vandermonde determinant and

$$D_1 = \begin{vmatrix} x_0 & 1 & 1 & \cdots & 1 \\ x_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_r \\ \vdots & & & & \\ x_{r-1} & \lambda_2^{r-1} & \lambda_3^{r-1} & \cdots & \lambda_r^{r-1} \end{vmatrix}.$$

We need to show that $D_1 \neq 0$. Let δ_i be the minor of x_i so that

$$D_1 = x_0\delta_0 - x_1\delta_1 + x_2\delta_2 - x_3\delta_3 + \cdots + (-1)^{r-1}x_{r-1}\delta_{r-1},$$

and define $Q(\lambda)$ as

$$Q(\lambda) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda & \lambda_2 & \cdots & \lambda_r \\ \lambda^2 & \lambda_2^2 & \cdots & \lambda_r^2 \\ \vdots & & & \\ \lambda^{r-1} & \lambda_2^{r-1} & \cdots & \lambda_r^{r-1} \end{vmatrix}.$$

Expanding along the first column, we have

$$Q(\lambda) = \delta_0 - \delta_1\lambda + \delta_2\lambda^2 - \cdots + (-1)^{r-1}\delta_{r-1}\lambda^{r-1}.$$

Since $\delta_{r-1} = V(\lambda_2, \lambda_3, \ldots, \lambda_r)$, we see that $\delta_{r-1} \neq 0$ and $Q(\lambda)$ is a polynomial of degree $r-1$ whose leading coefficient is $(-1)^{r+1}V(\lambda_2, \lambda_3, \ldots, \lambda_r)$. Moreover $\lambda_2, \lambda_3, \ldots, \lambda_r$ are the zeros of $Q(\lambda)$, so

$$Q(\lambda) = (-1)^{r+1}V(\lambda_2, \lambda_3, \ldots, \lambda_r)(\lambda - \lambda_2)(\lambda - \lambda_3) \cdots (\lambda - \lambda_r).$$

Also $P(\lambda)$ is a monic polynomial whose zeros are $\lambda_1, \lambda_2, \ldots, \lambda_r$ so

$$P(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_r) = k(\lambda - \lambda_1)Q(\lambda)$$

where

$$k = \frac{1}{(-1)^{r+1}V(\lambda_2, \lambda_3, \ldots, \lambda_r)}.$$

Therefore

$$\begin{aligned} &\lambda^r - \alpha_{r-1}\lambda^{r-1} - \alpha_{r-2}\lambda^{r-2} - \cdots - \alpha_0 \\ &\quad = k\left[-\lambda_1\delta_0 + (\delta_0 + \delta_1\lambda_1)\lambda - (\delta_1 + \delta_2\lambda_1)\lambda^2 + \cdots + (-1)^{r-1}\delta_{r-1}\lambda^r\right]. \end{aligned}$$

Equating coefficients, we see that

$$\begin{aligned} k\delta_0\lambda_1 &= \alpha_0 \\ k(\delta_0 + \delta_1\lambda_1) &= -\alpha_1 \\ k(\delta_1 + \delta_2\lambda_1) &= \alpha_2 \\ &\vdots \\ k\delta_{r-1} &= (-1)^{r-1}. \end{aligned}$$

Beginning with the first of these equations and working down the list, it is possible to show (using the fact that the α_i's are positive and $\lambda_1 > 0$) that the products $k\delta_0, k\delta_1, k\delta_2, \ldots$ are alternately positive and negative. Since

$$kD_1 = x_0k\delta_0 - x_1k\delta_1 + x_2k\delta_2 - \cdots + (-1)^{r-1}x_{r-1}k\delta_{r-1}$$

and $x_0, x_1, \ldots, x_{r-1}$ are all positive, we see that $kD_1 > 0$. Hence $D_1 \neq 0$.

Now that we have established that $c_1 \neq 0$ and that λ_1 has larger absolute value than any of the other zeros of $P(\lambda)$, we have

$$\lim_{i \to \infty} \frac{x_{i+1}}{x_1} = \lim_{i \to \infty} \frac{c_1\lambda_1^{i+1} + c_2\lambda_2^{i+1} + \cdots + c_r\lambda_r^{i+1}}{c_1\lambda_1^i + c_2\lambda_2^i + \cdots + c_r\lambda_r^i}$$

and the result follows by dividing the numerator and denominator by λ_1^i and letting $i \to \infty$. This proves Theorem 3.

We can now prove Theorem 1. Let $\{x_i\}$ be the sequence of numbers found in the first column of our ladder, that is, $x_i = a_{i0}$. Then by (5),

$$x_i = \binom{r}{1} x_{i-1} + \binom{r}{2}(n-1) x_{i-2} + \cdots + \binom{r}{r}(n-1)^{r-1} x_{i-r}$$

which may be rewritten as, if we replace i by $i+r$,

$$x_{i+r} = \binom{r}{1} x_{i+r-1} + \binom{r}{2}(n-1) x_{i+r-2} + \cdots + \binom{r}{r}(n-1)^{r-1} x_i. \qquad (6)'$$

But this is a difference equation of the form (6) where the α's are all positive. Moreover, the characteristic polynomial is

$$P(\lambda) = \lambda^r - \binom{r}{1}\lambda^{r-1} - \binom{r}{2}(n-1)\lambda^{r-2} - \cdots - \binom{r}{r}(n-1)^{r-1} \qquad (7)'$$

and its characteristic roots must satisfy

$$\lambda^r = \binom{r}{1}\lambda^{r-1} + \binom{r}{2}(n-1)^{r-2}\lambda^{r-1} + \cdots + \binom{r}{r}(n-1)^{r-1}$$
$$= \frac{1}{n-1}\left[(\lambda+n-1)^r - \lambda^r\right].$$

Consequently,

$$\left(\frac{\lambda}{y+n-1}\right)^r = \frac{1}{n}.$$

Let $R_1, R_2, \ldots, R_r$ be the rth roots of $1/n$. Then the zeros of $P(\lambda)$ satisfy the equations

$$\frac{k}{\lambda_k + n - 1} = R_k$$

so

$$\lambda_k = \frac{(n-1)R_k}{1-R_k}.$$

Let R_1 be the positive rth root of $1/n$. Then $\lambda_1 = (n-1)R_1/(1-R_1)$ is the positive real root of the characteristic polynomial given in (7)$'$. Clearly $x_0, x_1, \ldots, x_{r-1}$ are all positive numbers, since we start with 1's in the first row. We may therefore apply Theorem 3 to conclude that

$$\lim_{i\to\infty} \frac{x_{i+1}}{x_i} = \lim_{i\to\infty} \frac{a_{i+1,0}}{a_{i0}} = \lambda_1. \qquad (10)'$$

Also,

$$\lim_{i\to\infty} \frac{a_{i-k,0}}{a_{i0}} = \lim_{i\to\infty} \frac{a_{i-k,0}}{a_{i-k+1,0}} \cdot \frac{a_{i-k+1,0}}{a_{i-k+2,0}} \cdots \frac{a_{i-1,0}}{a_{i0}} = \left(\frac{1}{\lambda_1}\right)^k.$$

Dividing (4) by a_{i0} and letting $i \to \infty$ we have

$$\lim_{i\to\infty} \frac{a_{ij}}{a_{i0}} = \lim_{i\to\infty} \sum_{k=0}^{j} \binom{j}{k}(n-1)^k \frac{a_{i-k,0}}{a_{i0}}$$
$$= \sum_{k=0}^{j} \binom{j}{k}(n-1)^k \left(\frac{1}{\lambda_1}\right)^k$$
$$= \left(1 + \frac{n-1}{\lambda_1}\right)^j = \frac{1}{R_1^j} = n^{j/r}.$$

This completes the proof of Theorem 1.

Rate of convergence

An efficient method of obtaining a rational approximation to $n^{j/r}$ is to use the Newton-Raphson method to find zeros of $f(x) = x^r - n^j$ with a rational starting value. TABLE 4 compares the first 10 iterates given by the Newton-Raphson method with $x_0 = 1$ to the ladder approximations given in TABLE 3 for $3^{3/4} \approx 2.279507057$. From these data we see that, for this starting value, the ladder approximations move immediately with steadily diminishing absolute error toward the correct value, whereas the Newton-Raphson method, because of the poor starting value, requires several iterates to approach the root. However, the rate of convergence for the ladder approximations is not rapid, as it sometimes takes more than one iteration to gain an additional significant digit of accuracy. Since the rate of convergence for the Newton-Raphson method is quadratic, once the iterates get fairly close to the root, the number of significant figures of accuracy roughly doubles with each iteration providing very rapid convergence. A reasonable combination of the two methods would be to use the ladder approximation given by the second row of the ladder as a starting value for the Newton-Raphson method. This value is

$$\frac{a_{2j}}{a_{20}} = \frac{r + j(n-1)}{r}.$$

Ladder Approx. a_{i3}/a_{i0}	Newton-Raphson Iterates
1.0	1.0
2.5	7.5
2.285714286	5.641
2.272727273	4.268354124
2.280575539	3.288066136
2.279545455	2.655930877
2.279454023	2.352239204
2.279518619	2.282811417
2.279506226	2.279514225
2.279506839	2.279507057

TABLE 4

One point in favor of the ladder approach is that we approximate all the roots $n^{j/r}$; $j = 1, 2, 3, \ldots, r-1$ with just one table. Additionally, the only operations involved in constructing the table are addition and multiplication by $n-1$, whereas the Newton-Raphson method requires exponentiation.

References

[1] Daniel A. Cohen, Basic Techniques of Combinatorial Theory, John Wiley & Sons, New York, 1978.

[2] Thomas Heath, A History of Greek Mathematics, Oxford University Press, London, 1980.

[3] ______, The Works of Archimedes, Dover, New York, 1953.

[4] Peter Henrici, Discrete Variable Methods in Ordinary Differential Equations, John Wiley & Sons, New York, 1962.

[5] Morris Marden, Geometry of Polynomials, 2nd ed., Mathematical Surveys 3, American Mathematical Society, Providence, 1966.

[6] Ivor Thomas, Greek Mathematical Works, Vol. 1, Harvard University Press, Cambridge, 1967.

[7] G. C. Vedora, Notes on Theon of Smyrna, Amer. Math. Monthly, 58 (1951) 576–683.

[8] F. V. Waugh & M. W. Wakefield, Side-and-diagonal numbers, this MAGAZINE, 40 (1969) 74–83.

Primitive Pythagorean Triples of Gaussian Integers

JAMES T. CROSS
The University of the South
Sewanee, TN 37375

Our story can be described roughly as follows: we display a method for generating all solutions in the Gaussian integers of the equation,

$$\alpha^2 + \beta^2 = \gamma^2,$$

keeping one eye on the well-known scheme for generating all solutions in the positive integers of the equation,

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

(generating all right triangles with sides having positive integer lengths). Our exposition was inspired by undergraduates, and, with an occasional assist, is accessible to them.

Background

Let Z denote the set of rational (ordinary) integers. A Pythagorean Triple in Z is a set $\{x, y, z\}$ of positive integers satisfying $x^2 + y^2 = z^2$. A Primitive Pythagorean Triple (PPT) in Z is a Pythagorean Triple such that $(x, y) = 1$ (x and y are relatively prime). Any Pythagorean Triple can be reduced to a PPT by dividing its members through by a positive integer. Every carpenter knows one PPT: $\{3,4,5\}$; some know another: $\{5,12,13\}$; a few know more. Now the condition that $(x, y) = 1$ implies that the members of a PPT are relatively prime in pairs. Exactly two of them must be odd. Considering the congruence, $x^2 + y^2 \equiv z^2 \pmod 4$, we find that z is odd. We arrange the nomenclature so that x is odd and regard a PPT as an ordered triple, (x, y, z). Now let C denote the set of all ordered pairs, (c, d), of positive integers with c odd, d even, and $(c, d) = 1$. It is well known that C can be used to generate all PPT's through the one-one mapping of C onto the set of all PPT's:

$$(c, d) \to (\pm(c^2 - d^2), 2cd, c^2 + d^2),$$

where the sign is chosen to make the first member of the triple positive. For examples, $(1,2) \to (3,4,5)$ and $(3,2) \to (5,12,13)$.

One attractive way (see, for example, [**3**, p. 1]) to show the existence of this mapping begins with the factorization,

$$(x + yi)(x - yi) = z^2,$$

in the Gaussian integers. (The Gaussian integers, G, are the subring of the complex numbers consisting of all $x + yi$, where x and y are members of Z.) In this article we broaden our horizons to solve the more general problem: Find all PPT's in G. We delay a precise definition of a G PPT until we have reviewed a few pertinent facts about G. It is easy to see that there is room for different interpretations of the problem; for example, are $\{3,4,5\}$ and $\{3,5i,4i\}$ both PPT's and if so, are they different? (This type of question is easily disposed of in the Z case by requiring that the members of the triple be positive and that x be odd.)

The Gaussian integers

For a more detailed look at the Gaussian integers than we are going to take, the reader is referred to [**2**, chapters 12, 14–15], [**4**, pp. 245–259], or the delightful paper [**1**] by Harley Flanders. Now G is an integral domain which "enjoys" unique factorization of its nonzero nonunit members into primes (except for the order of factors and the multiplication of factors by units). If

$\eta = x + yi$ is in G, we define its norm, $N(\eta)$, to be $\eta\bar{\eta} = (x+yi)(x-yi) = x^2 + y^2$. It is simple to check that $N(\eta)$ is a nonnegative member of Z and that the norm of a product is the product of the norms. Then $N(\eta)|N(\zeta)$ in Z if $\eta|\zeta$ in G.

What are the units in G? Well, if $u = x + yi$ is a unit, then $uu^* = 1$ for some u^* in G. Then $N(u) = x^2 + y^2 = 1$. It follows that if U denotes the group of units in G, then

$$U = \{1, i, -1, -i\}.$$

When we reflect a bit on the PPT problem in Z, we wonder whether there is a natural way to define "even" and "odd" Gaussian integers and whether this concept may be significant in the G PPT problem, as it is in the Z case; there is, and it will be. Let $\delta = 1 + i$. We shall see that δ plays a role in G rather like that played by 2 in Z. First we check that δ is prime in G. If $\eta\zeta = \delta$ in G, then $N(\eta)N(\zeta) = N(\delta) = 2$, so that $N(\eta)$ or $N(\zeta)$ is 1. This means that one of the factors is a unit, the factorization is trivial, and δ is prime. The argument applies as well to $\bar{\delta}$; both are primes. They are associates, however, because $\delta = i\bar{\delta}$. (Two members of G are associates if one is a unit times the other, or, equivalently, if each divides the other in G.) We observe that 2 factors in this way: $2 = \delta\bar{\delta} = -i\delta^2 = i\bar{\delta}^2$. Next, we observe that δ divides the Gaussian integer, $x + iy$, if and only if $x \equiv y \pmod 2$:

$$\frac{x+yi}{\delta} = \frac{(x+yi)\bar{\delta}}{\delta\bar{\delta}} = \frac{x+y}{2} + \frac{(y-x)}{2}i.$$

This complex number quotient is in G if and only if $x \equiv y \pmod 2$. Finally, we prove the following lemma, in which $[a]$ denotes a's class in the indicated quotient ring.

LEMMA 1. $G/(\delta) = \{[0],[1]\}$.

Proof. Let $\eta = x + yi$ be in G. If $x \equiv y \pmod 2$, then $\eta \equiv 0 \pmod \delta$. Otherwise $\eta - 1 = x - 1 + yi \equiv 0 \pmod \delta$ and $\eta \equiv 1 \pmod \delta$. Moreover, since δ does not divide 1, [0] and [1] are distinct.

Of course, $G/(\delta)$ is a field isomorphic to $Z/(2)$, and as the reader has probably conjectured, we say that η is even or odd according as η is congruent $(\text{mod } \delta)$ to 0 or 1. It follows that the sum of two even or two odd Gaussian integers is even, the sum of an even one and an odd one is odd, the product of two odd ones is odd, and the product of an even one with one of either type is even.

Interpretation of the problem

If $\{\alpha, \beta, \gamma\}$ is a triple of nonzero members of G with α and β relatively prime and $\alpha^2 + \beta^2 = \gamma^2$, then the members of the triple are relatively prime in pairs. Checking through for even and odd possibilities, we find again that exactly two of them are odd; we define a Gaussian integer PPT to be an ordered triple, (α, β, γ), of nonzero members of G that are relatively prime in pairs, α and γ are odd while β is even, and $\alpha^2 + \beta^2 = \gamma^2$. This definition excludes $(3, 5i, 4i)$, for example, from the set of PPT's. However, we still have difficulties: for instance, $(3,4,5)$, $(-3,4,5)$, and $(3i, 4i, 5i)$ all meet the requirements of the definition, but we don't want to regard them as different. To address this problem, we partition the set of all PPT's into equivalence classes. Let $(\alpha_1, \beta_1, \gamma_1) \sim (\alpha, \beta, \gamma)$ mean that both triples are PPT's and that there exist units u_1, u_2, and u_3 in G such that $\alpha_1 = u_1\alpha$, $\beta_1 = u_2\beta$, and $\gamma_1 = u_3\gamma$. Then $\sim$ denotes an equivalence relation; we let $[(\alpha, \beta, \gamma)]$ denote the class containing (α, β, γ). We now interpret the problem as that of finding a way to generate all such equivalence classes.

Our solution

Let (α, β, γ) be a PPT in G. Then

$$(\alpha + \beta i)(\alpha - \beta i) = \gamma^2.$$

We emphasize that α and β are in G; the members of this product are generally not conjugate. Now the following lemma is important for us.

LEMMA 2. *Let (α, β, γ) be a PPT in G. Then $\alpha + \beta i$ and $\alpha - \beta i$ are relatively prime.*

Proof. Suppose π is a common prime factor of $\alpha + \beta i$ and $\alpha - \beta i$. Then π divides both 2α and $2\beta i$. Since i is a unit and α and β are relatively prime, π divides 2. Then we may take π to be δ. But then since β is even and $\alpha + \beta i$ is even, α is even, so that $(\alpha, \beta) \neq 1$.

Now since γ^2 is the product of two relatively prime factors, each is a unit times a square: $\alpha + \beta i = u_1 a^2$ and $\alpha - \beta i = u_2 b^2$, for some units u_1 and u_2 and some members a and b in G. Each of a and b is odd because γ is odd. We would prefer that each of these units be 1, and the following lemma assures us that we may assume that it is.

LEMMA 3. *Let (α, β, γ) be a PPT in G. Then there exist a member, $(\alpha_1, \beta_1, \gamma_1)$, in $[(\alpha, \beta, \gamma)]$ and integers A and B in G such that $\alpha_1 + \beta_1 i = A^2$ and $\alpha_1 - \beta_1 i = B^2$.*

Proof. Let $\alpha + \beta i = u_1 a^2$ and $\alpha - \beta i = u_2 b^2$. Then $u_1 u_2 a^2 b^2 = \gamma^2$ implies that $u_2 = \pm u_1$ because $u_1 u_2$ is a square in U. Then

$$\left(\frac{\alpha}{u_1} + \frac{\beta}{u_1} i\right)\left(\frac{\alpha}{u_1} - \frac{\beta}{u_1} i\right) = \pm a^2 b^2 = \pm \gamma^2.$$

Letting $A = a$; $B = b$ or ib according as $u_2 = u_1$ or $u_2 = -u_1$; $\alpha_1 = \alpha/u_1$; $\beta_1 = \beta/u_1$; and $\gamma_1 = \gamma/u_1$, we get the result.

We have seen now that if (α, β, γ) is a PPT in G and $\alpha + \beta i = a^2$ and $\alpha - \beta i = b^2$, then a and b are relatively prime and both are odd. The reader has probably conjectured that we intend to use pairs of relatively prime odd members of G to generate PPT's. This is substantially correct; without further restrictions on these integers, however, we would not achieve uniqueness in the generating process. First, recall that the Gaussian integer $\eta = x + yi$ is odd if $x \not\equiv y \pmod 2$. Let us say that the odd Gaussian integer, $x + yi$, is $(\text{odd})_r$ if x is odd and $(\text{odd})_i$ if y is odd. The following lemma shows that we may restrict our generators to $(\text{odd})_r$ integers.

LEMMA 4. *Let (α, β, γ) be a PPT in G such that $\alpha + \beta i = a^2$ and $\alpha - \beta i = b^2$. Then there exist a member $(\alpha_1, \beta_1, \gamma_1)$ in $[(\alpha, \beta, \gamma)]$ and integers A and B in G that are $(\text{odd})_r$ with $\alpha_1 + \beta_1 i = A^2$ and $\alpha_1 - \beta_1 i = B^2$.*

Proof. If both a and b in the hypothesis are $(\text{odd})_i$, then ia and ib are $(\text{odd})_r$; letting $A = ia$ and $B = ib$ and letting $(\alpha_1, \beta_1, \gamma_1) = (-\alpha, -\beta, \gamma)$, we get the desired result. Consideration of $a^2 + b^2$ and $a^2 - b^2 \pmod 4$ shows that α is odd and β even if and only if both a and b are $(\text{odd})_r$ or $(\text{odd})_i$.

The remaining restriction that we intend to impose on our generators is that the real part of each be positive. Since a and b are $(\text{odd})_r$, their real parts are nonzero and since the factors $\alpha + \beta i = a^2$ and $\alpha - \beta i = b^2$ don't know the difference between a and $-a$ or between b and $-b$, it is clear that we may require that their real parts be positive.

EXAMPLE 1. Let $(\alpha, \beta, \gamma) = (-4 + i, 4 + 8i, 4 + 7i)$. If π is a common prime factor of α and β, then $\pi | \alpha | 17$ and $\pi | \beta | 80$ (α and β divide their norms). But $(17, 80) = 1$. We can check that $\alpha^2 + \beta^2 = \gamma^2$. Here $(\alpha + \beta i) = ia^2$, where $a = 3 + 2i$, and $(\alpha - \beta i) = -ib^2$, where $b = 2 + i$. Illustrating Lemma 3, we let $(\alpha_1, \beta_1, \gamma_1) = (\alpha/i, \beta/i, \gamma/i) = (1 + 4i, 8 - 4i, 7 - 4i)$. Then $(\alpha_1, \beta_1, \gamma_1)$ is in $[(\alpha, \beta, \gamma)]$, and $\alpha_1 + \beta_1 i = (3 + 2i)^2$ while $\alpha_1 - \beta_1 i = (1 - 2i)^2$.

We recapitulate. If (α, β, γ) is a PPT in G, there exist relatively prime $(\text{odd})_r$ integers, a and b, in G with real parts positive and a member $(\alpha_1, \beta_1, \gamma_1)$ in $[(\alpha, \beta, \gamma)]$ such that $\alpha_1 + \beta_1 i = a^2$ and $\alpha_1 - \beta_1 i = b^2$. (Note that $\alpha_1 = (a^2 + b^2)/2$, and $\beta_1 = (a^2 - b^2)/2i$.) We are now ready to go in the other direction. Let P denote the set of all equivalence classes of PPT's in G and let S denote the set of all pairs, $\{a, b\}$ of relatively prime $(\text{odd})_r$ members of G with positive real parts. We can state our main result as the following theorem.

THEOREM 1. *The mapping,*

$$\{a, b\} \to [((a^2+b^2)/2, (a^2-b^2)/2i, ab)],$$

from S is onto P and is one-one.

Proof. Let $\{a, b\}$ be in S and let $(\alpha, \beta, \gamma) = ((a^2+b^2)/2, (a^2-b^2)/2i, ab)$. Then α, β, and γ are in G ($a^2 \equiv b^2 \equiv 1 \pmod 2$) and $\alpha^2 + \beta^2 = \gamma^2$. If π is a common factor of α and β, then π divides both $\alpha + \beta i = a^2$ and $\alpha - \beta i = b^2$. But a and b are relatively prime; no such π exists, and α and β are relatively prime. Since a and b are $(\text{odd})_r$, it follows that α is odd and β even. Thus (α, β, γ) meets the requirements for a PPT. The discussion above shows that our mapping is onto P. To see that it is one-one, suppose both $\{a, b\}$ and $\{A, B\}$ are in S and that they map to the same class in P. Then

$$((a^2+b^2)/2, (a^2-b^2)/2i, ab) = (\alpha, \beta, \gamma) \sim$$
$$((A^2+B^2)/2, (A^2-B^2)/2i, AB) = (\alpha_1, \beta_1, \gamma_1).$$

By definition, then, there exist units u_1, u_2, and u_3 so that $(\alpha_1, \beta_1, \gamma_1) = (u_1\alpha, u_2\beta, u_3\gamma)$. Then noting that 1 and -1 are the only unit squares and using the fact that both (α, β, γ) and $(\alpha_1, \beta_1, \gamma_1)$ are PPT's we get

$$\alpha^2 + \beta^2 = \gamma^2 \quad \text{and} \quad \pm\alpha^2 \pm \beta^2 = \pm\gamma^2.$$

These two equations taken together imply that the signs across the latter are identical, whence $u_2 = \pm u_1$. Then

$$A^2 + B^2 = u_1(a^2 + b^2) \quad \text{and} \quad A^2 - B^2 = u_2(a^2 - b^2)$$

yield

$$A^2 = u_1 a^2 \quad \text{and} \quad B^2 = u_1 b^2$$

or

$$A^2 = u_1 b^2 \quad \text{and} \quad B^2 = u_1 a^2,$$

according as $u_2 = u_1$ or $u_2 = -u_1$. Then u_1 is a square so that each of u_1 and u_2 is 1 or -1. If $u_2 = u_1 = 1$, then $A^2 = a^2$ and $B^2 = b^2$, implying that $A = a$ and $B = b$ because each of these four integers has positive real part. The choice, $u_1 = 1$ and $u_2 = -1$, leads to the conclusion that $A = b$ and $B = a$; then $\{A, B\} = \{a, b\}$. Trying $u_1 = -1$ and $u_2 = 1$ we find that $A = \pm bi$; this is not possible because bi is $(\text{odd})_i$. Finally, $u_1 = u_2 = -1$ would require that $A = \pm ai$, which is impossible for the same reason. The mapping is one-one.

EXAMPLE 2. The pair $\{a, b\} = \{1+2i, 1-2i\}$ generates the PPT class [(3,4,5)], while $\{3+2i, 3-2i\}$ generates [(5,12,13)]. The pair $\{3+2i, 1-2i\}$ generates $[(1+4i, 8-4i, 7-4i)]$. (In the next paragraph we check that the members of each of these generating pairs are relatively prime.)

How do we get hold of relatively prime pairs to use for generators? We need to know more about the primes in G. For proofs of the following facts, the reader is advised to see the texts or paper mentioned above. We know that δ is an even prime (of course, so are $-\delta$, $i\delta$, and $-i\delta$, just as -2 is prime along with 2 in Z). How about odd primes in G? If p is a positive prime in Z and $p \equiv 3 \pmod 4$, then p is prime in G. On the other hand, if q is a positive prime in Z and $q \equiv 1 \pmod 4$, then $q = \pi\bar{\pi}$ for some conjugate nonassociate primes π and $\bar{\pi}$ in G. Now δ and primes of the types p and π are, together with their associates, the only primes in G. For examples, $5 = (1+2i)(1-2i)$, $13 = (3+2i)(3-2i)$, and $41 = (5+4i)(5-4i)$. A listing of the odd primes goes like this:

$$3, 7, 11, 19, 23, \ldots,$$
$$1+2i, 1-2i, 3+2i, 3-2i, 1+4i, 1-4i, \ldots, .$$

We get our generating pairs, $\{a, b\}$, as products of powers of these odd primes so that no prime divides both a and b. In Example 2, we see that in all the given pairs, the generators are relatively prime. (For another method of testing for common factors, one can convince himself that the Euclidean Algorithm can find GCD's (greatest common divisors) in G.)

Finding the Z PPT's among the G PPT's

Each PPT in Z is also one in G; we should be able to find in our methods the well known scheme for generating them. Recalling that C is the set of all ordered pairs, (c, d), of relatively prime positive integers with c odd and d even, let K denote the subset of S consisting of all $\{a, b\}$, where a and b are conjugate. Then the mapping,

$$(c, d) \to \{c+di, c-di\},$$

is one-one from C onto K. Letting $a = c + di$ and $b = c - di$, we see that the conjugate pair, $\{a, b\}$ in K generates the PPT class

$$\left[\left((a^2+b^2)/2, (a^2-b^2)/2i, ab\right)\right] = \left[(c^2-d^2, 2cd, c^2+d^2)\right].$$

The conjugate pairs in S generate PPT classes having Z PPT representatives. For example, $(1, 2) \to \{1+2i, 1-2i\}$, and the latter pair generates $[(3, 4, 5)]$.

A caveat

The careful reader has observed that our definition of equivalence classes does not place $(3, 4, 5)$ and $(5, 4i, 3)$, for example, in the same class. More generally, (α, β, γ) and $(\gamma, \beta i, \alpha)$ are in different classes. We could have avoided this unpleasantness, but we were not willing to pay the price. The class $[(3, 4, 5)]$ is generated by $\{1+2i, 1-2i\}$, while the class $[(5, 4i, 3)]$ is generated by $\{1, 3\}$. Should we insist on consolidating the two classes, our mapping from S to P would no longer be one-one. As was mentioned earlier, there are other (and maybe better) ways to interpret the problem.

My student, Charles Yeomans, had fun solving a version of this problem as a project in an elementary number theory class. The "even" and "odd" terminology is due to him. He then turned to other worlds to conquer and never got around to writing up his work. I am grateful to him and to other students who have helped to make teaching fun for me.

Thanks are due two of my colleagues, John Koehl and William M. Priestley, the former of whom read the present paper while the latter read an earlier paper concerned with another problem in the Gaussian integers; both made valuable suggestions. Thanks also go to the referees and editors.

References

[1] Harley Flanders, A tale of two squares—and two rings, this MAGAZINE, 58 (1985) 3–11.

[2] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, second edition, Oxford University Press, 1945.

[3] Daniel A. Marcus, Number Fields, Springer, 1977.

[4] Ivan Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, fourth edition, Wiley, 1980.

PROBLEMS

LOREN C. LARSON, Editor
BRUCE HANSON, Associate Editor
St. Olaf College

LEROY F. MEYERS, PAST EDITOR
The Ohio State University

Proposals

To be considered for publication, solutions should be received by September 1, 1986.

1237. *Proposed by David Gale, University of California, Berkeley.*

A two-person game is played in the following manner. Start with n points in the plane situated so that no three are collinear. Players take turns drawing line segments between these points. The only stipulation is that line segments are not permitted to intersect the interior of previously drawn line segments (they may share common endpoints). The first person who is unable to draw a line segment in this manner is the loser of the game. Show that the outcome of this game depends only on the given configuration of points and is independent of the strategies of the two players. Find a "formula" for determining the winner.

1238. *Proposed by Clark Kimberling, University of Evansville.*

a. Prove that the interior of a triangle ABC contains a point P for which the three triangles APB, BPC, CPA have congruent incircles.

*b. Is P uniquely determined? Can the radii be determined? What can you say about the properties of P?

1239. *Proposed by Bruce Hanson, St. Olaf College.*

Let $r_1, r_2, r_3, \ldots$ be any enumeration of the rationals in the interval $(0,1)$. For each $j \geqslant 1$, let $r_j = .a_{j,1}a_{j,2}a_{j,3}\ldots$ be a decimal representation of r_j.

a. Prove that the "main diagonal" $.a_{1,1}a_{2,2}a_{3,3}\ldots$ is irrational.

ASSISTANT EDITORS: CLIFTON CORZATT 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 appropriate for Mathematics Magazine. Each solution should begin on a separate sheet containing the solver's name and full address.

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

b. For arbitrary positive integers j and k, prove that the "diagonal" $.a_{j,k}a_{j+1,k+1}a_{j+2,k+2}\cdots$ is irrational.

1240. *Proposed by Stephen Wayne Coffman (student), Western Maryland College.*

Evaluate

$$\sum_{n=1}^{\infty} \frac{H_n}{n \cdot 2^n}, \quad \text{where} \quad H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

1241. *Proposed by Richard Johnsonbaugh, DePaul University, and Sadahiro Saeki, Kansas State University.*

Does there exist a compact metric space M with an isometry that is into, but not onto, M?

Quickies

Answers to the Quickies are on p. 120.

Q708. *Submitted by Allen Schwenk, Western Michigan University, Kalamazoo, Michigan.*

Show that

$$\sum_{j=0}^{t} \binom{s+j}{j} 2^{t-j} + \sum_{j=0}^{s} \binom{t+j}{j} 2^{s-j} = 2^{t+s+1}.$$

Q709. *Submitted by William P. Wardlaw, U. S. Naval Academy, Annapolis.*

Let A be a real 3×3 matrix with negative determinant. Show that there is no matrix B over the real numbers such that A is the classical adjoint of B.

Q710. *Submitted by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

If n is any positive integer, show that the number $T = (1/8)n(n+1)(n+2)(n+3)$ is a triangular number.

Solutions

Isoptic of an Ellipse **March 1985**

1211. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*
Find the locus of points under which an ellipse is seen under a constant angle.

Solution by Volkhard Schindler, Berlin, East Germany.
We consider the ellipse $x^2/a^2 + y^2/b^2 = 1$ in a rectangular (x, y) coordinate system. It is well known that the tangent to the ellipse at the point (x_1, y_1) has equation $x_1x/a^2 + y_1y/b^2 = 1$. Since the tangent has x-intercept a^2/x_1 and y-intercept b^2/y_1, the slope m of the tangent from a point (x, y) outside the ellipse is given by

$$m = \frac{y}{x - a^2/x_1} = \frac{y - b^2/y_1}{x},$$

so that

$$\frac{x_1}{a} = \frac{ma}{mx - y} \quad \text{and} \quad \frac{y_1}{b} = \frac{b}{y - mx}. \tag{1}$$

Since (x_1, y_1) lies on the ellipse, we have $[ma/(mx - y)]^2 + [b/(y - mx)]^2 = 1$, which after simplification becomes

$$(x^2 - a^2)m^2 - 2xym + (y^2 - b^2) = 0. \tag{2}$$

If α is the constant angle subtended by the ellipse, then we can number the roots m_1, m_2 of (2) so that $\tan\alpha = (m_1 - m_2)/(1 + m_1m_2)$. Hence

$$\tan^2\alpha = \frac{(m_1 - m_2)^2}{(1 + m_1m_2)^2} = \frac{(m_1 + m_2)^2 - 4m_1m_2}{(1 + m_1m_2)^2},$$

which remains valid if α is replaced by $180° - \alpha$. Since $m_1 + m_2 = 2xy/(x^2 - a^2)$ and $m_1m_2 = (y^2 - b^2)/(x^2 - a^2)$, we obtain

$$\tan^2\alpha = 4\frac{b^2x^2 + a^2y^2 - a^2b^2}{(x^2 + y^2 - a^2 - b^2)^2}. \tag{3}$$

In particular, if $\alpha = 180°$, then (3) reduces to the equation of the original ellipse, as it should. If $\alpha = 90°$, then (3) reduces to $x^2 + y^2 = a^2 + b^2$, which is the equation of a circle of radius $\sqrt{a^2 + b^2}$.

Since equation (3) is not convenient for plotting, we introduce polar coordinates ($x = r\cos\theta$, $y = r\sin\theta$). Then (3) becomes $r^4 - 2Ar^2 + B = 0$, where

$$A = a^2 + b^2 + 2(b^2\cos^2\theta + a^2\sin^2\theta)\cot^2\alpha,$$

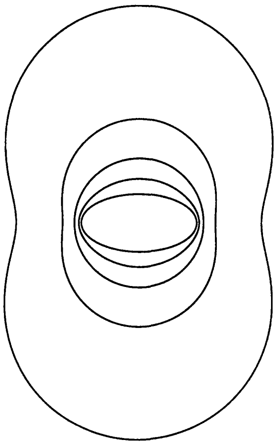
$$B = (a^2 + b^2)^2 + 4a^2b^2\cot^2\alpha,$$

from which we obtain

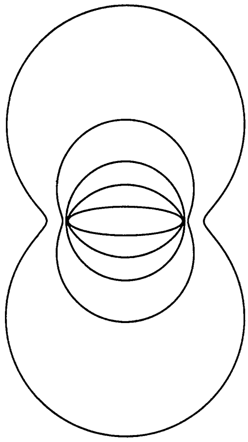
$$r^2 = A \pm \sqrt{A^2 - B}. \tag{4}$$

Since for fixed θ, r^2 decreases as α increases, we see that the plus sign in (4) is used when

$0^\circ < \alpha \leqq 90^\circ$, and the minus sign when $90^\circ \leqq \alpha < 180^\circ$. As seen from the figures, the loci are near-ellipses when $90^\circ < \alpha < 180^\circ$, and are nearly ovals of Cassini or lemniscates of Booth when $0^\circ < \alpha < 90^\circ$. Of course, if $a = b$, all loci are circles.



$a/b = 2.$



$a/b = 4.$

In each figure the values of α for the five curves, starting from the outermost, are 30°, 60°, 90°, 120°, and 180°.

Also solved by Michael V. Finn, J. T. Groenman (The Netherlands), L. Kuipers (Switzerland), Vania Mascioni (student, Switzerland), William A. Newcomb, Richard Parris, Stephanie Sloyan, and Robert L. Young. Solved partially by Zachary Franco (student) and the proposer.

M. S. Klamkin (Canada) found the result in R. C. Yates, *A Handbook on Curves and their Properties*, J. W. Edwards, Ann Arbor, 1947 (reprinted as *Curves and their Properties*, NCTM, 1974), pp. 138–140, where the terms *isoptic* and *orthoptic* are defined. None of the solvers considered the exceptional cases arising when, for example, x, x_1, y, y_1, or m is zero or m is infinite in (1).

An Inequality for the Logarithm — March 1985

1212. *Proposed by L. Bass and R. Výborný, The University of Queensland, Australia; and V. Thomée, Chalmers Institute of Technology, Sweden.*

Prove that if $x > 1$ and $0 < u < 1 < v$, then

$$\frac{v(x-1)(x^{v-1}-1)}{(v-1)(x^v-1)} < \log x < \frac{u(x-1)(1-x^{u-1})}{(1-u)(x^u-1)}.$$

Solution by M. S. Klamkin, University of Alberta, Canada.

The inequality can be rewritten as

$$\frac{F(v-2)}{F(v-1)} < \frac{F(-1)}{F(0)} < \frac{F(u-2)}{F(u-1)} \quad \text{for } 0 < x \neq 1 \quad \text{and } u < 1 < v,$$

where

$$F(\lambda) = \int_1^x t^\lambda \, dt \quad \text{for } x > 0 \text{ and any real } \lambda.$$

Hence it suffices to show that $F(\lambda)/F(\lambda+1)$ decreases as λ increases. This follows from the continuity of F and the fact that

$$\frac{d}{d\lambda}\left(\frac{F(\lambda)}{F(\lambda+1)}\right)=\frac{1}{(F(\lambda+1))^2}\left(\int_1^x t^{\lambda+1}\,dt\int_1^x t^{\lambda}(\log t)\,dt-\int_1^x t^{\lambda}\,dt\int_1^x t^{\lambda+1}(\log t)\,dt\right)$$

$$=\frac{1}{(F(\lambda+1))^2}\int_1^x\int_1^x s^{\lambda}t^{\lambda}(\log t)(s-t)\,ds\,dt$$

$$=\frac{1}{2(F(\lambda+1))^2}\int_1^x\int_1^x s^{\lambda}t^{\lambda}((s-t)\log t+(t-s)\log s)\,ds\,dt$$

$$=-\frac{1}{2(F(\lambda+1))^2}\int_1^x\int_1^x s^{\lambda}t^{\lambda}(s-t)(\log s-\log t)\,ds\,dt<0 \quad \text{if } \lambda\neq-1.$$

Also solved by L. Kuipers (Switzerland), Syrous Marivani, Vania D. Mascioni (student, Switzerland), William A. Newcomb, Bjorn Poonen (student), Heinz-Jürgen Seiffert (student, West Germany), J. M. Stark, and the proposers.

Marlow Sholander noted that the result follows from formula 3.12 of Leach and Sholander, "Extended mean values," *Amer. Math. Monthly* 85 (1978), pp. 84–90, in the form

$$E(u-1,u)<E(0,1)<E(v-1,v) \quad \text{for } u<1<v,$$

where

$$E(r,s)=\left(\frac{x^s-1}{x^r-1}\cdot\frac{r}{s}\right)^{1/(s-r)},$$

and the limiting case $E(0,1)=\lim_{r\to 0}E(r,1)$ is the logarithmic mean of 1 and x. In his paper, "Logarithmische Konvexität und Ungleichungsscharen," to appear in *Elemente der Mathematik*, *Mascioni* derives further inequalities from the fact that for fixed positive r, $F(\lambda)/F(\lambda+r)$ decreases as λ increases.

A Sum of Products — March 1985

1213. *Proposed by Nicholas K. Krier, Colorado State University, and Frank R. Bernhart, Rochester Institute of Technology.*

For each positive integer n let $S_n=\sum a_1a_2\dots a_n$, where the sum is taken over all n-tuples $(a_1,a_2,\dots,a_n)$ of positive integers such that $a_1=1$ and $a_{i+1}\leqq a_i+1$ for $1\leqq i<n$.

(a) How many terms are there in the sum?

(b) What is the value of S_n?

I. *Solution by Allen J. Schwenk, Western Michigan University.*

We shall show that the number of terms is the Catalan number $\binom{2n}{n}/(n+1)$ and that the sum is $(2n)!/(n!2^n)=(2n-1)!!$. The value $(2n-1)!!$ is well known as the number of ways to schedule matches for the first round of a tennis tournament with $2n$ players. We shall solve the problem by defining an equivalence relation having $\binom{2n}{n}/(n+1)$ classes such that each term in S_n gives the size of one class.

Let the players be numbered 1 through $2n$, and suppose that the lower-numbered player serves the first game. Define two schedules to be equivalent if they possess the same collection of n servers. How many sets $X=\{x_1,x_2,\dots,x_n\}$ can be server sets? Set X with $x_1<x_2<\cdots<x_n$ can be a server set if and only if $x_i\leqq 2i-1$ for all i. To see this, observe that if $x_i\geqq 2i$, then there are at least i receivers in the interval $[1,2i-1]$. But each receiver is served by a lower numbered server, so x_i belongs to the same interval, contradicting $x_i\geqq 2i$. On the other hand, $x_i\leqq 2i-1$ guarantees that the complementary set $Y=\{y_1,y_2,\dots,y_n\}$ with $y_1<y_2<\cdots<y_n$ has $y_i\geqq 2i>x_i$. Thus, pairing x_i with y_i for each i gives one schedule realizing X as a serving set.

The condition $x_i\leqq 2i-1$ guarantees that no interval $[1,j]$ has more receivers than servers. Thinking of servers as left parentheses and receivers as right parentheses, we see that the number of serving sets X is the number of legal arrangements of n pairs of parentheses, that is, $\binom{2n}{n}/(n+1)$.

Now how many ways can receivers be assigned consistent with a given serving set X? The receiver for the server x_i can be chosen in $a_i = 2i - x_i$ ways for $i = n, n-1, \ldots, 1$. Thus, exactly $a_1 a_2 \cdots a_n$ schedules have X as serving set. But $x_{i+1} \geqq x_i + 1$ implies $a_{i+1} = 2i + 2 - x_{i+1} \leqq 2i + 1 - x_i = 1 + a_i$. In other words, we have generated sequences $(a_1, a_2, \ldots, a_n)$ satisfying the conditions of the problem. Conversely, each such sequence $(a_1, a_2, \ldots, a_n)$ defines a serving set X by setting $x_i = 2i - a_i$. Thus, the given sum counts all schedules with each possible serving set, and so equals $(2n)!/(n!2^n)$, as claimed.

II. *Editor's composite of independent solutions by Judith Biasotti, Norwalk State Technical College; Peter W. Lindstrom, St. Anselm College; Richard Parris, Phillips Exeter Academy; and William P. Wardlaw, U.S. Naval Academy.*

Let A_n be the set of all n-tuples $(a_1, a_2, \ldots, a_n)$ of positive integers such that $a_1 = 1$ and $a_{i+1} \leqq a_i + 1$ for $1 \leqq i < n$, and let $A_{n,k}$ be the subset of A_n for which $a_n = k$. Let N_n be the number of elements in A_n, and S_n the sum of the products $a_1 a_2 \cdots a_n$ of elements $(a_1, a_2, \ldots, a_n)$ of A_n, with similar definitions for $N_{n,k}$ and $S_{n,k}$. Then

$$N_n = \sum_{k=1}^{n} N_{n,k} = N_{n+1,1}; \qquad N_{n,k} = 0 \quad \text{if } k < 1 \text{ or } k > n;$$

$$S_n = \sum_{k=1}^{n} S_{n,k} = S_{n+1,1}; \quad \text{and } S_{n,k} = 0 \quad \text{if } k < 1 \text{ or } k > n.$$

Furthermore,

$$N_{n,n} = 1 \quad \text{and } N_{n+1,k} = N_{n+1,k+1} + N_{n,k-1} \quad \text{for } n \geqq 1 \text{ and } 1 \leqq k \leqq n; \tag{1}$$

$$S_{n,n} = n! \quad \text{and} \quad S_{n+1,k} = \frac{k}{k+1} S_{n+1,k+1} + kS_{n,k-1} \quad \text{for } n \geqq 1 \text{ and } 1 \leqq k \leqq n. \tag{2}$$

The recursions (1) and (2) follow from the fact that the correspondence between $(a_1, a_2, \ldots, a_n, k)$ in $A_{n+1,k}$ and $(a_1, a_2, \ldots, a_n, k+1)$ in $A_{n+1,k+1}$ is one-to-one except when $a_n = k - 1$. Now (1) and (2) determine $N_{n,k}$ and $S_{n,k}$ for $1 \leqq k \leqq n$ (for fixed n, use induction on k downward from n), and a direct check shows that (1) and (2) are satisfied if we set

$$N_{n,k} = \frac{k}{2n-k}\binom{2n-k}{n} \quad \text{and} \quad S_{n,k} = \frac{n!}{2^{n-k}} N_{n,k} \quad \text{for } 0 \leqq k \leqq n. \tag{3}$$

Hence the answers to (a) and (b) are, respectively,

$$N_n = N_{n+1,1} = \frac{1}{2n+1}\binom{2n+1}{n+1} = \frac{1}{n+1}\binom{2n}{n}$$

and

$$S_n = S_{n+1,1} = \frac{n!}{2^n}\binom{2n}{n} = 1 \cdot 3 \cdot 5 \cdots (2n-1).$$

[The solvers used various stratagems to guess the formulas (3).]

III. *Solution by Kenneth L. Bernstein, MITRE Corporation.*

The solution requires two well-known relations among binomial coefficients:

$$b\binom{a+b}{a} = (a+1)\binom{a+b}{a+1}$$

and

$$\sum_{j=1}^{d}\binom{c-1+j}{c} = \binom{c+d}{c+1}.$$

The required sum S_n may be written as:

$$S_n = \sum_{a_1=1}^{a_0+1} a_1 \sum_{a_2=1}^{a_1+1} a_2 \cdots \sum_{a_n=1}^{a_{n-1}+1} a_n,$$

where $a_0 = 0$. Now the innermost (rightmost) summation may be written as:

$$\sum_{a_n=1}^{a_{n-1}+1} a_n\binom{a_n}{0} = 1\sum_{a_n=1}^{a_{n-1}+1}\binom{1-1+a_n}{1} = 1\binom{2+a_{n-1}}{2}.$$

By continuing in this manner, replacing successively the innermost remaining summation by a binomial coefficient with an extra factor, we obtain:

$$S_n = 1\cdot 3\cdots(2m-1)\sum_{a_1=1}^{a_0+1} a_1 \sum_{a_2=1}^{a_1+1} a_2 \cdots \sum_{a_{n-m}=1}^{a_{n-m-1}+1} a_{n-m}\binom{2m+a_{n-m}}{2m} \text{ for } 0 \leqq m \leqq n.$$

When $m = n$ we obtain

$$S_n = 1\cdot 3\cdots(2n-1)\binom{2n+0}{2n} = 1\cdot 3\cdots(2n-1).$$

To calculate the number N_n of terms in the sum, we have

$$N_n = \sum_{a_1=1}^{a_0+1}\sum_{a_2=1}^{a_1+1}\cdots\sum_{a_n=1}^{a_{n-1}+1} 1.$$

To reduce this sum we use the following relation:

$$\sum_{j=1}^{a+1}\frac{j+1}{m}\binom{2m+j}{m-1} = \frac{a+1}{m+1}\binom{2m+2+a}{m}.$$

Repeated use of the above formula, beginning at the innermost summation, yields

$$N_n = \sum_{a_1=1}^{a_0+1}\sum_{a_2=1}^{a_1+1}\cdots\sum_{a_{n-k}=1}^{a_{n-k-1}+1}\frac{a_{n-k}+1}{k}\binom{2k+a_{n-k}}{k-1} \quad \text{for } 1 \leqq k \leqq n$$

and, finally,

$$N_n = \frac{0+1}{n}\binom{2n+0}{n-1} = \frac{(2n)!}{n!(n+1)!}.$$

Also solved by J. C. Binz (Switzerland), St. Olaf College Problem Solving Group, James Propp, James G. Raisbeck (student), William P. Wardlaw (second solution), and the proposers. There were four incomplete or incorrect solutions.

The St. Olaf Group determined S_n under the more general conditions $a_1 = k$ and $1 \leqq a_{i+1} \leqq a_i + 1$ for $1 \leqq i < n$, where k is a fixed positive integer. *Propp* showed that

$$\sum\binom{a_1}{k}\binom{a_2}{k}\cdots\binom{a_n}{k} = \binom{k}{k}\binom{2k+1}{k}\binom{3k+2}{k}\cdots\binom{n(k+1)-1}{k},$$

where the summation is extended over all n-tuples $(a_1, a_2, \ldots, a_n)$ such that $a_1 = k$ and $k \leqq a_{i+1} \leqq a_i + k$ for $1 \leqq i < n$.

A Disguised Trigonometric Limit — March 1985

1214. *Proposed by Paul G. Nevai and G. A. Edgar, The Ohio State University.*

Let $A_0 = 1$ and $A_{n+1} = A_n + \sqrt{1 + A_n^2}$ for $n \geqq 0$. Show that $\lim_{n\to\infty} A_n/2^n$ exists, and find its value.

I. *Solution by Howard Morris, Chatsworth, California.*

More generally, if A_0 and λ are any real numbers with $\lambda > 0$, and if $A_{n+1} = A_n + \sqrt{\lambda^2 + A_n^2}$ for $n \geqq 0$, then

$$A_n = \lambda \cot \frac{\theta_0}{2^n},$$

where $\theta_n = \text{Arccot}(A_n/\lambda)$ for $n \geqq 0$ with $0 < \theta_n < \pi$. This follows by induction from

$$\frac{A_{n+1}}{\lambda} = \frac{A_n}{\lambda} + \sqrt{1 + \left(\frac{A_n}{\lambda}\right)^2}$$

and

$$\cot\frac{\theta}{2} = \frac{1+\cos\theta}{\sin\theta} = \cot\theta + \csc\theta = \cot\theta + \sqrt{1+\cot^2\theta} \quad \text{for } 0 < \theta < \pi.$$

Hence

$$\lim_{n\to\infty} \frac{A_n}{2^n} = \lim_{n\to\infty} \frac{\lambda/2^n}{\tan(\theta_0/2^n)} = \frac{\lambda}{\theta_0} = \frac{\lambda}{\text{Arccot}(A_0/\lambda)}.$$

In particular, if $A_0 = \lambda = 1$, then the limit is $4/\pi$.

For the similar recursion with $A_{n+1} = A_n + \sqrt{A_n^2 - \lambda^2}$, where $A_0 > \lambda > 0$, we have $A_n = \lambda \coth \frac{u_0}{2^n}$ and $\lim_{n\to\infty}(A_n/2^n) = \lambda/u_0$, where $u_0 = \text{Arccoth}(A_0/\lambda) > 0$.

Remark. The original problem is effectively that of approximating the radius of a circle of circumference 8 by using a circle inscribed in a regular 2^{n+3}-gon of perimeter 8.

II. *Solution by J. C. Binz, University of Bern, Switzerland.*

The limit is $4/\pi$. We prove first the somewhat more general

LEMMA. *Let* $(z_n)_{n=0}^{\infty}$ *be the complex sequence given by* $z_0 = c$ *with* $\text{Im}\, c > 0$, *and* $z_{n+1} = \frac{1}{2}(z_n + |z_n|)$ *for* $n \geqq 0$. *Then* $\lim_{n\to\infty} z_n$ *exists, and its value is* $(\text{Im}\, c)/(\text{Arg}\, c)$.

Let $c = \rho \exp i\varphi$ and $z_n = \rho_n \exp i\varphi_n$. We obtain the recursive formulas $\varphi_{n+1} = \frac{1}{2}\varphi_n$ and $\rho_{n+1} = \rho_n \cos \varphi_{n+1}$, and by iteration we have

$$z_n = \rho(\exp(2^{-n}i\varphi)) \prod_{k=1}^{n} \cos(2^{-k}\varphi) = \rho(\exp(2^{-n}i\varphi)) \prod_{k=1}^{n} \frac{\sin(2^{-k+1}\varphi)}{2\sin(2^{-k}\varphi)}$$

$$= \frac{\rho \sin\varphi}{2^n \sin(2^{-n}\varphi)} \exp(2^{-n} i\varphi).$$

It follows from $\lim_{n\to\infty} \exp(2^{-n}i\varphi) = 1$ and $\lim_{n\to\infty} 2^n \sin(2^{-n}\varphi) = \varphi$ that

$$\lim_{n\to\infty} z_n = \frac{\rho\sin\varphi}{\varphi} = \frac{\text{Im}\, c}{\text{Arg}\, c}.$$

For the original problem, put $B_n = A_n/2^n$, so that the recursion becomes $B_0 = 1$ and $B_{n+1} = \frac{1}{2}(B_n + \sqrt{B_n^2 + (2^{-n})^2})$ for $n \geqq 0$. If we set $z_n = B_n + 2^{-n}i$, we see that $z_0 = 1 + i$ and $z_{n+1} = \frac{1}{2}(z_n + |z_n|)$ for $n \geqq 0$. Hence, by the lemma,

$$\lim_{n\to\infty} B_n = \lim_{n\to\infty} z_n = \frac{1}{\pi/4} = \frac{4}{\pi}.$$

Also solved by Adib Abrishami(far) (Iran), Anon (Erewhon-upon-Spanish River), Frank P. Battles, Rich Bauer, Kenneth L. Bernstein, George Crofts, Roger Cuculière (France), Sheldon L. Degenhardt (student), Robert Doucette, Stephen E. Eldridge (England), Zachary Franco (student), Leon Gerber, Víctor Hernández (Spain), Dragan S. Janković, Kee-wai Lau (Hong Kong), Peter W. Lindstrom, Graham Lord, R. B. McNeill, Andreas Müller (student, Switzerland), William A. Newcomb, Gene M. Ortner, David Paget (Australia), Richard Parris, Bjorn Poonen (student), R. E. Rogers, Volkhard Schindler (East Germany), Allen J. Schwenk, Eberhard L. Stark (West Germany), John S. Sumner, L. Van Hamme (Belgium), Michael Vowe (Switzerland), G. C. Wake (New Zealand), Jet Wimp, Eric J. Wingler, and the proposers; partial solutions (existence of limit only) by Aqil M. Azmi (Saudi Arabia), Paul Bracken, L. Kuipers (Switzerland), L. Kuipers & P. Szüsz, Michiel Smid (student, The Netherlands), and Jan Söderkvist (student, Sweden). There was one incorrect solution.

1215. *Proposed by Andrea Laforgia, Università di Torino, Italy.*

Prove that

$$\tan x < \frac{\pi x}{\pi - 2x} \quad \text{for } 0 < x < \frac{\pi}{2}.$$

I. *Solution by Steven R. Weston, student, Pacific Lutheran University.*

First, recall that $\sin x < x$ for $0 < x < \pi/2$. Now if we linearly interpolate the cosine function between 0 and $\pi/2$, we obtain the function f defined by $f(x) = 1 - 2x/\pi$. Since the cosine function is concave downward on the interval $[0, \pi/2]$, we have $\cos x > f(x)$ when $0 < x < \pi/2$. Hence if $0 < x < \pi/2$, then

$$\tan x = \frac{\sin x}{\cos x} < \frac{x}{1 - \frac{2x}{\pi}} = \frac{\pi x}{\pi - 2x}.$$

II. *Solution by Yan-loi Wong, student, University of California, Berkeley.*

Since the cosine function is concave downward in $[0, \pi/2]$, we have $\cos t > 1 - 2t/\pi > 0$ for $0 < t < \pi/2$. Hence $\sec^2 t < (1 - 2t/\pi)^{-2}$, and so

$$\int_0^x \sec^2 t\, dt < \int_0^x \left(1 - \frac{2t}{\pi}\right)^{-2} dt,$$

or

$$\tan x < \frac{\pi x}{\pi - 2x} \quad \text{for } 0 < x < \frac{\pi}{2}.$$

III. *Solution by Zachary Franco, student, Brooklyn, New York.*

First note that the inequality to be proved is equivalent to

$$\frac{g(x)}{g\left(\frac{\pi}{2} - x\right)} < \frac{\pi}{2} \quad \text{for} \quad 0 < x < \frac{\pi}{2},$$

where $g(t) = (\sin t)/t$. Hence it is sufficient to show that the ratio of $g(t_1)$ to $g(t_2)$, for t_1 and t_2 in $(0, \pi/2)$, is bounded above by $\pi/2$. But this follows from knowing that $2/\pi < g(t) < 1$ for $0 < t < \pi/2$, which in turn is a consequence of the fact that g is decreasing on $(0, \pi/2)$ and $\lim_{t \to 0} g(t) = 1$. (*Proof:* $g'(t) = (\cos t)(t - \tan t)/t^2 < 0$ for $0 < t < \pi/2$.)

IV. *Comment by Vania D. Mascioni, student, ETH Zürich, Switzerland.*

Steċkin ("Some remarks on trigonometric polynomials" (Russian), *Uspehi Mat. Nauk* (N.S.) 10 (1955), no. 1 (63), 159–166; see also formula 3.4.29 in Mitrinović, *Analytic Inequalities*, p. 246) proved in one line that

$$\frac{4}{\pi} \cdot \frac{x}{\pi - 2x} < \tan x \quad \text{for} \quad 0 < x < \frac{\pi}{2}.$$

To obtain the inequality of this problem it is sufficient to set $x := \pi/2 - x$.

Also solved by fifty-six others, including seven students and the proposer. There were two incorrect solutions.

Eberhard L. Stark (West Germany) reports that his joint paper with Michael Becker, "On a hierarchy of quolynomial [*sic*!] inequalities for tan *x*," *Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat. Fiz.*, No. 620 (1978), pp. 133–138, contains a strengthened inequality:

$$\frac{8x}{\pi^2 - 4x^2} \leqq \tan x < \frac{\pi^2 x}{\pi^2 - 4x^2} \quad \text{for } 0 \leqq x \leqq \frac{\pi}{2}.$$

Answers

Solutions to the Quickies on p. 112.

A708. Consider the 2^{t+s+1} binary sequences of length $t+s+1$. Each sequence either has more than s 1's or has more that t 0's. Among the sequences with more than s 1's, there are $\binom{s+j}{j}2^{t-j}$ having their $(s+1)$st 1 in position $s+1+j$ where $0 \leqslant j \leqslant t$. Thus, the first sum counts sequences with more than s 1's. Similarly, the second sum counts sequences with more than t 0's. Thus, their total counts all sequences of length $t+s+1$.

A709. Suppose $A = \operatorname{adj} B$. Then from $AB = (\det B)I$ it follows that $(\det B)^2 = \det A < 0$. But this is impossible if B is real.

A710. T is a triangular number if for some positive integer k, one has $(1/8)\, n(n+1)\,(n+2)\,(n+3) = (1/2)\, k(k+1)$. By considering parity and noting that $n(n+3) < (n+1)\,(n+2)$, one is led to try $k = n(n+3)/2$. Then

$$k+1 = \frac{n(n+3)}{2} + 1 = \frac{(n+1)(n+2)}{2},$$

and the result follows.

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.

Ronan, Colin, and Mohs, Mayo, Scientist of the year: a sage for all seasons; though he's remembered for a comet Edmond Halley shone in many other modern sciences, *Discover* 7:1 (January 1986) 52-62.

Astronomer, editor of Newton's works, actuary, mathematician, inventor of a diving bell, and translator of Apollonius (from Greek and Arabic)--Halley was a consummate scientist. Too bad the weak passage of his comet has not brought more attention to the splendid man himself.

Parnas, David Lorge, Software aspects of strategic defense systems, *Communications of the ACM* 28:12 (December 1985) 1326-1335; also in *American Scientist* 73:5 (September-October 1985) 432-440, with letters and response in 74:1 (January-February 1986) 12-15.

Why Reagan's "star war's" defense may not work, according to a computer scientist who resigned from the presidential advisory committee (and its $1000/day consulting fee).

Allman, William F., Staying alive in the 20th century, *Science 85* 6:8 (1985) 30-41.

Valuable examination of the psychology of probability and risk-taking. Data on the actual risks of death from assorted causes are provided, along with assessments by the public of those risks. "Most people overestimated the numbers of deaths from causes that were sensational and underestimated more common causes of death that were less dramatic." In general, inability to cope with uncertainty leads to overestimation of small probabilities and underestimation of high ones. "The big question ... is whether our worries and fears, which are sometimes the result of faulty logic and misinformation ... should be considered when making decisions about risk." The article concludes with some curious specious reasoning that puts down scientists--a clue that *Science 85*'s authors are all non-scientist professional writers.

Brams, Steven J., *Rational Politics: Decisions, Games, and Strategy*, Congressional Quarterly Press, 1985; xiv + 233 pp, $16.95 (P).

"The major innovation of this book lies in its unified presentation of politics as a rational human activity, from ancient times to the present." Brams uses models based on game theory, decision theory, and social choice theory to analyze real-life situations such as the Cuban missile crisis, the Watergate tape case, and the Polish Solidarity confrontation. Many examples and ideas derive from his previous topical books; this book is an attempt to reach the broad audience of all students of political science.

Scharlau, Winfried, and Opolka, Hans, *From Fermat to Minkowski: Lectures on the Theory of Numbers and Its Historical Development*, Springer-Verlag, 1985; xi + 184 pp, $24.

" This book is not meant as a systematic introduction to number theory but rather as a historically motivated invitation to the subject, designed to interest the audience in number-theoretical questions and developments.... [W]e try to show how these theorems are the necessary consequences of natural questions.... The book will be successful if the reader understands that the representation of natural numbers by quadratic forms--say $n = x^2 + dy^2$ -- necessarily leads to quadratic reciprocity, or that Dirichlet, in his proof of the theorem on primes in arithmetical progression, simply had to find the class number formula." An excellent book at an advanced undergraduate level.

Kline, Morris, *Mathematics and the Search for Knowledge*, Oxford U Pr, 1985; vii + 257 pp, $19.95.

With this book, Kline has returned to his strong suit: the role of mathematics in the history of ideas. He begins with a survey of epistemology ("Is there an external world?"), then recounts the history of physics to illustrate "how mathematics reveals and determines our knowledge of the physical world." Rather than presenting the actual mathematics, he describes what is known about the physical world *only* through the medium of mathematics. "Contrary to the impression students acquire in school, mathematics is not just a series of techniques. It tells us what we have never known or even suspected about notable phenomena and in some instances even contradicts perception."

Clark, Colin W., *Bioeconomic Modelling and Fisheries Management*, Wiley, 1985; xii + 291 pp, $39.95.

Discusses the management of commercial fisheries, particularly with regard to the conflict between economic forces and biological factors; also, criticizes the historic pattern of overcapitalization followed by overfishing and shows how to avoid it. Simple mathematical models are derived and analyzed thoroughly. Calculus and simple differential equations predominate, along with some elementary probability; the optimization techniques of dynamic programming and optimal control theory are described briefly and used extensively behind the scenes.

Lang, Serge, *The Beauty of Doing Mathematics: Three Public Dialogues*, Springer-Verlag, 1985; xi + 127 pp, $19.80 (P).

It is noteworthy when a mathematician tries to explain non-trivial mathematics to the public, and even rarer when the attempt is successful. This book contains the transcription of three such successes, conducted by Serge Lang, on prime numbers, diophantine equations, and geometry of space. One only wishes that a book like this could appear in an inexpensive paperback edition that would be distributed and promoted widely to non-mathematicians. (A few grammatical errors occur in this translation from French, and an update footnote to p. 55 should tell that Mordell's conjecture has been proved.)

Turkle, Sherry, *The Second Self: Computers and the Human Spirit*, Simon & Schuster, 1984; 362 pp, $8.95 (P).

"[The question of mind in relation to machine] is becoming for us what sex was to the Victorians--threat and obsession, taboo and fascination." So the author concludes, after a psychological study of child and adult computer users that features anecdotal interviews à la Gail Sheehy's *Passages*. Turkle tries to explain the "holding power" of computers, the hackers' love for the machine, and the attraction of artificial intelligence. How will people see themselves in the computer age? This is her key question, and she has given us a preview of the answers.

NEWS & LETTERS

46th PUTNAM COMPETITION: WINNERS AND SOLUTIONS

Teams from 264 schools competed in the 1985 William Lowell Putnam mathematical competition. The top five winning teams, in descending rank, are:

Harvard University
Glenn D. Ellison, Douglas S. Jungreis, Michael Reid

Princeton University
Michael A. Abramson, Douglas R. Davidson, James C. Yeh

University of California, Berkeley
Michael J. McGrath, David P. Moulton, Jonathan E. Shapiro

Rice University
Charles R. Ferenbaugh, Thomas M. Hyer, Thomas M. Zavist

University of Waterloo
David W. Ash, Yong Yao Du, Kenneth W. Shirriff

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

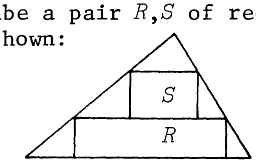
Martin V. Hildebrand	Williams College
Everett W. Howe	Cal. Inst. of Tech.
Douglas S. Jungreis	Harvard Univ.
Bjorn M. Poonen	Harvard Univ.
Keith A. Ramsay	Univ. of Chicago

Solutions to the 1985 Putnam problems were prepared for publication in this Magazine *by Loren Larson, St. Olaf College.*

A-1. Determine, with proof, the number of ordered triples (A_1, A_2, A_3) of sets which have the property that
(i) $A_1 \cup A_2 \cup A_3 = \{1,2,3,4,5,6,7,8,9,10\}$, and
(ii) $A_1 \cap A_2 \cap A_3 = \emptyset$,
where $\emptyset$ denotes the empty set. Express the answer in the form $2^a 3^b 5^c 7^d$, where a, b, c and d are nonnegative integers.

Sol. Every integer, $1 \le i \le 10$, falls into one of six mutually disjoint classes: $A_1 \cap \overline{A}_2 \cap \overline{A}_3$, $\overline{A}_1 \cap A_2 \cap \overline{A}_3$, $\overline{A}_1 \cap \overline{A}_2 \cap A_3$, $\overline{A}_1 \cap A_2 \cap A_3$, $A_1 \cap \overline{A}_2 \cap A_3$, and $A_1 \cap A_2 \cap \overline{A}_3$; hence there are $6^{10} = 2^{10}3^{10}$ different ordered triples.

A-2. Let T be an acute triangle. Inscribe a pair R,S of rectangles in T as shown:



Let $A(X)$ denote the area of polygon X. Find the maximum value, or show that no maximum exists, of $\dfrac{A(R) + A(S)}{A(T)}$, where T ranges over all triangles and R,S over all rectangles as above.

Sol. The region Q inside T but outside $R \cup S$ consists of five triangles. If we juxtapose the two triangles adjacent to R, and do the same to the triangles adjacent to S, we can think of Q as consisting of three small triangles similar to the given triangle. Let a,b,c denote the heights of these triangles (thus, $a+b+c$ is the height of the given triangle). The problem is equivalent to minimizing

$$\frac{A(Q)}{A(T)} = \frac{ka^2 + kb^2 + kc^2}{k(a+b+c)^2} = \frac{a^2 + b^2 + c^2}{(a+b+c)^2}.$$

By the Cauchy-Schwarz inequality,

$$(a + b + c)^2 \le (1 + 1 + 1)(a^2 + b^2 + c^2)$$
$$= 3(a^2 + b^2 + c^2)$$

with equality if and only if $a = b = c$. It follows that $\dfrac{A(Q)}{A(T)} \ge \dfrac{1}{3}$ and that the maximum value of $\dfrac{A(R) + A(S)}{A(T)}$ is $\dfrac{2}{3}$ (attained with $a = b = c$).

A-3. Let d be a real number. For each integer $m \geq 0$, define a sequence $\{a_m(j)\}$, $j = 0,1,2, \ldots$ by the conditions $a_m(0) = d/2^m$, and $a_m(j+1) = (a_m(j))^2 + 2a_m(j)$, $j \geq 0$. Evaluate $\lim_{n\to\infty} a_n(n)$.

Sol. We have $a_n(j+1) + 1 = (a_n(j) + 1)^2$, and hence by induction, $a_n(j) = (a_n(0) + 1)^{2^j}$. Therefore

$$\lim_{n\to\infty} a_n(n) = \lim_{n\to\infty} \left(1 + \frac{d}{2^n}\right)^{2^n} - 1 = e^d - 1.$$

A-4. Define a sequence $\{a_i\}$ by $a_1 = 3$ and $a_{i+1} = 3^{a_i}$ for $i \geq 1$. Which integers between 00 and 99 inclusive occur as the last two digits in the decimal expansion of infinitely many a_i?

Sol. We find that $a_3 \equiv 87 \pmod{100}$. If $a_i \equiv 87 \pmod{100}$ then $a_i = 87 + 100t$ and $a_{i+1} \equiv 3^{87+100t} \equiv 3^{7+20(4+5t)} \equiv 3^7 \equiv 87 \pmod{100}$. Thus, by induction, $a_i \equiv 87 \pmod{100}$ for all $i \geq 3$.

A-5. Let $I_m = \int_0^{2\pi} \cos(x)\cos(2x) \ldots \cos(mx)dx$. For which integers m, $1 \leq m \leq 10$, is $I_m \neq 0$?

Sol. Repeated use of the identity $\cos A_1 \cos A_2 = \frac{1}{2}(\cos(A_1+A_2) + \cos(A_1-A_2))$ shows that, in general, $\cos A_1 \cos A_2 \ldots \cos A_m =$

$$\frac{1}{2^{m-1}} \sum_{\varepsilon_k=\pm1} \cos(A_1 + \varepsilon_2 A_2 + \ldots + \varepsilon_m A_m).$$

It follows that $I_m =$

$$\frac{1}{2^{m-1}} \sum_{\varepsilon_k=\pm1} \int_0^{2\pi} \cos(1 + 2\varepsilon_2 + \ldots + m\varepsilon_m)x\, dx.$$

The integral $\int_0^{2\pi} \cos tx\, dx$ is zero if t is a nonzero integer and is 2π otherwise. Thus, $I_m \geq 0$, and $I_m \neq 0$ if and only if 0 can be written in the form $1 + 2\varepsilon_2 + \ldots + m\varepsilon_m$ for some $\varepsilon_2, \ldots, \varepsilon_m \in \{-1,1\}$. For a sum $1 + 2\varepsilon_2 + \ldots + m\varepsilon_m$, let r denote the sum of the positive terms and s the sum of the absolute values of the negative terms. Then $r-s = \frac{m(m+1)}{2}$. A necessary condition for $r=s$ is that $\frac{m(m+1)}{2}$ be even; that is, that $m \equiv 0$ or $3 \pmod 4$. Thus, the only candidates satisfying these conditions in $1 \leq m \leq 10$ are $m = 3,4,7$, and 8. We find that $I_m \neq 0$ for each of these because $1+2-3 = 0$, $1-2-3+4 = 0$, $(1+2-3) + (4-5-6+7) = 0$, and $(1-2-3+4) + (5-6-7+8) = 0$.

A-6. If $p(x) = a_0 + a_1x + \ldots + a_mx^m$ is a polynomial with real coefficients a_i, then set

$$\Gamma(p(x)) = a_0^2 + a_1^2 + \ldots a_m^2.$$

Let $f(x) = 3x^2 + 7x + 2$. Find, with proof, a polynomial $g(x)$ with real coefficients such that
(i) $g(0) = 1$, and
(ii) $\Gamma(f(x)^n) = \Gamma(g(x)^n)$,
for every integer $n \geq 1$.

Sol. Let $C(p(x))$ denote the constant term of $p(x)$. Observe that $\Gamma(p(x)) = C(p(x)p(1/x))$. Hence

$\Gamma((3x^2 + 7x + 2)^n) =$
$C((3x^2+7x+2)^n(3x^{-2}+7x^{-1}+2)^n)) =$
$C((3x+1)^n(x+2)^n(3x^{-1}+1)^n(x^{-1}+2)^n) =$
$C((3x+1)^n(3x^{-1}+1)^n(x+2)^n(x^{-1}+2)^n) =$
$C((3x+1)^n(3x^{-1}+1)^n(2x^{-1}+1)^n(2x+1)^n) =$
$C((3x+1)^n(2x+1)^n(3x^{-1}+1)^n(2x^{-1}+1)^n) =$
$C((6x^2+5x+1)^n(6x^{-2}+5x^{-1}+1)^n) =$
$\Gamma((6x^2 + 5x + 1)^n)$, and thus we can take $g(x) = 6x^2 + 5x + 1$.

B-1. Let k be the smallest positive integer with the following property:

There are distinct integers m_1, m_2, m_3, m_4, m_5 such that the polynomial

$$P(x) = (x-m_1)(x-m_2)(x-m_3)(x-m_4)(x-m_5)$$

has exactly k nonzero coefficients. Find, with proof, a set of integers m_1, m_2, m_3, m_4, m_5 for which this minimum k is achieved.

Sol. Write $p(x) = x^5 + ax^4 + bx^3 + cx^2 + dx + e$. By hypothesis, not both d and e can be zero. Also, $a^2 = \left(\sum_i m_i\right)^2 = \sum_i m_i^2 + 2 \sum_{i<j} m_i m_j = \sum_i m_i^2 + 2b$, and therefore, $a^2 - 2b = \sum_i m_i^2 > 0$. It follows that not both a and b can be zero. Thus $k \geq 3$.

Set $m_1 = -2$, $m_2 = -1$, $m_3 = 0$, $m_4 = 1$, $m_5 = 2$. Then $p(x) = x(x^2 - 1)(x^2 - 4) = x^5 - 5x + 4$. Hence $k = 3$, and this value of k is achieved for the given m_i's.

B-2. Define polynomials $f_n(x)$ for $n \geq 0$ by $f_0(x) = 1$, $f_n(0) = 0$ for $n \geq 1$, and

$$\frac{d}{dx}(f_{n+1}(x)) = (n + 1)f_n(x + 1)$$

for $n \geq 0$. Find, with proof, the explicit factorization of $f_{100}(1)$ into powers of distinct primes.

Sol. An examination of low order cases leads one to conjecture that $f_n(x) = x(x+n)^{n-1}$. Clearly this guess satisfies $f_0(x) = 1$, $f_n(0) = 0$ for $n \geq 1$. Now

$$f'_{n+1}(x) = (x+n+1)^n + nx(x+n+1)^{n-1}$$

$$= (n+1)(x+1)(x+n+1)^{n-1}$$

$$= (n+1)f_n(x+1) .$$

Hence $f_n(x) = x(x+n)^{n-1}$ as guessed. Therefore, $f_{100}(1) = 101^{99}$.

B-3. Let

$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$\cdots$
$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$\cdots$
$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$

be a doubly infinite array of positive integers, and suppose each positive integer appears exactly eight times in the array. Prove that $a_{m,n} > mn$ for some pair of positive integers (m,n).

Sol. Assume that $a_{m,n} \leq m \cdot n$ for all pairs of positive integers (m,n). Let $R(k) = \{(i,j) : i \cdot j \leq k^2\}$, and $F(k) = |R(k)|$. Observe that $F(k)$ is the number of lattice points lying under the graph of $y = k^2/x$. Hence

$$F(k) \geq \int_1^{k^2} \left(\frac{k^2}{x} - 1\right) dx =$$

$k^2(2\log k - 1) + 1$. When $k > e^5$, $F(k) > 9k^2$. Thus, for such k, there are more than $9k^2$ positions (i,j) which must be filled by elements $a_{i,j}$ less than or equal to k^2. But this is impossible because by hypothesis there are only $8k^2$ such numbers available.

B-4. Let C be the unit circle $x^2 + y^2 = 1$. A point p is chosen randomly on the circumference C and another point q is chosen randomly from the interior of C (these points are chosen independently and uniformly over their domains). Let R be the rectangle with sides parallel to the x- and y- axes with diagonal pq. What is the probability that no point of R lies outside of C?

Sol. Let $p = (\cos\theta, \sin\theta)$. It is not difficult to show that no point of R lies outside of C if and only if q lies entirely within the rectangle whose parallel sides are $x = \pm\cos\theta$ and $y = \pm\sin\theta$. For a given θ, the probability that q lies within this rectangle is

$\frac{2|\sin\theta|2|\cos\theta|}{\pi} = \frac{2}{\pi}|\sin 2\theta|$. Thus, the overall probability is

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{2}{\pi}|\sin 2\theta|d\theta = \frac{1}{2\pi}\cdot\frac{2}{\pi}\cdot 4 = \frac{4}{\pi^2} .$$

B-5. Evaluate $\int_0^\infty t^{-1/2}e^{-1985(t+t^{-1})}dt$. You may assume that $\int_{-\infty}^\infty e^{-x^2}dx = \sqrt{\pi}$.

Sol. Let $I = \int_0^\infty t^{-1/2}e^{-a(t+t^{-1})}dt$, where $a = 1985$. Let $u = 1/t$, and we find that $I = \int_0^\infty u^{-3/2}e^{-a(u+u^{-1})}du$.

Adding these last two integral equations we have $2I = \int_0^\infty (t^{-1/2}+t^{-3/2})e^{-a(t+t^{-1})}dt =$ $e^{-2a}\int_0^\infty (t^{-1/2}+t^{-3/2})e^{-a(t^{1/2}-t^{-1/2})^2}dt$.

Let $u = t^{1/2}-t^{-1/2}$, and the last equation is $I = e^{-2a}\int_{-\infty}^\infty e^{-au^2}du = \frac{\sqrt{\pi}}{\sqrt{a}}e^{-2a}$.

B-6. Let G be a finite set of real $n \times n$ matrices $\{M_i\}$, $1 \le i \le r$, which form a group under matrix multiplication. Suppose that $\sum_{i=1}^r tr(M_i) = 0$, where $tr(A)$ denotes the trace of the matrix A. Prove that $\sum_{i=1}^r M_i$ is the $n \times n$ zero matrix.

Sol. Let $S = \sum_{i=1}^r M_i$. For any j, $1 \le j \le r$, $M_jS = \sum_{i=1}^r M_jM_i = \sum_{i=1}^r M_i = S$, and hence $S^2 = \sum_{j=1}^r M_jS = rS$. Therefore the minimal polynomial $p(x)$ for S divides $x^2 - rx$ and every eigenvalue of S is zero. The only possible minimal polynomials are $p(x) = x$, $x-r$, $x(x-r)$. Since every zero of the minimal polynomial is an eigenvalue, the minimal polynomial is $p(x) = x$. Since S satisfies $p(S) = 0$, we have $S = 0$.

RAOUL HAILPERN HONORED

On January 10, 1986, at the Annual Business Meeting of the MAA in New Orleans, a Certificate of Merit was presented to Professor Raoul Hailpern of the State University of New York at Buffalo. Over the past 22 years Professor Hailpern has served the Association in a variety of ways, most recently as Editorial Director, in charge of production of the *Monthly*.

Professor Hailpern was born in Alexandria, Egypt and had a career in banking before giving up the world of business for mathematics. He took a bachelor's degree in London and master's and doctorate at Buffalo.

In the citation delivered in New Orleans Professor Paul Halmos said:

"Raoul Hailpern must surely be an instance of a uniqueness theorem. There cannot exist on this planet another person with his combination of mathematical knowledgeability, meticulous attention to detail, devotion to accuracy, and fanaticism for meeting deadlines. In all he has worked for fifteen tough editors who exhibited fifteen different styles of operations, but somehow he managed to get along with them all and to keep them all in line. I know that I couldn't have done half my work for the *Monthly* without his help, I know that my predecessor Ralph Boas would say the same, and I wasn't surprised to learn that in fact every one of the fifteen has extolled his virtues as an editorial colleague and helper.

"A Certificate of Merit is intended to recognize and to reward merit. According to my dictionary merit means value, excellence, superior performance. In his work for the Mathematical Association of America Raoul Hailpern has demonstrated all these qualities; in my opinion no one has ever deserved a Certificate of Merit more than he."

The Mathematical Association of America

CORPORATE MEMBERS

The Mathematical Association of America gratefully acknowledges the interest and support of the following:

CORPORATE MEMBERS

Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
Digital Equipment Corporation
International Business Machines Corporation
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Pergamon Press, Inc.
Prindle Weber and Schmidt, Inc.
Saunders College Publishing, Inc.
Scott Foresman & Company, Inc.
Springer-Verlag New York, Inc.
John Wiley & Sons, Inc.

CONTRIBUTING CORPORATE MEMBER

Bell Laboratories, Inc.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 59, NO. 2, April 1986