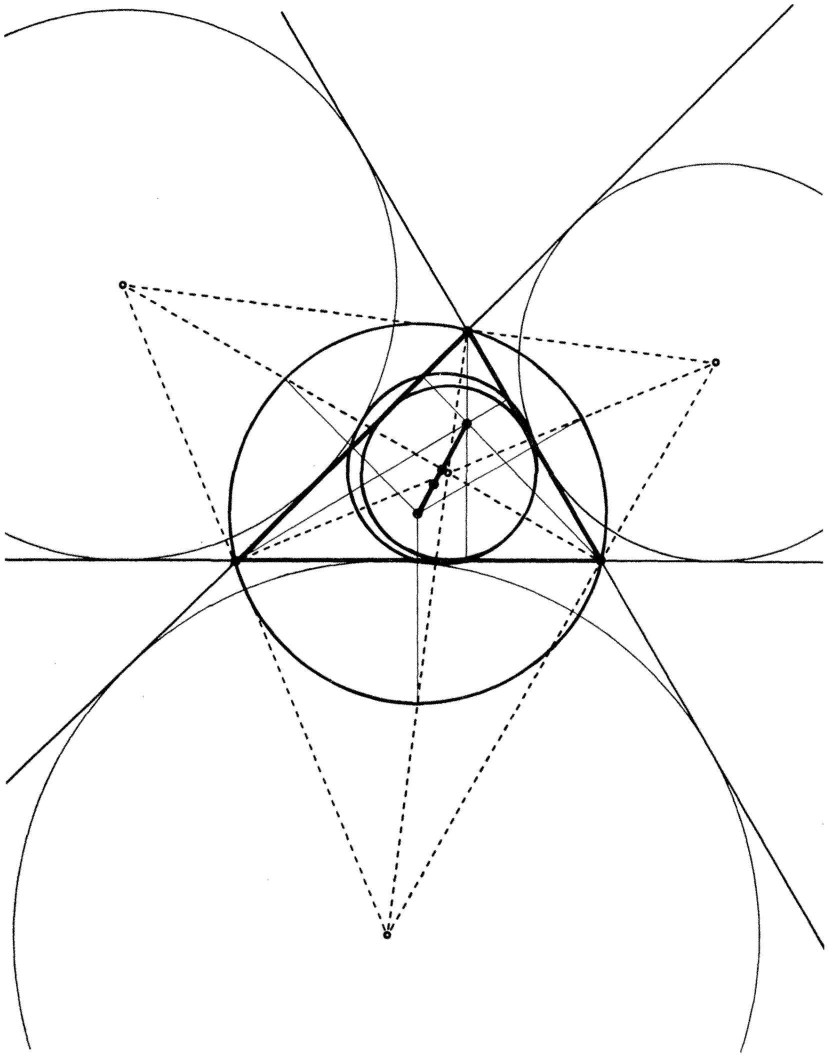
MATHEMATICS MAGAZINE



Vol. 56 No. 4
September, 1983

200 YEARS OF DERIVATIVE • VOTING PARADOXES
PROVING TRIG IDENTITIES • DERANGEMENTS

The Mathematical Association of America

CARUS MATHEMATICAL MONOGRAPHS

Students and teachers of mathematics as well as nonspecialists and workers in other fields, will all appreciate the expository excellence of these monographs. A wide range of topics in pure and applied mathematics are covered, each one written by a renowned expositor in the field. A glance at the list of titles of the books and the list of authors will convince you of the quality of this respected series of monographs. Twenty titles are currently available:

		List price	MAA Member
No. 1.	**Calculus of Variations,** by G. A. Bliss	$16.50	$12.00
No. 2.	**Analytic Functions of a Complex Variable,** by D. R. Curtiss	16.50	12.00
No. 3.	**Mathematical Statistics,** by H. L. Rietz	16.50	12.00
No. 4.	**Projective Geometry,** by J. W. Young	16.50	12.00
No. 6.	**Fourier Series and Orthogonal Polynomials,** by Dunham Jackson	16.50	12.00
No. 8.	**Rings and Ideals,** by N. H. McCoy	16.50	12.00
No. 9.	**The Theory of Algebraic Numbers** (Second edition) by Harry Pollard, and H. G. Diamond	16.50	12.00
No. 10.	**The Arithmetic Theory of Quadratic Forms,** by B. W. Jones	16.50	12.00
No. 11.	**Irrational Numbers,** by Ivan Niven	16.50	12.00
No. 12.	**Statistical Independence in Probability, Analysis and Number Theory,** by Mark Kac	16.50	12.00
No. 13.	**A Primer of Real Functions** (Third edition), by R. P. Boas, Jr.	16.50	12.00
No. 14.	**Combinatorial Mathematics,** by H. J. Ryser	16.50	12.00
No. 15.	**Noncommutative Rings,** by I. N. Herstein	16.50	12.00
No. 16.	**Dedekind Sums,** Hans Rademacher and Emil Grosswald	16.50	12.00
No. 17.	**The Schwarz Function and its Applications,** by P. J. Davis	16.50	12.00
No. 18.	**Celestial Mechanics,** by Harry Pollard	16.50	12.00
No. 19.	**Field Theory and its Classical Problems,** by Charles Hadlock	21.00	16.00
No. 20.	**The Generalized Rieman Integral,** by R. M. McLeod	18.00	13.50

THE MATHEMATICAL ASSOCIATION OF AMERICA

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

MATHEMATICS MAGAZINE

Volume 56, Number 4, September 1983
(ISSN— 0025- 570X)

COVER: Wonders of a triangle; see pp. 235–238.
Design by Dwight Paine.

ARTICLES

195 The Changing Concept of Change: The Derivative from Fermat to Weierstrass, *by Judith V. Grabiner.*

207 Paradoxes of Preferential Voting, *by Peter C. Fishburn and Steven J. Brams.*

NOTES

215 On Proving Trigonometric Identities, *by M. S. Klamkin.*

221 Empty or Infinitely Full?, *by Minja K. Paik.*

223 A Note on Linear Diophantine Equations, *by Kenneth J. Whitcomb.*

224 Matchings, Derangements, Rencontres, *by D. Hanson, K. Seyffarth, and J. H. Weston.*

230 A Hardy Old Problem, *by Martin D. Landau and William R. Jones.*

232 An Alternative to the Integral Test, *by Gerald Jungck.*

235 Triangle Rhyme, *by Dwight Paine.*

PROBLEMS

239 Proposals Number 1175–1177.

240 Quickie Number 686.

240 Solutions Number 1150–1153.

244 Answer to Quickie 686.

REVIEWS

245 Reviews of recent books and expository articles.

NEWS AND LETTERS

250 Comments on recent issues; MAA writing awards; Olympiad problems.

EDITORIAL POLICY

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

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

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

AUTHORS

Judith Victor Grabiner ("The Changing Concept of Change: The Derivative from Fermat to Weierstrass") received her Ph.D. (1966) in the History of Science from Harvard University, and since 1972 has taught the history of science at California State University, Dominguez Hills. She is author of *The Origins of Cauchy's Rigorous Calculus* (M.I.T., 1981), chairman of the Southern California Section of the MAA, and Book Review Editor of *Historia Mathematica.* In 1982-1983, she was Visiting Professor of History at U.C.L.A.

Peter Fishburn ("Paradoxes of Preferential Voting"), a member of the Mathematical Studies Department at Bell Labs, does research in combinatorics and graph theory, ordered sets, measurement theory, preference and utility-decision theory, and social choice. He is the author of a text on social choice theory, and co-author with Steven Brams of the book *Approval Voting* (Birkhäuser, 1983).

Steven J. Brams is Professor of Politics at New York University. He has written several books, the most recent of which are *The Presidential Election Game* (1978), *Biblical Games: A Strategic Analysis of Stories in the Old Testament* (1980), and *Superior Beings: If They Exist, How Would We Know?* (1983).

ILLUSTRATIONS

Victor Katz and Rodney Hood provided the stamps depicting mathematicians, pp. 204-205.
Gary Eichelsdorfer produced the computer-generated graphs on p. 196.

All other illustrations were provided by the authors.

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

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

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

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

ARTICLES

The Changing Concept of Change: The Derivative from Fermat to Weierstrass

First the derivative was used, then discovered, explored and developed, and only then, defined.

JUDITH V. GRABINER
Department of History
University of California, Los Angeles
Los Angeles, CA 90024

Some years ago while teaching the history of mathematics, I asked my students to read a discussion of maxima and minima by the seventeenth-century mathematician, Pierre Fermat. To start the discussion, I asked them, "Would you please define a relative maximum?" They told me it was a place where the derivative was zero. "If that's so," I asked, "then what is the definition of a relative minimum?" They told me, *that's* a place where the derivative is zero. "Well, in that case," I asked, "what is the difference between a maximum and a minimum?" They replied that in the case of a maximum, the second derivative is negative.

What can we learn from this apparent victory of calculus over common sense?

I used to think that this story showed that these students did not understand the calculus, but I have come to think the opposite: they understood it very well. The students' answers are a tribute to the power of the calculus in general, and the power of the concept of derivative in particular. Once one has been initiated into the calculus, it is hard to remember what it was like *not* to know what a derivative is and how to use it, and to realize that people like Fermat once had to cope with finding maxima and minima without knowing about derivatives at all.

Historically speaking, there were four steps in the development of today's concept of the derivative, which I list here in chronological order. The derivative was first *used*; it was then *discovered*; it was then *explored and developed*; and it was finally *defined*. That is, examples of what we now recognize as derivatives first were used on an ad hoc basis in solving particular problems; then the general concept lying behind these uses was identified (as part of the invention of the calculus); then many properties of the derivative were explained and developed in applications both to mathematics and to physics; and finally, a rigorous definition was given and the concept of derivative was embedded in a rigorous theory. I will describe the steps, and give one detailed mathematical example from each. We will then reflect on what it all means—for the teacher, for the historian, and for the mathematician.

The seventeenth-century background

Our story begins shortly after European mathematicians had become familiar once more with Greek mathematics, learned Islamic algebra, synthesized the two traditions, and struck out on their own. François Vieta invented symbolic algebra in 1591; Descartes and Fermat independently

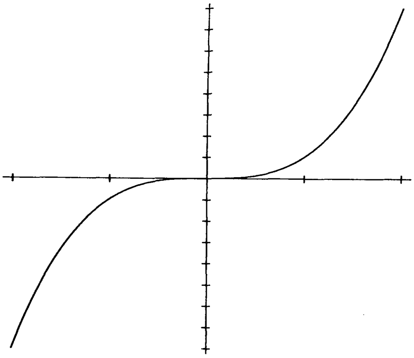


FIGURE 1

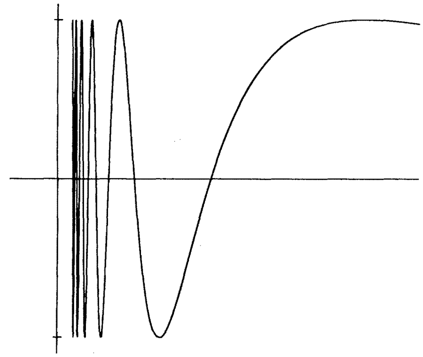


FIGURE 2

invented analytic geometry in the 1630's. Analytic geometry meant, first, that curves could be represented by equations; conversely, it meant also that every equation determined a curve. The Greeks and Muslims had studied curves, but not that many—principally the circle and the conic sections plus a few more defined as loci. Many problems had been solved for these, including finding their tangents and areas. But since any equation could now produce a new curve, students of the geometry of curves in the early seventeenth century were suddenly confronted with an explosion of curves to consider. With these new curves, the old Greek methods of synthetic geometry were no longer sufficient. The Greeks, of course, had known how to find the tangents to circles, conic sections, and some more sophisticated curves such as the spiral of Archimedes, using the methods of synthetic geometry. But how could one describe the properties of the tangent at an arbitrary point on a curve defined by a ninety-sixth degree polynomial? The Greeks had defined a tangent as a line which touches a curve without cutting it, and usually expected it to have only one point in common with the curve. How then was the tangent to be defined at the point $(0,0)$ for a curve like $y = x^3$ (FIGURE 1), or to a point on a curve with many turning points (FIGURE 2)?

The same new curves presented new problems to the student of areas and arc lengths. The Greeks had also studied a few cases of what they called "isoperimetric" problems. For example, they asked: of all plane figures with the same perimeter, which one has the greatest area? The circle, of course, but the Greeks had no general method for solving all such problems. Seventeenth-century mathematicians hoped that the new symbolic algebra might somehow help solve all problems of maxima and minima.

Thus, though a major part of the agenda for seventeenth-century mathematicians—tangents, areas, extrema—came from the Greeks, the subject matter had been vastly extended, and the solutions would come from using the new tools: symbolic algebra and analytic geometry.

Finding maxima, minima, and tangents

We turn to the first of our four steps in the history of the derivative: its *use*, and also illustrate some of the general statements we have made. We shall look at Pierre Fermat's method of finding maxima and minima, which dates from the 1630's [**8**]. Fermat illustrated his method first in solving a simple problem, whose solution was well known: *Given a line, to divide it into two parts so that the product of the parts will be a maximum.* Let the length of the line be designated B and the first part A (FIGURE 3). Then the second part is $B-A$ and the product of the two parts is

$$A(B-A)=AB-A^2. \tag{1}$$

Fermat had read in the writings of the Greek mathematician Pappus of Alexandria that a problem which has, in general, two solutions will have only one solution in the case of a maximum. This remark led him to his method of finding maxima and minima. Suppose in the problem just stated there is a second solution. For this solution, let the first part of the line be designated as $A+E$; the second part is then $B-(A+E)=B-A-E$. Multiplying the two parts together, we obtain

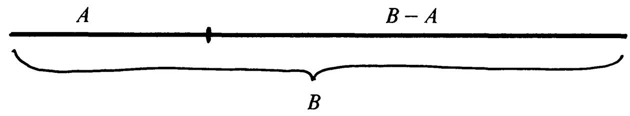


FIGURE 3

for the product

$$BA + BE - A^2 - AE - EA - E^2 = AB - A^2 - 2AE + BE - E^2. \tag{2}$$

Following Pappus' principle for the maximum, instead of two solutions, there is only one. So we set the two products (1) and (2) "sort of" equal; that is, we formulate what Fermat called the pseudo-equality:

$$AB - A^2 = AB - A^2 - 2AE + BE - E^2.$$

Simplifying, we obtain

$$2AE + E^2 = BE$$

and

$$2A + E = B.$$

Now Fermat said, with no justification and no ceremony, "suppress E." Thus he obtained

$$A = B/2,$$

which indeed gives the maximum sought. He concluded, "We can hardly expect a more general method." And, of course, he was right.

Notice that Fermat did not call E infinitely small, or vanishing, or a limit; he did not explain why he could first divide by E (treating it as nonzero) and then throw it out (treating it as zero). Furthermore, he did not explain what he was doing as a special case of a more general concept, be it derivative, rate of change, or even slope of tangent. He did not even understand the relationship between his maximum-minimum method and the way one found tangents; in fact he followed his treatment of maxima and minima by saying that the same method—that is, adding E, doing the algebra, then suppressing E—could be used to find tangents [**8**, p. 223].

Though the considerations that led Fermat to his method may seem surprising to us, he did devise a method of finding extrema that worked, and it gave results that were far from trivial. For instance, Fermat applied his method to optics. Assuming that a ray of light which goes from one medium to another always takes the quickest path (what we now call the Fermat least-time principle), he used his method to compute the path taking minimal time. Thus he showed that his least-time principle yields Snell's law of refraction [7] [**12**, pp. 387–390].

Though Fermat did not publish his method of maxima and minima, it became well known through correspondence and was widely used. After mathematicians had become familiar with a variety of examples, a pattern emerged from the solutions by Fermat's method to maximum-minimum problems. In 1659, Johann Hudde gave a general verbal formulation of this pattern [**3**, p. 186], which, in modern notation, states that, *given a polynomial of the form*

$$y = \sum_{k=0}^{n} a_k x^k,$$

there is a maximum or minimum when

$$\sum_{k=1}^{n} k a_k x^{k-1} = 0.$$

Of even greater interest than the problem of extrema in the seventeenth century was the finding of tangents. Here the tangent was usually thought of as a secant for which the two points came closer and closer together until they coincided. Precisely what it meant for a secant to "become" a

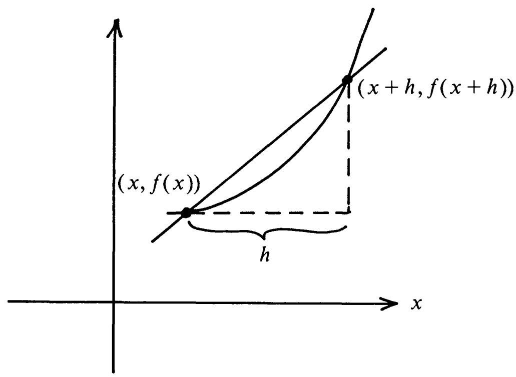


FIGURE 4

tangent was never completely explained. Nevertheless, methods based on this approach worked. Given the equation of a curve

$$y = f(x),$$

Fermat, Descartes, John Wallis, Isaac Barrow, and many other seventeenth-century mathematicians were able to find the tangent. The method involves considering, and computing, the slope of the secant,

$$\frac{f(x+h) - f(x)}{h},$$

doing the algebra required by the formula for $f(x+h)$ in the numerator, then dividing by h. The diagram in FIGURE 4 then suggests that when the quantity h vanishes, the secant becomes the tangent, so that neglecting h in the expression for the slope of the secant gives the slope of the tangent. Again, a general pattern for the equations of slopes of tangents soon became apparent, and a rule analogous to Hudde's rule for maxima and minima was stated by several people, including René Sluse, Hudde, and Christiaan Huygens [**3**, pp. 185–186].

By the year 1660, both the computational and the geometric relationships between the problem of extrema and the problem of tangents were clearly understood; that is, a maximum was found by computing the slope of the tangent, according to the rule, and asking when it was zero. While in 1660 there was not yet a general concept of derivative, there was a general method for solving one type of geometric problem. However, the relationship of the tangent to other geometric concepts—area, for instance—was not understood, and there was no completely satisfactory definition of tangent. Nevertheless, there was a wealth of methods for solving problems that we now solve by using the calculus, and in retrospect, it would seem to be possible to generalize those methods. Thus in this context it is natural to ask, how did the derivative as we know it come to be?

It is sometimes said that the idea of the derivative was motivated chiefly by physics. Newton, after all, invented both the calculus and a great deal of the physics of motion. Indeed, already in the Middle Ages, physicists, following Aristotle who had made "change" the central concept in his physics, logically analyzed and classified the different ways a variable could change. In particular, something could change uniformly or nonuniformly; if nonuniformly, it could change uniformly-nonuniformly or nonuniformly-nonuniformly, etc. [**3**, pp. 73–74]. These medieval classifications of variation helped to lead Galileo in 1638, without benefit of calculus, to his successful treatment of uniformly accelerated motion. Motion, then, could be studied scientifically. Were such studies the origin and purpose of the calculus? The answer is no. However plausible this suggestion may sound, and however important physics was in the later development of the calculus, physical questions were in fact neither the immediate motivation nor the first application of the calculus.

Certainly they prepared people's thoughts for some of the properties of the derivative, and for the introduction into mathematics of the concept of change. But the immediate motivation for the general concept of derivative—as opposed to specific examples like speed or slope of tangent—did not come from physics. The first problems to be solved, as well as the first applications, occurred in mathematics, especially geometry (see [**1**, chapter 7]; see also [**3**; chapters 4–5], and, for Newton, [**17**]). The concept of derivative then developed gradually, together with the ideas of extrema, tangent, area, limit, continuity, and function, and it interacted with these ideas in some unexpected ways.

Tangents, areas, and rates of change

In the latter third of the seventeenth century, Newton and Leibniz, each independently, invented the calculus. By "inventing the calculus" I mean that they did three things. First, they took the wealth of methods that already existed for finding tangents, extrema, and areas, and they subsumed all these methods under the heading of two general concepts, the concepts which we now call **derivative** and **integral**. Second, Newton and Leibniz each worked out a notation which made it easy, almost automatic, to use these general concepts. (We still use Newton's $\dot{x}$ and we still use Leibniz's dy/dx and $\int y\,dx$.) Third, Newton and Leibniz each gave an argument to prove what we now call the Fundamental Theorem of Calculus: the derivative and the integral are mutually inverse. Newton called our "derivative" a *fluxion*—a rate of flux or change; Leibniz saw the derivative as a ratio of infinitesimal differences and called it the *differential quotient*. But whatever terms were used, the concept of derivative was now embedded in a general subject—the calculus—and its relationship to the other basic concept, which Leibniz called the integral, was now understood. Thus we have reached the stage I have called *discovery*.

Let us look at an early Newtonian version of the Fundamental Theorem [**13**, sections 54–5, p. 23]. This will illustrate how Newton presented the calculus in 1669, and also illustrate both the strengths and weaknesses of the understanding of the derivative in this period.

Consider with Newton a curve under which the area up to the point $D=(x,y)$ is given by z (see FIGURE 5). His argument is general: "Assume any relation betwixt x and z that you please;" he then proceeded to find y. The example he used is

$$z=\frac{n}{m+n}ax^{(m+n)/n};$$

however, it will be sufficient to use $z=x^3$ to illustrate his argument.

In the diagram in FIGURE 5, the auxiliary line bd is chosen so that $Bb=o$, where o is not zero. Newton then specified that $BK=v$ should be chosen so that area $BbHK=$ area $BbdD$. Thus $ov=$ area $BbdD$. Now, as x increases to $x+o$, the change in the area z is given by

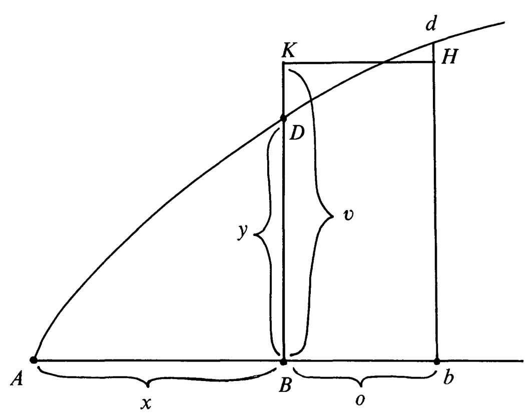


FIGURE 5

$$z(x+o)-z(x)=x^3+3x^2o+3xo^2+o^3-x^3=3x^2o+3xo^2+o^3,$$

which, by the definition of v, is equal to ov. Now since $3x^2o+3xo^2+o^3=ov$, dividing by o produces $3x^2+3ox+o^2=v$. Now, said Newton, "If we suppose *Bb* to be diminished infinitely and to vanish, or o to be nothing, v and y in that case will be equal and the terms which are multiplied by o will vanish: so that there will remain..."

$$3x^2=y.$$

What has he shown? Since $(z(x+o)-z(x))/o$ is the rate at which the area z changes, that rate is given by the ordinate y. Moreover, we recognize that $3x^2$ would be the slope of the tangent to the curve $z=x^3$. Newton went on to say that the argument can be reversed; thus the converse holds too. We see that derivatives are fundamentally involved in areas as well as tangents, so the concept of derivative helps us to see that these two problems are mutually inverse. Leibniz gave analogous arguments on this same point (see, e.g. [**16**, pp. 282–284]).

Newton and Leibniz did not, of course, have the last word on the concept of derivative. Though each man had the most useful properties of the concept, there were still many unanswered questions. In particular, what, exactly, is a differential quotient? Some disciples of Leibniz, notably Johann Bernoulli and his pupil the Marquis de l'Hospital, said a differential quotient was a ratio of infinitesimals; after all, that is the way it was calculated. But infinitesimals, as seventeenth-century mathematicians were well aware, do not obey the Archimedean axiom. Since the Archimedean axiom was the basis for the Greek theory of ratios, which was, in turn, the basis of arithmetic, algebra, and geometry for seventeenth-century mathematicians, non-Archimedean objects were viewed with some suspicion. Again, what is a fluxion? Though it can be understood intuitively as a velocity, the proofs Newton gave in his 1671 *Method of Fluxions* all involved an "indefinitely small quantity o," [**14**, pp. 32–33] which raises many of the same problems that the o which "vanishes" raised in the Newtonian example of 1669 we saw above. In particular, what is the status of that little o? Is it zero? If so, how can we divide by it? If it is not zero, aren't we making an error when we throw it away? These questions had already been posed in Newton's and Leibniz's time. To avoid such problems, Newton said in 1687 that quantities defined in the way that $3x^2$ was defined in our example were the *limit* of the ratio of vanishing increments. This sounds good, but Newton's understanding of the term "limit" was not ours. Newton in his *Principia* (1687) described limits as "ultimate ratios"—that is, the value of the ratio of those vanishing quantities just when they are vanishing. He said, "Those ultimate ratios with which quantities vanish are not truly the ratios of ultimate quantities, but limits towards which the ratios of quantities decreasing without limit do always converge; and to which they approach nearer than by any given difference, but never go beyond, nor in effect attain to, till the quantities are diminished in infinitum" [**15**, Book I, Scholium to Lemma XI, p. 39].

Notice the phrase "but never go beyond"—so a variable cannot oscillate about its limit. By "limit" Newton seems to have had in mind "bound," and mathematicians of his time often cite the particular example of the circle as the limit of inscribed polygons. Also, Newton said, "nor... attain to, till the quantities are diminished in infinitum." This raises a central issue: it was often asked whether a variable quantity ever actually reached its limit. If it did not, wasn't there an error? Newton did not help clarify this when he stated as a theorem that "Quantities and the ratios of quantities which in any finite time converge continually to equality, and before the end of that time approach nearer to each other than by any given difference, become ultimately equal" [**15**, Book I, Lemma I, p. 29]. What does "become ultimately equal" mean? It was not really clear in the eighteenth century, let alone the seventeenth.

In 1734, George Berkeley, Bishop of Cloyne, attacked the calculus on precisely this point. Scientists, he said, attack religion for being unreasonable; well, let them improve their own reasoning first. A quantity is either zero or not; there is nothing in between. And Berkeley characterized the mathematicians of his time as men "rather accustomed to compute, than to think" [**2**].

Perhaps Berkeley was right, but most mathematicians were not greatly concerned. The concepts of differential quotient and integral, concepts made more effective by Leibniz's notation and by the Fundamental Theorem, had enormous power. For eighteenth-century mathematicians, especially those on the Continent where the greatest achievements occurred, it was enough that the concepts of the calculus were understood sufficiently well to be applied to solve a large number of problems, both in mathematics and in physics. So, we come to our third stage: *exploration and development*.

Differential equations, Taylor series, and functions

Newton had stated his three laws of motion in words, and derived his physics from those laws by means of synthetic geometry [**15**]. Newton's second law stated: "*The change of motion* [*our 'momentum'*] *is proportional to the motive force impressed, and is made in the direction of the* [*straight*] *line in which that force is impressed*" [**15**, p. 13]. Once translated into the language of the calculus, this law provided physicists with an instrument of physical discovery of tremendous power—because of the power of the concept of the derivative.

To illustrate, if F is force and x distance (so $m\dot{x}$ is momentum and, for constant mass, $m\ddot{x}$ the rate of change of momentum), then Newton's second law takes the form $F = m\ddot{x}$. Hooke's law of elasticity (when an elastic body is distorted the restoring force is proportional to the distance [in the opposite direction] of the distortion) takes the algebraic form $F = -kx$. By equating these expressions for force, Euler in 1739 could easily both state and solve the differential equation $m\ddot{x} + kx = 0$ which describes the motion of a vibrating spring [**10**, p. 482]. It was mathematically surprising, and physically interesting, that the solution to that differential equation involves sines and cosines.

An analogous, but considerably more sophisticated problem, was the statement and solution of the partial differential equation for the vibrating string. In modern notation, this is

$$\frac{\partial^2 y}{\partial t^2} = \frac{T\partial^2 y}{\mu\partial x^2},$$

where T is the tension in the string and μ is its mass per unit length. The question of how the solutions to this partial differential equation behaved was investigated by such men as d'Alembert, Daniel Bernoulli, and Leonhard Euler, and led to extensive discussions about the nature of continuity, and to an expansion of the notion of function from formulas to more general dependence relations [**10**, pp. 502–514], [**16**, pp. 367–368]. Discussions surrounding the problem of the vibrating string illustrate the unexpected ways that discoveries in mathematics and physics can interact ([**16**, pp. 351–368] has good selections from the original papers). Numerous other examples could be cited, from the use of infinite-series approximations in celestial mechanics to the dynamics of rigid bodies, to show that by the mid-eighteenth century the differential equation had become the most useful mathematical tool in the history of physics.

Another useful tool was the Taylor series, developed in part to help solve differential equations. In 1715, Brook Taylor, arguing from the properties of finite differences, wrote an equation expressing what we would write as $f(x+h)$ in terms of $f(x)$ and its quotients of differences of various orders. He then let the differences get small, passed to the limit, and gave the formula that still bears his name: the Taylor series. (Actually, James Gregory and Newton had anticipated this discovery, but Taylor's work was more directly influential.) The importance of this property of derivatives was soon recognized, notably by Colin Maclaurin (who has a special case of it named after him), by Euler, and by Joseph-Louis Lagrange. In their hands, the Taylor series became a powerful tool in studying functions and in approximating the solution of equations.

But beyond this, the study of Taylor series provided new insights into the nature of the derivative. In 1755, Euler, in his study of power series, had said that for any power series,

$$a + bx + cx^2 + dx^3 + \cdots,$$

one could find x sufficiently small so that if one broke off the series after some particular term—say x^2—the x^2 term would exceed, in absolute value, the sum of the entire remainder of the series [**6**, section 122]. Though Euler did not prove this—he must have thought it obvious since he usually worked with series with finite coefficients—he applied it to great advantage. For instance, he could use it to analyze the nature of maxima and minima. Consider, for definiteness, the case of maxima. If $f(x)$ is a relative maximum, then by definition, for small h,

$$f(x-h)<f(x) \quad \text{and} \quad f(x+h)<f(x).$$

Taylor's theorem gives, for these inequalities,

$$f(x-h)=f(x)-h\frac{df(x)}{dx}+h^2\frac{d^2f(x)}{dx^2}-\cdots<f(x) \tag{3}$$

$$f(x+h)=f(x)+h\frac{df(x)}{dx}+h^2\frac{d^2f(x)}{dx^2}+\cdots<f(x). \tag{4}$$

Now if h is so small that $h\,df(x)/dx$ dominates the rest of the terms, the only way that both of the inequalities (3) and (4) can be satisfied is for $df(x)/dx$ to be zero. Thus the differential quotient is zero for a relative maximum. Furthermore, Euler argued, since h^2 is always positive, if $d^2f(x)/dx^2 \neq 0$, the only way both inequalities can be satisfied is for $d^2f(x)/dx^2$ to be negative. This is because the h^2 term dominates the rest of the series—unless $d^2f(x)/dx^2$ is itself zero, in which case we must go on and think about even higher-order differential quotients. This analysis, first given and demonstrated geometrically by Maclaurin, was worked out in full analytic detail by Euler [**6**, sections 253–254], [**9**, pp. 117–118]. It is typical of Euler's ability to choose computations that produce insight into fundamental concepts. It assumes, of course, that the function in question has a Taylor series, an assumption which Euler made without proof for many functions; it assumes also that the function is uniquely the sum of its Taylor series, which Euler took for granted. Nevertheless, this analysis is a beautiful example of the exploration and development of the concept of the differential quotient of first, second, and nth orders—a development which completely solves the problem of characterizing maxima and minima, a problem which goes back to the Greeks.

Lagrange and the derivative as a function

Though Euler did a good job analyzing maxima and minima, he brought little further understanding of the nature of the differential quotient. The new importance given to Taylor series meant that one had to be concerned not only about first and second differential quotients, but about differential quotients of any order.

The first person to take these questions seriously was Lagrange. In the 1770's, Lagrange was impressed with what Euler had been able to achieve by Taylor-series manipulations with differential quotients, but Lagrange soon became concerned about the logical inadequacy of all the existing justifications for the calculus. In particular, Lagrange wrote in 1797 that the Newtonian limit-concept was not clear enough to be the foundation for a branch of mathematics. Moreover, in not allowing variables to surpass their limits, Lagrange thought the limit-concept too restrictive. Instead, he said, the calculus should be reduced to algebra, a subject whose foundations in the eighteenth century were generally thought to be sound [**11**, pp. 15–16].

The algebra Lagrange had in mind was what he called the algebra of infinite series, because Lagrange was convinced that infinite series were part of algebra. Just as arithmetic deals with infinite decimal fractions without ceasing to be arithmetic, Lagrange thought, so algebra deals with infinite algebraic expressions without ceasing to be algebra. Lagrange believed that expanding $f(x+h)$ into a power series in h was always an algebraic process. It is obviously algebraic when one turns $1/(1-x)$ into a power series by dividing. And Euler had found, by manipulating formulas, infinite power-series expansions for functions like $\sin x, \cos x, e^x$. If functions like those have power-series expansions, perhaps everything could be reduced to algebra. Euler, in his book *Introduction to the analysis of the infinite* (*Introductio in analysin infinitorum*, 1748), had studied infinite series, infinite products, and infinite continued fractions by what he thought of as purely algebraic methods. For instance, he converted infinite series into infinite products by treating a series as a very long polynomial. Euler thought that this work was purely algebraic, and—what is crucial here—Lagrange also thought Euler's methods were purely algebraic. So Lagrange tried to make the calculus rigorous by reducing it to the algebra of infinite series.

Lagrange stated in 1797, and thought he had proved, that any function (that is, any analytic expression, finite or infinite) had a power-series expansion:

$$f(x+h)=f(x)+p(x)h+q(x)h^2+r(x)h^3+\cdots, \tag{5}$$

except, possibly, for a finite number of isolated values of x. He then defined a new function, the coefficient of the linear term in h which is $p(x)$ in the expansion shown in (5)) and called it the **first derived function** of $f(x)$. Lagrange's term "derived function" (*fonction dérivée*) is the origin of our term "derivative." Lagrange introduced a new notation, $f'(x)$, for that function. He defined $f''(x)$ to be the first derived function of $f'(x)$, and so on, recursively. Finally, using these definitions, he proved that, in the expansion (5) above, $q(x)=f''(x)/2$, $r(x)=f'''(x)/6$, and so on [**11**, chapter 2].

What was new about Lagrange's definition? The concept of *function*—whether simply an algebraic expression (possibly infinite) or, more generally, any dependence relation—helps free the concept of derivative from the earlier ill-defined notions. Newton's explanation of a fluxion as a rate of change appeared to involve the concept of motion in mathematics; moreover, a fluxion seemed to be a different kind of object than the flowing quantity whose fluxion it was. For Leibniz, the differential quotient had been the quotient of vanishingly small differences; the second differential quotient, of even smaller differences. Bishop Berkeley, in his attack on the calculus, had made fun of these earlier concepts, calling vanishing increments "ghosts of departed quantities" [**2**, section 35]. But since, for Lagrange, the derivative was a function, it was now the same sort of object as the original function. The second derivative is precisely the same sort of object as the first derivative; even the nth derivative is simply another function, defined as the coefficient of h in the Taylor series for $f^{(n-1)}(x+h)$. Lagrange's notation $f'(x)$ was designed precisely to make this point.

We cannot fully accept Lagrange's definition of the derivative, since it assumes that every differentiable function is the sum of a Taylor series and thus has infinitely many derivatives. Nevertheless, that definition led Lagrange to a number of important properties of the derivative. He used his definition together with Euler's criterion for using truncated power series in approximations to give a most useful characterization of the derivative of a function [**9**, p. 116, pp. 118–121]:

$$f(x+h)=f(x)+hf'(x)+hH, \text{ where } H \text{ goes to zero with } h.$$

(I call this the *Lagrange property of the derivative*.) Lagrange interpreted the phrase "H goes to zero with h" in terms of inequalities. That is, he wrote that,

$$\begin{array}{l}\textit{Given } D, h \textit{ can be chosen so that } f(x+h)-f(x)\\ \textit{lies between } h(f'(x)-D) \textit{ and } h(f'(x)+D).\end{array} \tag{6}$$

Formula (6) is recognizably close to the modern delta-epsilon definition of the derivative.

Lagrange used inequality (6) to prove theorems. For instance, he proved that a function with positive derivative on an interval is increasing there, and used that theorem to derive the Lagrange remainder of the Taylor series [**9**, pp. 122–127], [**11**, pp. 78–85]. Furthermore, he said, considerations like inequality (6) are what make possible applications of the differential calculus to a whole range of problems in mechanics, in geometry, and, as we have described, the problem of maxima and minima (which Lagrange solved using the Taylor series remainder which bears his name [**11**, pp. 233–237]).

In Lagrange's 1797 work, then, the derivative is defined by its position in the Taylor series—a strange definition to us. But the derivative is also *described* as satisfying what we recognize as the appropriate delta-epsilon inequality, and Lagrange applied this inequality and its nth-order analogue, the Lagrange remainder, to solve problems about tangents, orders of contact between

P. Fermat

R. Descartes

I. Newton

G.W. Leibniz

1637-38 1669 1684

Dates refer to these mathematician's major works which

curves, and extrema. Here the derivative was clearly a function, rather than a ratio or a speed.

Still, it is a lot to assume that a function has a Taylor series if one wants to define only *one* derivative. Further, Lagrange was wrong about the algebra of infinite series. As Cauchy pointed out in 1821, the algebra of finite quantities cannot automatically be extended to infinite processes. And, as Cauchy also pointed out, manipulating Taylor series is not foolproof. For instance, e^{-1/x^2} has a zero Taylor series about $x=0$, but the function is not identically zero. For these reasons, Cauchy rejected Lagrange's definition of derivative and substituted his own.

Definitions, rigor, and proofs

Now we come to the last stage in our chronological list: *definition*. In 1823, Cauchy defined the derivative of $f(x)$ as the limit, when it exists, of the quotient of differences $(f(x+h)-f(x))/h$ as h goes to zero [**4**, pp. 22–23]. But Cauchy understood "limit" differently than had his predecessors. Cauchy entirely avoided the question of whether a variable ever reached its limit; he just didn't discuss it. Also, knowing an absolute value when he saw one, Cauchy followed Simon l'Huilier and S.-F. Lacroix in abandoning the restriction that variables never surpass their limits. Finally, though Cauchy, like Newton and d'Alembert before him, gave his definition of limit in words, Cauchy's understanding of limit (most of the time, at least) was algebraic. By this, I mean that when Cauchy needed a limit property in a proof, he used the algebraic inequality-characterization of limit. Cauchy's proof of the mean value theorem for derivatives illustrates this. First he proved a theorem which states: *if* $f(x)$ *is continuous on* $[x, x+a]$, *then*

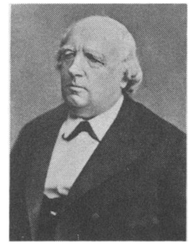
$$\min_{[x,x+a]} f'(x) \leqslant \frac{f(x+a)-f(x)}{a} \leqslant \max_{[x,x+a]} f'(x). \tag{7}$$

The first step in his proof is [**4**, p. 44]:

> Let δ, ε be two very small numbers; the first is chosen so that for all [absolute] values of h less than δ, and for any value of x [on the given interval], the ratio $(f(x+h)-f(x))/h$ will always be greater than $f'(x)-\varepsilon$ and less than $f'(x)+\varepsilon$.

(The notation in this quote is Cauchy's, except that I have substituted h for the i he used for the increment.) Assuming the intermediate-value theorem for continuous functions, which Cauchy had proved in 1821, the mean-value theorem is an easy corollary of (7) [**4**, pp. 44–45], [**9**, pp. 168–170].

Cauchy took the inequality-characterization of the derivative from Lagrange (possibly via an 1806 paper of A.-M. Ampère [**9**, pp. 127–132]). But Cauchy made that characterization into a definition of derivative. Cauchy also took from Lagrange the name derivative and the notation $f'(x)$, emphasizing the functional nature of the derivative. And, as I have shown in detail elsewhere [**9**, chapter 5], Cauchy adapted and improved Lagrange's inequality proof-methods to prove results like the mean-value theorem, proof-methods now justified by Cauchy's definition of derivative.



L. Euler **J.-L. Lagrange** **A.-L. Cauchy** **K. Weierstrass**

1755 **1797** **1823** **1861**

contributed to the evolution of the concept of the derivative.

But of course, with the new and more rigorous definition, Cauchy went far beyond Lagrange. For instance, using his concept of limit to define the integral as the limit of sums, Cauchy made a good first approximation to a real proof of the Fundamental Theorem of Calculus [**9**, pp. 171–175], [**4**, pp. 122–125, 151–152]. And it was Cauchy who not only raised the question, but gave the first proof, of the existence of a solution to a differential equation [**9**, pp. 158–159].

After Cauchy, the calculus itself was viewed differently. It was seen as a rigorous subject, with good definitions and with theorems whose proofs were based on those definitions, rather than merely as a set of powerful methods. Not only did Cauchy's new rigor establish the earlier results on a firm foundation, but it also provided a framework for a wealth of new results, some of which could not even be formulated before Cauchy's work.

Of course, Cauchy did not himself solve all the problems occasioned by his work. In particular, Cauchy's definition of the derivative suffers from one deficiency of which he was unaware. Given an ε, he chose a δ which he assumed would work for any x. That is, he assumed that the quotient of differences converged uniformly to its limit. It was not until the 1840's that G. G. Stokes, V. Seidel, K. Weierstrass, and Cauchy himself worked out the distinction between convergence and uniform convergence. After all, in order to make this distinction, one first needs a clear and algebraic understanding of what a limit is—the understanding Cauchy himself had provided.

In the 1850's, Karl Weierstrass began to lecture at the University of Berlin. In his lectures, Weierstrass made algebraic inequalities replace words in theorems in analysis, and used his own clear distinction between pointwise and uniform convergence along with Cauchy's delta-epsilon techniques to present a systematic and thoroughly rigorous treatment of the calculus. Though Weierstrass did not publish his lectures, his students—H. A. Schwartz, G. Mittag-Leffler, E. Heine, S. Pincherle, Sonya Kowalevsky, Georg Cantor, to name a few—disseminated Weierstrassian rigor to the mathematical centers of Europe. Thus although our modern delta-epsilon definition of derivative cannot be quoted from the *works* of Weierstrass, it is in fact the *work* of Weierstrass [**3**, pp. 284–287]. The rigorous understanding brought to the concept of the derivative by Weierstrass is signaled by his publication in 1872 of an example of an everywhere continuous, nowhere differentiable function. This is a far cry from merely acknowledging that derivatives might not always exist, and the example shows a complete mastery of the concepts of derivative, limit, and existence of limit [**3**, p. 285].

Historical development versus textbook exposition

The span of time from Fermat to Weierstrass is over two hundred years. How did the concept of derivative develop? Fermat implicitly used it; Newton and Liebniz discovered it; Taylor, Euler, Maclaurin developed it; Lagrange named and characterized it; and only at the end of this long period of development did Cauchy and Weierstrass define it. This is certainly a complete reversal of the usual order of textbook exposition in mathematics, where one starts with a definition, then explores some results, and only then suggests applications.

This point is important for the teacher of mathematics: the historical order of development of the derivative is the reverse of the usual order of textbook exposition. Knowing the history helps us as we teach about derivatives. We should put ourselves where mathematicians were before Fermat, and where our beginning students are now—back on the other side, before we had any concept of derivative, and also before we knew the many uses of derivatives. Seeing the historical origins of a concept helps motivate the concept, which we—along with Newton and Leibniz—want for the problems it helps to solve. Knowing the historical order also helps to motivate the rigorous definition—which we, like Cauchy and Weierstrass, want in order to justify the uses of the derivative, and to show precisely when derivatives exist and when they do not. We need to remember that the rigorous definition is often the end, rather than the beginning, of a subject.

The real historical development of mathematics—the order of discovery—reveals the creative mathematician at work, and it is creation that makes doing mathematics so exciting. The order of exposition, on the other hand, is what gives mathematics its characteristic logical structure and its incomparable deductive certainty. Unfortunately, once the classic exposition has been given, the order of discovery is often forgotten. The task of the historian is to recapture the order of discovery: not as we think it might have been, not as we think it should have been, but as it really was. And this is the purpose of the story we have just told of the derivative from Fermat to Weierstrass.

This article is based on a talk delivered at the Conference on the History of Modern Mathematics, Indiana Region of the Mathematical Association of America, Ball State University, April 1982; earlier versions were presented at the Southern California Section of the M. A. A. and at various mathematics colloquia. I thank the MATHEMATICS MAGAZINE referees for their helpful suggestions.

References

[1] Margaret Baron, Origins of the Infinitesimal Calculus, Pergamon, Oxford, 1969.

[2] George Berkeley, The Analyst, or a Discourse Addressed to an Infidel Mathematician, 1734. In A. A. Luce and T. R. Jessop, eds., The Works of George Berkeley, Nelson, London, 1951 (some excerpts appear in [**16**, pp. 333–338]).

[3] Carl Boyer, History of the Calculus and Its Conceptual Development, Dover, New York, 1959.

[4] A.-L. Cauchy, Résumé des leçons données à l'école royale polytechnique sur le calcul infinitésimal, Paris, 1823. In Oeuvres complètes d'Augustin Cauchy, Gauthier-Villars, Paris, 1882- , series 2, vol. 4.

[5] Pierre Dugac, Fondements d'analyse, in J. Dieudonné, Abrégé d'histoire des mathématiques, 1700–1900, 2 vols., Hermann, Paris, 1978.

[6] Leonhard Euler, Institutiones calculi differentialis, St. Petersburg, 1755. In Operia omnia, Teubner, Leipzig, Berlin, and Zurich, 1911- , series 1, vol. 10.

[7] Pierre Fermat, Analysis ad refractiones, 1661. In Oeuvres de Fermat, ed., C. Henry and P. Tannery, 4 vols., Paris, 1891–1912; Supplement, ed. C. de Waard, Paris, 1922, vol. 1, pp. 170–172.

[8] ______, Methodus ad disquirendam maximam et minimam et de tangentibus linearum curvarum, Oeuvres, vol. 1, pp. 133–136. Excerpted in English in [**16**, pp. 222–225].

[9] Judith V. Grabiner, The Origins of Cauchy's Rigorous Calculus, M. I. T. Press, Cambridge and London, 1981.

[10] Morris Kline, Mathematical Thought from Ancient to Modern Times, Oxford, New York, 1972.

[11] J.-L. Lagrange, Théorie des fonctions analytiques, Paris, 2nd edition, 1813. In Oeuvres de Lagrange, ed. M. Serret, Gauthier-Villars, Paris, 1867–1892, vol. 9.

[12] Michael S. Mahoney, The Mathematical Career of Pierre de Fermat, 1601–1665, Princeton University Press, Princeton, 1973.

[13] Isaac Newton, Of Analysis by Equations of an Infinite Number of Terms [1669], in D. T. Whiteside, ed., Mathematical Works of Isaac Newton, Johnson, New York and London, 1964, vol. 1, pp. 3–25.

[14] ______, Method of Fluxions [1671], in D. T. Whiteside, ed., Mathematical Works of Isaac Newton, vol. 1, pp. 29–139.

[15] ______, Mathematical Principles of Natural Philosophy, tr. A. Motte, ed. F. Cajori, University of California Press, Berkeley, 1934.

[16] D. J. Struik, Source Book in Mathematics, 1200–1800, Harvard University Press, Cambridge, MA, 1969.

[17] D. T. Whiteside, ed., The Mathematical Papers of Isaac Newton, Cambridge University Press, 1967–1982.

Paradoxes of Preferential Voting

What can go wrong with sophisticated voting systems designed to remedy problems of simpler systems.

PETER C. FISHBURN
Bell Laboratories
Murray Hill, NJ 07974

STEVEN J. BRAMS
New York University
New York, NY 10003

Preferential voting, developed by Thomas Hare [**12**] in the 1860's, is still used for major elections in Australia, Ireland and South Africa, as well as for local elections in many countries. From its inception, it has been touted as a way to promote full expression of electors' preferences and to ensure maximum and equitable consideration of each elector's vote. When used to fill several seats in a legislature, preferential voting provides representation for viable minorities and tends to distribute seats in proportion to the numbers of voters who favor the different parties.

It seeks to do all this on the basis of a single preferential (ranked) ballot by transferring votes, in part or in whole, from the most and least popular candidates to candidates with intermediate support. The most popular, elected first, don't need their "surpluses," and the least popular can never overcome their "deficits," so the transfers of both surpluses and deficits to the intermediate candidates determines which of these win. When there are n voters and c seats are to be filled, transfers are made sequentially until c candidates attain the vote quota needed for election. The quota is usually defined as

$$q=\left[\frac{n}{c+1}\right]+1,$$

where brackets signify the integer part of the argument. We shall use this concept later.

Despite its tendencies to promote individual interests and fair representation, preferential voting has several surprising and potentially damning defects. We shall begin by illustrating four of these through an apocryphal story of an election among three contenders for mayor of a small town. In this deliberately simple case, a candidate ranked first on more than 50 percent of the ballots is elected; if there is no such candidate, the one with the fewest first-place votes is scratched, then the one of the remaining two who ranks higher on more ballots is elected. Since this procedure is tantamount to plurality (vote-for-one) voting followed by a two-candidate runoff election, the defects or paradoxes developed in our story apply also to the common plurality-runoff scheme.

The story's four paradoxes are summarized here for reference and for readers who may wish to test them on their own.

NO-SHOW PARADOX: *The addition of identical ballots with candidate x ranked last may change the winner from another candidate to x.*

THWARTED-MAJORITIES PARADOX: *A candidate who can defeat every other candidate in direct-comparison majority votes may lose the election.*

MULTIPLE-DISTRICTS PARADOX: *A candidate can win in each district separately, yet lose the general election in the combined districts.*

MORE-IS-LESS PARADOX: *If the winner were ranked higher by some voters, all else unchanged, then another candidate might have won.*

Following the story, we shall discuss general problems confronting voting schemes and mention interesting mathematical work on the subject. We then return to preferential voting to illustrate two other paradoxes that arise only in more complex situations. We conclude with a note on paradox probabilities.

A funny thing happened on the way to the polls

Mr. and Mrs. Smith's car broke down on the way to the polls just before closing time. The Smiths were intensely interested in a tight race for mayor of their town among Mrs. Bitt, Mr. Huff and Dr. Wogg.

The ballot for mayor asked each voter to rank the three candidates from first choice to third choice. The townspeople knew that the election would be decided by the simple preferential voting method, which had been instituted by local referendum some years earlier. Everyone in town was pleased that they used such a sensible procedure for electing the head of their local government.

The Smiths were of one mind about the candidates. They favored Bitt to Huff to Wogg, and therefore both would have voted BHW. Although they liked Mrs. Bitt best, they were almost as fond of Mr. Huff but disliked and mistrusted Dr. Wogg. Much to their regret, the Smith's car problem prevented them from making it to the polls before closing time.

Many of their fellow townspeople did. When Mrs. Smith opened her newspaper the next morning, her eye was caught by a headline proclaiming "Huff Elected as 1,608 Go to Polls." She and her husband were delighted that Dr. Wogg had not won. They did feel a twinge of regret that their friend, Mrs. Bitt, was beaten. Perhaps their votes would have made a difference.

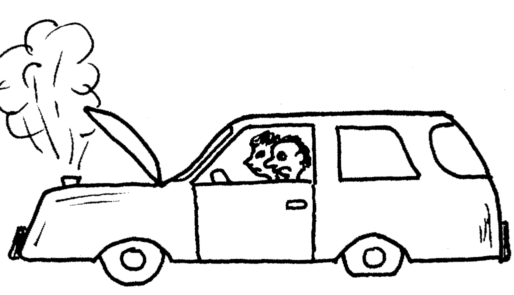
As Mrs. Smith read on, she noted that no candidate had gotten enough first-place votes to win outright. Mrs. Bitt had been scratched because she had the fewest first-place votes, and Mr. Huff went on to beat Dr. Wogg by a plurality of 917 to 691.

Toward the end of the article, on an inside page, Mrs. Smith read the tabulation of how the 1,608 voters cast their ballots shown in FIGURE 1.

Totals	Rankings	H over W	W over H
417	BHW	417	0
82	BWH	0	82
143	HBW	143	0
357	HWB	357	0
285	WBH	0	285
324	WHB	0	324
1608		917	691

FIGURE 1

It made her feel good that she and her husband would have voted with the largest of the six groups. As the article had noted earlier, Mrs. Bitt barely lost out on the initial count since Bitt, Huff and Wogg had first-place tallies of 499 (417 + 82), 500 (143 + 357), and 609 (285 + 324) respectively.



Mrs. Smith realized when she read this that Mr. Huff rather than Mrs. Bitt would have been scratched if she and her husband had voted. At least their friend, Mrs. Bitt, would have made the "runoff" if their car had not broken down.

Before leaving for her job as an actuary with an insurance company headquartered in the next town, Mrs. Smith decided to see what would have happened if she and her husband had voted. Her tabulation is shown in FIGURE 2.

Totals	Rankings	B over W	W over B
419	BHW	419	0
82	BWH	82	0
143	HBW	143	0
357	HWB	0	357
285	WBH	0	285
324	WHB	0	324
1610		644	966

FIGURE 2

To her chagrin, she saw that their votes would have made Dr. Wogg the winner even though he was ranked last on their ballots! This so shocked her that she checked her figures three times. When they refused to change, it hit her: *the whole thing depended on who was scratched after the initial count*. With Bitt out, Huff wins; with Huff out, Wogg wins. Even if 300 more people had voted BHW, Dr. Wogg would still have won.

Mrs. Smith was beginning to wonder if the town's procedure for electing a mayor was that sensible after all.

That evening, while reviewing her figures again, Mrs. Smith became aware of another curious fact. She realized that the winner, Mr. Huff, would have beaten either Mrs. Bitt (824 to 784) or Dr. Wogg (917 to 691) in a direct vote between the two. The "majority candidate"—that is, the candidate who could have beaten each of the others in direct pairwise votes—had indeed won. However, if the Smiths had voted, then not only would their last choice have won but the winner, Dr. Wogg, would not have been the candidate favored in separate pairwise contests to each of the other candidates.

At this point, Mrs. Smith suspected that their election procedure might be more than a little flawed and wondered if further probing might uncover other unusual possibilities. She vowed to make time for this over the weekend.

The Smith's town had two voting districts, called East and West. When the weekend came round, Mrs. Smith decided to compare the outcome with what might have happened in the separate districts. She suspected that the winner, Mr. Huff, might have lost in one if not both districts. The paper had reported that 588 people voted in the East and 1,020 had voted in the West. Moreover, it gave the East-West splits shown in FIGURE 3.

Totals	Rankings	East	West
417	BHW	160	257
82	BWH	0	82
143	HBW	143	0
357	HWB	0	357
285	WBH	0	285
324	WHB	285	39
1608		588	1020

FIGURE 3

Applying precisely the same election rule used for the general election to the East and the West separately, Mrs. Smith found that Mrs. Bitt would have won in both districts! She felt this was truly amazing since both Huff and Wogg had sizable majorities over Bitt in the overall electorate, so there was no way that Mrs. Bitt could have won in the combined districts.

Moreover, as Mrs. Smith noted, a multiple-district winner like Mrs. Bitt could be a "minority candidate" in the sense that this candidate would be defeated by every other candidate in direct-comparison votes. She also realized that this anomaly could arise only when different candidates were scratched on the first rounds in the several "district elections." In fact, Huff was scratched in the East, whereas Wogg was scratched in the West.

Mrs. Smith was now convinced that she had a very strong case against the supposedly sensible system used to elect the mayor of their town. At her request, the chairman of the local Election Board called a special meeting of the board to review her findings.

On the night before the board meeting, as she was going over her figures, Mrs. Smith discovered another irregularity. While pondering what would have occurred if she and her husband had voted (see FIGURE 2), she realized that if two or more of the 82 voters with ranking BWH had moved Wogg into first place (WBH), then Bitt rather than Huff would have been scratched and Huff rather than Wogg would have won. In other words, an increase in support for Dr. Wogg would have changed him from a winner to a loser! Extraordinary, thought Mrs. Smith, as she prepared her flip charts for her presentation to the Election Board.

The next day Mrs. Smith so impressed the board that they decided to appoint a select panel—chaired by Mrs. Smith, of course—to recommend a better election procedure. In particular, the board charged the panel with devising a system that would avoid all the paradoxes uncovered by Mrs. Smith.

At the panel's first meeting, one member suggested that they retain ranked voting but simply use the ballots to determine which of the several candidates was the majority candidate. He explained that this would directly resolve the Thwarted-Majorities Paradox and, moreover, would also take care of Mrs. Smith's other three paradoxes.

Mrs. Smith responded that this was a very good idea up to a point, but that it would not solve all their problems. She had been reading up on the subject and proceeded to tell the panel about the most famous paradox of them all, variously known as "Condorcet's phenomenon" [**4**], the "paradox of voting," and the "paradox of cyclical majorities."

Condorcet's phenomenon occurs when every candidate is beaten by some other candidate under direct-comparison voting. Mrs. Smith pointed out that this was not the case in their election, but it was certainly possible. For example, if 1,600 total ballots had been cast, with

400	for	BHW
500	for	WBH
700	for	HWB,

then Bitt beats Huff 900 to 700, Huff beats Wogg 1,100 to 500, and Wogg beats Bitt 1,200 to 400.

At this point, another panel member suggested that perhaps their problems would vanish if they used the method that his lodge used to choose its president. This method awards 2 points to a first-place vote, 1 point to a second-place vote, and 0 points to a third-place vote. The winner is the candidate with the most points. He noted that it could be extended in a straightforward way when there are more than three candidates.

The panel determined that this point-scoring system—sometimes referred to as Borda's "method of marks" [**2**], [**5**]—would resolve all of Mrs. Smith's paradoxes, with the possible exception of the Thwarted-Majorities Paradox. A quick review of the election data showed that the majority candidate, Mr. Huff, would have won under the point-scoring system. However, the panel also noticed that if 50 or so BHW voters had preferred Wogg to Huff and voted BWH, then despite the fact that Mr. Huff would remain the majority candidate, Dr. Wogg would win under the point-scoring system (see FIGURE 4).

Totals	Rankings	B Points	H Points	W Points
367	BHW	734	367	0
132	BWH	264	0	132
143	HBW	143	286	0
357	HWB	0	714	357
285	WBH	285	0	570
324	WHB	0	324	648
1608		1426	1691	1707

FIGURE 4

Confused and tired, the panel agreed that they had done enough for one meeting. Their next meeting was set for the following Wednesday.

Problems of voting systems

We end our story at this point because, in a sense, it has no end. The panel could meet forever without being able to fulfill its charge from the Election Board to avoid all four paradoxes. This is because there is a metaparadox lurking in the background which, in simplified form, says that *no* election procedure can simultaneously resolve Mrs. Smith's second and third paradoxes.

Let us elaborate. We assume, as before, that voters rank the candidates from most preferred to least preferred. With a fixed number of candidates, but any potential number of voters, Young [**21**] (see also [**22**]) showed in one of the most mathematically elegant papers on the subject that in order to avoid the Multiple-Districts Paradox as well as to satisfy fundamental equity conditions for voters and candidates, one must use a type of point-scoring system. In an attempt to avoid the Thwarted-Majorities Paradox, it is necessary to assign more points to a first-place vote than to a second-place vote, and so forth, which of course takes care of the No-Show and More-Is-Less Paradoxes.

However, given any set of decreasing point values for the various places, it is always possible to construct an example with a majority candidate who is not elected by the point-scoring system. In fact, nearly two hundred years ago Condorcet recognized that it is possible to construct examples with a majority candidate who is not elected by *any* point-scoring system with decreasing point values [**4**],[**9**]. For example, if there are seven voters such that

3	have	BHW
2	have	HWB
1	has	HBW
1	has	WBH,

then B has a 4-to-3 majority over each of H and W, but H beats B under *every* point-scoring system that assigns more points to a second-place vote than to a third-place vote.

Many other problems and paradoxes that plague preferential voting and other election systems seem to have surfaced only recently. The More-Is-Less Paradox, better known in the literature as the monotonicity paradox, was shown by Smith [**20**] to affect virtually all successive-elimination procedures based on point scoring. Further results on the monotonicity paradox appear in [**10**]. Within the specific context of preferential voting, the More-Is-Less Paradox and the Multiple-Districts Paradox are discussed in [**6**], [**7**].

As far as we know, the No-Show Paradox, which is closely related to the More-Is-Less Paradox, is not discussed elsewhere. However, another no-show paradox seems to have been discovered many years ago [**14**], [**17**]. This other paradox says that one of the candidates elected by preferential voting could have ended up a loser if additional people who ranked him in first place had actually voted. An example of this paradox appears in the next section.

The paradoxes discussed here and elsewhere [**9**], [**15**] reveal only the surface effects of deeper aspects of aggregation structures, such as those developed by Young [**21**]. Recent work on these structures was stimulated in large measure by Kenneth Arrow's classic impossibility theorem [**1**]. This theorem shows that a few simple and appealing conditions for aggregating diverse rankings into a consensus ranking are incompatible. Numerous variants of Arrow's theorem now exist [**8**], [**13**], [**19**], and these have been joined by related results [**11**], [13], [**16**], [**18**] which show that virtually every sensible election procedure for multicandidate elections is vulnerable to strategic manipulation by voters. In other words, there will be situations in which some voters can benefit by voting contrary to their true, or sincere, preferences.

An example of the latter phenomenon occurs in our story of the Smiths. If they had voted their true preference order, BHW, then Dr. Wogg would have won under preferential voting. However, if they had voted HBW, or any other order that did not have Mrs. Bitt in first place, then Mr. Huff would have won. Hence, by voting strategically (i.e., falsely), the Smiths would have helped to elect their second choice (H) rather than their last choice (W).

More paradoxes of preferential voting

Additional flaws in preferential voting can arise only when there are more than three contenders. We shall illustrate two of these after describing a general, and widely used, procedure for preferential voting.

In the general case, voters rank the candidates from most preferred to least preferred on their ballots. To be elected, a candidate must receive a quota q of weighted votes.

Each voter begins with voting weight 1. First-place votes are tallied for each candidate; those with q or more are elected. If c' are elected on this first round and $0 < c' < c$, then the weight of each voter whose first choice was elected is decreased from 1 to a nonnegative number (0 if there is no "surplus" over quota) so that the sum of all weights becomes $n - qc'$. Elected candidates are removed from the ballots, and new rounds follow until c candidates are elected, as described below.

After removal of the elected candidates, unelected candidates move up in the ballot rankings to fill in top places, and the process is repeated with a new, weighted tally of unelected candidates now in first place. Again, q is used as the quota for election. The process continues until either all c seats have been filled, or no unelected candidate gets at least q weighted votes in the latest tally. In the latter case, the candidate with the *smallest* weighted first-place tally is scratched, ballot rankings (but not voter weights) are revised accordingly, and the process continues until c seats are filled.

Instead of our earlier story, suppose now that Bitt, Foxx, Huff and Wogg are vying for two seats on the town council, and that 100 people vote as follows:

34	BHFW
25	FBHW
26	HWBF
9	WBFH
6	WHFB

The quota is 34, so Bitt is elected first. Since exactly 34 people voted for Bitt in first place, the weights of these voters are reduced to 0, leaving

25	FHW
26	HWF
9	WFH
6	WHF

Since none of the others reaches the quota, Wogg is scratched. Then Foxx, who has 34 (25 + 9) votes to 32 (26 + 6) for Huff, wins the second seat.

Now suppose five more Foxx supporters (FBHW) had voted, giving

34	BHFW
30	FBHW
26	HWBF
9	WBFH
6	WHFB

The new quota is $(105/3) + 1 = 36$. Since no candidate reaches the quota, Wogg is scratched:

34	BHF
9	BFH
30	FBH
26	HBF
6	HFB

At this point Bitt passes the quota with 43 (34 + 9) votes and is elected as before. Since Bitt exceeded the quota by 7 first-place votes, each of her 43 supporters retains 7/43 of a vote, giving aggregates of

5.5	HF	from Bitt's surplus
1.5	FH	
30	FH	
26	HF	
6	HF	

Since Huff now surpasses the quota with 37.5 (26 + 6 + 5.5), he becomes the second candidate elected. Thus Foxx, a winner in the first case, becomes a loser when five more voters show up with him in first place.

Our final paradox was suggested by a statement on a recent ballot of a professional society that listed eight candidates for four seats on the society's Nominating Committee [**3**]. The election was conducted by preferential voting. Society members were advised to mark candidates in order of preference until they were ignorant or indifferent concerning candidates whom they did not rank. The preferential voting system described earlier is easily modified to account for partial rankings: if a voter's marked candidates are removed or scratched before all seats are filled, that voter is then treated as if he never voted in the first place.

The ballot statement alluded to in the preceding paragraph claimed that "there is no tactical advantage to be gained by marking few candidates." FIGURE 2, suitably modified, shows that this is false. Suppose again that Foxx is in the race for two council seats along with Bitt, Huff, and Wogg, and that votes are precisely the same as shown in FIGURE 2, except that Foxx is the first choice of all 1,610 voters:

Totals	Rankings
419	FBHW
82	FBWH
143	FHBW
357	FHWB
285	FWBH
324	FWHB
1610	

Then Foxx wins a seat, and matters proceed as before when he is removed from the ballots, giving Dr. Wogg the other seat.

But suppose that Mr. and Mrs. Smith had voted just F instead of FBHW, i.e., had voted only for their first choice. Then, after Foxx is removed, we revert to FIGURE 1, where Mr. Huff wins the other seat. By voting only for their first choice, the Smiths prevent their last choice from winning the second seat.

This example provides a second instance of how some voters might induce preferred outcomes by misrepresenting their true preferences. In the present case, misrepresentation takes the form of a deliberate truncation of one's ranking rather than a false but complete ranking.

Paradox probabilities

Although virtually all voting systems for elections with three or more candidates can produce counterintuitive and disturbing outcomes, preferential voting is especially vulnerable because of its sequential elimination and vote-transfer provisions. Nevertheless, this system is still widely used in several countries.

Defenders of preferential voting—and there have been many over the past century—might argue that the paradoxes of preferential voting are not a problem because they occur so infrequently in practice. They would, we presume, claim that a few contrived examples should not deter us from using a carefully refined system that has proved its worth in countless elections.

Although probabilities of paradoxes have been estimated in other settings **[9]**, we know of no attempts to assess the likelihoods of the paradoxes of preferential voting discussed above, and would propose this as an interesting possibility for investigation. Is it indeed true that serious flaws in preferential voting such as the No-Show Paradox and the More-Is-Less Paradox are sufficiently rare as to cause no practical concern?

References

[1] K. J. Arrow, Social Choice and Individual Values, 2nd ed., Wiley, New York, 1963.

[2] Jean-Charles de Borda, Mémoire sur les élections au scrutin, Histoire de l'Académie Royale des Sciences, Paris, 1781.

[3] S. J. Brams, The AMS nominating system is vulnerable to truncation of preferences, Notices Amer. Math. Soc., 29 (1982) 136–138.

[4] Marquis de Condorcet, Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix, Paris, 1785.

[5] A. de Grazia, Mathematical derivation of an election system, Isis, 44 (1953) 42–51.

[6] G. Doron, The Hare voting system is inconsistent, Political Studies, 27 (1979) 283–286.

[7] G. Doron and R. Kronick, Single transferable vote: an example of a perverse social choice function, American Journal of Political Science, 21 (1977) 303–311.

[8] P. C. Fishburn, The Theory of Social Choice, Princeton University Press, 1973.

[9] ______, Paradoxes of voting, American Political Science Review, 68 (1974) 537–546.

[10] ______, Monotonicity paradoxes in the theory of elections, Discrete Applied Mathematics, 4 (1982) 119–134.

[11] A. Gibbard, Manipulation of voting schemes: a general result, Econometrica, 41 (1973) 587–601.

[12] T. Hare, The Election of Representatives, Parliamentary and Municipal: A Treatise, Longman, Green, London, 1861.

[13] J. S. Kelly, Arrow Impossibility Theorems, Academic Press, New York, 1978.

[14] J. C. Meredith, Proportional Representation in Ireland, Dublin, 1913.

[15] R. G. Niemi and W. H. Riker, The choice of voting systems, Scientific American, 234 (1976) 21–27.

[16] P. K. Pattanaik, Strategy and Group Choice, North-Holland, New York, 1978.

[17] Report of the Royal Commission Appointed to Enquire into Electoral Systems, HMSO, London, 1910.

[18] M. A. Satterthwaite, Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions, Journal of Economic Theory, 10 (1975) 187–217.

[19] A. K. Sen, Collective Choice and Social Welfare, Holden-Day, San Francisco, 1970.

[20] J. H. Smith, Aggregation of preferences with variable electorate, Econometrica, 41 (1973) 1027–1041.

[21] H. P. Young, Social choice scoring functions, SIAM J. Appl. Math., 28 (1975) 824–838.

[22] H. P. Young and A. Levenglick, A consistent extension of Condorcet's election principle, SIAM J. Appl. Math., 35 (1978) 285–300.

On Proving Trigonometric Identities

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

When I attended secondary school, one was not allowed to verify a trigonometric identity $A \equiv B$ by simply proving $A - B \equiv 0$, but one had to manipulate A into B or vice versa. This restriction was absurd since one could always write $A \equiv (A - B) + B$ and then prove $A - B \equiv 0$. Although I had a certain delight and challenge in establishing various trigonometric identities, there were many students who had difficulties with this type of problem since they were never given a general method of attack. Using the algorithmic method presented here, students should experience fewer difficulties with these kinds of problems. Additionally, these kinds of identities can be verified by a desk computer if it is provided with software for symbolic manipulation (and these now exist; see [**6**]).

The basis of the proof of $A - B \equiv 0$ is the fact that *every trigonometric polynomial identity* $P(\cos\theta, \sin\theta) \equiv 0$ *in one variable is a consequence of* $\cos^2\theta + \sin^2\theta \equiv 1$. Magid [**3**] gave an abstract algebraic proof of this which was designed to fit into an undergraduate course in abstract algebra. Subsequently, to make the result more accessible, Dobbs [**2**] gave a more elementary algorithmic proof. Here, we simplify Dobbs' proof, give a variation of his algorithmic method, and finally give several extensions.

Trigonometric identities of one variable

The essence of Dobbs' proof is that if a polynomial $P(C, S)$ in two variables C, S vanishes for all *real* C, S satisfying $C^2 + S^2 = 1$, then

$$P(C,S) = (C^2 + S^2 - 1)Q(C,S) \tag{1}$$

where Q is also a polynomial in C and S. To prove (1), Dobbs considers $P(C, S)$ as a polynomial in one variable, C, and divides it by $C^2 + S^2 - 1$. Then by the Remainder Theorem for a polynomial in C,

$$P(C,S) = (C^2 + S^2 - 1)Q(C,S) + A(S)C + B(S)$$

where A and B are polynomials in S. Thus $C^2 + S^2 - 1 = 0$ implies

$$A(S)C + B(S) \equiv 0. \tag{2}$$

We now simplify the remaining part of the proof by Dobbs which is to show that $A(S)$ and $B(S)$ vanish identically. If $C^2 + S^2 = 1$, then from (2) it follows that for all S satisfying $-1 \leqslant S \leqslant 1$,

$$A(S)\sqrt{1-S^2} + B(S) \equiv 0,$$
$$-A(S)\sqrt{1-S^2} + B(S) \equiv 0;$$

whence

$$A(S)\sqrt{1-S^2} \equiv B(S) \equiv 0.$$

Since $A(S)$ and $B(S)$ are polynomials which vanish for an infinite set of values of S, they must each vanish identically (by a corollary of the Fundamental Theorem of Algebra).

That $A(S)$ and $B(S)$ must vanish identically provides students with several algorithmic methods to verify polynomial trigonometric identities (by hand or by computer symbolic manipulation). We assume that our trigonometric identity to be proved has already been put in the form $P(C,S) \equiv 0$, where $C = \cos x$ and $S = \sin x$. (If the identity is a rational function of $\cos x$, $\sin x$, $\cot x$, $\tan x$, $\csc x$ and $\sec x$, we first make the replacements

$$\cot x = \frac{\cos x}{\sin x}, \qquad \tan x = \frac{\sin x}{\cos x}, \qquad \csc x = \frac{1}{\sin x}, \qquad \sec x = \frac{1}{\cos x},$$

and convert the resulting rational function in C and S into a polynomial function by multiplication by the least common denominator.)

Method I (Dobbs). Arrange the polynomial in the canonical form

$$P(C,S) = A_0(S)C^n + A_1(S)C^{n-1} + \cdots + A_n(S)$$

and carry out the division by $C^2 + S^2 - 1$ (with respect to C). If $P(C,S) = 0$ is an identity, there will be no remainder, and conversely.

Method II. One can always express $P(C,S)$ in the following polynomial form of odd and even functions of C and S:

$$P(C,S) = F(C^2,S^2) + G(C^2,S^2)C + H(C^2,S^2)S + I(C^2,S^2)CS.$$

As before, $P(C,S) \equiv 0$, if and only if,

$$F(C^2,S^2) \equiv 0, G(C^2,S^2) \equiv 0, H(C^2,S^2) \equiv 0 \text{ and } I(C^2,S^2) \equiv 0.$$

We now apply Method I to each of the last four functions. This separation of P in terms of functions of different parity is easy to do and simplifies the subsequent algebraic division to be done.

Method III. Here we replace C^2 by $1 - S^2$ and show that each of $F(1 - S^2, S^2)$, $G(1 - S^2, S^2)$, $H(1 - S^2, S^2)$, $I(1 - S^2, S^2)$ vanish identically. This is done by expanding out and arranging the terms of $F(1 - S^2, S^2)$, etc., in descending or ascending order and finding that all the coefficients of each power of S^2 vanish.

It is not necessary to carry out the complete reduction of $P(C,S)$ as above. One could also write

$$P(C,S) \equiv F_1(C^2,S) + G_1(C^2,S)C,$$

or

$$P(C,S) \equiv F_2(C,S^2) + G_2(C,S^2)S$$

and proceed as before.

As an example, consider the identity

$$\begin{aligned} &4\cos x + (1 + \sin x)^2 + \csc x \cos^2 x(1 + 2\csc x \cos x) \\ &+ \sec x \sin^2 x(\sec x \sin^2 x + \sec x \sin x + 2) \\ &= \sec x(2\csc^2 x + \sec x \csc x + \sec x). \end{aligned}$$

Expressing all the functions in terms of $\cos x$ and $\sin x$ and converting to a polynomial, the identity reduces to (with $C = \cos x$ and $S = \sin x$)

$$\{C^2S^2(1 + S^2) + S^6 - S^2\} + 2C\{C^4 + 2C^2S^2 + S^4 - 1\} + S\{C^4 + 2C^2S^2 + S^4 - 1\} = 0.$$

On replacing C^2 by $1 - S^2$ in each of the three expressions in the brackets, each vanishes identically, e.g.,

$$(1 - S^2)S^2(1 + S^2) + S^6 - S^2 = S^2(1 - S^4) + S^6 - S^2 = 0.$$

Trigonometric identities in several variables

As an extension of (1), we will show that if $P(\cos A, \sin A, \cos B, \sin B) = 0$ is a polynomial trigonometric identity, then the polynomial $P(C, S, C_1, S_1)$ will have the form

$$P(C,S,C_1,S_1) = (C^2+S^2-1)Q(C,S,C_1,S_1) + (C_1^2+S_1^2-1)R(C,S,C_1,S_1) \tag{3}$$

where Q and R are also polynomials in the four variables C, S, C_1, S_1. As before, by the Remainder Theorem,

$$P(C,S,C_1,S_1) = (C^2+S^2-1)Q(C,S,C_1,S_1) + A(S,C_1,S_1)C + B(S,C_1,S_1).$$

Since P vanishes if $C^2+S^2=1$ and $C_1^2+S_1^2=1$, so also must $A(S,C_1,S_1)C+B(S,C_1,S_1)$ vanish. Then, as before, so also must A and B vanish. Again as before, this implies that A and B must be divisible by $C_1^2+S_1^2=1$, giving the desired result (3). Thus one algorithmic procedure for establishing identities of the form $P(C,S,C_1,S_1)=0$ is to first divide P by C^2+S^2-1 to obtain a remainder

$$A(S,C_1,S_1)C + B(S,C_1,S_1).$$

Then in each of the polynomials A and B, the coefficients of each term S^r, $(r = 0, 1, \dots)$ which are polynomials in C_1, S_1 must vanish by virtue of $C_1^2+S_1^2=1$. These can be verified by the procedures treated previously for a polynomial identity in one variable. Alternatively, we can proceed in a fashion analogous to Method III.

Numerous examples of trigonometric identities of two variables arise from triangle identities, e.g.,

$$\tan x + \tan y + \tan z = \tan x \tan y \tan z, \tag{4}$$

$$\sin x + \sin y + \sin z = 4\cos\frac{x}{2}\cos\frac{y}{2}\cos\frac{z}{2}, \tag{5}$$

$$\sin^3 x\cos(y-z) + \sin^3 y\cos(z-x) + \sin^3 z\cos(x-z) = 3\sin x\sin y\sin z. \tag{6}$$

Replacing z by $\pi - x - y$, the above identities reduce to

$$\frac{\sin x}{\cos x} + \frac{\sin y}{\cos y} - \frac{\sin(x+y)}{\cos(x+y)} = -\frac{\sin x}{\cos x}\frac{\sin y}{\cos y}\frac{\sin(x+y)}{\cos(x+y)}, \tag*{(4)$'$}$$

$$\sin x + \sin y + \sin(x+y) = 4\cos\frac{x}{2}\cos\frac{y}{2}\sin\frac{x+y}{2}, \tag*{(5)$'$}$$

$$\sin^3(x+y)\cos(x-y) - \sin^3 x\cos(2y+x) - \sin^3 y\cos(2x+y) = 3\sin x\sin y\sin(x+y). \tag*{(6)$'$}$$

In order to use the algorithmic approach here, we must first reformulate the identities as polynomial equations in the variables $C = \cos kx$, $S = \sin kx$, $C_1 = \cos ky$, $S_1 = \sin ky$ where k is suitably chosen. To do this, we will also need the addition formulae

$$\sin(x+y) = \sin x\cos y + \sin y\cos x,$$

$$\cos(x+y) = \cos x\cos y - \sin x\sin y.$$

Often this will not be an efficient method but at least it will be direct. For example, it is much easier to establish (4)$'$ by a suitable grouping of terms, i.e.,

$$\frac{\sin x}{\cos x} + \frac{\sin y}{\cos y} = \frac{\sin(x+y)}{\cos(x+y)}\left\{1 - \frac{\sin x\sin y}{\cos x\cos y}\right\},$$

and then,

$$\frac{\sin(x+y)}{\cos x\cos y} = \frac{\sin(x+y)}{\cos(x+y)}\frac{\cos(x+y)}{\cos x\cos y}.$$

For (5)$'$, k is taken as $1/2$, leading to

$$2SC + 2S_1C_1 + 2(SC_1+S_1C)(CC_1-SS_1) = 4CC_1(SC_1+S_1C). \tag*{(5)$''$}$$

Then (5)″ simplifies to

$$SC\left(1-S_1^2-C_1^2\right)+S_1C_1\left(1-S^2-C^2\right)=0.$$

The third example (6)′ is more involved, and leads to (with $k=1$)

$$\begin{aligned}&S^3CC_1^4+3S^2C^2S_1C_1^3+3SC^3S_1^2C_1^2+C^4S_1^3C_1+S^4S_1C_1^3+3S^3CS_1^2C_1^2\\&+3S^2C^2S_1^3C_1+SC^3S_1^4-2S^3CC_1^2+S^3C+2S^4C_1S_1-2C^2S_1^3C_1\\&+S_1^3C_1+2SCS_1^4=3S^2S_1C_1+3SS_1^2C.\end{aligned}\tag{6''}$$

To establish (6)″, we merely replace C_1^2 by $1-S_1^2$ (but not C_1) and C^2 by $1-S^2$ (but not C) and then every term will cancel out. However, (6)″ can also be partitioned into several sub-identities by separating even and odd terms as was done in the previous section. This is justified by expressing $P(C,S,C_1,S_1)$ in the form

$$\begin{aligned}P(C,S,C_1,S_1)&=Q\left(C^2,S,C_1,S_1\right)+R\left(C^2,S,C_1,S_1\right)C\\&=Q_1\left(C^2,S,C_1^2,S_1\right)+Q_2\left(C^2,S,C_1^2,S_1\right)C_1\\&\quad+\left(R_1\left(C^2,S,C_1^2,S_1\right)+R_2\left(C^2,S,C_1^2,S_1\right)C_1\right)C.\end{aligned}$$

It then follows as before that each of Q_1, Q_2, R_1, and R_2 must vanish for $C^2+S^2=1$ and $C_1^2+S_1^2=1$.

In (6)″, there are no terms containing factors of the form $C^{2m}C_1^{2n}$ (even-even) or $C^{2m+1}C_1^{2n+1}$ (odd-odd). The odd-even terms and the even-odd terms are, after dividing out the common factor, respectively,

$$S^2C_1^4+3C^2S_1^2C_1^2+3S^2S_1^2C_1^2+C^2S_1^4-2S^2C_1^2+S^2+2S_1^4-3S_1^2,$$
$$S_1^2C^4+3C_1^2S^2C^2+3S_1^2S^2C^2+C_1^2S^4-2S_1^2C^2+S_1^2+2S^4-3S^2,$$

which are equivalent expressions. Both vanish identically after replacing C^2 by $1-S^2$ and C_1^2 by $1-S_1^2$.

Not only can we reduce triangle identities to polynomial identities of the form $P(C,S,C_1,S_1)=0$, but we can also reverse this process to obtain an endless source of triangle identities. First let $C_i=\cos x_i$, $S_i=\sin x_i$, $i=1,2,3$ where x_1,x_2,x_3 are angles of a triangle. Thus, $x_1+x_2+x_3=\pi$ and

$$\begin{aligned}&C_3=S_1S_2-C_1C_2, && C_1=S_2S_3-C_2C_3, && C_2=S_3S_1-C_3C_1,\\&S_3=S_1C_2+S_2C_1, && S_1=S_2C_3+S_3C_2, && S_2=S_3C_1+S_1C_3.\end{aligned}\tag{7}$$

We can now start with two arbitrary polynomials F and G to give the identity

$$P(C_1,S_1,C_2,S_2)=\left(C_1^2+S_1^2-1\right)F(C_1,S_1,C_2,S_2)+\left(C_2^2+S_2^2-1\right)G(C_1,S_1,C_2,S_2)\equiv 0.$$

Finally, we just replace some or all of the variables C_1,S_1,C_2,S_2 in $(C_1^2+S_1^2-1)$ and/or $(C_2^2+S_2^2-1)$ by their formulae in (7). These identities will not be cyclic but if cyclic ones are desired, we merely write

$$\begin{aligned}&\left(C_1^2+S_1^2-1\right)F_1(C_2,S_2,C_3,S_3)+\left(C_2^2+S_2^2-1\right)F_2(C_3,S_3,C_1,S_1)\\&\qquad+\left(C_3^2+S_3^2-1\right)F_3(C_1,S_1,C_2,S_2)\equiv 0\end{aligned}$$

and so on. As a particular example, we have

$$\sum_{\text{cyclic}}\left\{(S_2C_3+S_3C_2)^2+(S_2S_3-C_2C_3)^2-1\right\}\left\{(S_2S_3)^2+(C_2C_3)^2+1\right\}\equiv 0.$$

Extension of the factorization result

If $P(x,y)$ and $Q(x,y)$ are polynomials in two variables which are both zero for an infinite number of pairs of values (x_i,y_i), then they both must have a common polynomial factor. This follows from a known result [**3**, p. 210] which is proved by the algorithm for the greatest common divisor for polynomials of two variables,

THEOREM 1. *If two polynomials $P(x, y)$ and $Q(x, y)$ are relatively prime, then there is only a finite number of pairs of values (x_i, y_i) for which P and Q both vanish.*

Geometrically, this theorem tells us that two algebraic plane curves $P(x, y) = 0$, $Q(x, y) = 0$, can only intersect in an infinite number of points when they have an entire algebraic curve in common. As to the actual number of pairs of points (x_i, y_i), we have

BEZOUT'S THEOREM [**4**]. *If $P(x, y)$ and $Q(x, y)$ are relatively prime polynomials of degrees m and n, then the number of pairs of points (x_i, y_i), real or complex, for which both P and Q vanish is mn, counting multiplicities.*

Three results which follow from Bezout's Theorem are the following:

(i) *The equations $x + y = 0$ and $x^2 + y^2 = 0$ are both satisfied only by the point* $(0,0)$, *counted twice.*

To understand the double root, consider the double intersection of the line $x + y = 0$ with the circle $x^2 + y^2 = r^2$ as $r \to 0$.

(ii) *Two conics can intersect in at most four real points.*

(iii) *If a regular n-gon $(n > 4)$ is inscribed in (or circumscribed about) an ellipse, the ellipse must be the circumcircle (or the incircle) of the n-gon.*

If not, the number of intersections of the ellipse with the circumcircle (or the incircle) is at least n which is impossible by (ii). It can be shown that for $n = 4$, there exists a unique square inscribed in (or circumscribed about) any proper ellipse.

An immediate corollary to THEOREM 1 is the following.

COROLLARY. *Let $P(x, y)$ and $Q(x, y)$ be polynomials of two variables, with $Q(x, y)$ irreducible (in the complex number field). If P vanishes at all points (x_i, y_i) (real and complex) at which Q vanishes, then Q is a factor of P.*

This corollary also holds for polynomials in more than two variables even though Theorem 1 is invalid for polynomials of more than two variables [**3**, p. 213]. To see that Theorem 1 is invalid in these cases, just note that $P(x, y, z) = 0$ and $Q(x, y, z) = 0$ are two surfaces which intersect, in general, in a curve. Also, it should be noted that the condition of irreducibility is essential in the Corollary. To see this, just consider $Q(x, y) = xy^2$ and $P(x, y) = x^2y$.

It is to be noted that the vanishing points in the Corollary can be *real or complex*, whereas for the central identity $P(C, S) \equiv 0$, of our earlier discussion, we only considered all *real* points (C, S) such that $C^2 + S^2 = 1$. This restriction to only *real* points is an essential difference. For example, if $P(x, y, z)$ vanishes for all points, real or complex, where the irreducible polynomial

$$Q(x, y, z) = (x - y)^2 + (y - z)^2 + (z - x)^4$$

vanishes, then Q is a factor of P. However, if we restrict the vanishing points to only all real points, then P need not be divisible by Q (for example P could be the polynomial $x - y$). This leads to the following unsolved problem.

PROBLEM. *Characterize those irreducible polynomials $Q(x_1, x_2, \dots, x_n)$ (over the complex number field) such that if any polynomial $P(x_1, x_2, \dots, x_n)$ vanishes for all real $(x_1, x_2, \dots, x_n)$ for which Q vanishes, then Q must divide P.*

We now show that one such class of polynomials is

$$Q(x_1, \dots, x_n) = x_1^m + x_2^m + \cdots + x_n^m - 1 \text{ where } n \geq 2,\ m = 1, 2, \dots. \tag{8}$$

It is known that the polynomial $x_1^m + x_2^m + \cdots + x_r^m$ $(r \geq 3)$ is irreducible [**4**]. This easily implies that Q in (8) is also irreducible. Our proof is by induction and we first establish our result for $n = 2$ (this case includes the result of Magid and Dobbs).

Assume that $Q(x_1, x_2)$ is as in (8) and that some other polynomial $P(x_1, x_2)$ vanishes for all real (x_1, x_2) for which Q vanishes. Dividing P by Q, we get

$$P(x_1, x_2) = (x_1^m + x_2^m - 1)P_0(x_1, x_2) + x_1^{m-1}P_1(x_2) + x_1^{m-2}P_2(x_2) + \cdots + P_m(x_2).$$

We will show that all the polynomial terms $P_1, P_2, \ldots, P_m$ must vanish identically. For all x_2 satisfying $0 \leqslant x_2 \leqslant 1$ and $x_1 = (1 - x_2^m)^{1/m}$, we have

$$x_1^{m-1}P_1(x_2) + x_1^{m-2}P_2(x_2) + \cdots + P_m(x_2) = 0. \tag{9}$$

If we let $x_2 = 1$ in (9), then $x_1 = 0$, so $1 - x_2$ is a factor of $P_m(x_2)$. Now factor out all the $1 - x_2$ factors in each of the terms P_i. It is to be noted that the degrees of the factor $1 - x_2$ in each of the m terms must all be different, the first $m - 1$ of them being non-integral. We now factor out the lowest degree factor $(1 - x_2)^d$ from all the terms so that exactly one of the terms will have no $1 - x_2$ factor left. Thus (9) becomes

$$(1 - x_2)^d\left[(1 - x_2)^e R(x_2) + R_j(x_2)\right] \tag{10}$$

where

$$(1 - x_2)^d R_j(x_2) = (1 - x_2^m)^{(m-j)/m} P_j(x_2)$$

and d, e are positive rational numbers. Now let $x_2 = 1 - \varepsilon$ where $\varepsilon > 0$ is arbitrarily small. By continuity, the sum of terms in (10) cannot vanish. Consequently, that term R_j without the $1 - x_2$ factor must vanish identically (and hence P_j vanishes identically). Proceeding in a similar manner, it follows that all the remaining terms P_i must also vanish identically.

We now assume our result holds for $n = 2, 3, \ldots, k$ and establish the result for $n = k + 1$. Here,

$$P(\mathbf{X}) = Q_{k+1}P_0(\mathbf{X}) + x_1^{m-1}P_1(\mathbf{X}') + x_1^{m-2}P_2(\mathbf{X}') + \cdots + P_m(\mathbf{X}')$$

where $\mathbf{X}, \mathbf{X}'$ denote, respectively, the set of variables $x_1, x_2, \ldots, x_{k+1}$; $x_2, x_3, \ldots, x_{k+1}$, and $Q_{k+1}(\mathbf{X}) = x_1^m + x_2^m + \cdots + x_{k+1}^m - 1$. Then,

$$x_1^{m-1}P_1(\mathbf{X}') + x_1^{m-2}P_2(\mathbf{X}') + \cdots + P_m(\mathbf{X}') = 0$$

for all real $\mathbf{X}$ such that $Q_{k+1} = 0$. Setting $x_1 = 0$, it follows that $P_m(\mathbf{X}') = 0$ for all $\mathbf{X}'$ such that $x_2^m + x_3^m + \cdots + x_{k+1}^m = 1$. Thus by the inductive hypothesis, $P_m(\mathbf{X}')$ is divisible by $1 - x_2^m - x_3^m - \cdots - x_{k+1}^m$ or x_1^m. Now letting $x_1^m = \varepsilon$ (ε arbitrarily small) it follows, as in the case $n = 2$, that all the P_i must vanish identically. Whence, $P(\mathbf{X})$ is divisible by Q_{k+1}. Then by induction, our result holds for all $n \geqslant 2$.

Finally, we can show more generally in a similar inductive manner that if the polynomial $P(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ vanishes for all real values of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ such that simultaneously

$$Q_i(\mathbf{X}_i) \equiv x_{i1}^{\alpha(i)} + x_{i2}^{\alpha(i)} + \cdots + x_{im(i)}^{\alpha(i)} - 1 = 0 \qquad (i = 1, 2, \ldots, n)$$

where $m(i) \geqslant 2$ and $\alpha(i)$ are positive integers, then

$$P = Q_1P_1 + Q_2P_2 + \cdots + Q_nP_n$$

where the P_j's are polynomials in all the $\mathbf{X}_i$'s.

References

[1] M. Bocher, Introduction to Higher Algebra, Macmillan, New York, 1935.

[2] D. E. Dobbs, Proving trig. identities to fresh persons, The MATYC J., 14 (1980) 39–42.

[3] A. R. Magid, Trigonometric identities, this MAGAZINE, 47 (1974) 226–227.

[4] J. A. Tyrrell, Problem 475, this MAGAZINE, 35 (1962) 310.

[5] R. J. Walker, Algebraic Curves, Dover, New York, 1950, pp. 59, 111.

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

Empty or Infinitely Full?

Minja K. Paik
George Mason University
Fairfax, VA 22030

The concept of infinity and the inability to interchange two or more mathematical operations in the limit situation intrigue and baffle students of mathematics. One example that nicely illustrates these two concepts is found in the second chapter of Sheldon Ross's popular book, *A First Course in Probability* [**1**], under the subheading "Probability and a Paradox." Students of engineering and physics generally disavow the paradox quickly and say that the answer (in the second case) is simply wrong; mathematics students are more cautious in coming to a conclusion as they see a valid argument in the answers. In this note we will look into the nature of the problem and give a mathematical explanation of the paradox.

Quoting directly from the book, the paradox is as follows.

> Suppose that we possess an infinitely large urn and an infinite collection of balls labeled ball number 1, number 2, number 3, and so on. Consider an experiment performed as follows. At 1 minute to 12 P.M., balls numbered 1 through 10 are placed in the urn and ball number 10 is withdrawn. (Assume the withdrawal takes no time.) At $\frac{1}{2}$ minute to 12 P.M., balls numbered 11 through 20 are placed in the urn and ball number 20 is withdrawn. At $\frac{1}{4}$ minute to 12 P.M., balls numbered 21 through 30 are placed in the urn and ball number 30 is withdrawn. At $\frac{1}{8}$ minute to 12 P.M., and so on. The question of interest is, how many balls are in the urn at 12 P.M.?
>
> The answer to this question is clearly that there are an infinite number of balls in the urn at 12 P.M., since any ball whose number is not of the form $10n$, $n \geqslant 1$, will have been placed in the urn and will not have been withdrawn before 12 P.M. Hence the problem is solved when the experiment is performed as described.
>
> However, let us now change the experiment and suppose that at 1 minute to 12 P.M. balls numbered 1 through 10 are placed in the urn and ball number 1 is withdrawn; at $\frac{1}{2}$ minute to 12 P.M., balls numbered 11 through 20 are placed in the urn, and ball number 2 is withdrawn; at $\frac{1}{4}$ minute to 12 P.M., balls 21 through 30 are placed in the urn and ball number 3 is withdrawn; at $\frac{1}{8}$ minute to 12 P.M., balls numbered 31 through 40 are placed in the urn and ball number 4 is withdrawn, and so on. For this new experiment how many balls are in the urn at 12 P.M.?
>
> Surprisingly enough, the answer now is that the urn is *empty* at 12 P.M. For consider any ball—say, ball number n. At some time prior to 12 P.M., in particular, at $\left(\frac{1}{2}\right)^{n-1}$ minutes to 12 P.M., this ball would have been withdrawn from the urn. Hence, for each n, ball number n is not in the urn at 12 P.M.; therefore, the urn must be empty at this time.

The two experiments are identical from the *physical* standpoint. Then why the difference in the answer? A good deal of discussion is centered on the second experiment. Strictly speaking, of course, this experiment is impossible to carry out and the end result at 12 P.M. is never realizable. Some students think that it is absurd to talk about 12 P.M. in this context and that this feature alone is what the paradox is all about, it being somewhat related to Zeno's paradox. To eliminate this problem, we could rephrase the experiments in an equivalent form as follows.

Problem. *Suppose, on day 1, you put in an urn 10 balls numbered 1 through 10 and withdraw the ball numbered n_1. On day 2, 10 more balls numbered consecutively 11 through 20 are put in the urn with the ball numbered n_2 being withdrawn at the same time, and so on. Then, how many balls are there in the urn in the limit? Infinitely many or none?*

The paradox noted by Ross is that in the first experiment with $n_i = 10i$, the answer is infinitely many, while in the second experiment with $n_i = i$, the answer is none.

Still others say that this paradox has to do with the fact that an infinite set can be put into one-to-one correspondence with a proper subset of itself. However, scientists who understand this fact perfectly well still find the inconsistency of the two different answers disturbing since the number of balls remaining in the urn is increasing steadily and exactly in the same manner in each of the two situations. In the following discussion, we shall clarify the source of this paradox.

Let us call the first and the second experiments A and B respectively. Also, let A_n and B_n be the sets consisting of the integers representing the balls in the urns on day n in the respective experiments. Also let $\chi(A_n)$ be the cardinality of A_n. Clearly, in both cases $\chi(A_n)=\chi(B_n)=9n$, which increases without bound as the days go on. Of course, the cardinality is interpreted as the number of balls in the urn. This is why some say that there should be infinitely many balls in the urn in the limit in both cases. It is certainly true for experiment A, since

$$A_n=\{i|1\leqslant i<10n, i\not\equiv 0 \bmod 10\}$$

and $\lim A_n=\cup_{n=1}^{\infty}A_n$ is an infinite set. In the second experiment, however, since every ball is eventually withdrawn,

$$\limsup B_n=\liminf B_n$$

is the null set. Thus $\lim B_n$ exists and is equal to the null set. Basically, the paradox of the different answers in case A and case B results from the fact that the cardinality viewed as a set function is not continuous in the latter case, because $\lim_{n\to\infty}\chi(B_n)=\infty$, but $\chi(\lim_{n\to\infty}B_n)=0$. Disconcertingly, this function is discontinuous at infinity in the sense that the values of the function are ever increasing but its value at the limit point is zero.

A similar example of this type of situation is familiar in analysis. Consider a sequence of functions on the positive real line,

$$f_n(x)=\begin{cases}1 & \text{if } x\in[n+1,10n]\\ 0 & \text{otherwise.}\end{cases}$$

The limit of the integral $\int_0^{\infty}f_n(x)\,dx$ is infinity, but the integral of the limit, $f(x)=\lim_{n\to\infty}f_n(x)$, is zero.

Does it make sense to talk about a function which behaves singularly at infinity? Many scientists are accustomed to treating infinity as an approximation to a very large number. The use of the central limit theorem as a large sample approximation is an example. Therefore they are disturbed by a physical identity (that there are $9n$ balls in the urn on day n in both schemes) that holds true on day n for an arbitrarily large n but breaks down on day "infinity." We should remember, however, that the "day infinity" does not come, thereby the "paradox" is not a *real* paradox. It remains a fairy tale that all of Cinderella's glories suddenly disappear when 12 o'clock strikes.

It is interesting to ask what happens in our stated Problem if at each stage, the balls are withdrawn at random from those present, i.e., ball n_i which is withdrawn is any of the balls in the urn at stage i. Ross proves that in this case, the probability is 1 that in the limit the urn will be empty (in his formulation, at 12 P.M., the urn is empty). His proof can be briefly summarized as follows. Let E_n be the event that ball number 1 is still in the urn after the first n withdrawals. Clearly,

$$P(E_n)=\prod_{i=1}^{n}9i/(9i+1).$$

The event E that the number 1 ball will be in the urn in the limit is $\cap_{n=1}^{\infty}E_n=\lim_{n\to\infty}E_n$. Since probability is a continuous set function,

$$P(E)=P\Big(\lim_{n\to\infty}E_n\Big)=\lim_{n\to\infty}P(E_n).$$

But $\prod_{i=1}^{\infty}9i/(9i+1)=0$. Therefore, the probability that ball number 1 will eventually come out is 1. One can show similarly that *every* ball has a zero probability of remaining in the urn forever and that, by the covering theorem, all of the balls would come out of the urn by noon with probability 1.

Here again, one must not imagine the number of balls remaining in the urn to be decreasing to zero in some fashion. The unexpected conclusion is just a mathematical consequence of the way the random withdrawal process defines probability in an infinite dimensional product space.

Reference

[1] Sheldon Ross, A First Course in Probability, Macmillan, New York, 1976.

A Note on Linear Diophantine Equations

KENNETH J. WHITCOMB
Colorado State University
Fort Collins, CO 80523

Students in differential equations are always solving problems about oscillating springs with external impressed forces and getting answers like

$$y_1 = 3\sin 6t + 2\sin 4t \quad \text{or} \quad y_2 = 3\sin 7t + 2\sin 3t.$$

Clearly, neither y can ever be larger than 5, and some students (or maybe their teachers) may have wondered if y ever got that large. There is a simple way of determining this by inspection: y_2 makes it and y_1 does not. Here are two more: $3\sin 8t + 2\sin 5t$ never gets to 5, but $3\sin 6t + 2\sin 14t$ does. Give up? The answer is given by the following theorem.

THEOREM. *If a and b are integers ($\neq 0$) and A and B have the same sign, then*

$$A\sin at + B\sin bt \tag{1}$$

attains the value $|A| + |B|$ if and only if $a \equiv b \pmod 4$ and (a, b) divides $(a-b)/4$. When A and B have opposite sign, then (1) attains the value $|A| + |B|$ if and only if $a \equiv -b \pmod 4$ and (a, b) divides $(a+b)/4$.

Let us prove the case where A and B are both positive.

$$\sin at = 1 \text{ implies } at = \frac{\pi}{2} + k_1(2\pi) \quad \text{or} \quad t = \frac{\pi}{2a} + \frac{k_1(2\pi)}{a}$$

and

$$\sin bt = 1 \text{ implies } bt = \frac{\pi}{2} + k_2(2\pi) \quad \text{or} \quad t = \frac{\pi}{2b} + \frac{k_2(2\pi)}{b}$$

where k_1 and k_2 are integers. After equating these t results and simplifying, we have

$$b + 4bk_1 = a + 4ak_2 \quad \text{or} \quad 4(bk_1 - ak_2) = a - b.$$

To get any solution at all in integers for this last equation, 4 must divide $a - b$ [$a \equiv b \pmod 4$] and further, this linear Diophantine equation is solvable in integers k_1 and k_2 if and only if (a, b) divides $(a-b)/4$.

The interested reader can verify the other cases and perhaps see what happens when the sine terms are changed to cosines. Some readers might remark that the theorem is easily proved using calculus—which is true—but why use it if you don't have to? Some other interesting min-max problems of a trigonometric nature are solved by Niven, sans calculus, in [**1**, § 5].

Reference

[1] Ivan Niven, Maxima and Minima Without Calculus, MAA Dolciani Mathematical Expositions no. 6, 1981.

Here again, one must not imagine the number of balls remaining in the urn to be decreasing to zero in some fashion. The unexpected conclusion is just a mathematical consequence of the way the random withdrawal process defines probability in an infinite dimensional product space.

Reference

[1] Sheldon Ross, A First Course in Probability, Macmillan, New York, 1976.

A Note on Linear Diophantine Equations

Kenneth J. Whitcomb
Colorado State University
Fort Collins, CO 80523

Students in differential equations are always solving problems about oscillating springs with external impressed forces and getting answers like

$$y_1 = 3\sin 6t + 2\sin 4t \quad \text{or} \quad y_2 = 3\sin 7t + 2\sin 3t.$$

Clearly, neither y can ever be larger than 5, and some students (or maybe their teachers) may have wondered if y ever got that large. There is a simple way of determining this by inspection: y_2 makes it and y_1 does not. Here are two more: $3\sin 8t + 2\sin 5t$ never gets to 5, but $3\sin 6t + 2\sin 14t$ does. Give up? The answer is given by the following theorem.

Theorem. *If a and b are integers ($\neq 0$) and A and B have the same sign, then*

$$A\sin at + B\sin bt \tag{1}$$

attains the value $|A| + |B|$ if and only if $a \equiv b \pmod 4$ and (a, b) divides $(a-b)/4$. When A and B have opposite sign, then (1) *attains the value $|A| + |B|$ if and only if $a \equiv -b \pmod 4$ and (a, b) divides $(a+b)/4$.*

Let us prove the case where A and B are both positive.

$$\sin at = 1 \text{ implies } at = \frac{\pi}{2} + k_1(2\pi) \quad \text{or} \quad t = \frac{\pi}{2a} + \frac{k_1(2\pi)}{a}$$

and

$$\sin bt = 1 \text{ implies } bt = \frac{\pi}{2} + k_2(2\pi) \quad \text{or} \quad t = \frac{\pi}{2b} + \frac{k_2(2\pi)}{b}$$

where k_1 and k_2 are integers. After equating these t results and simplifying, we have

$$b + 4bk_1 = a + 4ak_2 \quad \text{or} \quad 4(bk_1 - ak_2) = a - b.$$

To get any solution at all in integers for this last equation, 4 must divide $a - b$ [$a \equiv b \pmod 4$] and further, this linear Diophantine equation is solvable in integers k_1 and k_2 if and only if (a, b) divides $(a-b)/4$.

The interested reader can verify the other cases and perhaps see what happens when the sine terms are changed to cosines. Some readers might remark that the theorem is easily proved using calculus—which is true—but why use it if you don't have to? Some other interesting min-max problems of a trigonometric nature are solved by Niven, sans calculus, in [**1**, § 5].

Reference

[1] Ivan Niven, Maxima and Minima Without Calculus, MAA Dolciani Mathematical Expositions no. 6, 1981.

Matchings, Derangements, Rencontres

D. Hanson
K. Seyffarth
J. H. Weston
University of Regina
Regina, Saskatchewan S4S 0A2, Canada

A standard problem introduced in many combinatorics texts (for example [1], [5]) when discussing the Principle of Inclusion and Exclusion is the **Derangement Problem.** It is often presented as follows:

At a restaurant n people check their hats, and when they leave their hats are returned in a random order. In how many ways can it happen that no one receives his own hat back, and further what is the probability of such an event?

We propose the following generalization:

Suppose there are initially $n - k$ unclaimed hats in the checkroom, and k people arrive and check their hats. When these k people leave, they are each given a hat, chosen at random from the n hats. What is the probability that exactly r of them receive their own hats?

This latter problem may be given a more realistic interpretation:

A matching question on an exam has k questions with n possible answers to choose from, each question having a unique answer. If a student guesses the answers at random, using each answer at most once, what is the probability of obtaining r of the k correct answers?

We will call such a question an (n, k)**-matching problem.**

In what follows we will see that it is almost pointless for the unprepared student to guess when answering such a matching problem. For example, the probability of obtaining a score of at most one correct on a $(10,7)$-matching problem is approximately .85, and the probability of scoring five right is about .0005. In fact the expected score on an (n, k)-matching problem is k/n, i.e., at most one.

By a k**-permutation**, σ, of a set $N = \{1,2,\dots,n\}$, we mean an ordered arrangement of k of the n elements of N. For $n \geqslant k \geqslant r \geqslant 0$, $n > 0$, let $D(n, k, r)$ denote the number of such k-permutations for which $\sigma(i) = i$ for exactly r values of i, $1 \leqslant r \leqslant k$; also let $P(n, k, r)$ denote the probability of such an event, i.e.,

$$P(n,k,r) = D(n,k,r)/(n!/(n-k)!). \tag{1}$$

Note that $D(n, k, r)$ is precisely the number of ways of obtaining exactly r correct answers on an (n, k)-matching problem (where we assume, without loss of generality, that the answer to question i is answer i), and that $P(n, k, r)$ is the corresponding probability.

Observe that

$$D(n,k,r) = \binom{k}{r} D(n-r, k-r, 0) \tag{2}$$

since we may first select the r questions to be answered correctly in $\binom{k}{r}$ ways. Consequently we may consider $D(n, k, 0)$ as one approach to evaluating the function $D(n, k, r)$. Recall that the **Principle of Inclusion and Exclusion** states that if $P_1, P_2, \dots, P_t$ are t properties which each element of a set S may or may not possess and if A_i is the subset of objects of S which have property P_i, $i = 1,2,\dots,t$, then the number of objects of S which have none of the properties $P_1, P_2, \dots, P_t$ is given by

$$\overline{|A_1 \cup A_2 \cup \cdots \cup A_t|} = |S| - \sum |A_i| + \sum |A_i \cap A_j|$$
$$- \sum |A_i \cap A_j \cap A_k| + \cdots + (-1)^t |A_1 \cap A_2 \cap \cdots \cap A_t|.$$

This Principle is simply a generalization of the 2-property identity $|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|$ (see **[1]** or **[5]** for a more detailed explanation). For our purposes let P_i, $i = 1, 2, \ldots, k$ be the property that a k-permutation, σ, of $N = \{1, 2, \ldots, n\}$ satisfies $\sigma(i) = i$. Then, in this case, S is the set of k-permutations of N and we have

$$|S| = n(n-1)(n-2)\cdots(n-k+1) = n!/(n-k)!.$$

Similarly $|A_i| = (n-1)!/(n-k)!$ since here we are counting those k-permutations with one fixed point and we have only $(n-1)$ choices for the first place, $(n-2)$ choices for the second, etc. In a like manner

$$|A_i \cap A_j| = (n-2)!/(n-k)!, \ldots, |A_1 \cap A_2 \cap \cdots \cap A_k| = (n-k)!/(n-k)!,$$

and therefore evaluating $\overline{|A_1 \cup A_2 \cup \cdots \cup A_k|}$ we obtain

THEOREM 1. $D(n,k,0) = \dfrac{1}{(n-k)!} \displaystyle\sum_{j=0}^{k} (-1)^j \binom{k}{j} (n-j)!, \ k \leqslant n.$

Then by (2)

COROLLARY. $D(n,k,r) = \dfrac{\binom{k}{r}}{(n-k)!} \displaystyle\sum_{j=0}^{k-r} (-1)^j \binom{k-r}{j} (n-r-j)!$

$$P(n,k,r) = \frac{\binom{k}{r}}{n!} \sum_{j=0}^{k-r} (-1)^j \binom{k-r}{j} (n-r-j)!.$$

TABLES 1 and 2 give values for $D(n,k,r)$ and $P(n,k,r)$ for $r \leqslant 1$ and $k, n \leqslant 10$.

In the Derangement Problem mentioned earlier, we wish to count the number of permutations of a set of n objects in which no object retains its rightful place, i.e., $D(n,n,0)$ in our terminology. It follows from Theorem 1 that

$$D(n,n,0) = \sum_{j=0}^{n} (-1)^j \frac{n!}{j!(n-j)!} (n-j)! = n! \sum_{j=0}^{n} \frac{(-1)^j}{j!} \qquad (3)$$

and

$$P(n,n,0)=\sum_{j=0}^{n}\frac{(-1)^j}{j!}. \tag{4}$$

It is interesting to note that $P(n,n,0)$ converges rapidly to $1/e$ and somewhat surprisingly is essentially independent of n. Consequently

$$D(n,n,0)=n!(1-1/1!+1/2!+\cdots+(-1)^n/n!)$$

is the closest integer to $n!/e$.

The classical "**Problème des Rencontres**," (see [3], [4]), of which the Derangement Problem is a special case, counts the number of permutations of a set of n objects in which r objects, $n \geqslant r \geqslant 0$, retain their rightful places, i.e., $D(n, n, r)$. This problem is due to the French mathematician Pierre Remond de Montmort (1678–1719), [3], and was investigated from a recurrence point of view by Euler and generalized by Laplace. It follows from (1), (2) and (3) that

$$P(n,n,r)\approx\frac{1}{n!}\binom{n}{r}\frac{(n-r)!}{e}=\frac{1}{r!e}. \tag{5}$$

A close look at TABLES 1 and 2 suggests that a number of recurrence relations exist for the numbers $D(n,k,r)$. For example, note the relationship between the jth column of $D(n,k,1)$ and the $(j-1)$st column of $D(n,k,0)$, or note that the second diagonal of the $D(n,k,1)$ table is the same as the first diagonal of the $D(n,k,0)$ table. We will look carefully at a few of these relations and illustrate their connection with the standard recurrences for the Derangement Problem. The derangements $D(n,n,0)$ in fact satisfy

$$D(n,n,0)=(n-1)(D(n-1,n-1,0)+D(n-2,n-2,0)) \tag{6}$$

$D(n,k,0)$

$n \backslash k$	1	2	3	4	5	6	7	8	9	10
1	0									
2	1	1								
3	2	3	2							
4	3	7	11	9						
5	4	13	32	53	44					
6	5	21	71	181	309	265				
7	6	31	134	465	1214	2119	1854			
8	7	43	227	1001	3539	9403	16687	14833		
9	8	57	356	1909	8544	30637	82508	148329	133496	
10	9	73	527	3333	18089	81901	296967	808393	1468457	1334961

$P(n,k,0)$

$n \backslash k$	1	2	3	4	5	6	7	8	9	10
1	0.0000									
2	.5000	.5000								
3	.6667	.5000	.3333							
4	.7500	.5833	.4583	.3750						
5	.8000	.6500	.5333	.4417	.3667					
6	.8333	.7000	.5917	.5028	.4292	.3681				
7	.8571	.7381	.6381	.5536	.4817	.4204	.3679			
8	.8750	.7679	.6756	.5958	.5266	.4664	.4139	.3679		
9	.8889	.7917	.7063	.6313	.5651	.5066	.4547	.4088	.3679	
10	.9000	.8111	.7319	.6613	.5982	.5417	.4910	.4455	.4047	.3679

TABLE 1

for $n=3,4,5,\ldots$, and a nice combinatorial verification of this result is due to Euler (see for example [**1**], p. 82). An analogous result is:

THEOREM 2. *For* $n \geqslant k \geqslant r \geqslant 0$, $n > 2$, *and* $D(n,k,-1)$ *defined to be* 0,

$$D(n,k,r) = D(n-1,k-1,r-1) + (n-1)D(n-1,k-1,r) + (k-1)(D(n-2,k-2,r) - D(n-2,k-2,r-1)). \tag{7}$$

Proof. Consider an (n,k)-matching problem with questions numbered $1,2,\ldots,k$ and answers numbered $1,2,\ldots,n$. Without loss of generality we assume the answer to question i is answer i. Of the k-permutations counted by $D(n,k,r)$

(i) there are $D(n-1,k-1,r-1)$ with a 1 in the first place,
(ii) there are $(n-k)D(n-1,k-1,r)$ with an integer $m > k$ in the first place, and
(iii) those with j, $1 < j \leqslant k$ in the first place may be partitioned into $(k-1)$ classes of equal size.

To examine those in (iii), we consider, without loss of generality, the case $j=2$. If 1 is in the second place, there are $D(n-2,k-2,r)$ k-permutations of the required type. If 1 is not in the second place, and there were no restrictions, we would have an $(n-1,k-1)$-matching problem requiring r correct and we can complete this problem in $D(n-1,k-1,r)$ ways. However, this count includes the possibility that question 2 is correctly answered—an impossibility since we have used answer 2 already. Thus we must subtract these possibilities of which there are $D(n-2,k-2,r-1)$. By (i), (ii) and (iii) we have

$$D(n,k,r) = D(n-1,k-1,r-1) + (n-k)D(n-1,k-1,r) + (k-1)(D(n-2,k-2,r) + (D(n-1,k-1,r) - D(n-2,k-2,r-1))),$$

$D(n,k,1)$

$n \backslash k$	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	0								
3	1	2	3							
4	1	4	9	8						
5	1	6	21	44	45					
6	1	8	39	128	265	264				
7	1	10	63	284	905	1854	1855			
8	1	12	93	536	2325	7284	14833	14832		
9	1	14	129	908	5005	21234	65821	133496	133497	
10	1	16	171	1424	9545	51264	214459	660064	1334961	1334960

$P(n,k,1)$

$n \backslash k$	1	2	3	4	5	6	7	8	9	10
1	1.0000									
2	.5000	0.0000								
3	.3333	.3333	.5000							
4	.2500	.3333	.3750	.3333						
5	.2000	.3000	.3500	.3667	.3750					
6	.1667	.2667	.3250	.3556	.3681	.3667				
7	.1429	.2381	.3000	.3381	.3591	.3679	.3681			
8	.1250	.2143	.2768	.3190	.3460	.3613	.3679	.3679		
9	.1111	.1944	.2560	.3003	.3310	.3511	.3628	.3679	.3679	
10	.1000	.1778	.2375	.2825	.3156	.3390	.3546	.3638	.3679	.3679

TABLE 2

which was to be shown.

A consequence of Theorem 2 is the following recurrence for the "rencontre" numbers.

COROLLARY: $D(n,n,r)=nD(n-1,n-1,r)+(-1)^{n-r}\binom{n}{r}$, $r \geqslant 0$.

Proof. Clearly the result holds for $r=n$. Also, if $r=n-1$ we have $D(n,n,n-1)= nD(n-1,n-1,n-1)+(-1)\binom{n}{n-1}=n-n=0$, and the result holds.

In what follows we assume $n>2$, $r<n-1$. Rearranging the terms of (7) with $k=n$ we have

$$[D(n,n,r)-nD(n-1,n-1,r)] = -[D(n-1,n-1,r)-(n-1)D(n-2,n-2,r)] + [D(n-1,n-1,r-1)-(n-1)D(n-2,n-2,r-1)]. \tag{8}$$

Set $D(n,n,r)-nD(n-1,n-1,r)=(-1)^{n-r}f(n,r)$. Then it follows that

$$(-1)^{n-r}f(n,r)=-(-1)^{n-r-1}f(n-1,r)+(-1)^{n-r}f(n-1,r-1),$$

i.e.,

$$f(n,r)=f(n-1,r)+f(n-1,r-1). \tag{9}$$

Recall Pascal's identity for the binomial coefficients

$$\binom{n}{r}=\binom{n-1}{r}+\binom{n-1}{r-1}. \tag{10}$$

We thus see that the recursion (9) for the $f(n,r)$'s is the same as that for the binomial coefficients, (10). Equation (9) implies $f(n,0)=f(n-1,0)=\cdots=f(3,0)$ which is easily seen to be equal to 1. On the other hand,

$$f(n,n-2)=D(n,n,n-2)-nD(n-1,n-1,n-2)=D(n,n,n-2)=\binom{n}{n-2},$$

since we can select $n-2$ questions to be answered correctly in $\binom{n}{n-2}$ ways. Thus $f(n,r)$ satisfies the same recursion as the binomial coefficients and has the same boundary conditions for $n<2$, $r \leqslant n-2$, so it follows that they are equal.

If we set $r=0$ in the Corollary we obtain

$$D(n,n,0)=nD(n-1,n-1,0)+(-1)^n, \tag{11}$$

the usual recurrence for the Derangement numbers.

Two final recurrences which we will illustrate are

$$\binom{r}{t}D(n,k,r)=\binom{k}{t}D(n-t,k-t,r-t),\ t \geqslant 0 \tag{12}$$

and

$$D(n,k,r)=kD(n-1,k-1,r)+D(n-1,k,r),\ n>k. \tag{13}$$

These are useful in obtaining a number of other recurrences. The first follows easily from (2) while (13) may be proved in a manner analogous to the combinatorial proof of (10). To see this, suppose answer j is one of the $n-k$ incorrect answers in an (n,k)-matching problem. Since answer j must either be used or not used in the matching we have two cases:

(a) If answer j is used, there are $\binom{k}{1}$ places for it and $D(n-1,k-1,r)$ ways to complete the problem.

(b) If answer j is not used, there are $D(n-1,k,r)$ ways to complete the problem.

Now (a) and (b) together imply (13).

We now return to the question of the score a student may expect to receive when guessing the answers to an (n, k)-matching problem. Let $E(n, k)$ denote this expected value, i.e.,

$$E(n,k) = \sum_{r=1}^{k} rP(n,k,r). \tag{14}$$

This expected value may be evaluated using (1), (2) and a rearranging of terms in the resulting double sum; however a more elegant approach is as follows (see [**2**], p. 217). Define a random variable X_j to be 1 or 0 according as answer j is in the jth spot or not. Set $S_k = X_1 + X_2 + \cdots + X_k$, $k \leqslant n$. Each answer has probability $1/n$ of being in the jth place hence the probability $P(X_j = 1) = 1/n$, $P(X_j = 0) = (n-1)/n$ and therefore the expected value $E(X_j) = 1/n$. In turn this implies $E(n,k) = \Sigma E(X_i) = E(S_k) = k/n$, since expectations are additive. In particular, $E(n,n) = 1$ (note that this latter value is independent of n). Interpreting the above results we see that the unprepared student can expect at most one correct answer in the long run when guessing, and should probably spend his time on other problems. With regard to the "hat check" problem, $E(S_n) = E(n,n) = 1$ implies that on the average we might expect one person to retrieve his own hat regardless of whether there are 10, 1,000, or 1,000,000 hats checked.

One final related problem. As in [**1**], let Q_n be the number of permutations of $\langle 1,2,\ldots,n \rangle$ in which none of the patterns $12, 23, \ldots, (n-1)n$ occur. It is known that

$$Q_n = \sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} (n-j)!, \tag{15}$$

(this may be seen using the Principle of Inclusion and Exclusion). The quantity Q_n is simply $D(n, n-1, 0)$ and in fact

$$Q_n = D(n, n-1, 0) = D(n,n,0) + D(n-1, n-1, 0), n \geqslant 1. \tag{16}$$

Equation (16) follows from the observation that if we answer an $(n, n-1)$-matching problem and fail to get any correct answers, this can be done in the same number of ways as answering an (n, n)-matching problem with none correct together with the number of ways of answering an (n, n)-matching problem where question n received the only correct answer. However, a combinatorial proof of the equality of Q_n and $D(n, n-1, 0)$, not using Inclusion and Exclusion, has avoided our detection and the equality appears coincidental. The recurrence in (16) can be generalized to

$$D(n, n-k, 0) = \sum_{j=0}^{k} \binom{k}{j} D(n-j, n-j, 0) \tag{17}$$

by a similar argument to the one given above, (see [**6**]).

For further study of permutation problems with restricted positions the reader is directed to [**4**], and in particular the relationship of $D(n, k, 0)$ to kth order differences of $n!/k$.

Work on this paper was supported in part by Natural Sciences and Engineering Research Council Canada operating grant A 8048.

References

[1] R. A. Brualdi, Introductory Combinatorics, North Holland, N.Y., 1979.
[2] W. Feller, An Introduction to Probability Theory and its Applications, vol. 1, Wiley, New York, 1966.
[3] P. R. Montmort, Essai d'analyse sur les jeux de hazard, Paris, 1708.
[4] J. Riordan, An Introduction to Combinatorial Analysis, Wiley, New York, 1958.
[5] A. Tucker, Applied Combinatorics, Wiley, New York, 1980.
[6] T. H. Wang, Elementary Problem 2947, Amer. Math. Monthly, 89 (1982) 334.

A Hardy Old Problem

Martin D. Landau
William R. Jones
Lafayette College
Easton, PA 18042

At least as early as 1908 in the classic work, *A Course of Pure Mathematics*, G. H. Hardy [2], [3] posed and discussed the problem "*If* $\phi(x)+\phi'(x)\to a$ *as* $x\to\infty$, *then* $\phi(x)\to a$ *and* $\phi'(x)\to 0$." Hardy's proof is rather "roundabout" and suffers in comparison with later solutions of this problem. W. Fulks includes this problem as an exercise in his *Advanced Calculus* [1] first published in 1961. A generalization of the problem is solved in the problems section of the Monthly in 1974 [5] and then again a solution of the "Hardy problem" appears in the Monthly in 1978 [4]. These latter proofs involve inequalities and epsilon arguments and differ greatly from Hardy's treatment. However, what is probably the best solution possible for this problem was provided to us by our colleague, Dr. L. Thomas Hill: Given that $\lim_{x\to+\infty}[f(x)+f'(x)]=L$, then using L'Hôpital's rule

$$\lim_{x\to+\infty} f(x)=\lim_{x\to+\infty}\frac{e^x f(x)}{e^x}=\lim_{x\to+\infty}\frac{e^x[f(x)+f'(x)]}{e^x}=L.$$

This elegant proof has the extra benefit that it allows an extension of the original problem in that L may be infinite. In this extended case, however, $f'(x)$ might not approach zero; indeed $f'(x)$ might have no limit at all.

This interesting problem has prompted us to consider all the possible values of $\lim_{x\to+\infty}[f(x)+f'(x)]$ in relation to the limits of $f(x)$ and $f'(x)$. If we let $\lim_{x\to+\infty}f(x)=L$ and $\lim_{x\to+\infty}f'(x)=L'$ where we permit L and L' to be any positive real number (Pos), negative real number (Neg), 0, $+\infty$, $-\infty$, or even fail to exist (FTE), then we can determine all possible values of $\lim_{x\to+\infty}[f(x)+f'(x)]$. These are shown in Table 1; an asterisk means that the prescribed combination of L and L' is not possible. We offer some examples which illustrate the information in the table.

It is easy to show that if $f'(x)$ has a nonzero or infinite limit as $x\to+\infty$, then $|f(x)|\to+\infty$. The first, second, fourth and fifth columns admit only one entry in each. Examples of functions for these columns in order are: e^x, x, $-x$, and $-e^x$.

For the third column ($L'=0$), examples of functions in order going down the column are: $\ln x$, $x^{-1}+1$, x^{-1}, $x^{-1}-1$, $-\ln x$ and $\sin\sqrt{x}$.

In the sixth column (L' fails to exist), examples of functions for the first row are $x+\sin x^2$ and $x+\sin x$. The functions $1+x^{-1}\sin x^2$, $x^{-1}\sin x^2$ and $-1+x^{-1}\sin x^2$ provide examples for the second, third and fourth rows, respectively. The functions $-x+\sin x^2$ and $-x+\sin x$ illustrate the fifth row, and finally, $\sin x$ illustrates the last row.

One way to interpret the result of Hardy's problem is to say that a real entry can occur in Table 1 only in the second, third and fourth rows of the third column and must agree with the L value of the row. Except for the last row-last column entry, all other entries in the table are readily determined from L and L'. However, when L and L' both fail to exist, it is not obvious that $f(x)+f'(x)$ must fail to have a limit. Dr. Hill's proof above allows easy disposal of this question. If $f+f'$ had a limit, even an infinite limit, then f itself would be required to have that limit; thus, FTE is the only possible entry for the last row-last column.

Let us consider some generalizations of the problem. Let r be a positive constant and suppose

$$\lim_{x\to+\infty}[f'(x)+rf(x)]=L, \tag{1}$$

finite or infinite. Applying L'Hôpital's rule to $e^{rx}f(x)/e^{rx}$ shows that $f(x)\to L/r$, $f'(x)\to 0$ if L is finite, and $f(x)\to L$ if L is infinite. When L is infinite, all sorts of possibilities remain for the limit of $f'(x)$, as can be seen by considering the functions x, $\ln x$, e^x and $x+\sin x$. If, in (1), the constant r is negative, no such general conclusions may be drawn. For instance, if $r=-1$, then the diversity of results for the functions x, $\ln x$, e^x, e^{-x}, e^{2x} and $-\sqrt{x}$ illustrate the fact that the limit of $f'(x)-f(x)$ does not, in general, determine the limits of $f(x)$ or $f'(x)$. However, with the additional hypothesis that $f(x)$ is *bounded* for sufficiently large x, we may again apply L'Hôpital's rule to the indeterminate form $e^{rx}f(x)/e^{rx}$. In this case we may conclude that $f(x)\to L/r$, and since f is bounded, L must be finite and $f'(x)\to 0$. In fact, if in (1), L is infinite when $r<0$, it can be shown that f has an infinite limit.

Next, let us consider a complex-valued function f of a real variable and determine what conclusions may be drawn if

$$\lim_{x\to+\infty}[f'(x)+\alpha f(x)]=\beta, \tag{2}$$

where α,β are complex and $\alpha\neq 0$. Without loss of generality we may restrict ourselves to the case $\beta=0$, for if $\beta\neq 0$, the function $g(x)=f(x)-\beta/\alpha$ has $g'(x)+\alpha g(x)\to 0$. Then if $g(x)\to\gamma$ we may conclude $f(x)\to\gamma+\beta/\alpha$.

We find that the sign of the real part of α is a determining factor in the results for complex-valued functions. We consider three separate cases.

(i) First, suppose $\mathrm{Re}(\alpha)=r>0$. In this case we can show that $\lim_{x\to+\infty}[f'(x)+\alpha f(x)]=0$ implies $\lim_{x\to+\infty}f(x)=0$. To see this, note that since $|e^{\alpha x}/e^{rx}|=1$, then

$$e^{\alpha x}[f'(x)+\alpha f(x)]/e^{rx}\to 0 \text{ as } x\to+\infty, \tag{3}$$

and since $|f(x)|=|e^{\alpha x}f(x)/e^{rx}|$, it suffices to show that $e^{\alpha x}f(x)/e^{rx}\to 0$. Now applying L'Hôpital's rule to the real (resp., imaginary) part of $e^{\alpha x}f(x)/e^{rx}$ and using the real (resp., imaginary) part from (3) completes the proof:

$$\lim_{x\to+\infty}\mathrm{Re}\left\{\frac{e^{\alpha x}f(x)}{e^{rx}}\right\}=\lim_{x\to+\infty}\frac{\mathrm{Re}\{e^{\alpha x}f(x)\}}{e^{rx}}=\lim_{x\to+\infty}\frac{\mathrm{Re}\{e^{\alpha x}[f'(x)+\alpha f(x)]\}}{re^{rx}}$$
$$=\frac{1}{r}\cdot\lim_{x\to+\infty}\mathrm{Re}\left\{\frac{e^{\alpha x}[f'(x)+\alpha f(x)]}{e^{rx}}\right\}=0.$$

This provides a new proof of a generalization of the Hardy problem given in [5]. Thus, we may say that

$$f'(x)+\alpha f(x)\to\beta \text{ as } x\to+\infty \text{ implies } f(x)\to\beta/\alpha \text{ and } f'(x)\to 0.$$

(ii) Next, suppose $\mathrm{Re}(\alpha)=r<0$. The examples in the real case with $r<0$ apply in this setting and we see that the limit of $f'(x)+\alpha f(x)$ does not, in general, determine the limits of $f(x)$ or $f'(x)$. With the restriction that $f(x)$ is bounded for sufficiently large x, however, the real and

L \ L'	$+\infty$	Pos	0	Neg	$-\infty$	FTE
$+\infty$	$+\infty$	$+\infty$	$+\infty$	*	*	FTE or $+\infty$
Pos	*	*	Pos	*	*	FTE
0	*	*	0	*	*	FTE
Neg	*	*	Neg	*	*	FTE
$-\infty$	*	*	$-\infty$	$-\infty$	$-\infty$	FTE or $-\infty$
FTE	*	*	FTE	*	*	FTE

TABLE 1. **Values of $\lim_{x\to+\infty}[f(x)+f'(x)]$, where $L=\lim_{x\to+\infty}f(x)$ and $L'=\lim_{x\to+\infty}f'(x)$. An asterisk indicates that combination of L and L' is not possible.**

imaginary parts of $e^{\alpha x}f(x)/e^{rx}$ become indeterminate forms to which one can apply L'Hôpital's rule. Proceeding in a manner similar to our previous proofs, we are able to show that

$$f'(x)+\alpha f(x)\to\beta \text{ as } x\to+\infty \text{ implies } f(x)\to\beta/\alpha \text{ and } f'(x)\to 0$$

even when $\mathrm{Re}(\alpha)<0$ if we also know that $f(x)$ is bounded for large x. An illustration of this case is given by the function $f(x)=ix/(x-i)$ and $\alpha=-1-i$; one may check that the limit β in (2) is $\beta=1-i$, and $\lim_{x\to+\infty}f(x)=\beta/\alpha=i$. Note the complete agreement of these results with those previously derived in the real case.

(iii) Last, consider $\mathrm{Re}(\alpha)=0$. Since $\alpha\neq 0$, then $\mathrm{Im}(\alpha)\neq 0$. In this case even requiring that f is eventually bounded and that the limit β in (2) exists does not ensure that $\lim_{x\to+\infty}f(x)$ exists. Consider the example $f(x)=e^{ix}$ and $\alpha=-i$; then $f'(x)+\alpha f(x)\equiv 0$, and the functions f, f' are bounded, but neither has a limit as $x\to+\infty$. Suppose we assume that *either* f or f' has a finite limit as $x\to+\infty$, and that $f'(x)+\alpha f(x)\to 0$ as $x\to+\infty$. Then both f and f' must have finite limits as $x\to\infty$, say γ, δ respectively, and $\delta+\alpha\gamma=0$. Then $\gamma=0$ if and only if $\delta=0$. However, if $\delta\neq 0$ then either $\mathrm{Re}(f')$ or $\mathrm{Im}(f')$ would have a nonzero limit which would force $|f(x)|\to+\infty$, contradicting $f(x)\to\gamma$. Therefore, the assumption that either f or f' has a finite limit leads to the conclusion that *both* of these limits are zero. More generally, with this assumption, we again may say

$$\text{if } f'(x)+\alpha f(x)\to\beta \text{ as } x\to+\infty, \text{ then } f(x)\to\beta/\alpha \text{ and } f'(x)\to 0.$$

An example of this last case is provided by the function $f(x)=1+i/x$ and $\alpha=i$; then $\beta=i$ and $\beta/\alpha=1$.

Reference [5] discusses conclusions one may draw about $\lim_{x\to+\infty}f(x)$ when certain linear combinations of f, f' and higher derivatives have finite limits. The reader might find it interesting to examine the relationship between $\lim_{x\to+\infty}[xf'(x)+rf(x)]$ and the limits of $f(x)$ and $f'(x)$. Replacing $e^{x}f(x)/e^{x}$ by $x^{r}f(x)/x^{r}$ allows one to use the methods of this paper to achieve similar results.

References

[1] W. Fulks, Advanced Calculus, Wiley, New York, 1961, p. 134.

[2] G. H. Hardy, A Course of Pure Mathematics, 1st ed., Cambridge, 1908, p. 246.

[3] ______, A Course of Pure Mathematics, 10th ed., Cambridge, 1955, p. 281.

[4] H. Kestelman, An old exercise, Amer. Math. Monthly, 85 (1978) 685–686.

[5] W. C. Waterhouse, Limit of solutions of a linear differential equation, Amer. Math. Monthly, 81 (1974) 92–93.

An Alternative to the Integral Test

GERALD JUNGCK

Bradley University

Peoria, IL 61625

As presented in Calculus texts, the Integral Test states that *if f is continuous, positive valued, and decreasing on* $[1,\infty)$, *then*

$$\sum_{n=1}^{\infty} f(n) \textit{ converges iff } \int_1^{\infty} f(x)\,dx \textit{ converges.}$$

Since $f(x)>0$ on $[1,\infty)$, $\int_1^\infty f(x)\,dx$ converges iff any antiderivative of f is bounded above. So the Integral Test is virtually an "Antiderivative Test." However, the Integral Test as stated fails to utilize the fact that we can be selective in our choice of antiderivatives.

imaginary parts of $e^{\alpha x}f(x)/e^{rx}$ become indeterminate forms to which one can apply L'Hôpital's rule. Proceeding in a manner similar to our previous proofs, we are able to show that

$$f'(x)+\alpha f(x)\to\beta \text{ as } x\to+\infty \text{ implies } f(x)\to\beta/\alpha \text{ and } f'(x)\to 0$$

even when $\operatorname{Re}(\alpha)<0$ if we also know that $f(x)$ is bounded for large x. An illustration of this case is given by the function $f(x)=ix/(x-i)$ and $\alpha=-1-i$; one may check that the limit β in (2) is $\beta=1-i$, and $\lim_{x\to+\infty}f(x)=\beta/\alpha=i$. Note the complete agreement of these results with those previously derived in the real case.

(iii) Last, consider $\operatorname{Re}(\alpha)=0$. Since $\alpha\neq 0$, then $\operatorname{Im}(\alpha)\neq 0$. In this case even requiring that f is eventually bounded and that the limit β in (2) exists does not ensure that $\lim_{x\to+\infty}f(x)$ exists. Consider the example $f(x)=e^{ix}$ and $\alpha=-i$; then $f'(x)+\alpha f(x)\equiv 0$, and the functions f, f' are bounded, but neither has a limit as $x\to+\infty$. Suppose we assume that *either* f or f' has a finite limit as $x\to+\infty$, and that $f'(x)+\alpha f(x)\to 0$ as $x\to+\infty$. Then both f and f' must have finite limits as $x\to\infty$, say γ, δ respectively, and $\delta+\alpha\gamma=0$. Then $\gamma=0$ if and only if $\delta=0$. However, if $\delta\neq 0$ then either $\operatorname{Re}(f')$ or $\operatorname{Im}(f')$ would have a nonzero limit which would force $|f(x)|\to+\infty$, contradicting $f(x)\to\gamma$. Therefore, the assumption that either f or f' has a finite limit leads to the conclusion that *both* of these limits are zero. More generally, with this assumption, we again may say

$$\text{if } f'(x)+\alpha f(x)\to\beta \text{ as } x\to+\infty, \text{ then } f(x)\to\beta/\alpha \text{ and } f'(x)\to 0.$$

An example of this last case is provided by the function $f(x)=1+i/x$ and $\alpha=i$; then $\beta=i$ and $\beta/\alpha=1$.

Reference [5] discusses conclusions one may draw about $\lim_{x\to+\infty}f(x)$ when certain linear combinations of f, f' and higher derivatives have finite limits. The reader might find it interesting to examine the relationship between $\lim_{x\to+\infty}[xf'(x)+rf(x)]$ and the limits of $f(x)$ and $f'(x)$. Replacing $e^{x}f(x)/e^{x}$ by $x^{r}f(x)/x^{r}$ allows one to use the methods of this paper to achieve similar results.

References

[1] W. Fulks, Advanced Calculus, Wiley, New York, 1961, p. 134.
[2] G. H. Hardy, A Course of Pure Mathematics, 1st ed., Cambridge, 1908, p. 246.
[3] ______, A Course of Pure Mathematics, 10th ed., Cambridge, 1955, p. 281.
[4] H. Kestelman, An old exercise, Amer. Math. Monthly, 85 (1978) 685–686.
[5] W. C. Waterhouse, Limit of solutions of a linear differential equation, Amer. Math. Monthly, 81 (1974) 92–93.

An Alternative to the Integral Test

Gerald Jungck
Bradley University
Peoria, IL 61625

As presented in Calculus texts, the Integral Test states that *if f is continuous, positive valued, and decreasing on* $[1,\infty)$, *then*

$$\sum_{n=1}^{\infty} f(n) \textit{ converges iff } \int_1^{\infty} f(x)\,dx \textit{ converges}.$$

Since $f(x)>0$ on $[1,\infty)$, $\int_1^{\infty}f(x)\,dx$ converges iff any antiderivative of f is bounded above. So the Integral Test is virtually an "Antiderivative Test." However, the Integral Test as stated fails to utilize the fact that we can be selective in our choice of antiderivatives.

In this note we propose as an alternative to the Integral Test an antiderivative test which can be proved and applied outside the context of the definite and improper integrals. This approach eliminates dependency on improper integrals, and also makes more feasible the option of introducing infinite series before the definite integral. Moreover, by choosing antiderivatives judiciously in this test, we obtain inequalities which prove useful in approximation and error analysis.

Our approach is based on the observation that an infinite series $\Sigma f(n)$ and its sequence $\{S_n\}$ of partial sums satisfies

$$S_{n+1} - S_n = f(n+1),$$

so it appears natural to investigate functions g such that $g(x+1) - g(x)$ approximates $f(x+1)$. Now if such a g is differentiable on $[x, x+1]$, then the Mean Value Theorem (MVT) implies that $g(x+1) - g(x) = g'(c)$ for some c between x and $x+1$; i.e., we want g' to approximate f. We are thus led to consider antiderivatives of f.

The **antiderivative test**, which we now state and prove, follows easily from the MVT and the fact that a series of positive terms converges if and only if its sequence of partial sums is bounded above.

THEOREM. *Let f be positive and nonincreasing on $[k,\infty)$, k a positive integer, and let g be any antiderivative of f. Then*

(A) *$\Sigma_{n=k}^{\infty} f(n)$ converges iff g is bounded above on $[k,\infty)$.*
(B) *If $\lim_{x\to\infty} g(x) = 0$, then $\Sigma_{n=k}^{\infty} f(n)$ converges.*

Furthermore, if S is the sum of $\Sigma_{n=k}^{\infty} f(n)$, then

$$-g(m+1) \leqslant \sum_{n=m+1}^{\infty} f(n) \leqslant -g(m) \text{ for } m \geqslant k, \tag{1}$$

$$f(k) - g(k+1) \leqslant S \leqslant f(k) - g(k), \tag{2}$$

and

$$0 \leqslant S - \left(\sum_{n=k}^{m} f(n) - g(m+1) \right) \leqslant f(m) \text{ for } m \geqslant k. \tag{3}$$

Proof of (A). For $n \geqslant k$ we can apply the MVT to g on $[n, n+1]$ to obtain

$$g(n+1) - g(n) = g'(x_n)((n+1) - n) = f(x_n)$$

for some $x_n \in (n, n+1)$. Since f is nonincreasing we have

$$f(n+1) \leqslant f(x_n) = g(n+1) - g(n) \leqslant f(n) \text{ for } n \geqslant k. \tag{4}$$

But $\Sigma_{n=k}^{m}(g(n+1) - g(n)) = g(m+1) - g(k)$ for $m \geqslant k$, so (4) implies

$$\sum_{n=k}^{m} f(n+1) \leqslant g(m+1) - g(k) \leqslant \sum_{n=k}^{m} f(n) \text{ for } m \geqslant k. \tag{5}$$

Now if $\Sigma_{n=k}^{\infty} f(n)$ converges, the right hand term in (5) is bounded above so that (5) implies that g is bounded on the set of integers $\geqslant k$. But $g'(x) = f(x) > 0$ and consequently g is increasing on $[k,\infty)$; therefore g is bounded on $[k,\infty)$.

Conversely, if g is bounded on $[k,\infty)$, (5) implies that the sequence of partial sums for the series $\Sigma_{n=k}^{\infty} f(n+1)$ is bounded above; thus $\Sigma_{n=k}^{\infty} f(n+1)$ and $\Sigma_{n=k}^{\infty} f(n)$ are convergent.

A few comments before we prove part (B) of the theorem. Observe that the requirement that $\lim_{x\to\infty} g(x) = 0$ can always be met if f has an antiderivative which is bounded above. For example, if $f(n) = 1/(1+n^2)$ we would let $g(x) = \tan^{-1} x - \pi/2$ instead of $\tan^{-1} x$. We introduce the following notation for the proof and subsequent examples:

$$S_m=\sum_{n=k}^{m} f(n),\ E_m=\sum_{n=m+1}^{\infty} f(n).$$

(Thus E_m is the error if we approximate S by S_m).

Proof of (B). If $\lim_{x\to\infty} g(x)=0$, then g is certainly bounded on $[k,\infty)$ so that the indicated series is convergent by part (A). Now if we let $m\to\infty$ in (5), we obtain

$$\sum_{n=k}^{\infty} f(n+1)\leqslant -g(k)\leqslant \sum_{n=k}^{\infty} f(n).$$

Since (5) is valid when any $k'\geqslant k$ is substituted for k, we can write

$$\sum_{n=m}^{\infty} f(n+1)\leqslant -g(m)\leqslant \sum_{n=m}^{\infty} f(n) \text{ for } m\geqslant k. \tag{6}$$

The left side of (6) implies that $E_m\leqslant -g(m)$, and the right side of (6) implies $-g(m+1)\leqslant E_m$; consequently, inequality (1) of (B) holds (for $m\geqslant k$). Inequality (2) follows from (1); just let $m=k$ in (1), and add $f(k)$ throughout (note that $S=E_k+f(k)$). Moreover, (1) implies that

$$0\leqslant E_m+g(m+1)=(S-S_m)+g(m+1)\leqslant g(m+1)-g(m)$$

for $m\geqslant k$, and hence (4) yields

$$0\leqslant S-(S_m-g(m+1))\leqslant f(m) \text{ for } m\geqslant k$$

which is inequality (3).

We now illustrate the use of the Theorem with several examples.

EXAMPLE 1. We consider the series

$$\sum_{n=1}^{\infty} n^{-2}, \tag{7}$$

and let $f(x)=x^{-2}$. If $g(x)=-x^{-1}$, then $g'=f$ and $g(x)\to 0$, so part (B) of the Theorem applies to g. By (2),

$$f(1)-g(2)=3/2\leqslant S\leqslant 2=f(1)-g(1),$$

which yields an initial approximation to S. Also, (1) implies that

$$-g(m+1)=(m+1)^{-1}\leqslant E_m\leqslant m^{-1}=-g(m)$$

is true for $m\geqslant 1$. Thus for S_m to approximate S accurately to three decimal places, it suffices that $m^{-1}<.5(10^{-3})$ or $m>2{,}000$. In fact, the above inequality shows that $m\geqslant 2{,}000$ is a necessary condition for 3 place accuracy. However, we can approximate S more economically by using (3), since

$$0\leqslant S-(S_m-g(m+1))\leqslant m^{-2}<1/2{,}000 \text{ when } m>\sqrt{20}\,10.$$

Thus $S_{45}-g(46)=1.622957+.021739=1.644696$ approximates S accurately to 3 decimal places.

EXAMPLE 2. $\sum_{n=1}^{\infty} n^{-p}$ converges if $p>1$ and diverges if $p<1$, since if $g(x)=x^{1-p}/(1-p)$ for $x\geqslant 1$, then $g(x)\to 0$ as $x\to+\infty$ if $p>1$, and $g(x)\to+\infty$ if $p<1$. So when $p>1$,

(1) implies $\quad (m+1)^{1-p}/(p-1)\leqslant E_m\leqslant m^{1-p}/(p-1)$ for $m\geqslant 1$;

(2) implies $\quad 1+2^{1-p}/(p-1)\leqslant S\leqslant 1+1/(p-1)$,

and

(3) implies that for $m\geqslant 1$, $\quad 0\leqslant S-(S_m-g(m+1))\leqslant m^{-p}$.

Consequently, if we use S_m to approximate S, we obtain n place accuracy if $m>$

$(2(10^n)/(p-1))^{1/(p-1)}$, and if we use $S_m - g(m+1)$ to approximate S, we obtain n place accuracy if $m > (2(10)^n)^{1/p}$.

The case $p = 1$ ignored in Example 2 can be dealt with even though the natural logorithm has not been introduced (see, e.g., [**1**], p. 566).

EXAMPLE 3. Given $\Sigma_{n=3}^{\infty} n(n^2 - 7)^{-5/2}$, let $g(x) = -(x^2 - 7)^{-3/2}/3$. Then if $x > \sqrt{83.3}$, $-g(x) < .5(10)^{-3}$, so that S_{10} approximates S accurately to three decimal places by (1).

In the above example a nice antiderivative g was readily apparent. If this is not the case, comparison techniques can sometimes be used by noting that if $0 < f_1(n) \leqslant f_2(n)$ for $n \geqslant k$ and $E_n{}^i$ is the error term for f_i $(i = 1,2)$, then $E_n{}^1 < E_n{}^2$ for $n \geqslant k$. Our next example illustrates this.

EXAMPLE 4. The sum of the first 41 terms, S_{41}, approximates $\Sigma_{n=2}^{\infty}(n^3 + 5n - 7)^{-4/3}$ accurately to at least 5 decimal places. To see this, note that $(n^3 + 5n - 7)^{-4/3} < n^{-4}$ for $n \geqslant 2$. Thus, if E_n denotes the error term for $\Sigma_{n=2}^{\infty} n^{-4}$, then by Example 2, $E_m < .5(10)^{-5}$ if $m^{-3}/3 < .5(10)^{-5}$ or $m \geqslant 41$. We thereby obtain 5-place accuracy for the original series.

None of the preceding—examples or theory—requires the concept of the definite integral. However, if the definite and improper integrals have been introduced, the Integral Test is an immediate corollary to Theorem (A). Since $\ln x$ and e^x would then be available, we could consider examples of the following type.

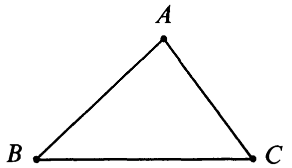
EXAMPLE 5. Inequality (1) assures us that S_4 approximates $\Sigma_{n=1}^{\infty} e^{-n^2} n$ accurately to 4 places, since if $g(x) = -e^{-x^2}/2$, $-g(m) < .5(10)^{-4}$ when $m \geqslant 4 > 2\sqrt{\ln 10}$.

Reference

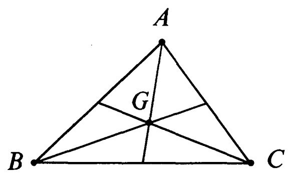
[1] Howard Anton, Calculus with Analytical Geometry, Wiley, New York, 1980.

Triangle Rhyme

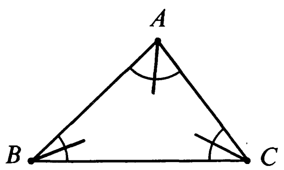
DWIGHT PAINE
Messiah College
Grantham, PA



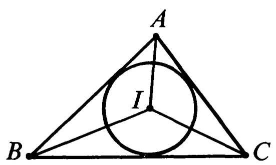
Let's talk about a **triangle**;
We'll call it ABC,



With **medians** converging
To the centroid, labeled G.



To bisect all the angles
Takes a dreadful steady eye,



But all the **angle bisectors**
Meet at the **incenter** I.

$(2(10^n)/(p-1))^{1/(p-1)}$, and if we use $S_m - g(m+1)$ to approximate S, we obtain n place accuracy if $m > (2(10)^n)^{1/p}$.

The case $p = 1$ ignored in Example 2 can be dealt with even though the natural logorithm has not been introduced (see, e.g., [**1**], p. 566).

EXAMPLE 3. Given $\Sigma_{n=3}^{\infty} n(n^2 - 7)^{-5/2}$, let $g(x) = -(x^2 - 7)^{-3/2}/3$. Then if $x > \sqrt{83.3}$, $-g(x) < .5(10)^{-3}$, so that S_{10} approximates S accurately to three decimal places by (1).

In the above example a nice antiderivative g was readily apparent. If this is not the case, comparison techniques can sometimes be used by noting that if $0 < f_1(n) \leqslant f_2(n)$ for $n \geqslant k$ and $E_n^{\,i}$ is the error term for f_i $(i = 1,2)$, then $E_n^{\,1} < E_n^{\,2}$ for $n \geqslant k$. Our next example illustrates this.

EXAMPLE 4. The sum of the first 41 terms, S_{41}, approximates $\Sigma_{n=2}^{\infty}(n^3 + 5n - 7)^{-4/3}$ accurately to at least 5 decimal places. To see this, note that $(n^3 + 5n - 7)^{-4/3} < n^{-4}$ for $n \geqslant 2$. Thus, if E_n denotes the error term for $\Sigma_{n=2}^{\infty} n^{-4}$, then by Example 2, $E_m < .5(10)^{-5}$ if $m^{-3}/3 < .5(10)^{-5}$ or $m \geqslant 41$. We thereby obtain 5-place accuracy for the original series.

None of the preceding—examples or theory—requires the concept of the definite integral. However, if the definite and improper integrals have been introduced, the Integral Test is an immediate corollary to Theorem (A). Since $\ln x$ and e^x would then be available, we could consider examples of the following type.

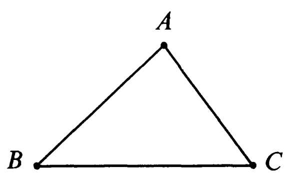
EXAMPLE 5. Inequality (1) assures us that S_4 approximates $\Sigma_{n=1}^{\infty} e^{-n^2} n$ accurately to 4 places, since if $g(x) = -e^{-x^2}/2$, $-g(m) < .5(10)^{-4}$ when $m \geqslant 4 > 2\sqrt{\ln 10}$.

Reference

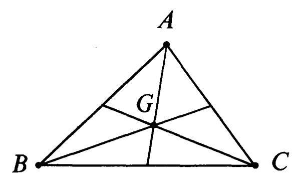
[1] Howard Anton, Calculus with Analytical Geometry, Wiley, New York, 1980.

Triangle Rhyme

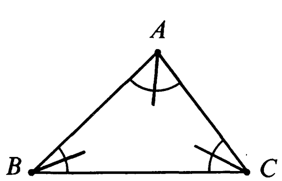
DWIGHT PAINE
Messiah College
Grantham, PA



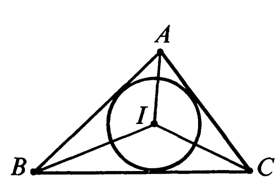
Let's talk about a **triangle**;
We'll call it ABC,



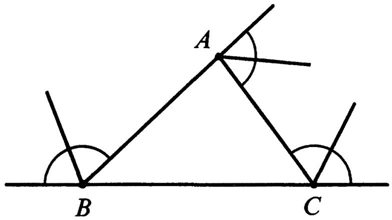
With **medians** converging
To the centroid, labeled G.



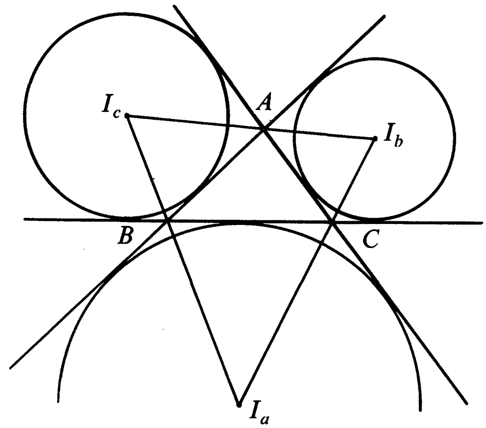
To bisect all the angles
Takes a dreadful steady eye,



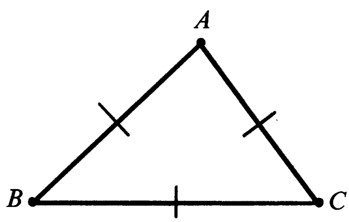
But all the **angle bisectors**
Meet at the **incenter** I.



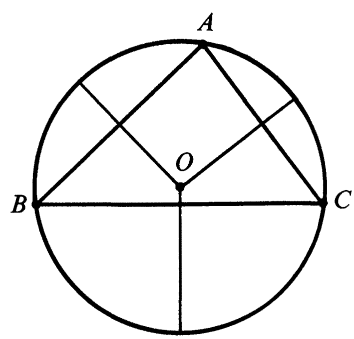
There are **external bisectors**,
Besides the other three,



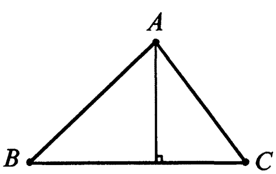
And they meet at the **excenters**:
I_a, I_b, I_c.



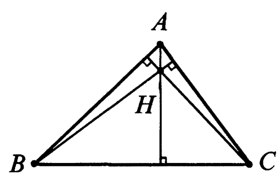
The **perpendicular bisectors**
Through the sides must go,



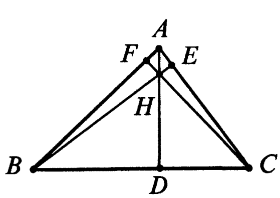
All radiating outward
From the **circumcenter** O.



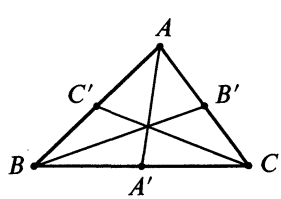
Although the **altitudes** are three,
Remarks my daughter Rachel,



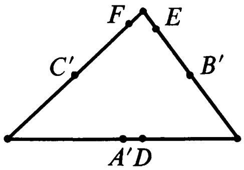
One point'll lie on all of them:
The **orthocenter** H'll.



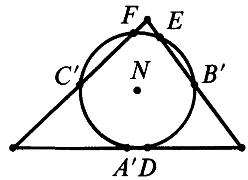
Their feet are often labeled
D, E, F (as in this rhyme),



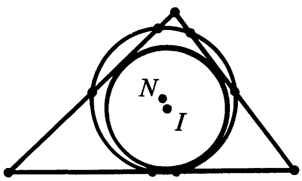
And **medians** too have feet
That we can call A, B, C prime.



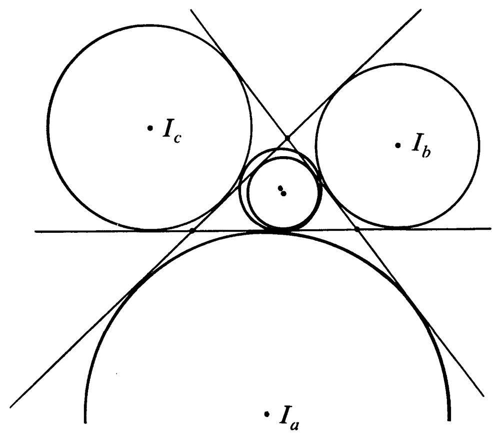
Those feet (I mean all six of them)
  Can all be found again



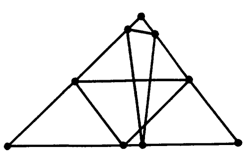
Around the **nine-point circle**
  With its nine-point center N.



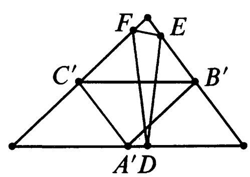
The nine-point circle, by the way,
  Does something *hard* to do:



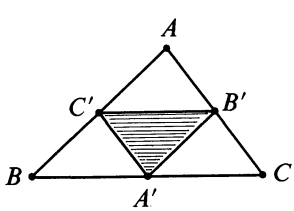
It *touches* the **incircle**
  And the three **excircles**, too!



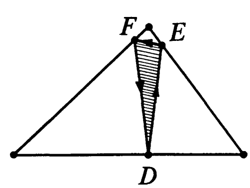
The **medial** and **orthic**
  Are triangles inscribed:



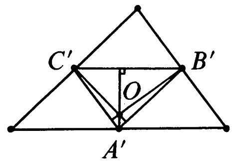
"ABC-prime and DEF"
  Is how they are described.



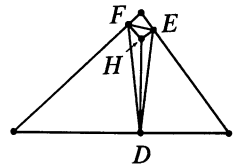
The medial is shaped just like
  Its parent ABC;



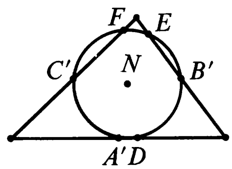
The orthic makes the shortest path
  To touch the sides (all three).



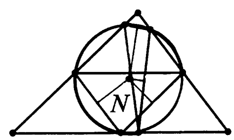
The medial has altitudes
That neatly meet at O;



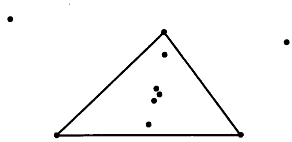
The orthic's angle bisectors
Through H are sure to go.



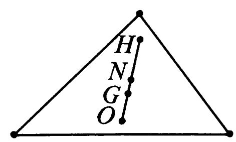
And since the corners of them both
Are nine-point-circle feet,



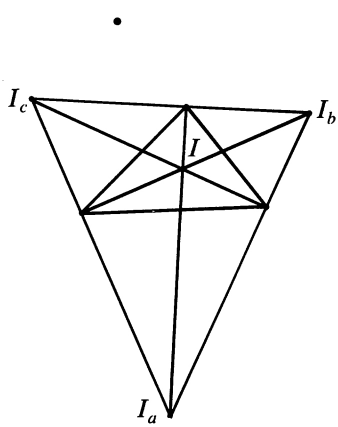
Their perpendicular bisectors
At N are sure to meet.



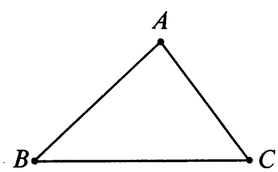
So triangles have centers;
Almost more than we can bear!
But not all helter-skelter—
There's amazing order there.



Four centers on the **Euler line**
Must lie in strict array,
With N halfway from O to H,
And G a third away.



Four others on an **orthocentric**
Quadrangle are found:
Incenter in the middle,
With excenters all around.



Who'd ever think a triangle—
Just lowly ABC—
Had all these beauties hidden
For us to find and see?

PROBLEMS

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

Proposals

To be considered for publication, solutions should be mailed before February 1, 1984.

1175. Suppose that $m = nq$, where n and q are positive integers. Prove that the sum of binomial coefficients

$$\sum_{k=0}^{n-1} \binom{(n,k)q}{(n,k)}$$

is divisible by m, where (x, y) denotes the greatest common divisor of x and y. [*Anon, Erewhon-upon-Spanish River.*]

1176. Let $f(t) = \det(tI - A)$ be the characteristic polynomial of the $n \times n$ matrix A. Find a formula for the characteristic polynomial of the adjoint of A in terms of f. (The adjoint of A, denoted $\operatorname{adj} A$, is the $n \times n$ matrix whose (r, s) element is the (s, r) cofactor of A.) [*H. Kestelman, University College London, England.*]

1177. Let $f_q(n)$ denote the probability that an $n \times n$ matrix with entries chosen at random from $\mathrm{GF}(q)$, the finite field of q elements, is invertible over $\mathrm{GF}(q)$.

(a) Show that $\lim_{n\to\infty} f_q(n) = \Pi_{k=1}^{\infty} (1 - q^{-k})$.

(b) Show that the value of the infinite product in (a) is irrational for every integer $q \geqslant 2$. [*Manuel Blum and J. O. Shallit, University of California.*]

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

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

Solutions should be written in a style appropriate for Mathematics Magazine. Each solution should begin on a separate sheet containing the solver's name and full address. It is not necessary to submit duplicate copies.

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

Quickies

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

Q686. It is known that for any triangle of side lengths a, b, and c:

$$3(ab+bc+ca) \leqslant (a+b+c)^2 \leqslant 4(ab+bc+ca).$$

Prove more generally that if $a_1, a_2, \dots, a_n$ are the sides of an n-gon, then

$$\frac{2n}{n-1}\sum_{i<j} a_i a_j \leqslant \left(\sum_{i=1}^{n} a_i\right)^2 \leqslant 4\sum_{i<j} a_i a_j,$$

and determine when there is equality. [*M. S. Klamkin*, *University of Alberta*.]

Solutions

Equatorial and Polar Mensuration **September 1982**

1150. To aid those uncertain of their fitness to join Mensa (a social organization for people of high intelligence), Mensa publishes tests, tests not infrequently flawed. A recent question was: "If you left your house and walked one mile east, three miles north, one mile west, and two miles south, how far would you then be from home?" What further conditions are needed to make the given answer, "One mile," correct? [*Marlow Sholander*, *Case Western Reserve University*.]

Solution: There is an infinite number of parallels which lead to the correct answer "One mile."

In fact, the longitude in arc corresponding to the mile west is equal to the longitude in arc corresponding to the mile east plus $2k\pi$, $k \in \mathbb{Z}$. Now $k=0$ if and only if the house lies 1.5 miles south of the equator. Let, then, $k \neq 0$ and let ψ be the latitude of the house. The three miles north displacement leads to a difference of $3/R$ in latitude, where R is the Earth's radius in miles. We then have to show that for every $k \neq 0$ the following equation has a solution ψ with $-\pi/2 < \psi < \pi/2 - 3/R$:

$$1/\cos\left(\psi + \frac{3}{R}\right) - 1/\cos\psi = 2k\pi R. \tag{1}$$

Denote by $f(\psi)$ the left side of equation (1); then f is continuous in the interval $(-\pi/2, \pi/2 - 3/R)$, so the existence of a solution to (1) is assured by the intermediate value theorem, since $\lim f(\psi)$ is $+\infty$ or $-\infty$ according as $\psi \nearrow \pi/2 - 3/R$ or $\psi \searrow -\pi/2$, respectively. Moreover, f is strictly increasing on the interval, since

$$f'(\psi) = \frac{\left(\sin\left(\psi+\frac{3}{R}\right) - \sin\psi\right)\left(1+\sin\left(\psi+\frac{3}{R}\right)\sin\psi\right)}{\cos^2\left(\psi+\frac{3}{R}\right)\cos^2\psi} > 0.$$

Hence the required solution is unique for each k. We have thus shown the existence of a (countable) infinity of parallels leading to the answer "One mile."

VANIA D. MASCIONI, student
Swiss Federal Institute of Technology

Also solved by Ragnar Dybvik (Norway), Jan Söderkvist (student), Harry Zaremba, and the proposer. Nine incomplete solutions giving only the near-equatorial answer were submitted. Several solvers considered the possibility of a flat or ellipsoidal (rather than spherical) Earth.

Hao-Nhien Vu (student) referred to a similar problem in Martin Gardner's *Aha! insight*, p. 41. *Mark Kantrowitz (student)* referred to the well-known problem which asks, "What color was the bear?"

Moving Normal Bisectors — September 1982

1151. Three points P, Q, R move on curves in the plane. At each instant, the normal at P to the curve on which P is moving coincides with the bisector of the angle RPQ. Corresponding conditions hold for the points Q and R.

(a) Show that the perimeter of the triangle PQR is constant.

(b) Find examples other than that of an equilateral triangle whose vertices move around a fixed circle. (If, say, P is fixed, then find examples where the bisector of the angle RPQ is a fixed line.)

(c)* Does the result in (a) have a dynamic interpretation in terms of three heavy masses moving on a smooth horizontal table with a light inextensible string looped over them? [*Peter J. Giblin, University of North Carolina at Chapel Hill and University of Liverpool.*]

Solution I: (a) Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be the position vectors of the points P, Q, and R, respectively, with respect to a fixed origin. On the curve on which P is moving we choose some parametrization with parameter s; hence $\mathbf{x} = \mathbf{x}(s)$ with $\mathbf{x}'(s) \neq \mathbf{0}$. The parametrizations on the other curves can be obtained by giving to Q and R the parameter s of P, so that $\mathbf{y} = \mathbf{y}(s)$ and $\mathbf{z} = \mathbf{z}(s)$ with $\mathbf{y}'(s) \neq \mathbf{0}$ and $\mathbf{z}'(s) \neq \mathbf{0}$. The properties of the bisector imply that

$$\left(\frac{\mathbf{y}-\mathbf{x}}{|\mathbf{y}-\mathbf{x}|}+\frac{\mathbf{z}-\mathbf{x}}{|\mathbf{z}-\mathbf{x}|}\right)\cdot\mathbf{x}'=0,\ \left(\frac{\mathbf{z}-\mathbf{y}}{|\mathbf{z}-\mathbf{y}|}+\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|}\right)\cdot\mathbf{y}'=0,\ \text{ and }\ \left(\frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}+\frac{\mathbf{y}-\mathbf{z}}{|\mathbf{y}-\mathbf{z}|}\right)\cdot\mathbf{z}'=0. \tag{1}$$

We have to show that

$$|\mathbf{y}-\mathbf{x}|+|\mathbf{z}-\mathbf{y}|+|\mathbf{x}-\mathbf{z}|=C \tag{2}$$

for some constant C. If we take the derivative of the left side of (2) with respect to s, we obtain

$$\frac{(\mathbf{y}-\mathbf{x})\cdot(\mathbf{y}'-\mathbf{x}')}{|\mathbf{y}-\mathbf{x}|}+\frac{(\mathbf{z}-\mathbf{y})\cdot(\mathbf{z}'-\mathbf{y}')}{|\mathbf{z}-\mathbf{y}|}+\frac{(\mathbf{x}-\mathbf{z})\cdot(\mathbf{x}'-\mathbf{z}')}{|\mathbf{x}-\mathbf{z}|}$$

$$=\left(-\frac{\mathbf{y}-\mathbf{x}}{|\mathbf{y}-\mathbf{x}|}+\frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\right)\cdot\mathbf{x}'+\left(-\frac{\mathbf{z}-\mathbf{y}}{|\mathbf{z}-\mathbf{y}|}+\frac{\mathbf{y}-\mathbf{x}}{|\mathbf{y}-\mathbf{x}|}\right)\cdot\mathbf{y}'+\left(-\frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}+\frac{\mathbf{z}-\mathbf{y}}{|\mathbf{z}-\mathbf{y}|}\right)\cdot\mathbf{z}',$$

and because of (1) this is equal to zero. Hence (2) is true.

The theorem is valid for curves in a space of arbitrary dimension.

(b) Let the bisector of angle RPQ be fixed. Apply rectangular coordinates. Let $PQ + QS = a$, where a is a positive constant, $P = (0, p)$, $Q = (x, y)$, and S is the midpoint of RQ. (See FIGURE 1.) Then $\sqrt{x^2 + (p-y)^2} = a - x$, so that Q lies on (a branch of) the parabola $(p-y)^2 = a^2 - 2ax$. The slope of the tangent at Q is

$$\frac{dy}{dx}=\frac{a}{p-y},$$

and the normal at Q has equation

$$\eta - y = -\frac{p-y}{a}(\xi - x).$$

Let the normal intersect the bisector PS at T. Then

$$PT=\frac{(p-y)(a-x)}{a}\quad\text{and}\quad TS=\frac{x(p-y)}{a}.$$

We have

$$PT : TS = PQ : SQ, \quad \text{or} \quad \sqrt{x^2 + (p-y)^2} = a - x.$$

Hence the normal to Q at the orbit of Q is the bisector of the angle PQR!

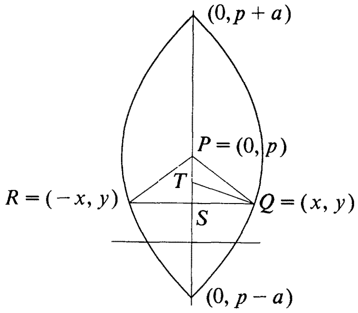


FIGURE 1

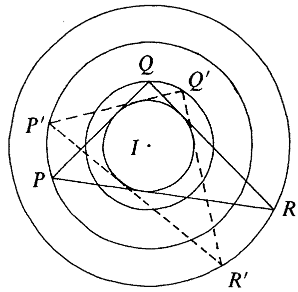


FIGURE 2

As a second example, let PQR be any triangle, and let I be its incenter. Let the triangle PQR rotate about I. (See FIGURE 2.) Then P, Q, and R describe three circles, and at each instant, the normals to the circles at P, Q, and R are the bisectors of the angles of the triangle.

M. KUIPERS
Groningen,
The Netherlands

and

L. KUIPERS
Sierre, Switzerland

Solution II: (b) An example with P fixed is given by taking two congruent parabolas with the same axis but facing opposite ways and having P as a common focus [as in Solution I(b)]. Then it's not hard to verify that the bisector of PQR is perpendicular to the tangent to the parabola at Q. (In fact, this is a well-known property of parabolic mirrors.) You can do similar tricks with an ellipse and a hyperbola. (See FIGURE 3.)

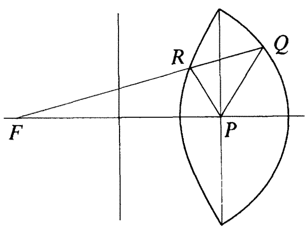


FIGURE 3
Arcs of confocal ellipse and hyperbola intersecting on their common latus rectum.

PETER J. GIBLIN
University of North Carolina at Chapel Hill
and University of Liverpool

Part (a) *was also solved by the proposer & I. R. Porteous (England). No solution to part* (c) *was received.*

For a related problem by the proposer, see the Unsolved Problems department of the *Monthly*, edited by Richard Guy, vol. 90 (1983), p. 121.

An Elevating Problem **September 1982**

1152. An elevator starts on the top floor of a 100-floor building and in its descent to the bottom (first) floor stops at at least 40 floors, counting both the top and bottom floors as stops. Show that somewhere in its travel the elevator had to stop at two floors that were exactly 9, 10, or 19 floors apart. [*Michael Gilpin and Robert Shelton, Michigan Technological University.*]

Solution I: Let $a_1, a_2, a_3, \ldots, a_{40}$ be the 40 floors stopped at by the elevator on its descent, where $100 = a_1 > a_2 > a_3 > \cdots > a_{40} = 1$. Adding 9 and 19 to each value yields two additional sequences with $109 = a_1 + 9 > a_2 + 9 > \cdots > a_{40} + 9 = 10$ and $119 = a_1 + 19 > a_2 + 19 > \cdots > a_{40} + 19 = 20$.

The 120 integers $a_1, \ldots, a_{40}$, $a_1 + 9, \ldots, a_{40} + 9$, $a_1 + 19, \ldots, a_{40} + 19$ are between 1 and 119 (inclusive). By the pigeonhole principle, at least two of these integers are equal: there are i and j such that $a_i = a_j + 9$ or $a_i + 9 = a_j + 19$ or $a_i = a_j + 19$, so that $a_i - a_j$ is 9, 10, or 19. Therefore the elevator stops at two floors that are exactly 9, 10, or 19 floors apart.

MIRIAM MCCANN
University of Wisconsin—Oshkosh

Solution II: A slightly stronger result will be proved:

Given 37 integers with $1 \leqslant a_1 < a_2 < a_3 < \cdots < a_{37} \leqslant 112$, there exist i and j such that $a_i - a_j$ is equal to 9, 10, or 19.

Proof. Let

$$A = \{a_i\colon 1 \leqslant a_i \leqslant 28\}, \quad B = \{a_i\colon 29 \leqslant a_i \leqslant 56\}, \quad C = \{a_i\colon 57 \leqslant a_i \leqslant 84\},$$

and

$$D = \{a_i\colon 85 \leqslant a_i \leqslant 112\}.$$

At least one of these four sets must contain at least 10 elements. In one such set, reduce all elements by 1 less than the smallest member of the set. This does not change the differences between elements. The new numbers satisfy

$$1 = b_1 < b_2 < b_3 < \cdots < b_n \leqslant 28, \quad \text{with } n \geqslant 10.$$

If one of the b_i's equals 10, then $b_i - b_1 = 9$ and we are done. Otherwise, partition the numbers $1, 2, 3, \ldots, 9, 11, 12, \ldots, 28$ in the following way:

$$\{1,11,20\}, \{2,12,21\}, \{3,13,22\}, \ldots, \{9,19,28\}.$$

Since there are at least 10 b_i's, two of them must occur in one of these nine triples and therefore have a difference of 9, 10, or 19.

CRAIG K. BAILEY
U.S. Naval Academy

Also solved by Sheldon B. Akers, University of Arizona Problem Solving Group, Robert E. Bernstein, Duane Broline, Curtis Cooper, Dennis Hamlin (student), David Hanson, Victor Hernández (Spain), G. A. Heuer & C. V. Heuer, Richard Johnsonbaugh, Henry S. Lieberman, Richard T. Mahoney, Vania D. Mascioni (student, Switzerland), J. M. Metzger, James Propp (student, England), Stanley Rabinowitz, St. Olaf College Problem Solving Group, Robert S. Stacy (West Germany), Michael Vowe (Switzerland), and the proposers. There was one incorrect solution.

The proposers' paper, *Forced differences between terms of subsequences of integer sequences* (to appear), contains several generalizations of the problem.

Sums of Relative Totients **September 1982**

1153. Let ϕ be Euler's totient function. Show that:

(a) if $n > 79$, then there are a and b such that $a + b = n$, $a > 1$, $b > 1$, and $\phi(a)/a + \phi(b)/b < 1$;
(b) if $n > 4$, then there are a and b such that $a + b = n$, $a > 1$, $b > 1$, and $\phi(a)/a + \phi(b)/b > 1$.
[*Charles R. Wall, Trident Technical College.*]

Solution: (a) Let $P(n)$ be the statement:

"There are a and b such that $a + b = n$, $a > 1$, $b > 1$, and $\phi(a)/a + \phi(b)/b < 1$."

We will prove the maximal extension of (a), namely:

(a′) $P(n)$ is true if and only if either n is even and greater than 7 or n is odd and greater than 20 but different from 23, 47, 49, 53, and 79.

Proof. (i) Let $n \geqslant 6$ be even and set $n-2 = 2^k d$ with $(2, d) = 1$. Then

$$\frac{\phi(2)}{2} + \frac{\phi(n-2)}{n-2} = \frac{1}{2} + \frac{2^{k-1}\phi(d)}{n-2} < \frac{1}{2} + \frac{2^{k-1}d}{n-2} = 1,$$

unless $d = 1$, that is, $n = 2^k + 2$. If $n = 2^k + 2$, then $n - 4 = 2(2^{k-1} - 1)$, and

$$\frac{\phi(4)}{4} + \frac{\phi(n-4)}{n-4} = \frac{1}{2} + \frac{\phi(2^{k-1}-1)}{n-4} < 1,$$

unless $k = 2$, in which case $n = 6$.

(ii) If m is even, then obviously $\phi(m)/m \leqslant \frac{1}{2}$. Since $\phi(105)/105 = 16/35 < \frac{1}{2}$, we see that $P(n)$ is true for all odd $n > 105$. (It suffices to take $a = 105$ and $b = n - 105$.)

(iii) For the odd numbers not exceeding 105, one finds by calculation that $P(n)$ is true for such n just when $n \geqslant 21$ and $n \notin \{23, 47, 49, 53, 79\}$. The following table lists suitable triples (n, a, b), one for each n not excluded.

(21,6,15)	(39,6,33)	(61,21,40)	(77,35,42)	(93,3,90)
(25,10,15)	(41,15,26)	(63,15,48)	(81,3,78)	(95,15,80)
(27,6,21)	(43,15,28)	(65,15,50)	(83,20,63)	(97,22,75)
(29,14,15)	(45,3,42)	(67,15,52)	(85,10,75)	(99,6,93)
(31,10,21)	(51,6,45)	(69,3,66)	(87,3,84)	(101,21,80)
(33,3,30)	(55,10,45)	(71,15,56)	(89,14,75)	(103,15,88)
(35,15,20)	(57,6,51)	(73,10,63)	(91,21,70)	(105,3,102)
(37,15,22)	(59,14,45)	(75,9,66)		

(b) Let $n > 4$, and let p be the largest prime not exceeding $n - 2$.

If $p = n - 2$, then $p \geqslant 3$ and

$$\frac{\phi(2)}{2} + \frac{\phi(n-2)}{n-2} = \frac{1}{2} + 1 - \frac{1}{p} > 1.$$

If $p < n - 2$, then

$$\frac{\phi(p)}{p} + \frac{\phi(n-p)}{n-p} = 1 - \frac{1}{p} + \frac{\phi(n-p)}{n-p} \geqslant 1 - \frac{1}{p} + \frac{2}{n-p}.$$

Bertrand's "postulate" assures us that $\frac{n-2}{2} < p < n-2$ (if $n > 4$), so that $n - p < p + 2$. Hence

$$\frac{2}{n-p} - \frac{1}{p} > \frac{2}{p+2} - \frac{1}{p} = \frac{p-2}{(p+2)p} > 0,$$

since $p \geqslant 3$. The desired inequality follows then with $a = p$ and $b = n - p$.

VANIA D. MASCIONI, student
Swiss Federal Institute of Technology

Also solved by Daniel A. Rawsthorne, J. M. Stark, and the proposer.

Answers

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

Q686. The first inequality is equivalent to $\Sigma_{i<j}(a_i - a_j)^2 \geqslant 0$, with equality if and only if the polygon is equilateral. The second inequality is equivalent to $\Sigma_{i=1}^n a_i(p - a_i) \geqslant \Sigma_{i=1}^n a_i^2$, where $p = \Sigma_{i=1}^n a_i$, with equality only for the degenerate n-gon in which $n - 2$ of the sides are of length zero.

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.

Winfree, Arthur T., *Sudden cardiac death: a problem in topology*, Scientific American 248:5 (May 1983) 144-161, 170.

Many sudden deaths are the result of fibrillation: a disruption of the coordinated contraction of the heart. Winfree, author of *The Geometry of Biological Time*, suggests that the nonretraction theorem of topology implies there must be a singular point in the state space of the heart motion and hence explains why a small mistimed stimulus to the heart can induce fibrillation. "If the topological mechanism proposed here is correct, its clinical implications could be somewhat bleak," because the heart is continually bombarded by electrical impulses from many courses; the challenge would be to preventive medicine to make the size of the "black holes" (singular regions) as small as possible.

Dauben, Joseph W., *Georg Cantor and the origins of transfinite set theory*, Scientific American 248:6 (June 1983) 122-131, 154.

Careful history of Cantor's development of the fundamental ideas of transfinite numbers, with appropriate explanations of the ideas. Dauben's 1979 book, *Georg Cantor: His Mathematics and Philosophy of the Infinite*, has the further details.

Diaconis, Persi, and Efron, Bradley, *Computer-intensive methods in statistics*, Scientific American 248:5 (May 1983) 116-130, 170.

Computer-intensive statistical methods can replace standard assumptions about data with massive calculations. Discussed in detail here is the bootstrap, a "resampling" method invented by Efron in 1977; other methods (jackknife, cross-validation, balanced repeated replications, random subsampling) are described only briefly.

Hofstadter, Douglas R., *Metamagical themas: the calculus of cooperation is tested through a lottery*, Scientific American 248:6 (June 1983) 14-28, 154.

Hofstadter has invented a one-shot n-person version of Prisoner's Dilemma, and he describes the result of trying it out on 20 friends (using $700 of cash payoffs from *Scientific American*!). "The answer I was hoping to receive from everyone was c[ooperate]." Shouldn't symmetry of the problem suggest symmetry (unanimity) of decision? Did cooperation or defection prevail? Read to find out! You may also want to participate in Hofstadter's "Luring Lottery," where the price is $1,000,000/N, where N is the number of entries submitted, and you may submit (on one postcard) as many entries as you like.

Hofstadter, Douglas R., *Metamagical themas: computer tournaments of the Prisoner's Dilemma suggest how cooperation evolves*, Scientific American 248:5 (May 1983) 16-26, 170.

Account of two tournaments among strategies for sequential play of Prisoner's Dilemma. The winner: "tit-for-tat," the shortest program submitted. Wider philosophical questions arise in the discussion: how can mutal cooperation emerge in society? how can it maintain itself in an evolving world? how can the results of this study be reconciled with human morality?

Fagin, Ronald, and Williams, John H., *A fair carpool scheduling algorithm*, IBM Journal of Research and Development 27 (1983) 133-139.

Simple carpool scheduling algorithm which is "fair" (in a precise sense), robust under contingencies (various working schedules, additions and deletions, vacations), and easy to implement. But will ordinary people use an algorithm whose justification they don't understand, merely on the say-so of a mathematician in the carpool?

Dembert, Lee, *In the world of numbers, he's No. 1: mathematician Erdös*, Los Angeles Times (26 March 1983) 1, 24-26.

Long, popular article about Paul Erdös on the occasion of his 70th birthday. Who else since Sylvester has published a thousand papers, and who in history has ever had several hundred co-authors? And who else has roamed the world for 50 years, never staying anywhere more than a month at a time? "[Erdös] is really driven to try to answer one more question, one more mountain to climb."

Kneale, Dennis, *Shaping ideas: a research topologist wows world of math by seeing the unseen*, Wall Street Journal (18 March 1983).

"Princeton's Thurston thinks about imaginary shapes, ignores the possible uses; tea, cookies, 'K derivatives." Popular look into the life of topologist Bill Thurston and his research on manifolds.

Davis, Philip J., The Thread: A Mathematical Yarn, Birkhäuser, 1983; vii + 127 pp, $12.95.

It is doubtful whether any writer could get more mileage (or satisfaction!) out of researching a single mathematician's name than Davis does out of Chebyshev's. But the name is only the thread that ties together this engaging entertainment. To tell more would be possible; the style--surprisingly coherent digressions--is the thing.

Hooke, Robert, How to Tell the Liars from the Statisticians, Dekker, 1983; xv + 173 pp, $15.95.

Occasionally a book's title alone compels a review of the book. "Liars, as portrayed on screen and in fiction, are often charming rogues, while statisticians are always persons of infinite dullness. In real life, this is not the way you tell one from the other." A large number (76) of succinct essays follow, devoid of formulas and nearly of numbers. Each essay has one thought-provoking idea, and the overall effect is to convey a splendid qualitative understanding of the important ideas of statistics.

Calinger, Ronald, Classics of Mathematics, Moore Publ. Co., 1982; xxv + 742 pp, (P).

Compendium of 133 selections from writings of leading mathematicians from antiquity through the early 20th century. Also included are introductory essays to periods and biographical sketches. Almost all selections are reprinted from other well-known collections, but it is convenient to have them available once more, and together.

Kennedy, Don H., Little Sparrow: A Portrait of Sophia Kovalevsky, Ohio U. Pr., 1983; ix + 341, $2.95, $12.95 (P).

First full biography of Kovalevsky, based on thorough research of all available sources (though assertions are not footnoted to source). "The biography is not about a mathematician, as such, but about an unusual woman who happens to have a secure place in the history of science as she does in Russian literature. Only superficially is her scientific work examined."

Bucciarelli, Louis L., and Dworsky, Nancy, Sophie Germain: An Essay in the History of the Theory of Elasticity, Reidel, 1980; xi + 147 pp.

"Why should the story of a woman's role in the development of a scientific theory be written.... [T]o celebrate...the heroism of a woman's struggle in a man's world? ...[T]o demonstrate that gender is irrelevant to the march of scientific ideas? ...[N]either. ...[T]o do justice both to the professional life of a woman in science and to the development of the theory with which she was engaged." A book-length scientific biography of Germain has been long overdue. Quotations from original sources have been translated into English.

Brams, Steven J., and Fishburn, Peter C., Approval Voting, Birkhäuser, 1983; xix + 198 pp, $24.95, $14.95 (P).

Here is definitive justification for the widespread use of approval voting, in which each voter can vote for ("approve of") as many candidates as he or she wishes, and the candidate with the most votes wins. No voting system is perfect, but the authors detail the superiority of approval voting to other systems currently in use. (In an article earlier in this issue they discuss the drawbacks of another voting scheme, preferential voting.)

Riker, William H., Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice, Freeman, 1982; xix + 311 pp (P).

The paradoxes of voting, after "a score of years of reflection on Black's and Arrow's discoveries," led Riker to reject the populism he had initially espoused, on grounds that elections cannot reveal unequivocally what the people want. (He dismisses the approval voting espoused by Brams and Fishburn on grounds: it can defeat a Condorcet winner [one who would win head-to-head contests against each other candidate]; permuting the preference orders among the voters can change the result [because of differing proclivity on how many votes to cast]; and it undermines the two-party system.) Despite the mathematically-revealed downfall of populism, and Riker's belief that there may be no protection against manipulation of the social agenda to generate disequilibrium and hence possibly shifts of power, he feels "the fundamental method to preserve liberty is to preserve ardently our traditional [U.S.] constitutional restraints--decentralized parties and multicameral government."

Sackson, Sid, A Gamut of Games, 2nd ed., Random House, 1982; xiv + 222 pp, $5.95 (P).

Collection of 38 original indoor games, played with the common equipment of cards, dice, checkers, etc. Many of the games are susceptible to some mathematical analysis; alumni of Berlekamp, Conway & Guy's *Winning Ways*, take notice!

Shubik, Martin, Game Theory in the Social Sciences: Concepts and Solutions, MIT Pr., ix + 514 pp, $35.

Compendium of methodology and concepts in game theory developed during the last 25 years. A subsequent volume will apply game theory to economic problems, in particular the role of money and the theory of oligopoly.

Schwartz, Richard H., Mathematics and Global Survival: Scarcity, Hunger, Population Growth, Pollution, Waste, (from the author: College of Staten Island, 715 Ocean Terrace, Staten Island, NY 10301); x + 385 pp, $12.50.

"Do you have liberal arts students who wonder why they have to study mathematics? Weak students who need instruction and practice at basic mathematical operations? Are you looking for applications of mathematics that can challenge and interest your students?" If so, this is the book for you. It covers basic operations, ratio and proportion, percent, graphical methods (graphs, pie charts, bar charts, histograms), sequences, simple population models, limits to growth, logarithms, scientific notation, and statistics (sampling, standard deviation, probability, the normal curve, chi-square, and correlation)--plus five chapters of case studies.

Eigen, Manfred, and Winkler, Ruthild, Laws of the Game: How the Principles of Nature Govern Chance, Harper & Row, 1983; xiv + 347 pp, $8.95 (P).

Startlingly imaginative book exploring play as the expression of creativity in nature and culture. Reminiscent of Hofstadter's *Gödel Escher Bach*--but written earlier and only now translated--this book dwells on the dichotomies of chance vs. necessity and creativity vs. rules. The authors use classical and new games to illustrate models of physical, biological, and social phenomena; they analyze Bach's music, as well as that of Schönberg. One only wishes for more elaboration in this excursion in the realm of ideas.

Sedgewick, Robert, Algorithms, Addison-Wesley, 1983; viii + 552 pp.

Splendid book chock-full of what math and computer science students should learn after a course or two in programming. Covered are algorithms for mathematics (quadrature, elimination, etc.), sorting, searching, string processing, geometry, graphs, and applications (fast Fourier transform, simplex method). Some minor slips do occur (no reference for Brent on p. 210; top of p. 97 should refer to bubble, not insertion, sort). Students embarking on this journey should have studied calculus and linear algebra.

Baber, Robert L., Software Reflected, North-Holland, 1982; ix + 192 pp, $29.95.

Thorough indictment of current practice in software design and development, together with sketches of futures that may grow out of the present, and an outline of what must be done to bring about a real improvement. One speculative future the author does not consider: error-free Japanese software may capture the world software systems market.

Pollack, Seymour V. (ed.), Studies in Computer Science, MAA Studies in Mathematics Vol. 22, 1982; xvii + 388 pp.

Nine extended essays on aspects of computer science, at an introductory level: development of computer science, programming languages, formal languages, analysis of programs, complexity, artificial intelligence, numerical analysis, simulation, and statistical data analysis.

Ralston, Anthony, and Young, Gail S., The Future of College Mathematics: Proceedings of a Conference/Workshop on the First Two Years of College Mathematics, Springer-Verlag, 1983; ix + 278 pp, $18.

Should the first two years of college mathematics be restructured to balance traditional calculus and linear algebra with discrete mathematics? If so, how? Arguments pro and con, and model curricula, make up this book, which should be the focus of curricular discussion in all mathematics departments.

Schoenfeld, Alan H., Problem Solving in the Mathematics Curriculum: A Report, Recommendations, and an Annotated Bibliography, MAA Notes Number 1, 1983; ii + 137 pp, $5 (P).

Stimulating report from an MAA task force on teaching problem-solving. The advice is directed mainly to the setting of a separate problem-solving course. Over half the book is devoted to an annotated bibliography of articles, books, and journals featuring problem-solving.

Lester, Frank K., Jr., and Garafalo, Joe, Mathematical Problem Solving: Issues in Research, Franklin Institute Pr., 1982; xii + 139 pp, $14.50.

What new ideas and perspectives does cognitive psychology, and particularly information-processing, suggest to the study of mathematics learning and problem-solving? The essays here by Schoenfeld and by Lesh & Akerstrom are especially provacative. Schoenfeld suggests that most instruction in mathematics is "deceptive and possibly fraudulent": teaching of "powerful mathematical techniques" boils down to training students in rote learning and mechanical use of technique, while students continue to lack some rudimentary thinking skills (such as how to read and restate word problems). Lesh and Akerstrom contend that real problems are quite unlike word problems, and content-independent heuristics are both unteachable and of dubious value.

Berry, John, *et al.* ("The Spode Group"), Solving Real Problems With Mathematics, Vols. 1, 2; Solving Real Problems With C.S.E. Mathematics, CIT (Cranfield Institute of Technology) Press, 1981-1983; vii + 105 pp, vii + 134 pp, vii + 139 pp, £ 4.50, £ 7.50, £ 7.50.

Collections of case studies (15, 21, 28 respectively), each consisting of a problem statement, teaching notes, and possible solutions. Most are imaginatively drawn from real-life experiences and will interest high school and college students. The elevator problem (Vol. 2, p. 38) is particularly ingenious and innovatively presented.

Hudspeth, Mary, and Hirsch, Lewis R., Studying Mathematics, Kendall/Hunt, 1982; xi + 51 pp, $4.95 (P).

"But I do okay in other subjects..." Starting from what makes mathematics different, this booklet discusses: note-taking, studying notes, reading the text, and working assignments; preparations for quizzes and tests; and dealing with test anxiety. Desirable as it would be for every mathematics student to have and read this book (and reread it, even), it would be a wise teacher who devoted a few minutes every day or two to reminding students of some of the advice presented here.

Smith, Douglas, *et al.*, A Transition to Advanced Mathematics, Brooks/Cole, 1983; xii + 177 pp, $22.95.

This text is intended to bridge the gap from (the now usually almost entirely intuitive) calculus to advanced courses relying on mathematical argument. Four chapters treat logic, set theory, relations, and functions, with three "application" chapters on cardinality, group theory, and real numbers. A novel feature is the inclusion of "Proofs to Grade" in the exercises, where students are called upon to criticize alleged proofs.

Artino, Ralph A., *et al.*, The Contest Problem Book IV: Annual High School Mathematics Examinations 1973-1982, MAA New Mathematical Library 29, 1983; xiv + 184 pp.

This volume completes the publication of contests prior to the new scheme (easier first round, new intermediate round) instituted in 1983.

NEWS & LETTERS

"NICE" ORTHOGONAL MATRICES

K. J. Heuvers (Symmetric matrices with prescribed eigenvalues," this *Magazine*, March, 1982, 106-111) has suggested some ways to produce a real symmetric matrix of order n with "nice" (rational) eigenvalues and eigenvectors. But the most difficult step was not discussed: how to find a "nice" orthogonal matrix X. This can be done as follows.

Let u be any nonzero vector in R^n, and define:

$$X_u = I - \frac{2}{u^T u} uu^T; \qquad (1)$$

X_u is a reflection in R^n whose mirror is the $(n-1)$-hyperplane $\{y \mid u^T y = 0\}$. The delightful properties of X_u are extolled by B. N. Parlett [1], where he also points out the important fact that *any* orthogonal matrix can be written as a product of reflections. (Reflections are sometimes called Householder transformations.) If the vector u has rational components, then X_u also has rational components, so is "nice." For example, if $n = 4$ and $u = (-1,1,2,-2)^T$, we obtain:

$$X_u = \frac{1}{5}\begin{pmatrix} 4 & 1 & 2 & -2 \\ 1 & 4 & -2 & 2 \\ 2 & -2 & 1 & 4 \\ -2 & 2 & 4 & 1 \end{pmatrix}.$$

The matrix X_u is always symmetric, which may not be desirable. More sophistication can be introduced using the product of several reflections. For two reflections X_u and X_v, defined as in (1), one can show

$$X_v X_u = X_u + \frac{2v}{v^T v}\left(2\frac{v^T u}{u^T u} u^T - v^T\right). \qquad (2)$$

It is simpler to use equation (2) than to form the product $X_v X_u$ when one computes by hand. For example, for $n = 5$ and $u = (2,1,-1,1,1)^T$, $v = (-1,1,1,-1,2)^T$ we obtain the orthogonal matrix

$$X_u X_v = \frac{1}{16}\begin{pmatrix} -2 & -3 & 11 & -11 & 1 \\ -6 & 7 & 1 & -1 & -13 \\ 10 & -1 & 9 & 7 & -5 \\ -10 & 1 & 7 & 9 & 5 \\ -4 & -14 & -2 & 2 & -6 \end{pmatrix}.$$

Once an orthogonal X is available, a matrix A with eigenvalues $\lambda_1,\ldots,\lambda_n$ can be obtained by the procedures suggested by Professor Heuvers, or from $A = XDX^T$ with $D = \text{diag}\,(\lambda_1,\ldots,\lambda_n)$. For the generalized eigenproblem $AX = BXD$ with A and B symmetric and B positive definite, let

$$X = I - (2/u^T v)uv^T$$

with $u \neq v$ and $u^T v \neq 0$. Then $X^{-1} = X$, $A = X^T DX$ and $B = X^T X$.

[1] B.N. Parlett, The Symmetric Eigenvalue Problem, Prentice-Hall, 1980.

Sergio Pissanetzky
Centro Atómico Bariloche and
Universidad Nacional de Cuyo
Argentina

STATISTICS AND LIFE INSURANCE

I enjoyed "Statistics and the Law" (this *Magazine*, March, 1983), but I feel the section entitled life expectancy may lead the readers to the wrong conclusion.

The comparison of "individual rights" and "group rights" in insurance arrangements must be approached with caution. Insurance programs work because there is a pooling of risks, and the claims of those who suffer the contingency insured against are paid by those who do not. Since pooling is required in insurance systems, the challenge is to find a fair classification system and a fair premium for each class.

The overlap theory described in the paper cannot be applied to insurance systems, because it could be used to prove any risk classification system unfair, an absurd conclusion. Consider age, for example. Suppose 1,000 60-year-olds and 1,000 70-year-olds each purchase a life annuity with $10,000 of retirement savings. Should they receive the same benefit? The overlap theory would say that since a substantial portion of the deaths are "matched," it is not fair to the 60-year-olds to give them a smaller pension benefit than the 70-year-olds. Yet, as a group, the cost per dollar of monthly benefit will be higher for the 60-year-olds. I don't believe lower benefits for the 60-year-olds is unfair, and I believe that most people would agree with me. Whether the situation is fair may depend on who is paying for the benefit. If an employer supplied the $10,000 and the employee had no option to take cash instead of a life income, then I believe the benefits should be equal. In any event, the issue is not as simple as presented in the paper.

Forrest A. Richen
Standard Insurance Company
Portland, OR 97207

SPECIAL NOVEMBER ISSUE

The November 1983 *Mathematics Magazine* will be a tribute to the work of the mathematical genius Leonhard Euler, marking 200 years since his death in 1783. Six articles, a glossary of "Euler" terms, and numerous illustrations will provide an overview of Euler's mathematical contributions as well as some close examination of his work from a contemporary viewpoint.

The usual Solutions to Problems, REVIEWS, and NEWS AND LETTERS sections will be omitted from the extra-large issue. Since the issue should be of interest to many who do not subscribe to the *Magazine*, the MAA will sell individual copies through its publication office in Washington.

PÓLYA, ALLENDOERFER 1982 AWARDS

At the business meeting on August 9, 1983, in Albany, the MAA honored five authors for excellence in expository writing. The awards, in the amount of $200 each, were for articles published in 1982 in *Mathematics Magazine* and the *Two-Year College Mathematics Journal*.

Recipients of the Carl B. Allendoerfer awards were:

Donald O. Koehler, "Mathematics and Literature," *Math. Magazine*, 55 (1982) 81-95.
Clifford H. Wagner, "A Generic Approach to Iterative Methods," *Math. Magazine*, 55 (1982) 259-273.

Recipients of the George Pólya awards were:

Douglas R. Hofstadter, "Analogies and Metaphors to Explain Gödel's Theorem," *TYCMJ*, 13 (1982) 98-113.
Paul R. Halmos, "The Thrills of Abstraction," *TYCMJ*, 13 (1982) 243-251.
Warren Page and Vedula N. Murty, "Nearness Relations Among Measures of Central Tendency and Dispersion," *TYCMJ*, 13 (1982) 315-327.

The award to Donald Koehler, who died suddenly in February, 1983, was received by his wife.

OLYMPIAD NEWS AND PROBLEMS

This year, the American High School Mathematics Examination (AHSME) and the American Invitational Mathematics Examination (AIME) involved 400,000 high school students throughout the United States and Canada. The 54 top performers in the two-stage contest competed in the Twelfth USA Mathematical Olympiad (USAMO) on May 3, 1983. The eight USAMO winners were:

*John M. Steinke, San Antonio, TX;
*Michael Reid, Woodhaven, NY;
*C. James Yeh, Mountain Brook, AL;
*Jeremy A. Kahn, New York, NY;
*Steven Newman, Ann Arbor, MI;
*Douglas S. Jungreis, Woodmere, NY;
Douglas R. Davidson, McLean, VA;
John T. O'Neil III, Princeton, NJ.

The winners were honored on June 7 in Washington, D.C., at the Twelfth USA Olympiad Awards Ceremony held in the National Academy of Sciences and the Diplomatic Reception Rooms of the U.S. Department of State.

Following the Awards Ceremony, the eight winners and sixteen other students who did well in the USAMO Examination participated in a seminar to train a U.S. team of six students for the 1983 International Mathematical Olympiad (IMO) held in Paris, France, on July 6-7. Names of those chosen for the U.S. team are starred in the list.

Teams from 32 countries competed in the IMO. The top five teams and their scores were:

West Germany	212
USA	171
Hungary	170
USSR	169
Romania	161

All U.S. team members were medalists; Michael Reid took a first-place gold.

Here are the problems from the two national contests -- USA and Canada -- how many can you solve? Solutions to these will appear in our January 1984 issue, as well as the IMO problems.

12th USA MATHEMATICAL OLYMPIAD

1. On a given circle, six points A, B, C, D, E and F are chosen at random, independently and uniformly with respect to arc length. Determine the probability that the two triangles ABC and DEF are disjoint, i.e., having no common points.

2. Prove that the roots of

$$x^5 + ax^4 + bx^3 + cx^2 + dx + e = 0$$

cannot *all* be real if $2a^2 < 5b$.

3. Each set of a finite family of subsets of a line is a union of two closed intervals. Moreover, any three of the sets of the family have a point in common. Prove that there is a point which is common to at least half of the sets of the given family.

4. Six segments S_1, S_2, S_3, S_4, S_5 and S_6 are given in a plane. These are congruent to the edges AB, AC, AD, BC, BD and CD, respectively, of a tetrahedron $ABCD$. Show how to construct a segment congruent to the altitude of the tetrahedron from vertex A with straight-edge and compass.

5. Consider an open interval of length $1/n$ on the real number line where n is a positive integer. Prove that the number of irreducible fractions p/q, with $1 \le q \le n$, contained in the given interval is at most $(n + 1)/2$.

15th CANADIAN OLYMPIAD, 1983

1. Find all positive integers w, x, y, and z which satisfy

$$w! = x! + y! + z! \; .$$

2. For each real number r, let T_r be the transformation of the plane that takes the point (x,y) into the point $(2^r x, r2^r x + 2^r y)$. Let F be the family of all such transformations, i.e., $F = \{T_r | r \text{ a real number}\}$. Find all curves $y = f(x)$ whose graphs remain unchanged by every transformation in F.

3. The area of a triangle is determined by the lengths of its sides. Is the volume of a tetrahedron determined by the areas of its faces?

4. Prove that for every prime number p, there are infinitely many positive integers n such that p divides $2^n - n$.

5. The geometric mean (G.M.) of k positive numbers is defined to be the positive kth root of their product. For example, the G.M. of 3, 4, and 18 is 6. Show that the G.M. of a set S of n positive numbers is equal to the G.M. of the G.M.'s of all non-empty subsets of S.

Eminent Mathematicians and Mathematical Expositors speak to

STUDENTS and TEACHERS in . . .

The NEW MATHEMATICAL LIBRARY

An internationally acclaimed paperback series providing:

- stimulating excursions for students beyond traditional school mathematics.
- supplementary reading for school and college classrooms.
- valuable background reading for teachers.
- challenging problems for solvers of all ages from high school competitions in the US and abroad.

The New Mathematical Library is published by the MATHEMATICAL ASSOCIATION OF AMERICA. The volumes are paperbound.

For information regarding the price of these publications, please contact The Mathematical Association of America at the address listed below.

NUMBERS: RATIONAL AND IRRATIONAL by Ivan Niven, NML-01

WHAT IS CALCULUS ABOUT? By W. W. Sawyer, NML-02

AN INTRODUCTION TO INEQUALITIES, by E. F. Beckenbach, and R. Bellman, NML-03

GEOMETRIC INEQUALITIES, by N. D. Kazarinoff, NML-04

THE CONTEST PROBLEM BOOK. Problems from the Annual High School Mathematics Contests sponsored by the MAA, NCTM, Mu Alpha Theta, The Society of Actuaries, and the Casualty Actuarial Society. Covers the period 1950-1960. Compiled and with solutions by C. T. Salkind. NML-05

THE LORE OF LARGE NUMBERS, by P. J. Davis, NML-06

USES OF INFINITY, by Leo Zippin, NML-07

GEOMETRIC TRANSFORMATIONS, by I. M. Yaglom, translated by Allen Shields, NML-08

CONTINUED FRACTIONS, by C. D. Olds, NML-09

GRAPHS AND THEIR USES, by Oystein Ore, NML-10

HUNGARIAN PROBLEM BOOKS I and II, based on the Eötvos Competitions 1894-1905 and 1906-1928. Translated by E. Rapaport, NML-11 and NML-12

EPISODES FROM THE EARLY HISTORY OF MATHEMATICS, by A. Aaboe, NML-13

GROUPS AND THEIR GRAPHS, by I. Grossman and W. Magnus, NML-14

THE MATHEMATICS OF CHOICE, by Ivan Niven, NML-15

FROM PYTHAGORAS TO EINSTEIN, by K. O. Friedrichs, NML-16

THE CONTEST PROBLEM BOOK II. A continuation of NML-05 containing problems and solutions from the Annual High School Mathematics Contests for the period 1961-1965. NML-17

FIRST CONCEPTS OF TOPOLOGY, by W. G. Chinn and N. E. Steenrod, NML-18

GEOMETRY REVISITED, by H. S. M. Coxeter, and S. L. Greitzer, NML-19

INVITATION TO NUMBER THEORY, by Oystein Ore, NML-20

GEOMETRIC TRANSFORMATIONS II, by I. M. Yaglom, translated by Allen Shields, NML-21

ELEMENTARY CRYPTANALYSIS — A Mathematical Approach, by Abraham Sinkov, NML-22

INGENUITY BY MATHEMATICS, by Ross Honsberger, NML-23

GEOMETRIC TRANSFORMATIONS III, by I. M. Yaglom, translated by Abe Shenitzer, NML-24

THE CONTEST PROBLEM BOOK III. A continuation of NML-05 and NML-17, containing problems and solutions from the Annual High School Mathematics Contests for the period 1966-1972. NML-25

MATHEMATICAL METHODS IN SCIENCE, by George Pólya, NML-26

INTERNATIONAL MATHEMATICAL OLYMPIADS, 1959-1977. Problems, with solutions, from the first nineteen International Mathematical Olympiads. Compiled and with solutions by S. L. Greitzer. NML-27

THE MATHEMATICS OF GAMES AND GAMBLING, by Edward W. Packel, NML-28

THE CONTEST PROGRAM BOOK IV, Annual High School Mathematics Examinations 1973-1982. Compiled and with solutions by R. A. Artino, A. M. Gaglione and Niel Shell. NML-29

THE MATHEMATICAL ASSOCIATION OF AMERICA

Send Orders to:

The Mathematical Association of America
1529 Eighteenth St., N.W.
Washington, D.C. 20036

TWO-YEAR COLLEGE MATHEMATICS READINGS

Edited by Warren Page

viii + 304 pages. Paperbound
List $19.50   MAA Member: $14.50

This work surveys the two-year college mathematics landscape. Stops are made at points of interest that cover areas from algebra to recreational mathematics.

There are many things of interest for everyone in this book. Some are expository articles that introduce the reader to new advances in mathematics. Other articles can be used to enrich course work and to stimulate both the student's and instructor's mathematical intellect.

This book will serve as an excellent source of enrichment for the talented student. Teachers of undergraduates will find it is an excellent reference source that they will want to consult when preparing lectures.

Topics covered are:

- **Algebra**
- **Geometry**
- **Number Theory**
- **Calculus**
- **Probability and Statistics**
- **Calculators and Computers**
- **Mathematics Education**
- **Recreational Mathematics**

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA

new

from the MAA Studies
in Mathematics . . .

Studies in Computer Science

Edited by Seymour Pollack
Volume 22, MAA Studies in Mathematics
408 pp. Cloth. List: $29.00 MAA Member: $22.00

Studies in Computer Science continues the tradition of excellence of the respected series MAA Studies in Mathematics. Written by computer scientists for mathematicians, the book presents a readable and balanced discussion of the role of mathematics in computer science, and the contributions of computing to mathematics.

Opening with an historical overview of the development of computer science, the narrative continues with three articles on the nature of computer programs and the programming process, a discussion of computational complexity, and a concluding group of articles on the impact of computer science on artificial intelligence, numerical analysis, statistics, and simulation.

In the introduction the editor expresses his hope that the book will provide *"an interesting look at an explosive field in the process of becoming."* What an understatement! The process is well advanced; the impact of computer science on mathematics and its teaching is already immense. And this book is not just interesting, it is exciting and important as background reading for anyone interested in the interactions between mathematics and computer science. Teachers will find that it enriches their teaching and students will find it rewarding supplementary reading. Order your copy today!

ORDER FROM: **The Mathematical Association of America**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

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

MATHEMATICS MAGAZINE VOL. 56, NO. 4, SEPTEMBER 1983