Vol. 63, No. 3 June 1990

MATHEMATICS MAGAZINE

Mathematicians should honor Thomas Hobbes, who came late in life to their subject, and understood it poorly, but loved it much, and staked on the supposed sureness of its methods his hopes for the peace and good government of mankind.

Hardy Grant

- Geometry and Politics: Thomas Hobbes
- Indeterminate Forms Revisited
- Drawing Pie Charts

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

Send new manuscripts to: Martha Siegel, Editor-elect, Mathematics Magazine, Towson State University, Towson, MD 21204. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and one copy and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added.

AUTHORS

Hardy Grant holds degrees from Queen's University (Kingston, Ont.) and McGill University (Montreal). His formal training was in mathematics and physics, but his main scholarly interest now is the impact of mathematical ideas on other areas of culture. He has for many years taught a "Humanities" course in this subject at York University, Toronto. His "desert island" writers are Will Durant, Joseph Needham, Donald Culross Peattie and George Steiner; his other passions include baseball, birdwatching, computer programming and pre-1950s popular music.

Vol. 63 No. 3, June 1990

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025–570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student, unemployed and emeritus members receive a 50% discount; new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $64 per year. Bulk subscriptions (5 or more copies) are available to colleges and universities for classroom distribution to undergraduate students at $5.40 per copy (minimum order 5 copies).

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, The Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

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

Postmaster: Send address changes to Mathematics Magazine Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Geometry and Politics: Mathematics in the Thought of Thomas Hobbes

HARDY GRANT
York University
North York, Ontario, Canada M3J 1P3

It is a commonplace of intellectual history that the 17th century's explosive development of mathematized science offered a seductive example in other spheres of thought. The successes of this "Scientific Revolution" seemed to rest on conceptual foundations and to use methods whose adoption promised similar triumphs wherever tried. The physicists' isolation, in thought and experiment, of categories amenable to quantification—mass, velocity, acceleration, force—had allowed the derivation of rigorous "laws" (Galileo's account of free fall, Newton's of universal gravitation); why should not other areas of study achieve so much? The clarity of ideas, the certainty of inference, characteristic of mathematical thinking became beacons. The "geometrical" manner of presenting a subject, stemming from Euclid's *Elements* and adopted by Galileo and Newton alike—the step-by-step deduction of results from explicit definitions and axioms—gave a model to those who would organize and expound their own realms to best advantage (Spinoza's *Ethics* is perhaps the most striking example). Few thinkers felt more deeply the lure of the new science and its mathematical mystique than the great British philosopher Thomas Hobbes, the author of *Leviathan*. Though his primary concerns—politics, morals, the law—might seem far from physical science in subject and spirit, Hobbes came to believe that mathematical categories and methods might be brought to bear even here, and might bring understanding and agreement where confusion and discord notoriously prevailed.

Like all of us he was shaped as much by his personal history as by the spirit of his age. The son of a disreputable clergyman, he made his way up the social scale by attaching himself, aged nineteen (1608), to a noble family as tutor, only to find his employer in shaky financial straits; his ensuing sense of insecurity may have been an impulse toward the certainties of mathematics and science[1]. Moreover his career had for background a painful and protracted time of social strife: the Puritan Rebellion, the "Long Parliament" (1640–53), civil war, the beheading of a king (Charles I, 1649), the ascendancy of Cromwell, the eventual restoration of the monarchy (1660). The royalist philosopher found it prudent to spend eleven of these turbulent years (1640–51) in exile in Paris.* In his eyes the conflicts tearing his homeland seemed to typify the worst of social ills, and to give practical urgency to a rational reconstruction of political life. Meanwhile his European travels brought personal encounters with some of the makers of the Scientific Revolution. In Florence he sought out Galileo, then (1636) an old man; what passed between them is not known, but Hobbes

*"The first," he declared, "of all that fled"[2]—rather like the Duke of Plaza-Toro in *The Gondoliers*.

revered the pioneering Italian as "the first that opened to us the gate of natural philosophy universal"—namely, the scientific understanding of motion[3]. In Paris he was made welcome in the circle of the Abbé Marin Mersenne, who acted as a sort of human post office for many of the leading wits of that exhilarating age. Through this valuable intermediary Hobbes bandied ideas at long range with Descartes, until the two great thinkers at last met face to face (1648). Hobbes (if we may believe John Aubrey's delectable "brief life" of him) said of Descartes that "had he kept himself to Geometry he had been the best Geometer in the world but that his head did not lye for Philosophy"[4]—though we shall see that he actually had no use for Descartes' mathematics either.

His own early training was in the humanities—the classics, history, philosophy; he came to mathematics relatively late in life, but memorably. Will Aubrey's much-quoted tale bear recycling one more time?

> He was 40 yeares old before he looked on Geometry; which happened accidentally. Being in a Gentleman's Library, Euclid's Elements lay open, and 'twas the 47 *El. libri* I [i.e., Book I, Prop. 47—the "Pythagorean" theorem]. He read the Proposition. *By G*—, sayd he (he would now and then sweare an emphaticall Oath by way of emphasis) *this is impossible*! So he reads the Demonstration of it, which referred him back to such a Proposition; which proposition he read. That referred him back to another, which he also read. *Et sic deinceps* [and so on] that at last he was demonstratively convinced of that trueth. This made him in love with Geometry.

Subsequently he indulged his new passion for the Queen of the Sciences by making diagrams and calculations on his thighs or on his bedsheets.[5]

So spurred, he went on to write many pages on mathematics, but usually his enthusiasm outran his insight. Much of the mathematical progress going on around him passed him by. A century after the work on cubic equations that forced even "imaginaries" into the realm of number, Hobbes restricted the latter concept to positive integers, and sharply contrasted its discrete character with the many continuous magnitudes of geometry and physics[6]; so lingering was a dichotomy that had fatefully colored the mathematics of classic Greece. Hobbes looked balefully on the introduction, by Viète, Descartes and others, of algebraic symbolism; he conceded that these new marks seemed vital as "scaffolds of demonstration," but "they ought no more to appear in public, than the most deformed necessary business which you do in your chambers."[7] Most notoriously, he claimed with complete confidence the duplication of the cube, "hitherto sought in vain", and the squaring of the circle; we may judge his grasp of this latter problem by his declaration that an "ordinary" man might accomplish it better than any geometer, by simply "winding a small thread about a given cylinder"[8]. Not an inspiring picture; but might one at least conjecture, on the ledger's other side, that Hobbes' preoccupation with mathematics helped shape the superb clarity and vigor of his prose?

In any case he needed no great technical competence to declare the importance of mathematics and mathematized science as models in other realms. He was convinced that there can be no true knowledge anywhere without proper method, and that here the geometers and physicists held the key. Increasingly he felt the inadequacy, even in social and political inquiry, of other purported paths to truth. He abandoned an early belief that universal verities about men and states can be reached by induction from a study of history; indeed, he urged, in *no* inquiry will the mere amassing of observations, however regular and consistent, yield general laws—"experience concludeth nothing universally."[9] Hobbes also denied (what Romanticism would later assert) that wisdom can come from a "sudden acuteness," from a leap of intuitive insight; and he poured scorn on thinkers who claimed divine inspiration for their profundities, who "take their own Dreams, and extravagant Fancies, and Madnesse, for testimonies of Gods Spirit"[10]. No, only the mathematicians' method, only strict deduction from sure premises, would serve. But this approach, so successful in geometry and physics, had never (Hobbes urged) been applied outside those fields. Geometry is "the onely Science that it hath pleased God hitherto to bestow on mankind"[11]. Thus he did not blush to claim his own application of its method as historic. He saw in his work a parallel to the scientific breakthroughs that loomed so dramatically in his time. Astronomy, he declared, had matured only with Copernicus, biology with William Harvey (the discoverer of the circulation of the blood), physics with Galileo, political science with—himself.[12]

But how does the path to true knowledge work in practice? Hobbes insisted on the vital preliminary role of the precise definition of terms. All fruitful reasoning, he urged, so begins.[13] Proper definitions evoke "perfect and clear ideas of... things" (this of course echoes Descartes); they allow a precision of discourse that banishes the verbal ambiguity and muddle, the "snare of words," for which Hobbes felt a lifelong contempt and horror.[14] When definitions have been made rigorous, inference (said Hobbes) may proceed in a manner similar to addition in arithmetic. The "sum", so to say, of terms may be another term ("body" plus "animate" plus "rational" equals "man"); or the "sum" of two terms may be an affirmation ("man is a living creature"), and then the "sum" of—the logical inference from—two such affirmations is a third ("every man is a living creature" plus "every living creature is a body" equals "every man is a body"), and the "sum" of many such inferences is a demonstration.[15] Thus Hobbes' theory of demonstration is essentially the theory of the syllogism, codified by Aristotle long before; but—perhaps because he despised Aristotle, on other grounds —he advised beginners to study the method not in the treatises of logicians but in its actual use by mathematicians, just as "little children learn to go, not by precepts, but by exercising their feet."[16]

Now any conclusion reached by this process may seem merely conditional—a mere statement that C holds *if* A and B are true. Indeed Hobbes himself so asserted. Scientific knowledge, he said, has just this conditional, this "if... then" character, in contrast to the "absolute" knowledge that our senses provide. And yet—crucially—the conclusions of scientific reasoning are (Hobbes repeatedly insisted) "eternal" and "immutable" truths.[17] For just as the axioms of Euclid's *Elements* were, to the Greeks, not mere assumptions but self-evident statements of physical fact, so for Hobbes all reasoning is anchored ultimately in the sure testimony of the senses. Right reasoning is, precisely, that which proceeds "from principles that are found indubitable by experience." Such principles are (or can be made) self-evident and can win the agreement of any man "that will but examine his own mind."[18] If only, he urged, elemental concepts and principles were made sufficiently clear, all men would see them alike; and this is true because in fact these fundamentals are already present, however obscurely, in all men's (essentially similar) minds. Like Socrates—whose famous illustration of the technique was, significantly, geometrical—Hobbes claimed only to evoke in his hearers what they already knew.[19]

Hence the possibility of a rational foundation for political science. But if all must spring from precise definitions and sure axioms, from the unambiguous understanding and use of basic concepts, how were these vital starting points to be attained? Here Hobbes placed himself in the mainstream of a methodological tradition that had figured crucially in Western thought since ancient times. It had two aspects—one a cardinal technique of Greek mathematicians, the other a similar strategy in philosophy and science. In the mathematical development that culminated in Euclid, the method of "analysis" assumed the truth of a conjectured theorem, or the achievement of a desired construction, and tried to argue "backward" from this starting point to (respectively) a theorem already proved or a construction already effected, perhaps ultimately to definitions and axioms. Sometimes, of course, the chain of inference ended in a contradiction, which exposed the original assumption as untenable. But when, more commonly, the argument reached some foundation already established or assumed, then a reversal of its steps, called "synthesis", supplied for the desired theorem or construction a rigorous proof. Analysis was thus a tool for *discovering* proofs of things already suspected (or at any rate hoped for). The "synthetic" half of the double procedure came to serve as the classic form for *expounding* a unified system of mathematics from first principles, the most familiar example being of course the *Elements*. The definitive *discussion* of the whole technique was given by Pappus in the fourth century A.D.[20] Meanwhile a philosophical methodology with strong ties to the mathematicians' analysis and synthesis had been articulated by one of the giants of Greek thought. Plausibly, though not certainly, taking his cue from the geometers, Aristotle saw an analogous double procedure as the path to all truly scientific knowledge. Such knowledge, he declared, is of *causes*: we attain it only when we give rigorous demonstrations of the necessary connections between natural occurrences and the hidden principles that engender them. Our immediate experience is of a bewildering mass of sense impressions; these we must dissect and analyze in thought, seeking to reduce complex phenomena to their simple constituents, to see the universal and essential in the particular and accidental, finally to identify the elementary factors which produce and explain our observations. This movement of thought corresponded, for Aristotle and for the tradition that he here founded, to the passage in geometrical analysis from the uncertainly conjectured to the definitely known. And similarly the geometers' synthetic proof, which reversed the steps of analysis and deduced theorems from axioms, was paralleled in Aristotelian science by rigorous (syllogistic) demonstrations of necessary causal connections between simple theoretical principles and the complex effects that we experience.[21] The influence of these ideas was enormous. Analysis and synthesis in this Aristotelian sense—in Latin translation as *resolutio* and *compositio*, respectively—entered a hundred medieval and Renaissance tracts on philosophical and scientific method. The legacy was especially strong at the University of Padua,[22] and here Harvey studied and Galileo taught, both of them seminal influences on Hobbes. Galileo, and after him Newton, would make history by abandoning the Aristotelian search for causes; for them analysis aimed rather at establishing experimentally, and expressing mathematically, fundamental relations among the variables involved in a physical phenomenon, and casting these as axioms for further deduction—Galileo's $s \propto t^2$ for free fall plays just this role in his thought. But otherwise these leaders of the Scientific Revolution preserved much of the logic, and the psychological basis, of the durable Aristotelian scheme.[23]

For his part Hobbes laid it down that all proper philosophy must use "resolution" or "composition" or a mixture of the two.[24] He applied the method himself at several different levels of inquiry. The search for clear definitions became an "analysis" of familiar, specific terms (e.g., "gold") into constituent concepts of greater universality, like "heavy," "visible," "solid," which might be further dissected in their turn. The process was conceived as a passage from things known to our senses to things apprehended by reason, from things "more known to us" to things "more known to nature"—distinctions voiced already by Aristotle and repeated constantly in the ensuing centuries.[25] In physics, according to Hobbes, "analysis" lays bare the respective contributions of individual factors, while "synthesis" explains how these join in the total effect. He offered, as an example, an account of our perception of light. Here (he said) we may isolate, as constituent elements, the motions of particles in some luminous object, the transmission of those motions through an intervening medium, finally the reception of the motions by the "fitting disposition" of our bodily organs; these, taken together, are then seen to form a necessary and sufficient explanation.[26]

But Hobbes' most ambitious application of this ancient twofold path to knowledge amounted essentially to the program of his whole life's work. For resolution and composition, he declared, were the keys to his ultimate goal, the elucidation of fundamental principles of political organization and justice. The affairs of men in society he would reduce, by "analysis", to the passions of individual people—and these in turn to the motions of those individuals' physical particles.[27] In the first half of this program he again took a clue from the mathematicians and physicists. For just as Euclid's points and lines are abstractions, which physical incarnation on paper or blackboard cannot really represent, and as Galileo idealized the motions of objects by neglecting factors like friction and air resistance, so Hobbes sought to analyze human behaviour in a *hypothetical* early collection of unorganized individuals. (The famous description of primitive life as "solitary, poore, nasty, brutish and short"[28] refers to this imagined scenario, not to history as actually recorded.) The second half of his grand design, the reduction of human mental states to the motions of material particles, may seem drastic (or even preposterous, according to taste); but in fact it reflects one of Hobbes' deepest convictions, that in such motions is the ultimate explanation of all things. And here again our philosopher merely shared a leading preoccupation of his age. The ancient doctrine of atomism, that pictured matter as composed of invisible, indivisible particles moving in a void, had been revived and modernized by various Renaissance thinkers (notably Pierre Gassendi, 1592–1655). Mathematically treatable as point masses, these corpuscles offered a fruitful conceptual foundation for the Scientific Revolution—and to some extremists a sufficient basis for all philosophy. To Hobbes their motion served to account for all desire, will, love and hatred, all psychological and spiritual phenomena of whatever sort; he shocked his readers by refusing to conceive even God as wholly immaterial.[29] Rigorous mathematical elaboration of the corpuscular philosophy awaited Newton's *Principia* (1687); it was of course far beyond Hobbes' powers. Yet he had no doubt how, in theory, the description of those moving atoms should proceed: by application of geometry. "Nature worketh by Motion; the Wayes, and Degrees whereof cannot be known, without the knowledge of the Proportions and Properties of Lines, and Figures."[30] Through such knowledge, then, we understand the motions of material particles, which cause and explain individuals' thoughts and actions, and these in turn are the source of political attitudes and institutions. So—to summarize—geometry is central to the great philosopher's thought in two quite distinct ways: as methodological guide and example, and as the most basic of all branches of knowledge, from which "synthesis" might deduce, step by step, the immutable laws of social justice.

Now it is well known that Euclid's geometry, so important as a model for Hobbes, is wholly synthetic—a steady deductive passage from simple to complex, from supposedly evident definitions and postulates to deep theorems and difficult constructions. Of the preliminary "analysis," the possibly laborious elucidation of the basic concepts, the discoveries of the proofs and the constructions, no hint remains; the dust of the workshop has been cleared away. And so with Hobbes. He too proposed to expound his ideas only in the "synthetical" order of their demonstration from "primary" propositions "manifest of themselves," omitting the earlier discovery of those propositions from "the sense of things". He set out a program of sweeping grandeur, that would pass from "universal definitions" to a geometrical account of simple motion, then to the "internal passions" of men, finally to "civil philosophy; which takes up the last place".[31] He planned treatises expounding each of these levels in turn, the higher to depend and build on the lower as the later books of Euclid on the earlier.[32]

Such, at least, was the theory and the intent. But external events deflected the tidy execution of the scheme. The factional conflicts around him provoked Hobbes to a political statement (*De cive*, 1642) that preceded any exposition of its supposed underpinnings in physics; "what was last in order", he admitted, "is yet come forth first in time".[33] But in fact, he now conceded, political science can stand alone, "its own principles sufficiently known by experience." Those who "have not learned the first part of philosophy, namely, *geometry* and *physics*, may, notwithstanding, attain the principles of civil philosophy, by the *analytical method*"—and he went on to give an example of an axiom thus reachable, that the appetites and passions of men, if unchecked, make constant war inevitable.[34] Moreover this suddenly granted independence of political theory from physics took elsewhere another, and more surprising, twist. Hobbes claimed that civil philosophy may attain the kind of *certainty* that mathematics enjoys—whereas physics may not. For, he explained, true knowledge is of the causes of things, and in studying nature we can say only that our conjectured causes may—never that they must—produce the effects that we observe; but in geometry we understand fully the cause, the "generation," of (say) a circle, because we draw it by a known procedure, and likewise we can grasp the laws and principles of civil society precisely because it is we who frame them.[35] Thus it seems that for Hobbes mathematics and political science are in principle equally accessible to the understanding; he seems never to consider that the complexities and perversities of human beings may make them less scrutable than points and lines. If (as he repeatedly insisted[36]), advances toward a rational politics had in his time lagged scandalously behind the spectacular progress of geometry, the cause was not intellectual but moral. The study of men and states is clouded by passions which geometry does not arouse:

> The doctrine of Right and Wrong, is perpetually disputed, both by the Pen and the Sword: Whereas the doctrine of Lines, and Figures, is not so; because men care not, in that subject what be truth, as a thing that crosses no mans ambition, profit, or lust. For I doubt not, but if it had been a thing contrary to any mans right of dominion, or to the interest of men that have dominion, *That the three Angles of a Triangle should be equall to two Angles of a Square*; that doctrine should have been, if not disputed, yet by the burning of all books of Geometry, suppressed, as farre as he whom it concerned was able.[37]

From this perspective one goal of a politics based on right reason was the finding of first principles so secure, and hence so worthy of trust, that no such self-interest would seek to "displace" them.[38]

We need not follow his philosophy into the elaborations that make his greatest book a leviathan in volume as in name. Sometimes from his materialistic psychology, more often from observation of the world around him, he drew the "axioms" on which all would be made to rest: that all men are moved by "appetites and aversions", that all seek power continually, that every man's power resists and hinders others'. . . . From these in turn came his specific proposals for civil order—for example a passionate preference for monarchy to democracy, a strict limiting of property rights, an unequivocal subordination of church to state. Many of his conclusions met bitter hostility in his time, and can still raise eyebrows in ours—like his call for a sovereign authority of (many would say) disturbingly absolute powers.* But his method, as opposed to his particular doctrines, won some contemporary praise. He recorded with pardonable pride that a short summary of his *De cive*, published in France, bore the title *Ethics Demonstrated*; the translator, a certain François Bonneau, assured Louis XIV that the *only* two "demonstrative sciences" were this work of Hobbes and Euclid's *Elements*.[40] More strikingly, most of Hobbes' fiercest opponents adopted his rational methods even as they assailed his ideas.[41] Partly (not, of course, wholly) from Hobbes' example, the goals of deductive argument and logical exactness spread to every corner of Europe's mental life, so bringing the "Age of Reason" to birth.

Viewed in the way that most concerns us here, as an imitation of mathematical and scientific forerunners, Hobbes' system has weaknesses that leap readily enough to the eye. For example his exclusive emphasis on syllogistic arguments was misguided, as a wider knowledge of his models could have told him; already in antiquity the Stoic logicians had seen that these Aristotelian patterns do not fully mirror mathematical reasoning, and Galileo felt deeply their inadequacy for the new experimental science. But the problems with Hobbes' edifice begin at the very foundations, with the postulates on which all rests. Like Euclid he has been detected using undeclared assumptions.[42] More importantly, the axioms which do appear explicitly are, as we have seen, just the kind of inductive generalizations from experience that the great philosopher had expressly sought to discredit. Inevitably they betray the historical context, and the personal observations, of the man who framed them, so that Hobbes, like some other theorists (Freud is the classic case in psychology), can be suspected of mistaking limited local perceptions for global truths; indeed some of his postulates, seen from larger perspectives, are demonstrably false.[43] In short his assumptions are not remotely so self-evident as he fondly supposed, and his trust in their universal acceptance seems a gross delusion. Some experience of the cut and thrust and compromise of practical politics might have tempered his sureness; so argued his contemporary the Earl of Clarendon, sometime Chancellor to Charles II, who suggested urbanely that a stint in Parliament or the courts probably would have shown Hobbes that "his solitary cogitations, how deep soever, and his too peremptory adhering to some Philosophical Notions, and even Rules of Geometry, have misled him in the investigation of Policy."[44] Altogether, it is temptingly easy to dismiss Hobbes' whole endeavour as hopelessly naive, this dream that reasonable men could be brought to see human affairs with the clarity and consensus of geometers poring over a proof. But it is fairer, and in historical terms more correct, to see his system as an early expression of the Enlightenment belief that the extension of scientific methods and ideals to social problems would foster understanding, tolerance and harmony at the expense of error, prejudice and strife. In the days of *Leviathan*, when the Scientific Revolution itself was still so new, no one could yet say how valid that belief might prove. Mathematicians should honor Thomas Hobbes, who came late in life to their subject, and understood it poorly, but loved it much, and staked on the supposed sureness of its methods his hopes for the peace and good government of mankind.

*"A geometer's panacea for peace," says a (sympathetic) biographer.[39]

Notes

References to Hobbes' writings give volume and page numbers of *The English Works of Thomas Hobbes*, ed. W. Molesworth. For convenience I give also, where appropriate, page numbers of the widely available Penguin edition of *Leviathan*, ed. C. B. Macpherson (abbreviated as "Lev").

1. Cf. Macpherson, op. cit., p. 15; Richard Peters, *Hobbes* (Penguin, 1956), pp. 26, 28, 47.
2. Quoted in Peters, p. 27.
3. I, viii.
4. John Aubrey, *Brief Lives*, ed. O. L. Dick (Ann Arbor, 1962), p. 158.
5. Ibid., p. 150; italics in original.
6. I, 26, 96, 141.
7. VII, 248; cf. VII, 316.
8. VII, 3; I, 288. A good account of Hobbes' ventures into technical mathematics can be found in J. F. Scott, *The Mathematical Work of John Wallis* (London, 1938).
9. IV, 18.
10. II, iii; III, 379 = Lev. p. 426.
11. III, 23-4 = Lev. p. 105.
12. I, viii–ix.
13. III, 380 = Lev. p. 428.
14. I, 81; I, viii; III, 36 = Lev. p. 116; IV, 23, etc.
15. I, 44ff; cf. III, 29–30 = Lev. p. 110.
16. I, 54–5.
17. III, 52 = Lev. p. 131; III, 71 = Lev. pp. 147–8; II, 44–6; VI, 122; III, 664 = Lev. p. 682.
18. IV, 24; I, 84; I, 74.
19. II, xiv, 295; IV, 212; I, 304; IV, 1–2, 27; cf. Plato, *Meno* 8lff.
20. The best account of this method and its history is J. Hintikka and U. Remes, *The method of Analysis: Its Geometrical Origin and its General Significance* (D. Reidel, 1974); see pp. 8–10 for the passage from Pappus. (These authors, however, do not mention Hobbes.)
21. Aristotle, *Posterior Analytics*, *passim*; cf. *Physics* 184a17–20 and *Nichomachean Ethics* 1112b20 ff.
22. See the classic paper by J. H. Randall, "The Development of Scientific Method in the School of Padua", in his *The School of Padua and the Emergence of Modern Science* (Padua, 1961), pp. 13–68.
23. Galileo, *Dialogue Concerning the Two Chief World Systems*, tr. S. Drake (Berkeley and Los Angeles, 1970), p. 51; Newton, *Opticks*, Q. 30 (Dover ed. pp. 404–5); I. B. Cohen, *Introduction to Newton's 'Principia'* (Cambridge, Mass., 1971), pp. 294–5.
24. I, 66, 309–10.
25. I, 69, 67; cf. Aristotle, *Physics* 184a17–20.
26. I, 75–9.
27. See the notice "To the Reader", by one "F. B.", at the beginning of Vol. IV of the *English Works*.
28. III, 113 = Lev. p. 186.
29. III, 672 = Lev. p. 689.
30. III, 669 = Lev. p. 686.
31. I, 80–1, 87.
32. "To the Reader" (see n. 27 above).
33. II, xx.
34. I, 74; italics in original.
35. VII, 184.
36. II, iv; IV, ep. ded.
37. III, 91 = Lev. p. 166; cf. IV, ep. ded.
38. IV, ep. ded.
39. Peters, p. 33.
40. VII, 333; A. Rogow, *Thomas Hobbes* (New York and London, 1986), p. 145.
41. S. I. Mintz, *The Hunting of Leviathan* (Cambridge, 1970), pp. viii, 83, 149–51.
42. E.g., Macpherson, p. 30.
43. J. W. N. Watkins, *Hobbes' System of Ideas* (London, 1965), pp. 166ff, draws on Sir James Frazer's famous *The Golden Bough* to rebut Hobbes' axiom that all men fear and shun violent death.
44. Quoted in J. Bowle, *Hobbes and his Critics* (London, 1969), p. 152.

NOTES

Indeterminate Forms Revisited

R. P. BOAS
3540 NE 147th Street
Seattle, WA 98155

1. Introduction You must all have seen at least one calculus textbook. It may surprise some of you that three centuries ago no such book existed: the very first book that was in any sense a calculus text was published, anonymously, in 1696, under the rather forbidding title *Analysis of the Infinitely Small* [**5**]. It was well known in European mathematical circles that the author was a French marquis, Guillaume de L'Hôpital. (I give him the modern French spelling, which at least keeps students from pronouncing the silent s in L'Hospital.) The book was hardly easy reading. It began with propositions like this: "One can substitute, one for the other, two quantities which differ only by an infinitely small quantity; or (what amounts to the same thing) a quantity that is increased or decreased only by another quantity infinitely less than it, can be considered as remaining the same." This sort of presentation gave calculus a reputation, which has survived to modern times, of being unintelligible.

Sylvester, writing in about 1880 [**10**, vol. 2, pp. 716-17], said that when he was young (around 1830) "a boy of sixteen or seventeen who knew his infinitesimal calculus would have been almost pointed out in the streets as a prodigy like Dante, who had seen hell." (Here and now, we would very likely find students of the same age feeling much the same; but Sylvester, in 1870, was teaching students who dealt casually with topics that we would now describe as advanced calculus.) When I was in high school (somewhat later), calculus was thought of, by otherwise well-educated people, as being as deep and mysterious as (say) general relativity is thought of today. My parents knew somebody who was reputed to know calculus, but they had no idea what that was (and they were college teachers—of English). Nowadays there are perhaps too many calculus books, but some of the answers that students give to examination questions make me wonder whether the subject has even now become sufficiently intelligible.

In his own time, and for long afterwards, L'Hôpital had an impressive reputation. Today he is remembered only for "L'Hôpital's rule," which evaluates limits like

$$\lim_{x \to 1} \frac{(2x - x^4)^{1/2} - x^{1/3}}{1 - x^{3/4}}$$

(L'Hôpital's own example) by replacing the numerator and denominator by their derivatives and hoping for the best.

L'Hôpital's rule seems to have fallen somewhat out of favor; I have heard it claimed that all it is useful for is as an exercise in differentiation.

It has been known for some time that many of L'Hôpital's results, including the rule, were purchased (quite literally) from John (= Jean = Johann) Bernoulli. Immediately after L'Hôpital's death in 1704, Bernoulli published an article claiming that he had communicated the rule for 0/0 to L'Hôpital, along with other material, before L'Hôpital had published it. This claim was disbelieved for some two hundred years; sceptics wondered why Bernoulli had not advanced his claim earlier. The reason for the delay eventually became clear when Bernoulli's correspondence with L'Hôpital came to light in the early 1900s. Bernoulli gave the rule to L'Hôpital only after L'Hôpital had promised to pay for it, had repeatedly asked for it, and had finally come across with the first installment. We now also know that there are records of Bernoulli's having lectured on the rule before L'Hôpital's book was published.

In the preface to his book, L'Hôpital says, "I acknowledge my debt to the insights of MM Bernoulli, above all to those of the younger [John], now Professor at Groningen. I have unceremoniously made use of their discoveries and of those of M Leibnis [sic]. Consequently I invite them to claim whatever they wish, and will be satisfied with whatever they may leave me." Considering what we now know, this seems somewhat disingenuous, especially since L'Hôpital was clearly unable to discover for himself how to prove the rule of which, as Plancherel once said of his own theorem, he had "the honor of bearing the name."

You can find the whole story in the 1955 volume of Bernoulli's correspondence [7], or in Truesdell's review of the volume [11].

I used to wonder, from time to time, what kind of proof L'Hôpital had used, but never when I was where I could find a copy of his book. Recently I happened to mention this question to Professor Alexanderson—who promptly produced his own copy. Professor Underwood Dudley, who is more resourceful than I am, also found a copy, and has translated it into modern terminology [4], but retaining its geometric character (L'Hôpital thought of functions as curves). L'Hôpital actually considered only $\lim_{x\to a} f(x)/g(x)$, where a is finite, $f(a)=g(a)=0$, and both $f'(a)$ and $g'(a)$ exist, are finite, and not zero. In analytical language, what L'Hôpital did amounts to writing

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f(x)-f(a)}{g(x)-g(a)}=\lim_{x\to a}\frac{f'(a)+\varepsilon(x)}{g'(a)+\delta(x)}(\varepsilon\to 0,\delta\to 0)=\frac{f'(a)}{g'(a)}.$$

It is not trivial to extend such a proof to the cases when $f'(a)$ and $g'(a)$ do not exist (but have limits as $x\to a$), or are both zero, or $f(a)=g(a)=\infty$, or $a=\infty$. I do not know when or by whom these refinements were added, but the complete theory was in place by 1880 [8, 9].

2. A common modern proof If you saw a proof of L'Hôpital's rule in a modern calculus class, the probability is about 90% that it is Cauchy's proof. This proof appeals to mathematicians because it is elegant, but often fails to appeal to students because it is subtle. It depends on knowing Cauchy's refinement of the mean-value theorem, namely that (with appropriate hypotheses)

$$\frac{f(x)-f(a)}{g(x)-g(a)}=\frac{f'(c)}{g'(c)},\quad c \text{ between } a \text{ and } b. \tag{1}$$

Given this, L'Hôpital's rule becomes obvious.

In spite of its elegance, Cauchy's proof seems to me to be inappropriate for an elementary class. Any proof that begins with a lemma like Cauchy's mean value theorem, that says "Let us consider...," repels most students. Students are also uncomfortable with the nebulous point c. They want to know where it is, and feel that the instructor is deliberately keeping them in the dark. Of course, the exact location of c is completely irrelevant (although numerous papers have been written about it).

3. A caution Cauchy's proof tacitly assumes that there is a (one-sided) neighborhood of the point a in which $g'(x) \neq 0$. Strictly speaking, if there is no such neighborhood, the limit in (1) is not defined, and we would have no business talking about it. However, if f' and g' are given by explicit formulas, they may happen to share a common factor that is zero at a, and the temptation to cancel this factor is irresistible. One can obtain a spurious result in this way [**8**, **9**; **3**, p. 124, ex. 24; **1**].

Let me give you a specific example, just to emphasize that there is a reason for the requirement that $g'(x) \neq 0$. Let $f(x) = 2x + \sin 2x$, $g(x) = x \sin x + \cos x; a = +\infty$. Then $f'(x) = 4\cos^2 x$, $g'(x) = x \cos x$, and $f'(x)/g'(x) \to 0$, whereas $f(x)/g(x)$ does not approach a limit. The trouble comes from cancelling a factor that changes sign in every neighborhood of the point a; it would have been legitimate to cancel a quadratic factor.

Some writers think that the difficulty arises only in artificial cases that would never occur in practice. But then, what happens to our claim to be giving correct proofs?

You might not guess from Cauchy's proof that there is a discrete analog of L'Hôpital's rule; see, for example, [**6**]. This was known to Stolz in the 1890's, and has often been rediscovered.

4. A more satisfactory proof I want now to show you a proof of L'Hôpital's rule that avoids the difficulties of Cauchy's and establishes a good deal more. It may seem more complicated, but not if you include a proof of Cauchy's mean value theorem as part of Cauchy's proof. This proof is also quite old; Stolz knew it, but preferred Cauchy's proof, perhaps because of Cauchy's reputation. It has been published several times by people (including me) who failed to search the literature.

Let us suppose that $f(x)$ and $g(x)$ approach 0 as $x \to a$ from the left, where a might be $+\infty$; it does no harm to define (if necessary) $f(a) = g(a) = 0$. We may suppose that $g'(x) > 0$ (otherwise consider $-g(x)$). Now let $f'(x)/g'(x) \to L$, where $0 < L < \infty$. Then, given $\varepsilon > 0$, we have, if x is sufficiently near a, and a is finite,

$$L - \varepsilon \leqslant \frac{f'(x)}{g'(x)} \leqslant L + \varepsilon,$$

$$(L-\varepsilon)g'(x) \leqslant f'(x) \leqslant (L+\varepsilon)g'(x) \quad (\text{since } g'(x) > 0).$$

Integrate on (x, a) to get

$$-(L-\varepsilon)g(x) \leqslant -f(x) \leqslant -(L+\varepsilon)g(x)$$

(notice that since g increases to 0, we have $g(x) < 0$). Since $-g$ and $-f$ are positive near a,

$$L - \varepsilon \leqslant \frac{f(x)}{g(x)} \leqslant L + \varepsilon,$$

$$\lim_{x \to a} \frac{f(x)}{g(x)} = L.$$

Only formal changes are needed if $a = +\infty$ or if $L = 0$ or ∞.

For the ∞/∞ case, we get, in the same way, with $\delta > 0$,

$$L - \varepsilon < \frac{f(a-\delta) - f(x)}{g(a-\delta) - g(x)} < L + \varepsilon,$$

$$L - \varepsilon < \frac{\dfrac{f(a-\delta)}{f(x)} - 1}{\dfrac{g(a-\delta)}{g(x)} - 1} \cdot \frac{f(x)}{g(x)} < L + \varepsilon.$$

Letting $x \to a$, we obtain

$$L - \varepsilon \leqslant \liminf_{x \to a} \frac{f(x)}{g(x)} \leqslant \limsup_{x \to a} \frac{f(x)}{g(x)} \leqslant L + \varepsilon.$$

Letting $\varepsilon \to 0$, we obtain $f(x)/g(x) \to L$.

If it happens that $f'(a) = g'(a) = 0$, one repeats the procedure with f'/g', and so on. If $f^{(n)}(a) = g^{(n)}(a) = 0$ for every n (which can happen with f and g not identically zero), the procedure fails. Otherwise, the limit can be handled more simply in a single step, as we shall see below.

5. Generalizations If f and g are defined only on the positive integers, we can reason in a similar way with differences instead of derivatives to conclude that if the differences of g are positive, and $f(n)$ and $g(n)$ approach zero as $n \to \infty$, then if

$$\frac{f(n) - f(n-1)}{g(n) - g(n-1)} \to L \qquad \text{as } n \to \infty,$$

it follows that $f(n)/g(n) \to L$. This is sometimes called Cesàro's rule. For more detail, and illustrations, see [6]. A possibly more familiar version is as follows:

If $a_n \to 0$ and $b_n \to 0$, and $a_n/b_n \to L$, then

$$\frac{\sum_{k=1}^{n} a_k}{\sum_{k=1}^{n} b_k} \to L.$$

The key point in the proof of L'Hôpital's rule is the principle that the integral of a nonnegative function ($\not\equiv 0$) is positive. More precisely, if $f(x) \geqslant 0$ on $p \leqslant x \leqslant q$ then

$$\int_p^q f(t)\,dt > 0 \qquad \text{if } p < x < q \text{ and } f(t) \not\equiv 0.$$

Repeated integration with the same lower limit has the same property, as we see by rewriting the n-fold iterated integral as a single integral:

$$\frac{1}{(n-1)}!\int_p^x (t-p)^{n-1} f(t)\,dt.$$

This suggests the appropriate treatment of the case of L'Hôpital's rule when $f'(a) = g'(a) = 0$, or more generally when $f^{(k)}(a) = g^{(k)}(a) = 0, k = 1, 2, \ldots, n-1$, but not both of $f^{(n)}(a)$ and $g^{(n)}(a)$ are 0. The positivity of iterated integration then yields the conclusion of L'Hôpital's rule in a single step.

An operator that carries positive functions to positive functions is conventionally called a positive operator. If F is an invertible operator whose inverse is positive, we can conclude that if $F[f(x)]/F[g(x)] \to L$ and $F[g] > 0$, then $f(x)/g(x) \to L$.

As an example of the use of operators, consider $D + P(x)I$, where $D = d/dx$ and I is the identity operator. This is the operator that occurs in the theory of the linear first-order differential equation $y' + P(x)y = Q(x)$. The solution of this differential equation, with $y(a) = 0$, is

$$y = \exp\left\{\int_a^x P(t)\,dt\right\} \int_a^x Q(t)\exp\left\{\int_a^t P(u)\,du\right\} dt. \qquad (2)$$

In other words, (2) provides the inverse Λ of $D + P(x)I$.

The explicit formula shows that if $Q(x) \geqslant 0$ we have $\Lambda[Q] > 0$, so that Λ is a positive operator. Hence we may conclude that if

$$\frac{[D+PI]f}{[D+PI]g} \to L$$

and $[D + PI]g > 0$ then

$$(L-\varepsilon)[D+PI]g < [D+PI]f < (L+\varepsilon)[D+PI]g.$$

A positive linear operator evidently preserves inequalities. Consequently, if we apply Λ to both sides, we obtain

$$(L-\varepsilon)g(x) < f(x) < (L+\varepsilon)g(x)$$

and hence

$$f(x)/g(x) \to L.$$

Thus $D + P(x)I$ can play the same role as D in L'Hôpital's rule. It is at least possible that $D + P(x)I$ might be simpler than D.

Since some forms of fractional integrals and derivatives are defined by positive operators, one could also formulate a fractional L'Hôpital's rule.

This article is the text of an invited address to a joint session of the American Mathematical Society and the Mathematical Association of America, January 14, 1989.

REFERENCES

1. R. P. Boas, Counterexamples to L'Hôpital's rule, *Amer. Math. Monthly* 94 (1986), 644–645.
2. ______, *Indeterminate Forms Revisited*, videotape, Amer. Math. Soc., Providence, RI, 1989.
3. R. C. Buck and E. F. Buck, *Advanced calculus*, 3rd ed., McGraw-Hill, New York, etc., 1978.
4. U. Dudley, Review of Calculus with Analytic Geometry, *Amer. Math. Monthly* 95 (1988), 888–892.
5. [G. de L'Hôpital], *Analyse des infiniment petits, pour l'intelligence des lignes courbes*, Paris, 1696; later editions under the name of le Marquis de L'Hôpital.
6. Xun-Cheng Hwang, A discrete L'Hôpital's rule, *College Math. J.* 19 (1988), 321–329.
7. [O. Spiess, ed.], *Die Briefwechsel von Johann Bernoulli*, vol. 1, Birkhäuser, Basel, 1955.
8. O. Stolz, Über die Grenzwerthe der Quotienten, *Math. Ann.* 15 (1889), 556–559.
9. ______, *Grundzüge der Differential- und Integral-rechnung*, vol. 1, Teubner, Leipzig, 1893.
10. J. J. Sylvester, *The Collected Mathematical Papers*, Cambridge University Press, 1908.
11. C. Truesdell, The new Bernoulli edition, *Isis* 49 (1958), 54–62.

Lowering the Odds for "Even Up"

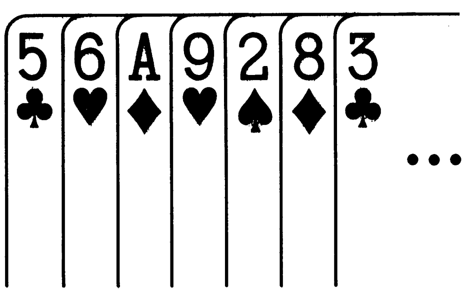
LEON HARKLEROAD
4417 Woodgate Lane
Louisville, KY 40220

What are the odds for me to win at solitaire? The answer to this, of course, depends on the particular solitaire game in question. For example, the chances of winning at Clock Solitaire have been shown [1] to be 1 in 13. In this paper we will take a look at another standard solitaire, Even Up, and examine both the role of strategy in playing the game and the odds of winning.

How to play Here are the rules of Even Up, as given in [2]:

> Discard from the deck all face cards—kings, queens and jacks.
> Deal cards one at a time in a single overlapping row. Remove and discard any two adjacent cards whose numerical total is an even number. After each discard, close the gap in the row and look to see if there is an added play at the junction. You win the game if you succeed in discarding the entire deck.

For example, say that the layout after dealing looks like this:



You can start by discarding the ace and nine, and then the two and eight. Or, after discarding the ace and nine, you can discard the six and two, which become adjacent after the removal of the first pair. Or you can begin by discarding the two and eight, and then the nine and three, and so on.

Luck or skill? In general, as the above example shows, you may have several courses of action available to you for a given layout. How should you choose your moves in order to discard all of the cards? Or does your choice of moves matter at all? In other words, is the outcome of the game already determined by the deal, regardless of your discarding strategy?

To help answer this, first notice that adding the values of adjacent cards is unnecessary. The sum will be even exactly when the two cards share the same parity —either both odd or both even. So you could just as well replace all the aces, threes, fives, sevens, and nines with cards marked a and all the rest with cards marked b. You would then play by removing pairs of adjacent cards that are marked with the same letter. Again, if you get to discard all of the cards, you win; otherwise, at the end of play you will have a pattern of alternating a's and b's.

For a given starting layout of a's and b's, the pattern left at the finish will *not* depend on how you choose your moves. Whatever your strategy, the starting layout will determine the string of a's and b's that you end up with.

Why is this? The explanation can be given in terms of group theory. We may think of all the possible strings of a's and b's as the elements of a group. To multiply two strings, simply concatenate them—place them end to end. The group's identity element will be the empty string. Saying that two consecutive a's may be removed is thus the same as saying that aa equals the identity, so a is its own inverse. Likewise, b will serve as its own inverse, and therefore the inverse of a string of a's and b's will just be that string's reversal.

In the terminology generally used, what we have is the group generated by a and b subject to the relations $a^2 = e$, $b^2 = e$. In this group every string equals *one and only one* of the following:

$$\text{the empty string}, a, b, ab, ba, aba, bab, \ldots \qquad (*)$$

(See, for example, Theorem 1.4 of [3].) So for a given starting string, regardless of the sequence of moves, there can be only one final outcome, namely, the string in $(*)$ equal to the starting string. This means that luck, not skill, totally determines what happens in a game of Even Up: if the original layout of a's and b's equals the empty string, you will win; otherwise, you will lose.

So what are the odds? According to [2], your chances of winning Even Up are approximately 1 in 3. But the estimated odds in [2] need to be taken with a grain of salt—the chances of winning Clock Solitaire are pegged at 1 in 100 there! What we need to find is the proportion of strings consisting of 20 a's and 20 b's that equal the empty string in the group described above. The total number of strings with 20 a's and 20 b's is $\binom{40}{20}$. Computing how many of those strings can be reduced to the empty string takes a little more work. We will look at both a recurrence relation and a direct counting argument that yield the answer.

It will be useful to consider a general deck consisting of x a's and y b's, where $x + y$ is even. With the total number of cards even, since you discard the cards in pairs, an even number of cards will remain at the end of the game. So the final string will be empty or look like $abab \ldots ab$ or $baba \ldots ba$. These three cases may all be represented in the group by $(ab)^n$, with n being zero, positive, or negative, respectively (remember that $ba = (ab)^{-1}$.)

In general, then, we may ask: how many strings of x a's and y b's equal $(ab)^z$ in the group? Let $f(x, y, z)$ denote the answer to this question. For example, the number of winning deals in Even Up is $f(20, 20, 0)$ (and so $f(20, 20, 0)/\binom{40}{20}$ represents the probability of winning.)

The strings of x a's and y b's that equal $(ab)^z$ come in four types: those beginning with aa, ab, ba, and bb. If the string begins with aa, then the remaining $x - 2$ a's and y b's must form another string equaling $(ab)^z$, as aa equals the identity. Similarly, if the string starts with ab, the remaining $x - 1$ a's and $y - 1$ b's must form a string equal to $(ab)^{z-1}$, and so on. Thus

$$\begin{aligned} f(x, y, z) = {} & f(x-2, y, z) + f(x-1, y-1, z-1) \\ & + f(x-1, y-1, z+1) + f(x, y-2, z) \end{aligned}$$

for $x, y \geqslant 2$. This recurrence relation builds up $f(x, y, z)$ in terms of values of f for a deck with two fewer cards. When the number of a's or of b's has been reduced to zero or one, the value of f is simple to figure out. For example, $f(0,y,0)=1$, $f(0,y,z)=0$ for $z\neq 0$, and so on.

Armed with the recurrence relation and the values of f for x or $y<2$, you can calculate $f(x,y,z)$ in general. A very short Pascal program can be written to do the dirty work for you. Examining the values leads to the conjecture that

$$f(x,y,z)=\binom{\frac{x+y}{2}}{\frac{x+z}{2}}\binom{\frac{x+y}{2}}{\frac{y+z}{2}}.$$

This formula is easily verified by an inductive proof based on the recurrence relation.

Another approach Is there a direct way to generate the formula for $f(x,y,z)$, as opposed to conjecture-and-verify? Here is one such way. Think of the starting layout as broken up into pairs: the first two cards, the next two cards, etc. Those pairs which are *aa* or *bb* may be discarded, leaving you with just *ab*'s and *ba*'s. The string will equal $(ab)^z$, where z is the difference between the number of *ab*'s and the number of *ba*'s. So if there are k *ab*'s, there must be $k-z$ *ba*'s. Between them, the k *ab*'s and $k-z$ *ba*'s account for $2k-z$ of the x *a*'s. Thus the other $x+z-2k$ *a*'s must form $(1/2)(x+z)-k$ *aa* pairs; similarly, there will be $(1/2)(y+z)-k$ *bb*'s. Now these *aa*, *ab*, *ba*, and *bb* pairs can be arranged in

$$\binom{\frac{x+y}{2}}{\frac{x+z}{2}-k,k,k-z,\frac{y+z}{2}-k}$$

different ways. So, allowing for all possible choices† of k,

$$\sum_k\binom{\frac{x+y}{2}}{\frac{x+z}{2}-k,k,k-z,\frac{y+z}{2}-k}$$

different strings of x *a*'s and y *b*'s will equal $(ab)^z$. But this quantity is

$$\sum_k\frac{\left(\frac{x+y}{2}\right)!}{\left(\frac{x+z}{2}-k\right)!k!(k-z)!\left(\frac{y+z}{2}-k\right)!}$$

$$=\sum_k\frac{\left(\frac{x+y}{2}\right)!}{\left(\frac{x+z}{2}\right)!\left(\frac{y-z}{2}\right)!}\frac{\left(\frac{x+z}{2}\right)!}{\left(\frac{x+z}{2}-k\right)!k!}\frac{\left(\frac{y-z}{2}\right)!}{(k-z)!\left(\frac{y+z}{2}-k\right)!}$$

$$=\sum_k\binom{\frac{x+y}{2}}{\frac{x+z}{2}}\binom{\frac{x+z}{2}}{k}\binom{\frac{y-z}{2}}{\frac{y+z}{2}-k}$$

$$=\binom{\frac{x+y}{2}}{\frac{x+z}{2}}\sum_k\binom{\frac{x+z}{2}}{k}\binom{\frac{y-z}{2}}{\frac{y+z}{2}-k}$$

†We may take k to range from 0 to $(1/2)(x+y)$ in the summations that follow, but, in fact, the multinomial coefficients will equal 0 except when $\max(0,z)\leqslant k\leqslant(1/2)(z+\min(x,y))$.

$$= \begin{pmatrix} \frac{x+y}{2} \\ \frac{x+z}{2} \end{pmatrix} \begin{pmatrix} \frac{x+z}{2} + \frac{y-z}{2} \\ \frac{y+z}{2} \end{pmatrix} \qquad (**)$$

$$= \begin{pmatrix} \frac{x+y}{2} \\ \frac{x+z}{2} \end{pmatrix} \begin{pmatrix} \frac{x+y}{2} \\ \frac{y+z}{2} \end{pmatrix}.$$

The second factor in $(**)$ comes from the general formula

$$\sum_k \binom{m}{k}\binom{n}{r-k} = \binom{m+n}{r}$$

—to choose r objects out of $m+n$, you choose some out of the first m and the rest out of the last n.

Summing up As we have seen, your chances of winning a game of Even Up are $f(20,20,0)/\binom{40}{20}$. By the formula for $f(x,y,z)$, this equals $\binom{20}{10}\binom{20}{10}/\binom{40}{20}$, approximately 0.2476. So you have almost exactly a 1 in 4 chance at Even Up. Your odds may be a little lower than [2] claimed, but they still beat the odds for Clock Solitaire! And even if you lose, you will probably come close to winning. Your chances of having exactly four cards left at the end (the fewest you can leave in a losing deal) are over 40 percent.

This paper was written with support from the Charles A. Dana Foundation while I was at Cornell University on sabbatical leave from Bellarmine College. I would like to thank all three of the above institutions.

REFERENCES

1. David Kleiner, The last 1 (solution to Problem 1066), this MAGAZINE 53 (1980), 184–185.
2. Albert Morehead and Geoffrey Mott-Smith, *The Complete Book of Solitaire and Patience Games*, Bantam, 1966.
3. Wilhelm Magnus, Abraham Karrass, and Donald Solitar, *Combinatorial Group Theory*, Dover, 1976.

Euclid's Algorithm = Reverse Gaussian Elimination

KUNG-WEI YANG
Western Michigan University
Kalamazoo, MI 49008

It is well known that a useful consequence of Euclid's algorithm is "Bézout's identity," which says:

If the greatest common divisor of the integers a and b is d, then there exist integers x and y such that $ax+by=d$.

$$= \begin{pmatrix} \dfrac{x+y}{2} \\ \dfrac{x+z}{2} \end{pmatrix} \begin{pmatrix} \dfrac{x+z}{2} + \dfrac{y-z}{2} \\ \dfrac{y+z}{2} \end{pmatrix} \qquad (**)$$

$$= \begin{pmatrix} \dfrac{x+y}{2} \\ \dfrac{x+z}{2} \end{pmatrix} \begin{pmatrix} \dfrac{x+y}{2} \\ \dfrac{y+z}{2} \end{pmatrix}.$$

The second factor in $(**)$ comes from the general formula

$$\sum_k \binom{m}{k}\binom{n}{r-k} = \binom{m+n}{r}$$

—to choose r objects out of $m+n$, you choose some out of the first m and the rest out of the last n.

Summing up As we have seen, your chances of winning a game of Even Up are $f(20,20,0)/\binom{40}{20}$. By the formula for $f(x,y,z)$, this equals $\binom{20}{10}\binom{20}{10}/\binom{40}{20}$, approximately 0.2476. So you have almost exactly a 1 in 4 chance at Even Up. Your odds may be a little lower than [2] claimed, but they still beat the odds for Clock Solitaire! And even if you lose, you will probably come close to winning. Your chances of having exactly four cards left at the end (the fewest you can leave in a losing deal) are over 40 percent.

This paper was written with support from the Charles A. Dana Foundation while I was at Cornell University on sabbatical leave from Bellarmine College. I would like to thank all three of the above institutions.

REFERENCES

1. David Kleiner, The last 1 (solution to Problem 1066), this MAGAZINE 53 (1980), 184–185.
2. Albert Morehead and Geoffrey Mott-Smith, *The Complete Book of Solitaire and Patience Games*, Bantam, 1966.
3. Wilhelm Magnus, Abraham Karrass, and Donald Solitar, *Combinatorial Group Theory*, Dover, 1976.

Euclid's Algorithm = Reverse Gaussian Elimination

KUNG-WEI YANG
Western Michigan University
Kalamazoo, MI 49008

It is well known that a useful consequence of Euclid's algorithm is "Bézout's identity," which says:

If the greatest common divisor of the integers a and b is d, then there exist integers x and y such that $ax + by = d$.

The traditional way of computing x and y by Euclid's algorithm and successive back-substitution [**1**, p. 18] can be quite tedious. It may help to realize that if we express Euclid's algorithm as a sequence of row operations, the reversal of the familiar Gaussian elimination (division by an integer $\neq \pm 1$ is not permissible, however) has the same effect as that of successive back-substitution.

An example is worth a thousand explanations. Let's look at a solution of a linear system in (a slightly unorthodox) sequential form ($r_i =$ the ith row).

r_1	$-3x$	$+$	$5y =$	1
r_2	$-2x$	$+$	$3y =$	-2
$r_3(=-2r_2+r_1)$	x	$-$	$y =$	5
$r_4(=2r_3+r_2)$			$y =$	8
$r_5(=r_4+r_3)$	x		$=$	13

If we reverse this solution and suppress the variables for brevity, we have, for $a = 13$ and $b = 8$,

r_1	1		$=$	13
r_2		1	$=$	8
$r_3(=-2r_2+r_1)$	1	-2	$=$	-3
$r_4(=3r_3+r_2)$	3	-5	$=$	-1

We now obtain from the last row the desired result

$$3 \cdot 13 \quad + \quad -5 \cdot 8 \quad = \quad -1.$$

The strategy is to decrease the absolute value of the right-hand integer by repeated application of the division algorithm (which after all *is* Euclid's algorithm). Tactically, for every row operation, one can use on the right-hand side the remainder with the smallest absolute value, which is what we have done here, or one can use the positive remainder. The integer with the smallest absolute value, i.e., the one right before the remainder becomes 0, is then, up to a factor of 1 (± 1), the greatest common divisor.

It is worth noting that the same procedure applies to any number of integers $a_1, a_2, \dots, a_n$ simultaneously. It also applies to finding greatest common divisors in any Euclidean ring (see [**3**, p. 189] for a definition and [**2**, pp. 212–217] for more details) such as a polynomial ring or the Gaussian integers.

REFERENCES

1. G. E. Andrews, *Number Theory*, W. B. Saunders, Philadelphia, 1971.
2. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Clarendon Press, Oxford, 5th ed., 1979.
3. I. N. Herstein, *Abstract Algebra*, Macmillan Publishing Co., New York, 1986.

How Hard Can It Be to Draw a Pie Chart?

DIANE L. SOUVAINE*†
Rutgers University
New Brunswick, NJ 08903

CHRISTOPHER J. VAN WYK†
AT&T Bell Laboratories
Murray Hill, NJ 07974

Pie charts are a familiar way to display how a total is composed of or divided into parts. For example, FIGURE 1 displays the distribution of science and engineering doctorates awarded in 1984. (The source of this data is Table 1015 in Reference 7.) There is a lively debate about the merits of pie charts as a means to display data [**1**, **2**, **6**]. We are not concerned here with the accuracy of people's perceptions of pie charts, however, but rather with the difficulty of laying out a pie chart.

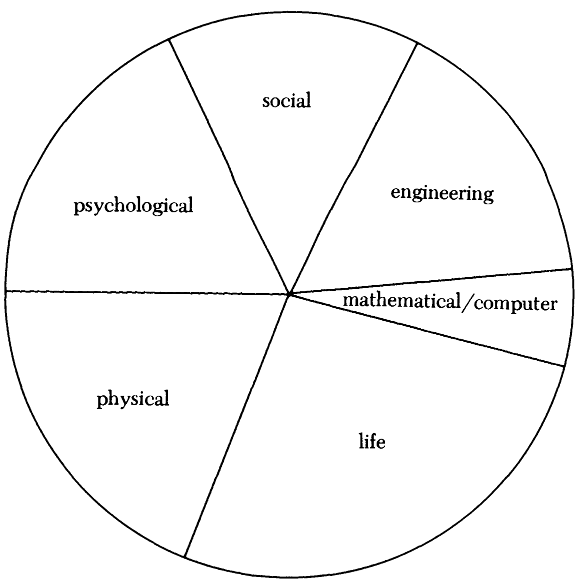


FIGURE 1
Distribution of science and engineering doctorates conferred in 1984.

The elements of a *pie-chart problem* are angle-label pairs (θ_i, Λ_i) and a radius r. To solve the problem, we must compose a circle of radius r out of sectors so that

(i) sector i subtends an angle θ_i;
(ii) for $i \neq j$, the ith sector does not overlap the jth sector; and,
(iii) for each i, Λ_i fits inside the ith sector.

We can express a solution to a pie-chart problem by reporting, for each i, the angle ϕ_i which the angle bisector of sector i makes with respect to the horizontal.

Requirements (i) and (ii) can be met simply by checking that the angles θ_i sum to 2π. It is requirement (iii), that all labels in a pie chart must lie inside the

*This author was supported in part by a Henry Rutgers Research Fellowship.
†This work was performed in the spring of 1987, when both authors were guests of the Department of Computer Science at Princeton University.

corresponding pie slices, that provides the interest in pie-chart problems. For example, in FIGURE 1, the slice for mathematical and computer science doctorates subtends a small angle of the circle, but has a long label; if the long label is going to fit into the small slice, then that slice must lie close to the horizontal. In the notation of the preceding paragraph, when θ_i is small and Λ_i is large in a pie-chart problem, then there is a very restricted range of possible values of ϕ_i in any solution.

In this paper we address two versions of the pie-chart problem. In both versions, we reduce the number of variables that must be considered by taking the radius of the pie, r, to be one. In both versions, we also consider labels Λ_i that have the particularly simple form of horizontal line segments; we use λ_i as a reminder that we are solving this simplified problem. In practice the labels will be more complicated; nevertheless, horizontal line segments provide enough challenge, as we shall see.

The first pie-chart problem we consider is called *Sectors in Fixed Order*. The input to this problem is a *sequence* $((\theta_i, \lambda_i)|1 \leqslant i \leqslant n)$, together with the requirement that the pie slices must appear in order by i around the pie. To solve the problem, we can imagine rotating the entire pie until all labels fit into their sectors. The actual execution of this rotation involves some interesting mathematics.

The second pie-chart problem we consider is called *Unspecified Sector Order*. The input to this problem is a set $\{(\theta_i, \lambda_i)|1 \leqslant i \leqslant n\}$. In a solution to this problem, the sectors can appear in any order around the pie. This problem turns out to be much harder to solve: we can show that it is at least as difficult as any of a well-known class of problems that are thought to be intractable.

Sectors in Fixed Order

As a circle sector rotates, the length of the longest horizontal line segment that fits inside it changes. Thus, our first step in solving the pie-chart problem given sectors in fixed order is to characterize the change in maximum horizontal extent of a rotating sector. Consider a sector of the unit circle that subtends an angle θ and whose bisector lies at angle ϕ with respect to the positive x-axis. Let $\mu_\theta(\phi)$ be the length of the longest horizontal line segment that fits into this sector. (See FIGURE 2.)

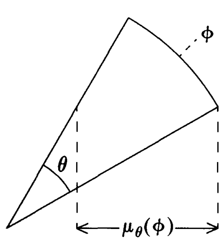


FIGURE 2
The parameters θ and ϕ in the definition of $\mu_\theta(\phi)$. Note that ϕ is the angle that the sector bisector makes with respect to the horizontal.

Given a sector in some orientation, its maximum horizontal extent does not change when the sector is reflected horizontally or vertically (or both):

$$\mu_\theta(\phi) = \mu_\theta(\pi - \phi) = \mu_\theta(\pi + \phi) = \mu_\theta(-\phi). \tag{1}$$

We can use (1) to compute $\mu_\theta(\phi)$ for any fixed ϕ given the value of $\mu_\theta(\phi)$ for ϕ in the first quadrant; therefore we plot the function $\mu_\theta(\phi)$ only for $0 \leqslant \phi \leqslant \pi/2$.

Case I: Nonobtuse sectors If a sector is nonobtuse $(0 \leqslant \theta \leqslant \pi/2)$, and it has an edge in each of the first and fourth quadrants, then it contains the circle radius along the positive x-axis, so its maximum horizontal extent is one, which occurs at the center of the circle. On the other hand, if a nonobtuse sector has both edges in the upper half plane, then its maximum horizontal extent occurs at the lower of the two endpoints of the radii that define the sector. (FIGURE 2 suggests strongly that the maximum occurs at this endpoint, but does not prove it. This fact follows from some of the results below, but you might want to verify it for yourself now.) Formula (2) is valid for $0 \leqslant \theta \leqslant \pi/2$:

$$\mu_\theta(\phi) = \begin{cases} 1, & 0 \leqslant \phi \leqslant \theta/2 \\ \dfrac{\sin\theta}{\sin(\phi + \theta/2)}, & \theta/2 \leqslant \phi \leqslant \pi/2. \end{cases} \tag{2}$$

FIGURE 3 shows $\mu_\theta(\phi)$ for $\theta = \pi/6$, $\pi/4$, and $\pi/3$, and $0 \leqslant \phi \leqslant \pi/2$. We can use equation (1) to complete the graphs to include all values of ϕ, by first reflecting the curve about the line $\phi = \pi/2$, then reflecting both curves about the line $\phi = \pi$.

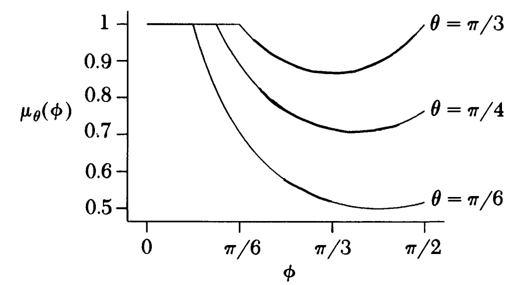


FIGURE 3

EXERCISES. For what value of ϕ is $\mu_\theta(\phi)$ minimized? Is there a value of ϕ such that $\mu_\theta(\pi/2)$ is the minimum value of $\mu_\theta(\phi)$?

As long as the sector subtends an angle of at most $\pi/3$, its maximum horizontal extent is one, the sector radius. When the sector subtends a larger angle, however, its maximum horizontal extent can exceed one. For example, FIGURE 4 shows the graph of $\mu_{\pi/2}(\phi)$: the maximum horizontal extent of a quarter-circle is $\sqrt{2}$, which occurs when the sector opens straight upwards ($\phi = \pi/2$). (Note that formula (2) applies to this case.)

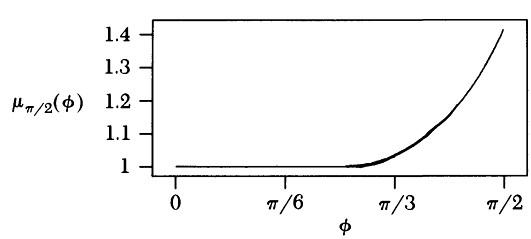


FIGURE 4

Case II: Obtuse, convex sectors The behavior of $\mu_\theta(\phi)$ for $\theta > \pi/2$ is more complicated. We begin to study it by computing the maximum horizontal extent of a semicircle. Let $l_\phi(y)$ be the horizontal extent at height y of a semicircle whose bisector lies at angle ϕ to the positive x-axis. To find $\mu_\pi(\phi) = \max_y l_\phi(y)$, we need consider only $0 \leqslant y \leqslant \sin(\phi + \pi/2) = \cos\phi$ (see FIGURE 5). In the interval $0 \leqslant y \leqslant \cos\phi$, the function $l_\phi(y)$ is given by $\sqrt{1-y^2} + y\tan\phi$.

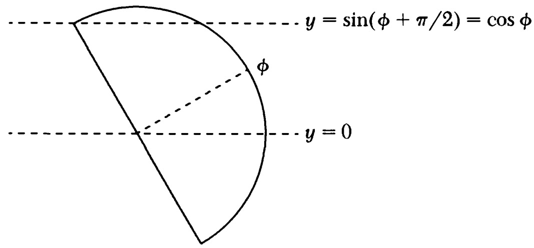


FIGURE 5
To compute $\mu_\pi(\phi)$, we need consider only $0 \leqslant y \leqslant \cos\phi$. If $y < 0$, then $l_\phi(y) < 1 = l_\phi(0)$. If $y > \cos\phi$, then $l_\phi(y) < l_\phi(\cos\phi) = 2\sin\phi$.

When $0 \leqslant \phi \leqslant \pi/4$, the maximum value of $l_\phi(y)$ occurs at $y = \sin\phi$. One way to see this is to observe that the semicircle lies between two parallel lines through the endpoints of the radius at angle ϕ; the horizontal distance between these parallel lines gives an upper bound on $l_\phi(y)$, which is actually realized at $y = \sin\phi$. On the other hand, when $\pi/4 \leqslant \phi \leqslant \pi/2$, $l_\phi(y)$ is maximized at an endpoint of the interval $0 \leqslant y \leqslant \cos\phi$. We conclude that

$$\mu_\pi(\phi) = \begin{cases} 1/\cos\phi, & 0 \leqslant \phi \leqslant \pi/4 \\ 2\sin\theta, & \pi/4 \leqslant \phi \leqslant \pi/2. \end{cases} \tag{3}$$

FIGURE 6 shows $\mu_\pi(\phi)$.

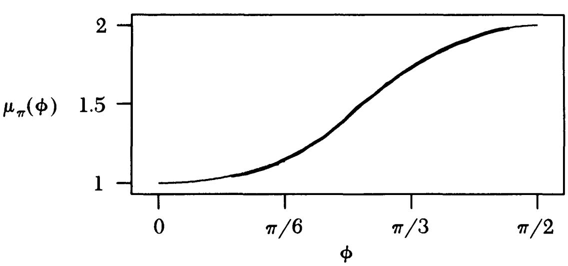


FIGURE 6

EXERCISE. It is tempting (though incorrect) to generalize from the case of nonobtuse sectors and assert that $\mu_\pi(\phi)$ is always attained at an endpoint of one of the radii that defines a sector. This incorrect generalization would lead us to calculate $\mu'_\pi(\phi) = \max(1, 2\sin\phi)$. For what value of ϕ does $\mu'_\pi(\phi)$ deviate most from $\mu_\theta(\phi)$? How large is the deviation?

We can use equations (2) and (3) to calculate $\mu_\theta(\phi)$ when $\pi/2 \leqslant \theta \leqslant \pi$. To express the three cases that arise, let $\psi = \pi/2 - \theta/2$, so that $\phi - \psi$ is normal to the leading edge of the sector. The first case arises when the edges of the sector lie in the first and fourth quadrants ($0 \leqslant \phi \leqslant \psi$): the maximum horizontal extent of one occurs at the center of the circle. In both the second and third cases, one edge of the sector lies in the second quadrant. When $0 \leqslant \phi - \psi \leqslant \pi/4$, the non-corner maximum from the semicircle case applies. When $\pi/4 \leqslant \phi - \psi \leqslant \pi/2$, the maximum occurs at the endpoint of the edge in the second quadrant. Formula (4) is valid for $\pi/2 \leqslant \theta \leqslant \pi$.

$$\mu_\theta(\phi) = \begin{cases} 1, & 0 \leqslant \phi \leqslant \psi \\ 1/\cos(\phi - \psi), & \psi \leqslant \phi \leqslant \pi/4 + \psi \\ 2\sin(\phi - \psi), & \pi/4 + \psi \leqslant \phi \leqslant \pi/2. \end{cases} \tag{4}$$

FIGURE 7 shows $\mu_{3\pi/4}(\phi)$.

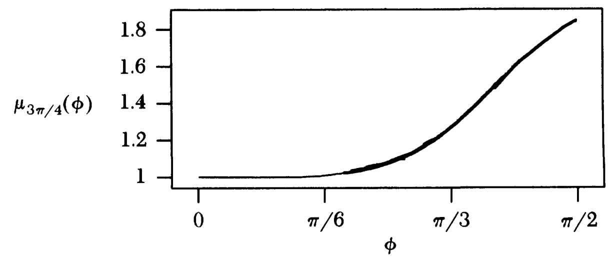


FIGURE 7

Case III: Non-convex sectors Formulas (1–4) can be used to compute $\mu_\theta(\phi)$ for $\theta > \pi$. Any sector that subtends an angle of more than π can be chopped by a horizontal line through the circle center into two or three sectors. The maximum horizontal extent of the entire arc is just the maximum of the maximum horizontal extents of each of the component sectors.

Now that we can compute $\mu_\theta(\phi)$ for all values of θ and ϕ, we can solve the pie-chart problem Sectors in Fixed Order. Suppose we are given a fixed order of pairs of sector angles and associated label lengths, $((\theta_i, \lambda_i)|1 \leqslant i \leqslant n)$. Once we fix the angle for any one sector, the values of all other angles are forced, so let Φ be ϕ_1, the *master angle* for the entire fixed order. Define $\sigma_i(\Phi)$ by the equation

$$\sigma_i(\Phi) = -\lambda_i + \mu_{\theta_i}\left(\Phi + (\theta_i - \theta_1)/2 + \sum_{1 \leqslant j < i} \theta_j\right).$$

If $\sigma_i(\Phi) \geqslant 0$, then when the master angle is Φ, sector i can contain its label. Thus, if we can find a value of Φ for which $\sigma_i(\Phi) \geqslant 0$ for $1 \leqslant i \leqslant n$, we can draw the pie chart specified by $((\theta_i, \lambda_i)|1 \leqslant i \leqslant n)$.

As an example, consider this fixed order of sector angles and associated label lengths:

$$((\theta_i, \lambda_i)) = ((\pi/4, 0.95), (\pi/2, 1.4)(\pi/3, 0.9), (3\pi/4, 1.6), (\pi/6, 0.85)).$$

The graph in FIGURE 8 shows that it is possible to draw the pie chart by choosing Φ to be near $\pi/8$.

EXERCISE. In general it does not suffice to look only at values of Φ in the first quadrant: we might miss a solution. What range of values of Φ must one consider to be sure of finding a solution if one exists?

We have solved the Fixed Sector Order problem for a unit circle, assuming implicitly that label lengths were scaled to fit inside the circle. Here is a related problem: suppose the order of the sectors were determined, and the label lengths were given as absolute lengths. How would you determine the maximum possible radius that permits all labels to fit inside their slices?

Unspecified Sector Order

Now suppose that we are given just the set of slice angles and associated horizontal segment lengths, and no information about the order of the slices. At first glance this

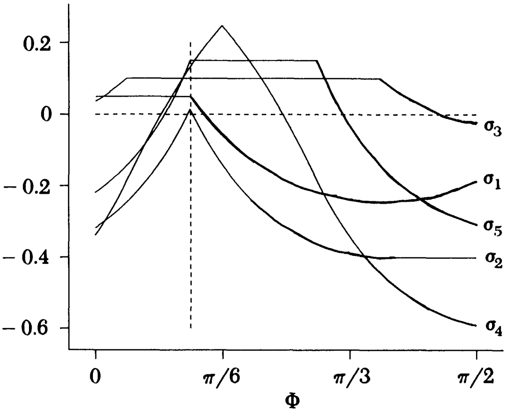


FIGURE 8

problem seems substantially harder. An obvious solution is to perform the computation described above on all $(n-1)!$ possible orders of the n slices, but the amount of work involved does not make this algorithm very appealing. Unfortunately, even at second and third glances the problem of laying out a pie chart remains hard.

There are at least two possible reasons for our failure to devise a practical algorithm that solves the pie-chart layout problem: we are not very clever, or the problem is very hard. Fortunately for our problem-solving ego, there is a way to certify that pie-chart layout is a hard problem, using the theory of **NP**-completeness [3].

This theory is concerned with the *existence* of solutions, rather than their construction. In our case, given the angle-label pairs, $\{(\theta_i, \lambda_i) | 1 \leqslant i \leqslant n\}$, we would ask whether there exists any orientation of the slices so that each slice can contain its label, and the slices fit together to form a circle. Although proving existence is sometimes easier than actually constructing a solution, that does not appear to be the case here. Indeed, the natural way to prove the existence of a pie-chart layout is to exhibit the layout.

An existence problem is in the class of problems **NP** if an alleged solution to any particular instance of the existence problem can be verified using a number of steps that is polynomial in the size of the problem instance. The **P** in **NP** stands for polynomial time; in general an algorithm that runs in polynomial time is considered to be feasible. The **N** in **NP**, which stands for "nondeterministic," refers to the unspecified source of the alleged solution, which is usually called a "guessing module," a handy but non-existent accessory for modern digital computers. A problem in **NP** can be solved in polynomial time if we have a module that makes good guesses.

Problems in **NP** exhibit an important asymmetry. If we verify that a guess actually solves a problem in **NP**, then we know that the answer to the existence problem is "yes," and we also have a constructive proof of this positive answer. On the other hand, if a guess does not check out, we do *not* have a negative answer to the existence problem, but only evidence that the module made a bad guess. In the case of pie-chart layout, this corresponds to the plausible observation that checking the validity of a guess is easy, but proving that no solution exists is harder.

Certain problems in **NP** are distinguished by the property that if they can be solved in polynomial time without the use of a guessing module (i.e., in *deterministic* polynomial time), then any other problem in **NP** can also be solved in deterministic polynomial time. These problems are said to be **NP**-complete, because a solution to any of them gives a polynomial-time solution to all problems in **NP**. One such

NP-complete problem is PARTITION: Can a given set of integers $\{x_i | 1 \leqslant i \leqslant n\}$ be partitioned into two sets $A \cup B$ such that the sums of both sets are the same, $\Sigma_{x \in A} x = \Sigma_{x \in B} x$? (Note that PARTITION might be considered to be easier than pie chart layout, since the obvious brute-force solution need consider only 2^{n-1} combinatorially different configurations, rather than $(n-1)!$. However, 2^{n-1} is still not polynomial in n.)

It might appear that the pie-chart layout problem is in **NP**. Given $\{(\theta_i, \lambda_i) | 1 \leqslant i \leqslant n\}$, an ordering of the slices, and a master angle Φ, that were alleged to yield a valid pie chart, we could use the formulas above to check whether all labels indeed fit inside their slices. However, this would mean testing whether n transcendental functions are positive, a calculation that is not known to be in **NP**.

Nevertheless, we can show that the pie-chart layout problem is at least as hard as PARTITION. Given any instance of PARTITION, $\{x_i | 1 \leqslant i \leqslant n\}$, we construct the following instance of the pie-chart problem Unspecified Sector Order:

$$\{(1,1),(1,1),(1,1),(1,1)\} \cup \left\{ \left(\frac{x_i}{\sum\limits_{1 \leqslant i \leqslant n} x_i} (2\pi - 4), 0 \right) \right\}$$

This derived pie-chart problem has two kinds of pieces: four pie slices that each subtend one radian and have a label of unit length, called *enforcers*; and n pie slices whose angles are related to the original instance of PARTITION, called *surrogates*.

We know that in any solution to this pie-chart problem, the placement of the enforcers is completely determined. Each must have a horizontal edge to permit its label to fit, so each must have one edge against the x-axis. This placement leaves two empty sectors of size $\pi - 2$, one on each side of the horizontal diameter of the pie.

If there is to be a solution to this pie-chart problem, each of the surrogates must be placed in one of the two empty sectors left by the enforcers. Note that the surrogates can lie at any angle, because their label lengths are zero. Thus, we can solve this pie-chart problem if and only if there is a way to divide the surrogates into two sets that each subtend the same total angle. In turn, such a solution would tell us how to solve the original instance of PARTITION. FIGURE 9 shows an example.

This construction shows that if we could solve the pie-chart layout problem, we could solve the PARTITION problem. Since PARTITION is **NP**-complete, it follows that pie-chart layout is at least as difficult as any problem in **NP**. A problem that is at least as difficult as any problem in **NP**, but is not known itself to be in **NP**, is said to be **NP**-hard. We do not need to feel bad about failing to find an efficient algorithm to solve an **NP**-hard problem.

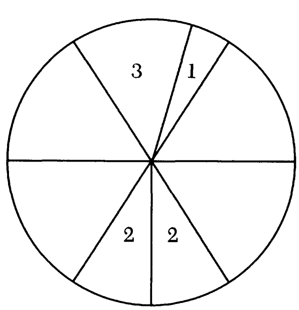


FIGURE 9
Construction for the PARTITION problem $\{1,2,2,3\}$.

Our discussion has barely skimmed the surface of the theory of **NP**-completeness. There are several other problems related to the complexity of the pie-chart layout problem. If input to the **NP**-complete problem PARTITION is written in unary notation, so that its size is $\sum_{1 \leqslant i \leqslant n} x_i$, the problem can be solved in time polynomial in the input size. PARTITION is said to have a *pseudo-polynomial* time algorithm because the algorithm is not polynomial when the input is presented in a more sensible fashion, such as binary notation, of size $\sum_{1 \leqslant i \leqslant n} \log_2 x_i$. Is there a pseudo-polynomial time algorithm for the pie-chart layout problem?

Many other problems in drawing are challenging. Two examples are trees [5] and graphs [4].

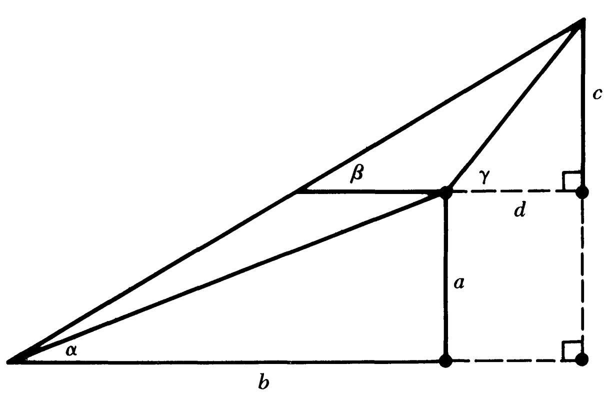
Thanks to Bernard Chazelle, who suggested an idea that led to the **NP**-hardness proof given above.

REFERENCES

1. J. Bertin, *Semiology of Graphics*, Univ. of Wisconsin, 1983. (Translation of Semiologie graphique, 1967.)
2. W. S. Cleveland and R. McGill, Graphical perception: theory, experimentation, and the development of graphical methods, *J. Amer. Stat. Assoc.* 79 (1984), 531–554.
3. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
4. R. J. Lipton, S. C. North, and J. S. Sandberg, A method for drawing graphs, *Proc. Symp. Comput. Geom.*, Baltimore, Maryland (1985), 153–160.
5. K. J. Supowit and E. M. Reingold, The complexity of drawing trees nicely, *Acta Inform.* 18 (1983), 377–392.
6. E. R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT, 1983.
7. U.S. Bureau of the Census, *Statistical Abstract of the United States: 1986* (106th edition), Washington, D.C., 1985.

Proof without Words: The Mediant Property

$$\alpha < \beta < \gamma \rightarrow \frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$$



—Richard A. Gibbs
Fort Lewis College
Durango, CO 81301

Our discussion has barely skimmed the surface of the theory of **NP**-completeness. There are several other problems related to the complexity of the pie-chart layout problem. If input to the **NP**-complete problem PARTITION is written in unary notation, so that its size is $\sum_{1 \leqslant i \leqslant n} x_i$, the problem can be solved in time polynomial in the input size. PARTITION is said to have a *pseudo-polynomial* time algorithm because the algorithm is not polynomial when the input is presented in a more sensible fashion, such as binary notation, of size $\sum_{1 \leqslant i \leqslant n} \log_2 x_i$. Is there a pseudo-polynomial time algorithm for the pie-chart layout problem?

Many other problems in drawing are challenging. Two examples are trees [5] and graphs [4].

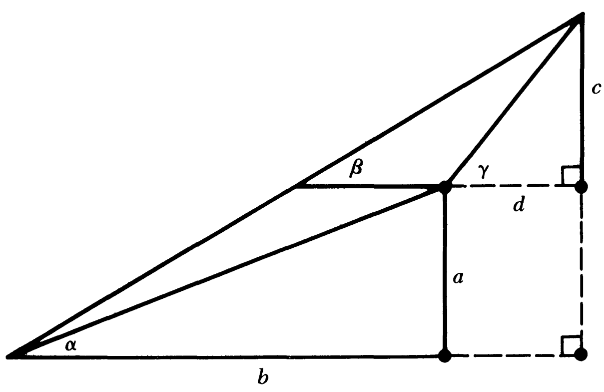
Thanks to Bernard Chazelle, who suggested an idea that led to the **NP**-hardness proof given above.

REFERENCES

1. J. Bertin, *Semiology of Graphics*, Univ. of Wisconsin, 1983. (Translation of Semiologie graphique, 1967.)
2. W. S. Cleveland and R. McGill, Graphical perception: theory, experimentation, and the development of graphical methods, *J. Amer. Stat. Assoc.* 79 (1984), 531–554.
3. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
4. R. J. Lipton, S. C. North, and J. S. Sandberg, A method for drawing graphs, *Proc. Symp. Comput. Geom.*, Baltimore, Maryland (1985), 153–160.
5. K. J. Supowit and E. M. Reingold, The complexity of drawing trees nicely, *Acta Inform.* 18 (1983), 377–392.
6. E. R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT, 1983.
7. U.S. Bureau of the Census, *Statistical Abstract of the United States: 1986* (106th edition), Washington, D.C., 1985.

Proof without Words: The Mediant Property

$$\alpha < \beta < \gamma \rightarrow \frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$$



—Richard A. Gibbs
Fort Lewis College
Durango, CO 81301

Easy Algorithms for Finding Eigenvalues

CLIFFORD REITER
Lafayette College
Easton, PA 18042

Many students in an introductory linear algebra course realize that writing down the characteristic equation for a modest size matrix would be an algebraic nightmare. Even if the characteristic equation is known, finding its roots may seem hard. This leads many students to believe that it is not practical to try to compute the eigenvalues for large or modest size matrices. A brief introduction to appropriate numerical techniques can dispel that conviction.

The simplest numerical technique for estimating eigenvectors is the power method. Unfortunately convergence of the power method can be quite slow. However, a second simple technique, the Rayleigh quotient iteration, converges at a dazzling rate. We will see an example where we are able to estimate accurately an eigenvalue of a 3 by 3 matrix to 16 significant figures by solving four linear systems of equations and by using a little vector arithmetic. Both the power method and the Rayleigh quotient iteration are easy to experiment with in an interactive matrix arithmetic environment on a computer. This note collects the necessary results so that students will have an easy time running such experiments on a computer. We will be able to compare the power method and the Rayleigh quotient iteration with *a posteriori* error estimates. All of this is a nice glimpse of numerical linear algebra. Moreover, students get a chance to experiment with algorithms for finding eigenvalues that converge quickly and will generalize to large matrices with little difficulty.

The methods The *power method* applied to the matrix A is the iterative process

$$X_1 = A\hat{X}_0, X_2 = A\hat{X}_1, \ldots, X_{k+1} = A\hat{X}_k,$$

where $\hat{X}_i$ is the unit vector in the X_i direction and $\hat{X}_0$ is an initial guess for an eigenvector. Suppose A has a complete set of eigenvectors $V_1, \ldots, V_n$ corresponding to eigenvalues $\lambda_1, \cdots, \lambda_n$ with $|\lambda_1| > |\lambda_2| \geqslant \cdots \geqslant |\lambda_n|$. Call λ_1 the *dominant* eigenvalue. Since $V_1, \ldots, V_n$ are a basis for $\mathbb{R}^n$, $\hat{X}_0$ has the form $\hat{X}_0 = c_1V_1 + \cdots + c_nV_n$. It is easy to check that

$$\hat{X}_k = \frac{c_1V_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k V_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k V_n}{\left\| c_1V_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k V_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k V_n \right\|} \cdot \left(\frac{\lambda_1}{|\lambda_1|}\right)^k.$$

Since

$$\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0 \text{ as } k \to \infty \text{ for } i \geqslant 2$$

we see

$$\hat{X}_k \to \frac{c_1V_1}{\|c_1V_1\|}\left(\frac{\lambda_1}{|\lambda_1|}\right)^k,$$

provided $c_1 \neq 0$. That is, $\hat{X}_k$ converges to the dominant eigenvector if $\lambda_1 > 0$ or oscillates if $\lambda_1 < 0$; but $\hat{X}_{2k}$ converges in any case.

Notice that $c_1 \neq 0$ is required of the initial guess. Oddly, because of roundoff errors in machine computation this tends not to be a difficulty in practice. See [5] for some remarks about that. In all our examples we will choose $\hat{X}_0 = (1/\sqrt{3})(1\ 1\ 1)$, so that the component in the direction of the standard basis vectors will be nonzero. Notice that the rate of convergence depends on how near $|\lambda_2/\lambda_1|$ is to one. Once we have observed convergence of an approximate eigenvector we still need to know how to estimate the associated eigenvalue. The following lemma provides a nice answer.

LEMMA 1. *Given a square matrix A and a vector $X \neq 0$, the μ which minimizes $\|AX - \mu X\|$ is a Rayleigh quotient* $\mu = \dfrac{X \cdot (AX)}{X \cdot X}$.

The lemma can easily be checked using the second derivative test on $f(\mu) = \|AX - \mu X\|^2$. It says that $\mu_k = \hat{X}_k \cdot (A\hat{X}_k)$ is the best estimate in the above sense for an eigenvalue associated with the unit vector $\hat{X}_k$. We now consider two examples of the power method.

Example 1. Let

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 8 & 3 \\ 1 & 3 & 15 \end{pmatrix} \quad \text{and} \quad \hat{X}_0 = \frac{1}{\sqrt{3}}(1\ 1\ 1).$$

We use the power method to get estimates for the eigenvector associated with the dominant eigenvalue and the Rayleigh quotient to estimate the associated eigenvalue.

TABLE 1. An example of the power method converging.

Iteration		Approximate eigenvector		Approximate eigenvalue
0	0.57735	0.57735	0.57735	12.666667
1	0.25220	0.54643	0.79863	15.719081
2	0.16672	0.45781	0.87328	16.194590
3	0.14106	0.40770	0.90215	16.289097
4	0.13137	0.38355	0.91413	16.308542
5	0.12724	0.37238	0.91932	16.312536
6	0.12541	0.36729	0.92161	16.313356
7	0.12459	0.36497	0.92264	16.313524
8	0.12421	0.36392	0.92311	16.313559

The correct digits of the estimated eigenvalues are underlined. Notice the approximate eigenvalues and eigenvectors seem to be converging. A more precise analysis of the convergence is presented following another example.

In the next example there is not a dominant eigenvalue, although this is not obvious ahead of time. Accordingly, we do not expect the power method to converge.

Example 2. Let

$$A = \begin{pmatrix} \cos(2) & -\sin(2) & 0 \\ \sin(2) & \cos(2) & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \approx \begin{pmatrix} -.41615 & -.90930 & 0 \\ .90930 & -.41615 & 0 \\ 0 & 0 & .5 \end{pmatrix}$$

and $X_0 = (1/\sqrt{3})(1\ 1\ 1)$. The matrix A is a rotation in the x-y plane with a contraction along the z-axis. We present the results of several iterations of the power method in TABLE 2.

TABLE 2. An example of the power method not converging.

Iteration		Approximate eigenvector		Approximate eigenvalue
0	0.57735	0.57735	0.57735	−0.110765
1	−0.88363	0.32877	0.33333	−0.314353
2	0.071831	−0.98211	0.17408	−0.388385
3	0.87312	0.47950	0.088045	−0.409045
4	−0.80168	0.59612	0.044151	−0.414361
5	−0.20858	−0.97776	0.022092	−0.415700
6	0.97605	0.21727	0.011048	−0.416035
7	−0.60377	0.79714	0.0055242	−0.416119
8	−0.47359	−0.88074	0.0027621	−0.416140

From the results, it appears that the approximate eigenvectors are not converging, but their z-components are approaching zero. The approximate eigenvalues are converging. However, the actual eigenvalues are $1/2$ and $\cos(2) \pm i \sin(2) \approx -.41615 \pm i.90930$. The approximate eigenvalues are quite far off.

These examples lead to several questions: How do we recognize convergence? How do we estimate nondominant eigenvectors? How can we speed the rate of convergence?

The previous examples demonstrated that simply observing the convergence of approximate eigenvalues can be dangerous. We will get error bounds on our approximate eigenvalues that are simple to compute and reasonably precise. This will allow us to verify the accuracy of our estimated eigenvalues explicitly.

Several things can be done to estimate nondominant eigenvalues. We will be able to do that with the Rayleigh quotient iteration and the inverse power method. The reader may also want to read about deflation, see [3, p. 288, ex. 18] or the block power method, see [4, p. 289]. The Rayleigh quotient iteration and the inverse power method have the advantage that they also speed up the rate of convergence.

The inverse power method is based on the following observation.

LEMMA 2. *If A has an eigenvalue λ_k with corresponding eigenvector V_k and μ is not an eigenvalue of A then $(A-\mu I)^{-1}$ has eigenvalue $(\lambda_k-\mu)^{-1}$ corresponding to eigenvector V_k.*

The proof is an easy exercise. The *inverse power method* uses the power method on $(A-\mu I)^{-1}$. The point is that if μ is near λ_k then $(\lambda_k-\mu)^{-1}$ is huge while $(\lambda_j-\mu)^{-1}$ is small for $j \neq k$. For example, if a matrix A has eigenvalues 21, 20, and 3 then the power method on A will converge at a very slow rate since $20/21 \approx 1$. However, if we guess that there is an eigenvalue near $\mu = 19.9$, then $(A - 19.9I)^{-1}$ has eigenvalues near .91, 10, and −.06. Therefore, the power method on $(A - 19.9I)^{-1}$ will converge fairly quickly to the eigenvector corresponding to 10 which is the eigenvector of A corresponding to 20. Notice we were able to improve the rate of convergence and estimate a nondominant eigenvector of A.

However, we were required to make an initial guess μ for the approximate eigenvalue. Gerschgorin's theorem makes precise how the eigenvalues of a matrix are guaranteed to be "near" to the diagonal elements of the matrix.

GERSCHGORIN'S CIRCLE THEOREM. Every eigenvalue λ of an n by n matrix $A = (a_{ij})$ satisfies at least one of the inequalities:

$$|\lambda - a_{ii}| \leqslant r_i \quad \text{where } r_i = \sum_{\substack{j \neq i \\ 1 \leqslant j \leqslant n}} |a_{ij}|.$$

This says that every eigenvalue of A falls into one of the circles in the complex plane with center a_{ii} and radius $r_i =$ the sum of the magnitudes of the off diagonal entries of row i. As noted by [**3**, p. 302] and [**4**, p. 304] the theorem follows from the observation that if λ is an eigenvalue of A corresponding to the eigenvector X, then $AX = \lambda X$, so

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i \quad \text{and thus} \quad (\lambda - a_{ii}) x_i = \sum_{\substack{j \neq i \\ 1 \leqslant j \leqslant n}} a_{ij} x_j \quad \text{for all} \quad 1 \leqslant i \leqslant n.$$

We then get the bound that we want if we choose i with $|x_i|$ as large as possible so that $|x_j/x_i| \leqslant 1$:

$$|\lambda - a_{ii}| \leqslant \sum_{\substack{j \neq i \\ 1 \leqslant j \leqslant n}} |a_{ij}| \left| \frac{x_j}{x_i} \right| \leqslant \sum_{\substack{j \neq i \\ 1 \leqslant j \leqslant n}} |a_{ij}| = r_i.$$

Gerschgorin's Theorem explains why we will "guess" that the eigenvalues of a matrix are near to the diagonal entries of the matrix.

The Rayleigh quotient iteration is just the inverse power method with guesses for the eigenvalue μ updated in the natural way: via the Rayleigh quotient.

Rayleigh quotient iteration:

(0) Guess an initial eigenvalue μ_0 and unit eigenvector $\hat{X}_0$.

(1) Find X_{k+1} so $(A - \mu_k I) X_{k+1} = \hat{X}_k$ (this is one step of the inverse power method).

(2) Let $\mu_{k+1} = \hat{X}_{k+1} \cdot A\hat{X}_{k+1}$ (use the Rayleigh quotient to update the approximate eigenvalue).

(3) Repeat (1) and (2) until μ_{k+1} is exactly an eigenvalue or the desired accuracy is reached.

We will see that for symmetric matrices it is easy to check whether the desired accuracy has been reached.

Error estimates and an example One can use the approximate eigenvalues and eigenvectors to get error bounds on the eigenvalues. The estimate when A is symmetric is particularly simple so we will give it here. See [**2**, p. 448] for the details and an estimate which does not assume the symmetry of A.

THEOREM. *If A is symmetric, $\hat{X}$ is a unit vector and λ_i is the eigenvalue of A nearest μ. Then*

$$|\mu - \lambda_i| \leqslant \|A\hat{X} - \mu\hat{X}\|.$$

So the length of the *residue* $A\hat{X} - \mu\hat{X}$ gives a bound on the error of μ approximating the true eigenvalue λ_i. When $\hat{X}$ is approximately an eigenvector associated with μ this bound will be small. In Example 1, the length of the residue for the eighth iteration is $|A\hat{X}_8 - \mu_8\hat{X}_8| \approx .0089$. This gives an explicit bound on the error of the estimate of the eigenvalue. The power method had given us about four significant figures. In Example 2, the length of the residue for the eighth iteration is $|A\hat{X}_8 - \mu_8\hat{X}_8| \approx .9093$. While the bound of the Theorem does not apply since $A \neq A^T$

it is clear that $\mu_8, \hat{X}_8$ are not an eigenvalue, eigenvector pair.

We are ready for an example of the Rayleigh quotient iteration:

Example 3. Let

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 8 & 3 \\ 1 & 3 & 15 \end{pmatrix} \quad \text{and} \quad \hat{X}_0 = (1/\sqrt{3})(1\ 1\ 1).$$

We compare four iterations of the power method on A with initial guess X_0 with four iterations of the Rayleigh quotient iteration for the same A and $\hat{X}_0$ along with $\mu_0 = 15$ as Gerschgorin's theorem might suggest as a guess for the dominant eigenvalue.

TABLE 3. An example of estimated eigenvalues and error bounds.

	Power method		Rayleigh quotient iteration	
Iteration	Approximate eigenvalue	Error bound	Approximate eigenvalue	Error bound
0	12.666666666666666	5.3E0	15.000000000000000	5.8E0
1	15.719081272084803	2.3E0	16.201807549175964	1.1E0
2	16.194589560077870	1.0E0	16.313552340303307	1.2E-2
3	16.289097072280768	4.7E-1	16.313567783095323	2.0E-8
4	16.308541550755940	2.1E-1	16.313567783095326	2.1E-15

The error bounds given in the table are provided by the above theorem. The correct digits of the estimated eigenvalues are underlined. Notice that in the fourth iteration of the Rayleigh quotient iteration the estimated eigenvalue did not change much yet there was a considerable improvement in the error bound. This was possible since the estimated eigenvector was considerably improved at this step.

Classroom experience The material of this paper plus the "easy" proofs can be presented in about 2 hours of lecture. My students were given access to APL on microcomputers. Each student was assigned a different matrix and was asked to make comparisons like those in the table. Those comparisons can be done interactively in APL or could be done using a matrix arithmetic package. Nearly all of the students did the correct computation and spent about an hour on the assignment. (They had a couple of previous computer assignments using APL.)

After their experiments with the Rayleigh quotient iteration most of my students are very impressed by its rapid convergence. If the students are not forewarned they often reach a state of disbelief: "I am getting an unbelievable difference in the rate of convergence." It is a pleasure to assure these students: "Your work is right."

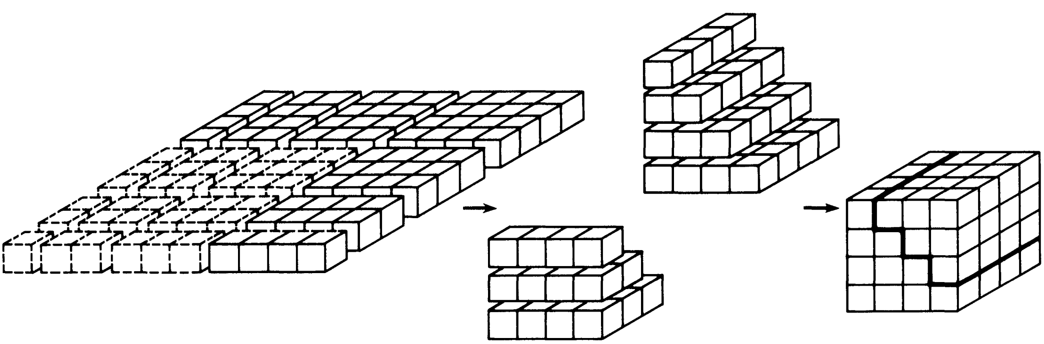
In review of their assignment I remark that the Rayleigh quotient iteration on nondeficient matrices has quadratic convergence and it has cubic convergence for symmetric matrices. See [3]. The convergence is local in the sense that if an approximate eigenvector with nonzero component in the direction of an actual eigenvalue is fixed, then convergence will occur for sufficiently close guesses of the eigenvalue. This is true for complex eigenvalues as well so long as vector length and the Rayleigh quotient are computed with the appropriate complex conjugates. It is worth mentioning that more "robust" algorithms exist. The QR algorithm converges rapidly and it finds all the eigenvalues at once. However, it is more complicated, see [3] or [4]. Modern work on parallel computing seems likely to have a major impact on "state of the art" algorithms for computing eigenvalues, see [1]. Once in a while a good student will notice that $A - \mu_k I$ is nearly singular and wonder if that causes the Rayleigh quotient iteration to be unstable. Stewart [3] explains why this is not a problem.

The Rayleigh quotient iteration is a simple algorithm which rapidly estimates eigenvalues. The accuracy can even be guaranteed! There are easy algorithms for finding eigenvalues.

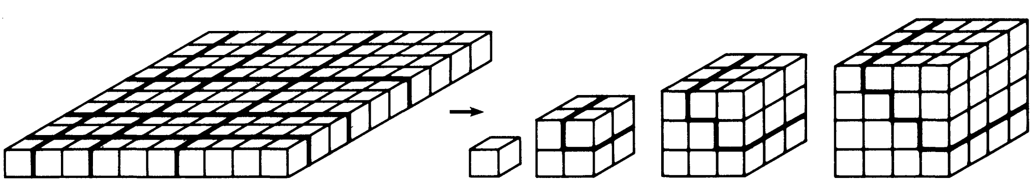
REFERENCES

1. J. J. Dongarra and D. C. Sorensen, A fast algorithm for the symmetric eigenvalue problem. *Proceedings of the 7th Symposium on Computer Arithmetic*, Univ. of Illinois, IEEE Computer Society, June 1985.
2. Ben Noble, *Applied Linear Algebra*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1969.
3. G. W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York, 1973.
4. Gilbert Strang, *Linear Algebra and Its Applications*, 2nd edition, Academic Press, New York, 1976.
5. Gareth Williams and Donna Williams, The power method for finding eigenvalues on a microcomputer, *Amer. Math. Monthly* 93 (1986), 562–565.

Proof without Words: Squares of Triangular Numbers



$$t_n = 1 + 2 + \cdots + n \Rightarrow t_n^2 - t_{n-1}^2 = n^3.$$



$$t_n^2 = (1 + 2 + \cdots + n)^2 = 1^3 + 2^3 + \cdots + n^3.$$

—ROGER B. NELSEN
LEWIS AND CLARK COLLEGE

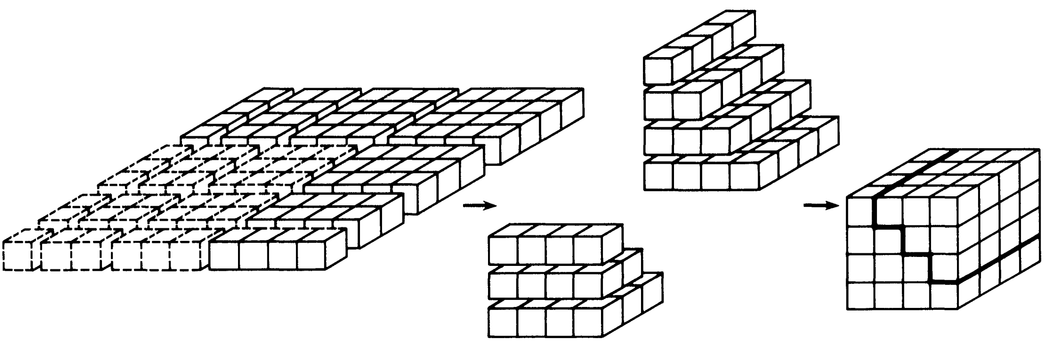
good student will notice that $A - \mu_k I$ is nearly singular and wonder if that causes the Rayleigh quotient iteration to be unstable. Stewart [3] explains why this is not a problem.

The Rayleigh quotient iteration is a simple algorithm which rapidly estimates eigenvalues. The accuracy can even be guaranteed! There are easy algorithms for finding eigenvalues.

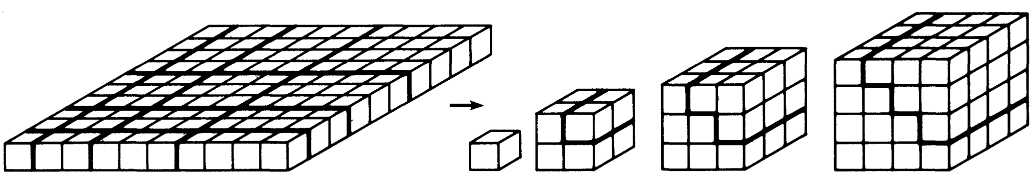
REFERENCES

1. J. J. Dongarra and D. C. Sorensen, A fast algorithm for the symmetric eigenvalue problem. *Proceedings of the 7th Symposium on Computer Arithmetic*, Univ. of Illinois, IEEE Computer Society, June 1985.
2. Ben Noble, *Applied Linear Algebra*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1969.
3. G. W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York, 1973.
4. Gilbert Strang, *Linear Algebra and Its Applications*, 2nd edition, Academic Press, New York, 1976.
5. Gareth Williams and Donna Williams, The power method for finding eigenvalues on a microcomputer, *Amer. Math. Monthly* 93 (1986), 562–565.

Proof without Words: Squares of Triangular Numbers



$$t_n = 1 + 2 + \cdots + n \Rightarrow t_n^2 - t_{n-1}^2 = n^3.$$



$$t_n^2 = (1 + 2 + \cdots + n)^2 = 1^3 + 2^3 + \cdots + n^3.$$

—Roger B. Nelsen
Lewis and Clark College

Hyperbolic and Trigonometric Crossing Points

P. J. LEAH
J. B. WILKER
University of Toronto
Scarborough, Ontario
Canada M1C 1A4

1. Introduction The intimate relationships between the six trigonometric functions and the six hyperbolic functions are evident in the expressions of all twelve by exponentials; still, the restriction of the domain to the real variable hides the periodicity of the hyperbolic functions and emphasizes that they are monotone on $[0, \infty)$, and conversely conceals the unboundedness of sine and cosine and leaves their periodicity intact. The trigonometric-hyperbolic relationship seems weakened.

Now an examination of the crossing points of the graphs of the hyperbolic functions on the domain $x > 0$ reveals the curious coincidence that the height at which the graphs of cosh and csch cross is the same as the height at which the graphs of sinh and coth cross; both crossings occur on the line $y = \sqrt{\tau}$ where $\tau = (1+\sqrt{5})/2$ is the golden ratio.

In an effort to give an elegant explanation for this and to find possible generalizations, we considered the analogous crossing problem for the trigonometric functions and discovered, somewhat to our surprise, that there is a corresponding coincidence. Restricted to $[0, \pi/2]$, the graphs of sec and cot cross once as do the graphs of tan and csc. Moreover these crossings also occur on the line $y = \sqrt{\tau}$.

It turns out that there is a pleasant explanation for these coincidences. There is a close connection between the graphs of the hyperbolic functions on $\mathbb{R}$ and the graphs of the trigonometric functions on $(-\pi/2, \pi/2)$. By exploiting this connection we can relate the hyperbolic crossing phenomenon to the trigonometric one. While our original observation about the hyperbolic crossings was based on messy algebra, the corresponding fact about trigonometric crossings can be deduced from the existence of a unique right-angled triangle whose sides are in geometric progression. It is this triangle that is responsible for bringing the golden ratio τ into the story.

As a final remark before turning to the details, we mention that a biproduct of our discussion is a novel derivation of the formula

$$\int_0^x \sec t\,dt = \log(\sec x + \tan x)$$

not altogether unrelated to its appearance in connection with the Mercator map projection ([**5**, p. 153] or [**3**, p. 438]).

We wish to thank M. J. Evans, B. R. Monson and the referees of an earlier draft for their interest and helpful comments.

2. The connection between the graphs We present a sketch of the graphs of the trigonometric functions on $(-\pi/2, \pi/2)$ (Figure 1) and a sketch of the graphs of the hyperbolic functions on $\mathbb{R}$ (Figure 2).

The hyperbolic picture looks exactly like a stretched version of the trigonometric one and forces upon us the unorthodox pairing

$$\tan \sim \sinh \tag{1}$$
$$\cot \sim \operatorname{csch} \tag{1'}$$
$$\sec \sim \cosh \tag{2}$$
$$\cos \sim \operatorname{sech} \tag{2'}$$
$$\sin \sim \tanh \tag{3}$$
$$\csc \sim \coth. \tag{3'}$$

Since tan and sinh are monotone on their respective domains we can define

$$f\colon \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$$

unambiguously by writing

$$\tan x = \sinh f(x). \tag{1}$$

This overlaps with an idea of G. H. Hardy ([4, page 415, Examples LXXXVIII number 26]) though by starting with (1) rather than (2) we avoid all ambiguity of sign. Later we shall give a number of derivations for the explicit form of f. But the observation we want now is that the same function f links each trigonometric function with its hyperbolic mate as defined by the list (1)–(3′). This seems to have gone unnoticed at least in textbook literature although it is perhaps implicit in some of the older books (see, for example, Chrystal [1, pp. 312, 313] and the references cited there). In any event the proof is quite straightforward.

The primed relations such as

$$\cot x = \operatorname{csch} f(x) \tag{1'}$$

follow from the unprimed by taking reciprocals on both sides of the equation.

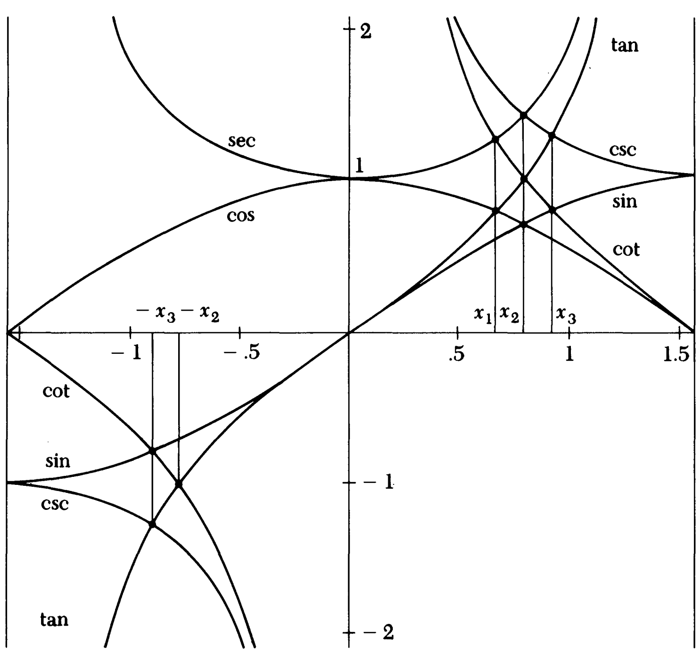


FIGURE 1
The graphs of the trigonometric functions.

Relation (2) follows from (1) because $\sec^2 = 1 + \tan^2$ is matched by $\cosh^2 = 1 + \sinh^2$ and relation (3) follows from (2′) because $\sin^2 = 1 - \cos^2$ is matched by $\tanh^2 = 1 - \operatorname{sech}^2$.

3. The function f We provide four derivations of the fact that the function f of the last section is given by

$$f(x) = \log(\sec x + \tan x).$$

First, it follows from (1) that

$$\tan x = \frac{e^{f(x)} - e^{-f(x)}}{2}.$$

This leads to the quadratic equation

$$\left(e^{f(x)}\right)^2 - 2(\tan x)e^{f(x)} - 1 = 0$$

with roots

$$e^{f(x)} = \frac{2\tan x \pm \sqrt{4\tan^2 x + 4}}{2} = \tan x \pm \sec x$$

and the correct sign follows from the fact that $e^{f(x)}$ must be positive.

As a second proof we can take f as given and compute

$$\begin{aligned}\sinh f(x) &= \frac{1}{2}\left(e^{f(x)} - e^{-f(x)}\right)\\ &= \frac{1}{2}\left((\sec x + \tan x) - (\sec x + \tan x)^{-1}\right)\\ &= \tan x.\end{aligned}$$

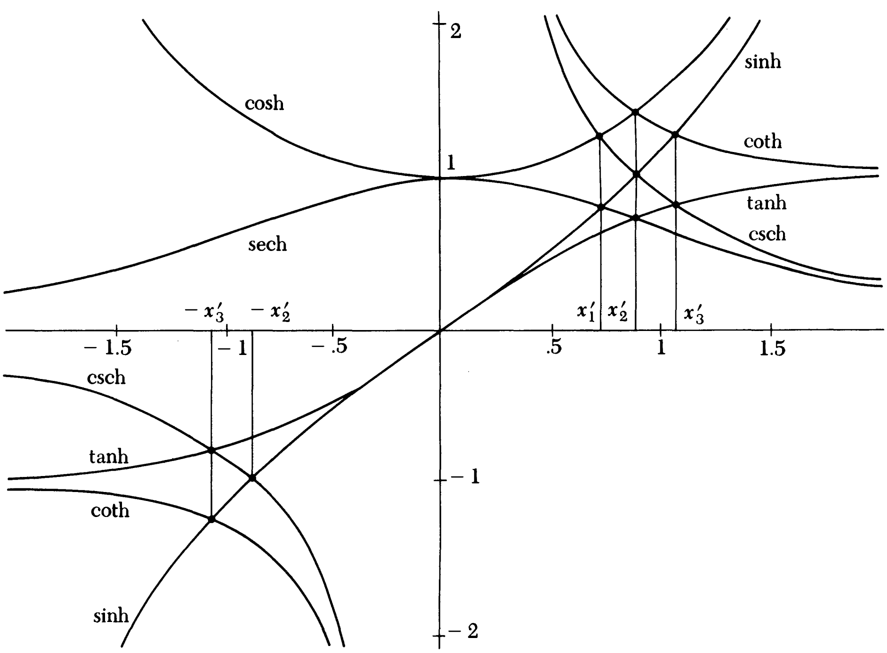


FIGURE 2
The graphs of the hyperbolic functions.

The last equality holds because the identity $\sec^2 x - \tan^2 x = 1$ implies that $(\sec x + \tan x)^{-1} = \sec x - \tan x$.

A third proof based on (1) and (2) is very fast indeed:

$$\sec x + \tan x = \cosh f(x) + \sinh f(x) = e^{f(x)}.$$

Finally, for a fourth proof that completes the novel derivation mentioned in the introduction, we differentiate (1) to obtain

$$\sec^2 x = \cosh f(x)\, f'(x).$$

If we use (2), this simplifies to

$$f'(x) = \sec x.$$

Since $f(0) = 0$ we must have

$$f(x) = \int_0^x f'(t)\, dt = \int_0^x \sec t\, dt.$$

Our function f is the inverse of the Gudermannian function

$$gd(x) = \arctan \sinh x$$

which is sometimes treated in connection with the hyperbolic functions. See, for example, [**3**, pp. 435–439], [**1**, pp. 311–315], and [**2**, pp. 318, 319].

4. The golden triangle Suppose we have an angle θ, $0 < \theta < \pi/2$, for which some pair of trigonometric ratios is equal. Then a right-angled triangle containing θ must have either two of its sides equal or all three of its sides in geometric progression.

The first case yields the familiar isosceles right-angled triangle with sides $1, 1, \sqrt{2}$ and angle $\theta = \pi/4$.

The second case is less familiar and hence more interesting. If we assume the sides are of lengths $1 < s < s^2$ then Pythagoras tells us that $s^4 = s^2 + 1$. It follows that $s^2 = \tau$ and hence $s = \sqrt{\tau}$. We are looking at the triangle with sides $1, \sqrt{\tau}, \tau$ and of course θ can be either of the minor angles.

Figure 3 summarizes, in order of increasing base angle x, $0 < x < \pi/2$, the various possibilities for the equality of two trigonometric functions. The golden triangle PQR with angles $x_1 = \sec^{-1}\sqrt{\tau}$ and $x_3 = \sec^{-1}\tau$ enables us to see immediately why the graphs of sec and cot cross at the same height as the graphs of tan and csc and why this height is $y = \sqrt{\tau}$.

5. Stretching for the hyperbolic facts The results of the last three sections can now be combined to explain the crossings of the hyperbolic functions in the domain $x > 0$.

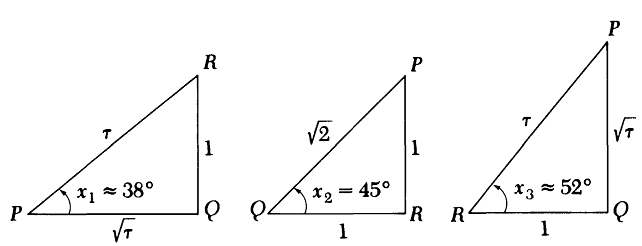


FIGURE 3
The three special angles.

Crossings over $x \leqslant 0$ can be deduced from the obvious symmetries of the functions.

At $x_1 = \sec^{-1}\sqrt{\tau}$, we have $\sec x_1 = \cot x_1 = \sqrt{\tau}$, as in §4. Accordingly at $x_1' = f(x_1)$ we have, from the pairings in §2,

$$\cosh x_1' = \operatorname{csch} x_1' = \sqrt{\tau}$$

and, of course, by taking reciprocals,

$$\operatorname{sech} x_1' = \sinh x_1' = \frac{1}{\sqrt{\tau}}.$$

Using the explicit formula for f derived in §3, we see that the value of x_1' is given by

$$x_1' = \log\left(\sqrt{\tau} + \frac{1}{\sqrt{\tau}}\right) = \log\frac{\tau+1}{\sqrt{\tau}} = \log\frac{\tau^2}{\sqrt{\tau}} = \frac{3}{2}\log\tau.$$

At $x_2 = \pi/4$ we have the familiar equalities:

$$\sec x_2 = \csc x_2 = \sqrt{2}$$
$$\tan x_2 = \cot x_2 = 1$$
$$\sin x_2 = \cos x_2 = \frac{1}{\sqrt{2}}.$$

Accordingly at $x_2' = f(x_2) = \log(1 + \sqrt{2})$ we have

$$\cosh x_2' = \coth x_2' = \sqrt{2}$$
$$\sinh x_2' = \operatorname{csch} x_2' = 1$$
$$\tanh x_2' = \operatorname{sech} x_2' = \frac{1}{\sqrt{2}}.$$

Incidentally, x_2' yields the inflection point in the graph of $y = \operatorname{sech} x$ because

$$\operatorname{sech}'' x = (\tanh^2 x - \operatorname{sech}^2 x)\operatorname{sech} x.$$

Finally, at $x_3 = \sec^{-1}\tau$ we have $\tan x_3 = \csc x_3 = \sqrt{\tau}$. Accordingly at $x_3' = f(x_3) = \log(\tau + \sqrt{\tau})$ we have

$$\sinh x_3' = \coth x_3' = \sqrt{\tau}$$

and, of course,

$$\operatorname{csch} x_3' = \tanh x_3' = \frac{1}{\sqrt{\tau}}.$$

REFERENCES

1. G. Chrystal, *Algebra Part II*, second edition, Black, London, 1931.
2. H. S. M. Coxeter, *Introduction to Geometry*, second edition, Wiley, New York, 1969.
3. W. A. Granville, P. F. Smith, W. R. Langley, *Elements of Differential and Integral Calculus*, revised edition, Ginn and Company, 1941.
4. G. H. Hardy, *A Course of Pure Mathematics*, Cambridge University Press, tenth edition, 1967.
5. D. Laugwitz, *Differential and Riemannian Geometry*, Academic Press, 1965.

On the Measure of Solid Angles

FOLKE ERIKSSON
Chalmers and the University
S-412 96 Gothenburg, Sweden

The measure of a solid angle in 3-space is defined to be the area E of the corresponding spherical triangle $T = ABC$ on the unit sphere, with the center at the vertex O of the angle (FIGURE 1). We denote the unit vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$ by **a**, **b**, and **c**, respectively. The measure E can then be expressed in terms of dot products and the triple product $[\mathbf{a}, \mathbf{b}, \mathbf{c}] = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$:

$$\tan\frac{E}{2} = \frac{|[\mathbf{a},\mathbf{b},\mathbf{c}]|}{1 + \mathbf{b}\cdot\mathbf{c} + \mathbf{c}\cdot\mathbf{a} + \mathbf{a}\cdot\mathbf{b}}. \tag{1}$$

One would think that such a nice formula should be old and well-known. But so far, I have not been able to find it in the literature. However, almost the same result was known to Euler [4, p. 215] and Lagrange [8, p. 340]. (The title above is a translation of the Latin title of [4].) We use the standard notations a, b, c for the sides of our spherical triangle T, and A, B, C for the angles. The area of T is its excess $E = A + B + C - \pi$. The old result can be written

$$\tan\frac{E}{2} = \frac{P}{1 + \cos a + \cos b + \cos c}, \tag{2}$$

where

$$P = (1 - \cos^2 a - \cos^2 b - \cos^2 c + 2\cos a \cos b \cos c)^{1/2}. \tag{3}$$

The denominators in (1) and (2) are obviously the same. Euler and Lagrange knew also that their numerator (3) was the volume of the parallelepiped with edges **a**, **b**, **c**. Let us prove this fact, starting from the modern expression for that volume

$$V = |[\mathbf{a},\mathbf{b},\mathbf{c}]| = \pm\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

in terms of components of the unit vectors. We then have by the product theorem for determinants

$$V^2 = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}^2 = \begin{vmatrix} \mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} & \mathbf{a}\cdot\mathbf{c} \\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{c} \\ \mathbf{c}\cdot\mathbf{a} & \mathbf{c}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c} \end{vmatrix}$$

$$= \begin{vmatrix} 1 & \cos c & \cos b \\ \cos c & 1 & \cos a \\ \cos b & \cos a & 1 \end{vmatrix} = 1 - \cos^2 a - \cos^2 b - \cos^2 c + 2\cos a \cos b \cos c.$$

Thus the equivalence of (1) and (2) is established.

The altitude from A of the parallelepiped with edges **a**, **b**, **c** is (FIGURE 2)

$$h = \sin b \sin C = \sin c \sin B.$$

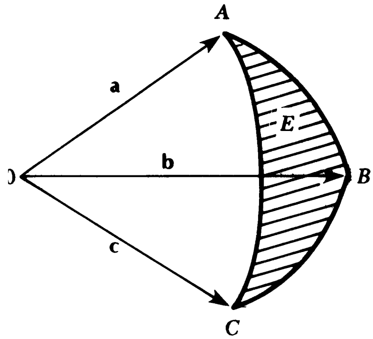


FIGURE 1

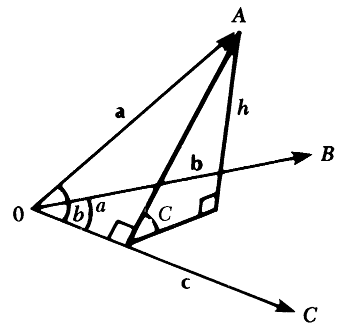


FIGURE 2

This gives, by the way, the Law of Sines:

$$\frac{\sin b}{\sin B}=\frac{\sin c}{\sin C}\left(=\frac{\sin a}{\sin A}\right). \tag{4}$$

Since the volume = (area of base) · altitude, we have

$$P=V=\sin a \sin b \sin C. \tag{5}$$

Proofs of (2) can be found in [4], [8] and Todhunter-Leathem [**9**, p. 104, eq. (30) and p. 28, eq. (17)]. I was not able to find a simpler proof until the referee called my attention to an old theorem, which gives *one* angle representing $(1/2)(\pi - E)$. This theorem [**9**, p. 116, eq. (20)] is obtained from the following construction (FIGURE 3). Let L and M be the midpoints of BC and AC, respectively, in our spherical triangle ABC. The great circles AB and LM intersect in antipodal points P and Q. Draw arcs AA', BB' and CC' perpendicular to LM. Then triangles AMA' and CMC' are congruent, as are BLB' and CLC'. Thus the angle[1] $(A, MA') = (C, MC') = u$, say, and the angle $(B, B'L) = (C, C'L) = v$. Furthermore $AA' = CC' = BB'$. Thus triangles $AA'P$ and $BB'Q$ are congruent, because the angles at P and Q are equal and the angles at A' and B' are right angles. Hence angle $(A, A'P) = (B, QB')$. Call this angle t. We have now (FIGURE 3):

$$A+u+t=\pi,\ B+v+t=\pi,\ C=u+v,$$

and by addition

$$A+B+C+2t=2\pi$$

or

$$2t=2\pi-(A+B+C)=\pi-E.$$

[1] It is convenient to denote the signed angle at A directed from AM to AA' by (A, MA'). In some cases with obtuse angle A, u is to be considered negative, and likewise for v and t (below).

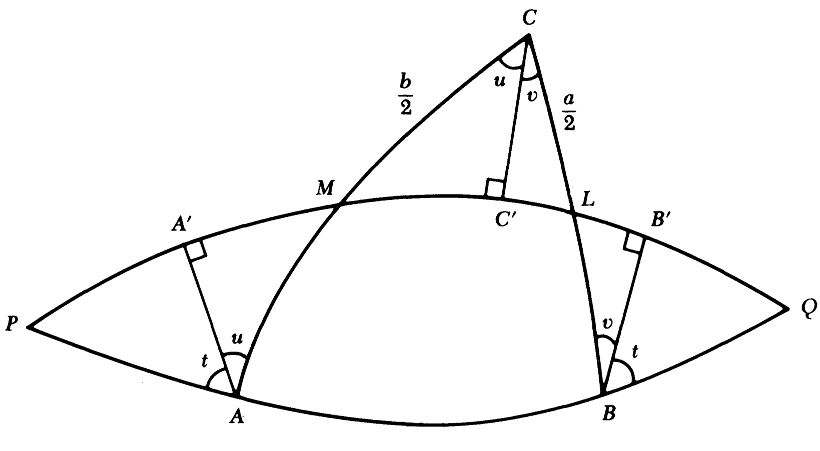


FIGURE 3

Thus

$$\frac{1}{2}(\pi - E) = t = (A, A'P), \tag{6}$$

which is the theorem[2] we need.

Proof of (2). From (6) and FIGURE 3 we get now by elementary spherical trigonometry the following expressions for $\tan E/2$:

$$\tan\frac{E}{2} = \frac{1}{\tan t} = \frac{\sin AA'}{\tan PA'} = \frac{\sin AA'}{\cot ML} = \frac{\sin CC' \sin ML}{\cos ML}$$
$$= \frac{\sin\frac{b}{2} \sin M \sin ML}{\cos ML} = \frac{\sin\frac{b}{2} \sin C \sin\frac{a}{2}}{\cos ML}. \tag{7}$$

(In the second equality we expressed $\tan t$ by a well-known formula for the right-angled triangle $AA'P$. The third equality comes from $PA' + ML = (1/2)PQ = \pi/2$, and the last from the Law of Sines for triangle CML.)

For the denominator $\cos ML$ we get by the Law of Cosines for triangle CML (and in the third equality for ABC)

$$\cos ML = \cos\frac{a}{2}\cos\frac{b}{2} + \cos C \sin\frac{a}{2}\sin\frac{b}{2}$$
$$= \frac{4\cos^2\frac{a}{2}\cos^2\frac{b}{2} + \cos C \sin a \sin b}{4\cos\frac{a}{2}\cos\frac{b}{2}}$$
$$= \frac{(1+\cos a)(1+\cos b) + \cos c - \cos a \cos b}{4\cos\frac{a}{2}\cos\frac{b}{2}}$$
$$= \frac{1 + \cos a + \cos b + \cos c}{4\cos\frac{a}{2}\cos\frac{b}{2}}.$$

[2] I don't know the origin of this theorem. Since [9] gives no reference, and the result is not found in earlier editions of the book (by Todhunter alone), I might guess that it is due to Leathem himself.

Inserting this in the last member of (7) we get

$$\tan\frac{E}{2}=\frac{\sin a\sin b\sin C}{1+\cos a+\cos b+\cos c},$$

which is (2) with the numerator in the form (5).

Remark 1. The importance of the quantity P was recognized by Lagrange in [8]. P, now called the *polar sine* of the solid angle, and its "dual," the 3-*dimensional sine*

$$S={}^3\sin T=\sin a\sin B\sin C$$

were extensively studied by G. Junghann [6, 7] around 1860. The most striking result in Junghann's "tetrahedrometry" is the Law of Sines: *The areas of the faces of a tetrahedron are proportional to the* 3-*dimensional sines of the opposite corners*. This law was given before in 1850 by Joachimsthal [5, p. 40]. It was examined in this MAGAZINE in 1965 by Allendoerfer [1]. Compare also [3]. The "dual" result of (2) is

$$\tan\frac{a+b+c}{2}=\frac{S}{\cos A+\cos B+\cos C-1}.$$

Remark 2. Another simple formula for the excess E of the spherical triangle T, given in Keogh's theorem [9, p. 119], is worth restating here:

$$\sin\frac{E}{2}=\operatorname{polsin}T',$$

where the right member is the polar sine of the triangle $T'=LMN$ with the midpoints of the sides of T as vertices.

Remark 3. The radii R and r of the circumscribed and inscribed circles of the spherical triangle T can be expressed very simply in terms of our S and P:

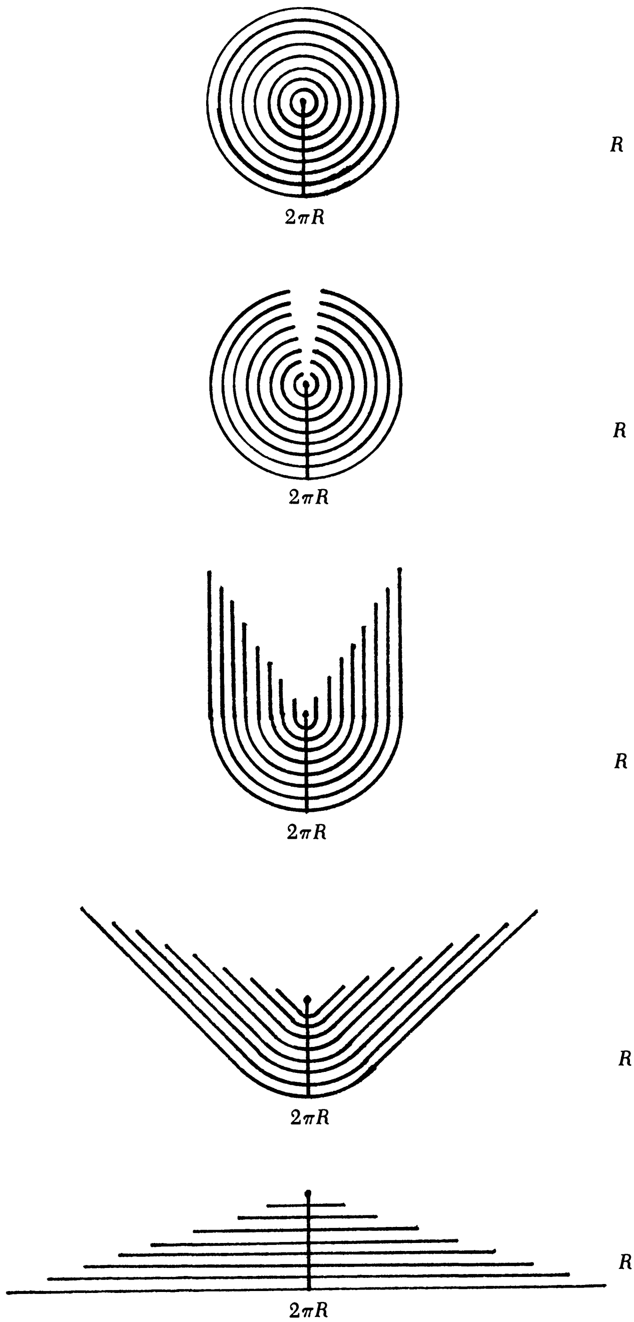
$$\tan R=\frac{2}{S}\sin\frac{E}{2},\quad \cot r=\frac{2}{P}\sin\frac{a+b+c}{2},$$

as in Coxeter's book [2, p. 236f].

REFERENCES

1. C. Allendoerfer, Generalizations of theorems about triangles, this MAGAZINE 38 (1965), 253–259.
2. H. S. M. Coxeter, *Non-Euclidean Geometry*, Toronto, 1942.
3. F. Eriksson, The law of sines for tetrahedra and n-simplices, *Geom. Dedicata* 7 (1978), 71–80.
4. L. Euler, De mensura angulorum solidorum, in *Opera omnia* 26, 204–223. (Orig. in *Acta acad. sc. Petrop.* 1778.)
5. F. Joachimsthal, Sur quelques applications des déterminants à la Géométrie, *J. Math.* (Crelle) 40 (1850), 21–47.
6. G. Junghann, Beiträge zur Tetraedrometrie, *Arch. Math. Phys.* (Grunert) 34 (1860), 369–396.
7. ______, *Tetraedrometrie I-II*, Gotha, 1862–63.
8. J.-L. Lagrange, Solutions de quelques problèmes relatifs au triangles sphériques, in *Oeuvres* 7, 331–359. (Orig. in *J. de l'Ecole Polyt.* 1798.)
9. I. Todhunter and J. G. Leathem, *Spherical Trigonometry*, Macmillan, London, 1901.

Proof without words: Area of a Disk is πR^2



$A = (1/2)(2\pi R)R = \pi R^2$

Russell Jay Hendel
Dowling College
Oakdale, NY 11772

PROBLEMS

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

GEORGE GILBERT, *associate editor*
St. Olaf College

Proposals

To be considered for publication, solutions should be received by November 1, 1990.

1348. *Proposed by David Callan, University of Bridgeport, Bridgeport, Connecticut.*

A *Monthly* problem (E 3194, Feb 1987; solution March 1989) reads: "Let S be the smallest set of rational functions containing x and y, and closed under subtraction and reciprocals [of nonzero functions]. Show that $1 \notin S$." Characterize those rational functions that are in S.

1349. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, Maryland.*

Let K be a field, n a positive integer, and $\mathbf{I}$ the $n \times n$ identity matrix. Give necessary and sufficient conditions on n and K such that for every $n \times n$ matrix $\mathbf{A}$ over K there is an element a in K such that $\mathbf{A} + a\mathbf{I}$ is invertible.

1350. *Proposed by Hugh Noland, Chico, California.*

In the well-known Tower of Hanoi puzzle one starts with three pegs, two of which are empty and one of which contains n disks, no two of the same size, stacked in order of size, with the smallest on top. It is required to move all the disks to one of the empty pegs, by moving one disk at a time, subject to the condition that no disk ever rest on a smaller one. It is easy to show that the number of moves required is $2^n - 1$.

Suppose instead that one has $2n$ disks, numbered according to size, with the smallest numbered 1. If all the odd-numbered disks occupy one peg and all the even numbered disks another, stacked according to size, how many moves are required to move all the disks onto the empty peg, the requirement again being that no disk ever rest on a smaller?

1351. *Proposed by Florin S. Pîrvănescu, Slatina, Romania.*

In the acute triangle ABC, let D be the foot of the perpendicular from A to BC, let E be the foot of the perpendicular from D to AC, and let F be a point on the line segment DE. Prove that AF is perpendicular to BE if and only if $FE/FD = BD/DC$.

1352. *Proposed by Mark Krusemeyer, Carleton College, Northfield, Minnesota.*

a. Suppose three lines are drawn independently and in random directions through the origin in the plane. (The lines will each extend in two opposite directions from the origin; "random" means that given two equal angles with vertex at the origin, each line is equally likely to be inside one as inside the other.) What is the probability that all the angles formed at the origin by adjacent pairs of lines will be acute? (For example, if the lines are $y = 0$, $y = x$, $y = 2x$, then the angle formed by $y = 2x$ and $y = 0$ as an adjacent pair of lines at the origin will not be acute. However, if the lines are $y = 0$, $y = 2x$, $y = -2x$, then all angles at the origin will be acute.)

b. Same question, with "three lines" replaced by "n lines."

Quickies

Answers to the Quickies are on page 198.

Q763. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Canada.*

Determine all real solutions of the simultaneous equations

$$2x(1+y+y^2) = 3(1+y^4),$$
$$2y(1+z+z^2) = 3(1+z^4),$$
$$2z(1+x+x^2) = 3(1+x^4).$$

Q764. *Proposed by José Heber Nieto, Universidad del Zulia, Venezuela.*

Prove that infinitely many 4-tuples of integers a, b, c, d exist such that $a > b > c > d > 1$ and $a!d! = b!c!$.

Q765. *Proposed by Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

Show that for any given integer a, the Diophantine equation $(a^2-1)(b^2-1) = c^2-1$ has at least two distinct solutions.

Solutions

All of the problems with solutions in this issue were posed by Murray Klamkin. These problems, from the June 1989 issue, were published in recognition of Professor Klamkin's lifelong contributions to problem solving.

Circumscribed polygons **June 1989**

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

Solution by the proposer.

If we denote the consecutive angles of the n-gon by $2A_1, 2A_2, \ldots, 2A_n$, then

$$P \equiv \prod a_i = \prod (\cot A_i + \cot A_{i+1}),$$

where the products and sums here and subsequently are cyclic over i and $\Sigma A_i = (n-2)\pi/2$ with $A_i < \pi/2$.

First, we consider $n > 4$. By taking two consecutive angles very close to π, we can make P arbitrarily small; that is to say, there is no minimum in this case.

Consider the case $n = 4$. Since $\cot x$ is convex in $(0, \pi/2)$,

$$\cot x + \cot y \geqslant 2 \cot \frac{(x+y)}{2}.$$

Hence,

$$P = a_1 a_2 a_3 a_4 \geqslant 16 \cot \frac{A_1 + A_2}{2} \cdot \cot \frac{A_2 + A_3}{2} \cdot \cot \frac{A_3 + A_4}{2} \cdot \cot \frac{A_4 + A_1}{2}.$$

Then, since

$$\cot \frac{A_3 + A_4}{2} = \frac{1}{\cot \frac{A_1 + A_2}{2}} \quad \text{and} \quad \cot \frac{A_4 + A_1}{2} = \frac{1}{\cot \frac{A_2 + A_3}{2}},$$

we get that

$$a_1 a_2 a_3 a_4 \geqslant 16.$$

This is a stronger result than the minimum perimeter circumscribed quadrilateral is a square.

Now consider the case $n = 3$. We start with the known inequality $(a_1 a_2 a_3)^2 \geqslant (4F/\sqrt{3})^3$, where F is the area of the triangle and equality holds if $a_1 = a_2 = a_3$. Since $a_1 a_2 a_3 = 4RF$, where R is the circumradius, a geometrical interpretation of this inequality is that the inscribed triangle of largest area in a circle of radius R is the equilateral one. Now, $F = \cot A_1 + \cot A_2 + \cot A_3$, and it is known that the minimum area circumscribed triangle is the equilateral one. This follows easily from the convexity of $\cot x$ for x in $(0, \pi/2)$. Again there is equality only for the equilateral problem. Thus $\min(a_1 a_2 a_3) = 24\sqrt{3}$, or equivalently,

$$\prod (\cot A_i + \cot A_{i+1})^2 \geqslant \left(4 \sum \cot A_i / \sqrt{3}\right)^3.$$

Also solved by Duane M. Broline, Francis M. Henderson, and L. Kuipers (Switzerland).

Characterization of parallelepiped **June 1989**

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

Solution by the proposer.

First, the polyhedron must be convex. If not, there would be a pair of reentrant faces and the area of cross sections parallel to one of these two faces could not be the same. We now show that the polyhedron must be a parallelepiped. Consider three parallel sections whose distances from a face F are x, x_1, and x_2, where $x = w_1 x_1 + w_2 x_2$, $w_1 + w_2 = 1$ and $w_1, w_2 > 0$. It then follows by the Brun-Minkowski inequality (L. A. Lyusternik, *Convex Figures and Polyhedra*, Dover, New York, 1963, pp. 117–118), that the areas of the three sections must satisfy

$$A(x) \geqslant w_1 A(x_1) + w_2 A(x_2),$$

and equality holds if and only if the region between the outer sections is a cylindrical solid. Since we have equality by hypothesis, the figure must be a prism with respect to each face and hence must be a parallelepiped.

Comment from the proposer: I set the same problem with *area* replaced by *perimeter* in the 1980 Canadian Mathematics Olympiad. In this case the figure could also be a regular octahedron. Whether or not there are any other solutions for this problem is still open.

Maximization with restraint **June 1989**

1324. Determine the maximum value of

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

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

I. *Solution by Eugene Lee, Boeing Commercial Airplanes, Seattle, Washington.*

The maximum is $(1/n)^{n+1}$. To prove this, we establish inductively the proposition P_n: The function $f_n(x_1, \dots, x_n) \equiv x_1 \cdots x_n(x_1^2 + \cdots + x_n^2)$ over the simplex $\sigma_{n-1}(\lambda) \equiv \{(x_1, \dots, x_n) : x_i \geqslant 0, \ \Sigma_i^n x_i = \lambda\}$, any $\lambda > 0$, attains its maximum uniquely when $x_1 = \cdots = x_n$.

One way to prove P_2 is to observe that

$$f_2(x, \lambda - x) = f_2(\lambda/2, \lambda/2) - 2(x - \lambda/2)^4.$$

Now let $n \geqslant 3$. Fix x_n, $0 < x_n < \lambda$, and rewrite f_n as

$$f_n(x_1, \dots, x_n) = x_n f_{n-1}(x_1, \dots, x_{n-1}) + x_n^3 \prod_{i=1}^{n-1} x_i.$$

The induction hypothesis says that $f_{n-1}(x_1, \dots, x_{n-1})$ has a unique maximum over $\sigma_{n-2}(\lambda - x_n)$ at $x_1 = \cdots = x_{n-1}$. But the same is true of $\Pi_{i=1}^{n-1} x_i$ (from the arithmetic mean–geometric mean inequality). Hence, for any fixed $0 < x_n < \lambda$, $f_n(x_1, \dots, x_{n-1}, x_n)$ attains its maximum uniquely when $x_1 = \cdots = x_{n-1}$.

The role of x_n being replaceable by any x_k, we have proved P_n. (For if $(x_1, \dots, x_n)$ lies in the interior of $\sigma_{n-1}(\lambda)$ with $x_i \neq x_j$ for some i, j, then, taking k different from i or j, and letting $y_k = x_k$, $y_r = (\lambda - x_k)/(n-1)$ for $r \neq k$, we see that $f_n(x_1, \dots, x_n) < f_n(y_1, \dots, y_n)$.)

II. *Solution by Professor Freidkin, University of the Witwatersrand, Johannesburg, Republic of South Africa.*

Let $F(x_1, x_2, \dots, x_n) \equiv x_1 x_2 \cdots x_n(x_1^2 + x_2^2 + \cdots + x_n^2)$. We will show that F attains its maximum value (of $n^{-(n+1)}$) when all of the arguments are equal.

Suppose that two of the arguments of F, say x_1 and x_2, are not equal. Then we have the following.

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= x_1 x_2 \prod_{i=3}^{n} x_i \left(x_1^2 + x_2^2 + \sum_{i=3}^{n} x_i^2 \right) \\ &= \left(\frac{x_1 + x_2}{2} + \frac{x_1 - x_2}{2} \right)\left(\frac{x_1 + x_2}{2} - \frac{x_1 - x_2}{2} \right) \\ &\quad \times \prod_{i=3}^{n} x_i \left(\frac{1}{2}(x_1 + x_2)^2 + \frac{1}{2}(x_1 - x_2)^2 + \sum_{i=3}^{n} x_i^2 \right) \end{aligned}$$

$$= \frac{1}{8}\left((x_1+x_2)^4 - (x_1-x_2)^4\right)\prod_{i=3}^{n} x_i$$
$$+ \frac{1}{4}\prod_{i=3}^{n} x_i\left((x_1+x_2)^2 - (x_1-x_2)^2\right)\sum_{i=3}^{n} x_i^2$$
$$< \frac{1}{8}(x_1+x_2)^4\prod_{i=3}^{n} x_i + \frac{1}{4}\prod_{i=3}^{n} x_i(x_1+x_2)^2\sum_{i=3}^{n} x_i^2$$
$$\equiv F\left(\frac{x_1+x_2}{2}, \frac{x_1+x_2}{2}, x_3, \dots, x_n\right).$$

Thus, the maximum must occur when all the x_i's are equal.

III. *Solution by the proposer.*

We will show that

$$\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)^{n+2} \geqslant x_1x_2\cdots x_n\left(\frac{x_1^2+x_2^2+\cdots+x_n^2}{n}\right) \qquad (1)$$

so that the desired maximum value is n^{-n-1} and is taken on when all the x_i are equal.

Our proof is by induction. First, (1) is valid for $n=2$, since it reduces to $(x_1-x_2)^4 \geqslant 0$. We now assume (1) is valid for $n=k$ and we will prove it valid for $n=k+1$. Let $A=(x_1+x_2+\cdots+x_k)/k$ and $P=x_1x_2\cdots x_k$. Then from (1), with $n=k$,

$$Px\left(x_1^2+x_2^2+\cdots+x_k^2\right)+Px^3 \leqslant kxA^{k+2}+Px^3 \leqslant kxA^{k+2}+x^3A^k.$$

It now suffices to show that

$$(k+1)\left(\frac{x_1+x_2+\cdots+x_k+x}{k+1}\right)^{k+3} = (k+1)\left(\frac{kA+x}{k+1}\right)^{k+3} \geqslant kxA^{k+2}+x^3A^k$$

or equivalently, that $kxA^{k+2}+x^3A^k \leqslant k+1$, where, without loss of generality, we have assumed that $kA+x=k+1$.

Using the standard calculus technique, we differentiate with respect to A and set it to zero,

$$D_A(kxA^{k+2}+x^3A^k) = k(k+2)xA^{k+1}+kx^3A^{k-1}+(kA^{k+2}+3x^2A^k)(-k)=0$$

or

$$\left(kt^3-(k+2)t^2+3t-1\right)t^{k-1}=0,$$

where $t=A/x$. The cubic factors into $(t-1)(kt^2-2t+1)$. The only real roots are 0 and 1. The maximum occurs for $A=x=1$ and we have proved the case for $n=k+1$. Thus the result is valid for all $n=2,3,4,\dots$.

Also solved by Seung-Jin Bang (Republic of Korea), Todd R. Bault, Paul Bracken (Canada), Duane M. Broline, Centre College Problem Solving Group, Robert Doucette, François Dubeau (Canada), Michael Golomb, Russell Jay Hendel, L. Kuipers (Switzerland), Kee-Wai Lau (Hong Kong), Patrick Dale McCray, Jean-Marie Monier (France), and Heinz-Jürgen Seiffert (West Germany). Most solutions were based on the method of Lagrange multipliers.

Generalization of a 1959 Putnam problem **June 1989**

1325. a. Determine the minimum value of

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

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

b. Determine the minimum value of

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

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

Solution by the proposer.

a. We will show that the minimum is $(n-1)/2n$. The first part of the proof is indirect. Assume that for any F_i and all x_i that

$$|F_1(x_1) + F_2(x_2) + \cdots + F_n(x_n) - x_1 x_2 \cdots x_n| < a \leqslant \frac{n-1}{2n}.$$

Let

$$S_0 = F_1(0) + F_2(0) + \cdots + F_n(0),$$
$$S_1 = F_1(1) + F_2(1) + \cdots + F_n(1).$$

We have

$$|S_0| < a, \qquad |1 - S_1| < a,$$

and for $j = 1, 2, \ldots n$,

$$|S_1 - F_j(1) + F_j(0)| < a.$$

Thus,

$$a + (n-1)a + na > |S_0| + (n-1)|1 - S_1| + \sum_{j=1}^{n} |S_1 - F_j(1) + F_j(0)|$$

and, therefore,

$$2na > |-S_0 + (n-1)(1 - S_1) + (n-1)S_1 + S_0| = n - 1.$$

Hence,

$$T_1 \equiv |F_1(x_1) + F_2(x_2) + \cdots + F_n(x_n) - x_1 x_2 \cdots x_n| \geqslant \frac{n-1}{2n}$$

for some choice of the x_i's. That the minimum is $(n-1)/2n$ will follow by the choice of functions

$$F_i(x_i) \equiv \frac{x_i}{n} - \frac{n-1}{2n^2}$$

for all i. Here,

$$T_1 = \left| \sum_{i=1}^{n} \left(\frac{x_i}{n} - \frac{n-1}{2n^2} \right) - x_1 x_2 \cdots x_n \right|$$

and all we need now show is that $T_{\max} = (n-1)/2n$. Since T is linear in the x_i's, its maximum will be taken on by each variable being 0 or 1. For all ones, $T = (n-1)/2n$. For $r(<n)$ ones and $n-r$ zeros,

$$T = \left|\frac{r}{n} - \frac{n-1}{2n}\right| \leqslant \frac{n-1}{2n}$$

and with equality only for $r = n-1$.

Comment: The special case of showing that the minimum is greater than or equal to $1/4$ for $n = 2$ was a 1959 Putnam problem (A. M. Gleason, R. E. Greenwood, L. M. Kelly, *The William Lowell Putnam Mathematical Competition, Problems and Solutions, 1938–1964*, MAA, Washington, D.C., 1980, p. 499).

b. First consider the even case; that is, replace n by $2n$. We will show that the minimum is $n/4$. The first part of the proof is indirect. Assume that for any F_i and all x_i that

$$|F_1(x_1)F_2(x_2) \cdots F_{2n}(x_{2n}) - x_1 - x_2 - \cdots - x_{2n}| < a \leqslant n/4.$$

Let

$$P_0 = |F_1(0)F_2(0) \cdots F_{2n}(0)|,$$
$$P_1 = |F_1(1)F_2(1) \cdots F_{2n}(1)|.$$

We then have,

$$-a < P_0 < a, \quad 2n - a < P_1 < 2n + a, \quad P_0P_1 < a(2n+a),$$
$$n - a < |F_1(0)F_2(0) \cdots F_n(0)F_{n+1}(1)F_{n+2}(1) \cdots F_{2n}(1)| < n + a.$$

We now take all combinations similar to the last inequality with n zeros and n ones and multiply them to give

$$(n-a)^\alpha < (P_0P_1)^{\alpha/2} < (n+a)^\alpha,$$

where $\alpha = \binom{2n}{n}$. Equivalently,

$$(n-a)^2 < P_0P_1 < (n+a)^2.$$

Also,

$$(n-a)^2 < P_0P_1 < a(2n+a)$$

and from this we get $a > n/4$. Hence,

$$T_2 \equiv |F_1(x_1)F_2(x_2) \cdots F_{2n}(x_{2n}) - x_1 - x_2 - \cdots - x_{2n}| \geqslant n/4$$

for some choice of the x_i's. That the minimum is $n/4$ will follow by the choice of functions $F_i(x_i) \equiv ax_i + b$ for all i where a, b are taken to satisfy

$$(a+b)^{2n} = 2n + n/4 \quad \text{and} \quad a^{2n} = n/4.$$

Here,

$$T_2 = \left|\prod_{i=1}^{2n}(ax_i + b) - x_1 - x_2 - \cdots - x_{2n}\right|.$$

Since T_2 is linear in the x_i's, it takes on its extreme values when all the x_i's are 0 or 1. Thus it now remains to show that

$$(a+b)^r a^{2n-r} \leqslant (n+4r)/4$$

or, equivalently, that

$$(9n)^r (n)^{2n-r} \leqslant (n+4r)^{2n}$$

for $r = 0, 1, \ldots, 2n$. Letting

$$\varphi(r) \equiv 2n\ln(n+4r) - r\ln 9n - (2n-r)\ln n,$$

we find that $\varphi(0) = \varphi(2n) = 0$ and $\varphi''(r) < 0$ which imply that $\varphi(r) \geqslant 0$.

Now consider the odd case; that is, replace n by $2n-1$. We will show that the minimum is $n(n-1)/2(2n-1)$. As before, the first part of the proof is indirect. Assume that for any F_i and all x_i that

$$|F_1(x_1)F_2(x_2)\cdots F_{2n-1}(x_{2n-1} - x_1 - x_2 - \cdots - x_{2n-1})| < a \leqslant \frac{n(n-1)}{2(2n-1)}.$$

Let

$$P_0 = |F_1(0)F_2(0)\cdots F_{2n-1}(0)|,$$

$$P_1 = |F_1(1)F_2(1)\cdots F_{2n-1}(1)|.$$

We then have

$$-a < P_0 < a, \quad 2n-1-a < P_1 < 2n-1+a, \quad P_0P_1 < a(2n-1+a),$$

$$n-a < |F_1(0)F_2(0)\cdots F_{n-1}(0)F_n(1)F_{n+1}(1)\cdots F_{2n-1}(1)| < n+a$$

$$n-1-a < |F_1(0)F_2(0)\cdots F_n(0)F_{n+1}(1)F_{n+2}(1)\cdots F_{2n-1}(1)| < n-1+a.$$

We now take all combinations similar to the last inequality with n zeros and $n-1$ ones, and all combinations similar to the next to the last inequality with $n-1$ zeros and n ones, and multiply all of them together to give (after taking a root)

$$(n-a)(n-1-a) < P_0P_1 < (n+a)(n-1+a).$$

Then

$$(n-a)(n-1-a) < a(2n-1+a)$$

or

$$a > \frac{n(n-1)}{2(2n-1)}.$$

Hence,

$$T_2 \equiv |F_1(x_1)F_2(x_2)\cdots F_{2n-1}(x_{2n-1}) - x_1 - x_2 - \cdots - x_{2n-1}| \geqslant \frac{n(n-1)}{2(2n-1)}$$

for some choice of the x_i's. That the minimum is $n(n-1)/2(2n-1)$ will follow by the choice of functions $F_i(x_i) \equiv ax_i + b$ for all i where a, b satisfy

$$(a+b)^{2n-1}=\frac{n(n-1)}{2(2n-1)} \quad \text{and} \quad a^{2n-1}=\frac{n(n-1)}{2(2n-1)}.$$

Since T_2 is linear in the x_i's, it takes on its extreme values when all the x_i's are 0 or 1. Thus it now remains to show that

$$(a+b)^r a^{2n-1-r} < \frac{n(n-1)}{2(2n-1)} + r$$

or, equivalently, that

$$(9n^2-9r+2)^r (n^2-n)^{2n-1-r} \leqslant (n^2+n(4r-1)-2r)^{2n-1}$$

for $r = 0, 1, \ldots, 2n-1$. The proof goes through as before using concavity.

Also solved by Duane M. Broline and Michael Golomb.

Projectiles subject to drag **June 1989**

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

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

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

Solution by Michael Golomb, Purdue University, West Lafayette, Indiana.

We may assume the unit of mass is chosen so that $m = 1$. The equation of motion during ascent is

$$v\,dv/dx = -g - F(v^2/c), \qquad v(0) = v_0 > 0.$$

At the maximum height a the velocity is 0, thus

$$\int_0^v \frac{v\,dv}{g+F(v^2/c)} = -\int_a^x dx = a - x.$$

We set $v^2/c = u$ and obtain

$$\int_0^{v^2/c} \frac{du}{g+F(u)} = \frac{2}{c}(a-x).$$

In the same way, we obtain for the descent

$$\int_0^{v^2/c} \frac{du}{g-F(u)} = \frac{2}{c}(a-y).$$

Thus, if (1) is to hold then

$$\exp\left(\int_0^{v^2/c} \frac{du}{g+F(u)}\right) + \exp\left(\int_0^{v^2/c} \frac{du}{-g+F(u)}\right) = 2. \tag{2}$$

Set

$$w = v^2/c, \quad P(w) = \frac{1}{g + F(w)}, \quad Q(w) = \frac{1}{-g + F(w)}.$$

Then differentiating (2) twice with respect to w, we obtain

$$(P' + P^2)e^{\int_0^w P} + (Q' + Q^2)e^{\int_0^w Q} = 0. \tag{3}$$

But $P' = -F'P^2$, $Q' = -F'Q^2$, hence (3) becomes

$$(1 - F')(P^2 e^{\int_0^w P} + Q^2 e^{\int_0^w Q}) = 0.$$

Since $P^2 e^{\int P} + Q^2 e^{\int Q} > 0$, we conclude $F' = 1$, i.e., $F(w) = w + k$, where k is a constant. With this F, (2) becomes

$$\frac{g + v^2/c + k}{g + k} + \frac{-g + v^2/c + k}{-g + k} = 2.$$

It is readily seen that this equation implies $k = 0$, thus $F(w) = w$.

Also solved by A. Bhattacharya and the proposer.

Answers

Solutions to the Quickies on p. 190.

A763. Since $1 + x + x^2 > 0$, etc., it follows that x, y, z are all positive. Without loss of generality we may assume $x \geqslant y \geqslant z$. Then

$$2x(1 + x + x^2) \geqslant 3(1 + x^4)$$

so that

$$0 \geqslant (x - 1)^2(3x^2 + 4x + 3).$$

Thus $x = 1$ giving the one real solution $x = y = z = 1$.

A764. All integers $n > 2$ satisfy the identity

$$(n^2 + n)!(n - 1)! = (n^2 + n - 1)!(n + 1)!$$

and $n^2 + n > n^2 + n - 1 > n + 1 > n - 1$.

A765. Note first that the product of any four consecutive integers is always one less than a perfect square since

$$x(x + 1)(x + 2)(x + 3) + 1 = (x^2 + 3x)(x^2 + 3x + 2) + 1 = (x^2 + 3x + 1)^2.$$

Using this with $x = a - 1$ and $b = a + 1$, one readily verifies that

$$(a^2 - 1)(b^2 - 1) = (a - 1)a(a + 1)(a + 2) = (a^2 + a - 1)^2 - 1$$

yielding the solution $b = a + 1$ and $c = a^2 + a - 1$. Similarly, with $x = a - 2$ and $b = a - 1$ one obtains

$$(a^2 - 1)(b^2 - 1) = (a - 2)(a - 1)a(a + 1) = (a^2 - a - 1)^2 - 1$$

yielding the solution $b = a - 1$ and $c = a^2 - a - 1$.

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.

Dunham, William, *Journey Through Genius: The Great Theorems of Mathematics*, Wiley, 1990; xiii + 300 pp, $19.95. ISBN 0-471-50030-5

Explores for the general reader some of the enduring ideas of mathematics, with historical and biographical perspectives—an excellent text for a liberal arts mathematics course! The book examines the creativity involved, by presenting the proofs. The theorems are selected on the basis of several considerations, including being particularly insightful or ingenious, being important, and being accessible to readers with only high-school algebra and geometry. You may have made a different selection; here is Dunham's: Hippocrates' quadrature of the lune, Euclid's proofs of the Pythagorean theorem and of the infinitude of primes, Archimedes' area of the circle, Heron's formula, Cardano's solution of the cubic, Newton's binomial theorem, Bernoulli's divergence of the harmonic series, Euler's summation of the reciprocals of the squares of the integers and his refutation of Fermat's conjecture of primality of Fermat numbers, Cantor's non-denumerability of the continuum and his generation by power set of greater and greater cardinals.

Guy, Richard, *Fair Game: How to Play Impartial Combinatorial Games*, COMAP, 1989; vii + 113 pp + errata sheet, $12.95 (P). ISBN 0-912843-16-0.

"All two-person impartial games (ones in which the same moves are available no matter whose play it is) reduce to Nim." This delightful booklet builds up to (and beyond) this fact through a richly varied multitude of example games, accompanied by 62 exercises (with solutions). Only high-school mathematics is required.

Parker, Donn B., et al., *Ethical Conflicts in Information and Computer Science, Technology, and Business*, QED Information Sciences (QED Plaza, Box 82-181, Wellesley, MA 02181), 1990; 214 pp, (P). ISBN 0-89435-313-6

Reports on the opinions of 34 professionals who were presented with scenarios depicting ethical problems, and on the general principles that these individuals suggested. The 53 scenarios are given, together with comments by the surveyed individuals. Topics treated include professional standards, obligations, and accountability; property ownership; confidentiality, business practices; and employer-employee relationships. In an era in which the behavior of many computer science students is at variance with what society and the profession expect, we need to engage students from the first computer science course on with ethical issues and the attitudes of professionals.

Kolata, Gina, 1-in-a-trillion coincidence, you say? Not really, experts find, *New York Times* (25 February 1990), B5, B8.

What are the odds of someone, somewhere in the US, winning a lottery twice in a given four month period? (Before you read further, do a little back-of-the-envelope calculation yourself.) Statisticians Persi Diaconis and Frederick Mosteller of Harvard are trying to demystify coincidences. They reviewed and analyzed a large body of them, distilling their "law of the very large numbers": With a large enough sample, any outrageous thing is apt to happen. (The answer: Better than 1 in 30.)

Fukagawa, H., and D. Pedoe, *Japanese Temple Geometry Problems: San Gaku*, Charles Babbage Research Centre (Box 272, St. Norbert Postal Station, Winnipeg, Canada R3V 1L6), 1989; xvi + 206 pp (P).

"A selection from the hundreds of problems in Euclidean geometry displayed on the devotional mathematical tablets (SANGAKU) which were hung under the roof of shrines or temples in Japan during two centuries of schism from the west, with solutions and answers." This is the first publication in a Western language of a book on Japanese temple geometry, which includes (prior to their discovery in the West) the Malfatti theorem and the Soddy hexlet theorem, together with a large number of theorems involving ellipses and circles that have never appeared in the West.

Murray, J.D., *Mathematical Biology*, Springer-Verlag, 1989; xiv + 767 pp, $59. ISBN 0-387-19460-6

A rich and masterful treasure trove of deterministic biological models (stochastic models were excluded for lack of space). Interested readers need to be familiar with calculus and differential equations. Topics include discrete and continuous population models (for single populations and for interacting populations), reaction kinetics and biological oscillators, biological waves, animal coat patterns, and epidemic models.

Euler, Leonhard, *Introduction to Analysis of the Infinite*, Book II, transl. John D. Blanton, Springer-Verlag, 1990; xii + 504 pp, $59.

First English translation of the second volume of Euler's famous work, this one concerns "higher geometry": theory of curves, conics, cubic curves, tangents to curves, curvature, theory of solids, and surfaces of solids.

Davis, Ronald M. (ed.), *A Curriculum in Flux: Mathematics at Two-Year Colleges*, MAA, 1989; vi + 48 pp, $7.50 (P). ISBN 0-88386-065-6

Report of the Joint Subcommittee (of AMATYC and the MAA) on Mathematics Curriculum at Two-Year Colleges. Calculators and computers are now an essential part of all mathematics courses; statistics is an area now essential for all students; all courses must incorporate problem solving that involves higher-order reasoning and critical thinking; remedial instruction should be reconsidered; all courses should increase their content of discrete mathematics. Included are course outlines, including those for courses in technical and vocational mathematics.

Crossley, J.N., *The Emergence of Number*, 2nd ed., World Scientific, 1987; x + 222 pp, $45. ISBN 9971-50-413-8; $26 (P). ISBN 9971-50-414-6

Revision of a 1980 book that attempts "to allow the modern reader to see through the eyes of the mathematicians of the past." Natural numbers, complex numbers, and real numbers form the three parts of the book, which is replete with (translated) quotations from original sources.

Copp, Newton, *Vaccines: An Introduction to Risk* (from John G. Truxal, Mathematics Dept., SUNY, Stony Brook, NY 11794); 78 pp, (P).

Detailed history of the discovery of smallpox vaccine, followed by a very careful risk assessment of the value of vaccination for smallpox, swine flu, and whooping cough. Nine problems are included.

Borwein, Jonathan M., and Peter B. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, Wiley, 1986; 400 pp, $51.95. ISBN 0-471-83138-7

The AGM of the title is the arithmetic-geometric mean iteration of Gauss, the two term recursion

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_n b_n}\ .$$

The limit of this simple process is an elliptic integral; study leads quickly to theta functions, quadratic algorithms for pi, the Rogers-Ramanujan identities, the Mellin transform, and other delights.

Solow, Daniel, *How to Read and Do Proofs: An Introduction to Mathematical Thought Processes*, 2nd ed., Wiley, 1990; xx + 242 pp, $19.95 (P).

Every math student who ever needs to write a proof should have this book, and math majors should sleep with it under their pillows! Solow systematizes the process of writing a proof; after his book, a student never again has to wonder where or how to start on a proof. What's left for the student to concentrate on, of course, is the mathematics. This second edition includes a few refinements of the techniques, improves discussions, and twice as many exercises (with answers only to the odd-numbered ones). No mathematics beyond high school is required.

Cipra, Barry A., Mathematics untwists the double helix, *Science* (23 February 1990) 913-915; Molecular biologists team up with mathematicians to investigate DNA, *SIAM News* 23:2 (March 1990), 1, 16.

DNA, whether in linear or circular form, lends itself well to mathematical analysis, because mathematicians can disregard the biochemical details and consider it as a curve in space—"grist for the mill of differential geometers and topologists." As a curve, DNA links, twists, writhes, and knots as it wraps around itself ("supercoils"). Mathematical results, such as James White's 1968 theorem that Link = Twist + Writhe, and Vaughn Jones's 1984 discovery of new polynomial invariants of knots, have proven increasingly valuable to molecular biologists.

Mallios, William S., *Statistical Modeling: Applications in Contemporary Issues*, Iowa State U Pr, 1989; xiv + 233 pp, $38.95. ISBN 0-8138-0307-1

Case studies of statistical modeling based on regression and discriminant analysis (hence at an advanced level). Issues investigated include terrorism, drunk driving, discrimination, gambling, currency fluctuations, and the adequacy of the jury system.

Corcoran, Elizabeth, Profile: Fischer Black: Calculated risks enable mathematician to turn profit, *Scientific American* (March 1990), 78-79.

A Ph.D. mathematician, formerly on the faculty of MIT, now a partner with the Wall Street trading firm Goldman Sachs, successfully modeled the price of stock options and is now promoting a new model for a portfolio of international stocks and currencies.

Corcoran, Elizabeth, Profile: Not just a pretty face: Compressing pictures with fractals, *Scientific American* (March 1990), 77-78.

Michael Barnsley (of fractal fame) has discovered a method to compress data by a factor of 500 or more, using fractals. Barnsley, a former academic, refuses to disclose how the method works and is trying to secure patents to keep it proprietary. Barnsley's company is producing decoders that will plug into a microcomputer and are fast enough for real-time video at 30 frames per second.

Truxal, John G., *Feedback—Automation* (from John G. Truxal, Mathematics Dept., SUNY, Stony Brook, NY 11794); 84 pp, (P).

Introductory treatment of feedback, with interesting applications. An appendix treats elevator control theory and includes real data. A dozen project-problems follow. This is one in a series of monographs of the New Liberal Arts Program of the Sloan Foundation. Other titles include *Vaccines: An Introduction to Risk*, by Newton Copp (reviewed above); *Probability Examples*, by John G. Truxal; *Information Theory*, by Morton A. Tavel; and *Expert Systems: Basic Concepts*, by Joseph D. Bronzino and Ralph A. Morelli.

Stillwell, John, *Mathematics and Its History*, Springer-Verlag, 1989; x + 369 pp, $49.80. ISBN 0-387-96981-0

Gives "a unified view of undergraduate mathematics by approaching the subject through its history," for the benefit of senior undergraduates.

Lewis, P.A.W., and E.J. Orav, *Simulation Methodology for Statisticians, Operations Analysts, and Engineers*, Vol. I, Wadsworth & Brooks Cole, 1989; xvi + 416 pp, $51.95. ISBN 0-534-09450-3

Part I is an introduction to simulation, moving from basic modeling to Monte Carlo and generation of random variables with specific distributions. Part II, "Sophisticate Simulation," occupies the latter 75% of the book and covers numerical and graphical summaries of data, multifactor simulations, assessing variability (stratifying, jackknifing, bootstrapping), bivariate random variables, and variance reduction.

Krüger, Lorenz, et al. (eds.), *The Probabilistic Revolution: Vol. 1: Ideas in History; Vol. 2, Ideas in the Sciences*, MIT Pr, 1987; xv + 449 pp, $32.50; xvii + 457 pp, $32.50. ISBN 0-262-11118-7, 0-262-11119-5

"This monumental work traces the rise, the transformation, and the diffusion of probabilistic and statistical thinking in the nineteenth and twentieth centuries."

Castillo-Chavez, C. (ed.), *Mathematical and Statistical Approaches to AIDS Epidemiology*, Springer-Verlag, 1989; viii + 405 pp, $47.10 (P). ISBN 0-387-52174-7

Research papers on the modeling of AIDS.

Weiss, Rick, A flight of fancy mathematics: Chaos brings harmony to a birder's puzzle, *Science News* 137 (17 March 1990), 172.

"Birds of a feather are chaotic together." Swooping flocks do not rely upon a leader, but instead their coordinated chaotic behavior can be explained by a few simple mathematical axioms of attraction and repulsion.

NEWS AND LETTERS

1990 Canadian Mathematical Olympiad—Problems and Solutions

1. A competition involving $n \geq 2$ players was held over k days. On each day, the players received scores of 1, 2, 3, ..., n points with no two players receiving the same score. At the end of the k days, it was found that each player had exactly 26 points in total. Determine all pairs (n, k) for which this is possible.

Sol. The total number of points awarded to all players over all days is $\frac{1}{2}kn(n+1) = 26n$, so that $k(n+1) = 52$. For $(n,k) = (3,13)$ we have a possible allocation given by

$$(26,26,26) = (1,2,3) + 2(2,3,1) + 2(3,1,2) + 3(1,3,2) + 2(3,2,1) + 3(2,1,3)\,.$$

For $(n,k) = (12,4)$, we have

$$(26,26,\ldots,26,26) = 2(1,2,\ldots,11,12) + 2(12,11,\ldots,2,1).$$

For $(n,k) = (25,2)$, we have

$$(26,26,\ldots,26,26) = (1,2,\ldots 24,25) + (25,24,\ldots,2,1).$$

However, $(n,k) = (51,1)$ is not realizable. Hence (n,k) must be one of (3,13), (12,4), (25,2).

2. A set of $\frac{1}{2}n(n+1)$ distinct numbers is arranged at random in a triangular array:

```
              x
           x     x
        x     x     x
        .     .     .
        .     .     .
        .     .     .
x    x  .     .     .   x    x
```

Let M_k be the largest number in the k^{th} row from the top. Find the probability that

$$M_1 < M_2 < M_3 < \cdots < M_n\,.$$

Sol. Let p_n be the probability when there are n rows. Clearly, $p_1 = 1$ and $p_2 = \frac{2}{3}$. In general, the largest number must go in the last row, the probability for which is $\frac{n}{\frac{1}{2}n(n+1)} = \frac{2}{n+1}$. There is no restriction on the remaining members of the last row. The probability that the numbers of the first n - 1 rows are suitably placed is p_{n-1}. Hence $p_n = \frac{2}{n+1}p_{n-1} = \frac{2^n}{(n+1)!}$ for $n \geq 3$.

3. Let $ABCD$ be a convex quadrilateral inscribed in a circle, and let diagonals AC and BD meet at X. The perpendiculars from X meet the sides AB, BC, CD, DA at A', B', C', D' respectively. Prove that

$$|A'B'| + |C'D'| = |A'D'| + |B'C'|\,.$$

($|A'B'|$ is the length of line segment $A'B'$, etc.)

Sol. 1. Let d be the length of the diameter of the circle $ABCD$. $A'XB'B$ is concyclic and the angle subtended by $A'B'$ at B equals the angle subtended by AC in circle $ABCD$. Therefore

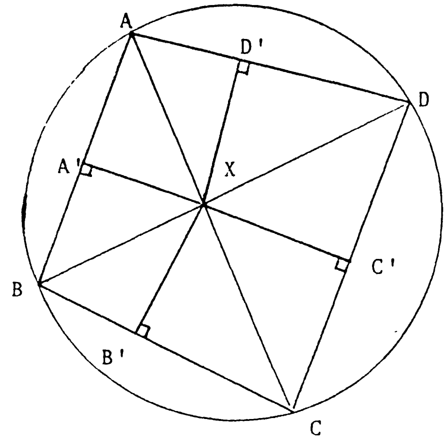
$$\frac{|A'B'|}{|BX|} = \frac{|AC|}{d}\,,$$

since BX is a diameter of circle $A'XB'B$. Similarly $\frac{|C'D'|}{|DX|} = \frac{|AC|}{d}$. Hence

$$|A'B'| + |C'D'| = \frac{|AC|}{d}[|BX| + |DX|] = \frac{|AC||BD|}{d}$$

Likewise,

$$|A'D'| + |B'C'| = \frac{|AC||BD|}{d}\,.$$

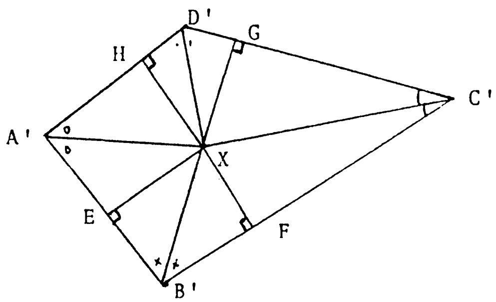


Sol. 2. Since $A'XB'B$ is concyclic, $\angle B'BX = \angle B'A'X$. Since $A'AD'X$ is concyclic, $\angle XA'D' = \angle XAD$. Since $ABCD$ is concyclic, $\angle CAD = \angle CBD$. Hence $\angle B'A'X = \angle CBD = \angle CAD = \angle XA'D'$ so that XA' bisects $\angle D'A'B'$. Similarly, XB', XC', XD' bisect $\angle A'B'C'$, $\angle B'C'D'$, $\angle C'D'A'$, respectively. Drop perpendiculars from X to meet $A'B'$, $B'C'$, $C'D'$, $D'A'$ at E, F, G, H, respectively. Then

$$|A'H| = |A'E|,\ |B'E| = |B'F|,$$
$$|C'F| = |C'G|,\ |D'G| = |D'H|$$

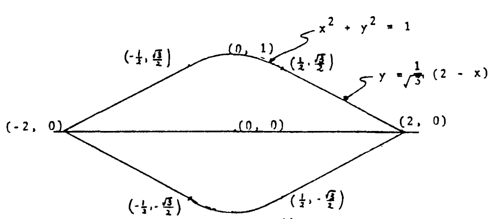
so that

$$|A'D'| + |B'C'| = |A'B'| + |C'D'|.$$



4. A particle can travel at speeds up to 2 metres per second along the x-axis, and up to 1 metre per second elsewhere in the plane. Provide a labelled sketch of the region which can be reached within one second by the particle starting at the origin.

Sol. Because of symmetry, we can restrict attention to the positive quadrant. The most efficient path to an accessible point follows the x-axis to a point $(t, 0)$ and then moves directly to the point (where $0 \le t \le 2$). We consider paths of this type.



Fix the abscissa of an accessible point at x and look for possible values of y. A particle leaving the x-axis at $(t,0)$ can move a further distance of at most $\frac{1}{2}(2 - t)$. Then $(x - t)^2 + y^2 \le \left(\frac{2-t}{2}\right)^2$, so that

$$y^2 \le \frac{3}{4}\left[\frac{4}{9}(2 - x)^2 - \left(t - \frac{2(2x - 1)}{3}\right)^2\right].$$

If $\frac{1}{2} \le x \le 2, y^2$ is maximized when $t = \frac{2}{3}(2x - 1)$ and so $y \le \frac{1}{\sqrt{3}}(2 - x)$. If $0 \le x \le \frac{1}{2}$, y^2 is maximized when $t = 0$ and so $y^2 \le 1 - x^2$.

Remark: The set of points accessible from $(t,0)$ on the x-axis lie in a circle of radius $\frac{1}{2}(2\text{-}t)$ and centre $(t,0)$. These circles are similar, under a dilatation with centre $(2,0)$ to the circle $x^2+y^2 = 1$. The lines $y = \pm\frac{1}{\sqrt{3}}(2 - x)$ are the envelopes of this family of circles.

5. Suppose that a function f defined on the positive integers satisfies

$$f(1) = 1,\ f(2) = 2,$$
$$f(n+2) = f(n+2\text{-}f(n+1)) + f(n+1\text{-}f(n))$$

$(n \ge 1)$.

(a) Show that

(i) $0 \le f(n+1) - f(n) \le 1$

(ii) if $f(n)$ is odd, then $f(n+1) = f(n) + 1$.

(b) Determine, with justification, all values of n for which

$$f(n) = 2^{10} + 1.$$

Sol. (a)(i) The result holds for $n = 1$. Suppose that we have shown for $n = 1,2,\ldots,m\text{-}1$, that $0 \le f(n+1) - f(n) \le 1$. Then, for each such n,

$$\left[(n+2) - f(n+1)\right] - \left[(n+1) - f(n)\right]$$
$$= 1 - \left[f(n+1) - f(n)\right] \in \{0,1\},$$

so that

$$f(n + 2 - f(n+1)) - f(n + 1 - f(n)) \in \{0,1\} \qquad (*)$$

by the induction hypothesis.

There are two cases:

First Case. $f(m) = f(m\text{-}1) + 1 \Rightarrow$

$$f(m+1) - f(m) = f(m+1\text{-}f(m)) - f(m\text{-}1\text{-}f(m\text{-}2))$$
$$= f(m\text{-}f(m\text{-}1) - f(m\text{-}1\text{-}f(m\text{-}2)) \in \{0,1\} \text{ by}$$

$(*)$.

Second Case. $f(m) = f(m - 1) \Rightarrow$

$$f(m\text{-}f(m\text{-}1)) = f(m\text{-}2\text{-}f(m\text{-}3)).$$

By $(*)$, each member of this equation equals $f(m\text{-}1\text{-}f(m\text{-}2))$. Hence

$$f(m+1) - f(m) = f(m+1\text{-}f(m)) - f(m\text{-}1\text{-}f(m\text{-}2))$$
$$= f(m+1\text{-}f(m)) - f(m\text{-}f(m\text{-}1)) \in \{0,1\} \text{ by } (*).$$

Thus, (a)(i) holds for $n = m$, and the required result follows by induction.

(a)(ii) Suppose the result has been established for $n = 1, 2, \ldots, m\text{-}1$. Let $f(m)$ be odd. Then $f(m\text{-}1)$ is even, so that $f(m) = f(m\text{-}1) + 1$. Hence

$$f(m+1) = f(m+1\text{-}f(m)) + f(m\text{-}f(m\text{-}1))$$
$$= 2f(m\text{-}f(m\text{-}1)).$$

Therefore, $f(m+1)$ is even and equals $f(m) + 1$, by (a)(i).

(b) We will prove by induction that, for each integer k, $n = 2^k$ is the unique solution of $f(n) = 2^{k-1} + 1$, so that, in particular, $n = 2^{11}$ is the unique solution for $f(n) = 2^{10} + 1$.

For $k = 2$, this is evident. Suppose we have shown that $n = 2^m$ is the unique solution for $f(n) = 2^{m-1} + 1$. From (a), we deduce that there is a unique number u such that $f(u) = 2^m + 1$. [If there is no such u, then $f(n)$ would assume a constant value from some point on and this would contradict the recursion equation defining $f(n+2)$ for large n.]

Now $2^m + 1 = f(u\text{-}f(u\text{-}1)) + f(u\text{-}1\text{-}f(u\text{-}2))$ and $f(u\text{-}1) = 2^m$. Since $f(u\text{-}f(u\text{-}1)) - f(u\text{-}1\text{-}f(u\text{-}2)) \in \{0,1\}$, we have $f(u\text{-}f(u\text{-}1)) = f(u\text{-}1\text{-}f(u\text{-}2)) + 1 = 2^{m-1}+1$. Therefore, by the induction hypothesis, $u - f(u\text{-}1) = 2^m$ and, hence, $u = 2^m + f(u\text{-}1) = 2^{m+1}$.

LETTER TO THE EDITOR

Editor:

In the April 1989 issue of this Journal, David J. Smith and Mavina K. Vamanamurthy discuss the volume of the n-dimensional ball of radius r. They find the volume first by cross-sections, then by Fubini's Theorem, and finally by using the transformation formula for multiple integrals in conjunction with a (polar-coordinate) transformation of $[0,r] \times [0,2\pi] \times [0,\pi]^{n-2}$ onto the n-dimensional ball $B^n(r,0)$ of radius r and center at 0. Using this particular transformation, it takes almost a full page to compute the Jacobian. ([1])

One obtains a vast simplification by using instead the transformation

$$x_1 = r \cos \theta_1$$

$$x_2 = r \sin \theta_1 \cos \theta_2$$

$$\vdots$$

$$x_n = r \sin \theta_1 \sin \theta_2 \ldots \sin \theta_{n-1} \cos \theta_n$$

which maps the interior of $[0,\pi]^n$ bijectively onto the interior of $B^n(r,0)$ and the boundary of $[0,\pi]^n$ onto the boundary of $B^n(r,0)$. ([2]) Since the Jacobian is triangular, it is equal to the product of the diagonal terms

$$(\partial x_1/\partial\theta_1)(\partial x_2/\partial\theta_2)\ldots(\partial x_n/\partial\theta_n)$$
$$= (-r \sin \theta_1)(-r \sin \theta_1 \sin \theta_2)\ldots(-r \sin \theta_1 \ldots \sin \theta_n)$$
$$= (-r)^n \sin^n\theta_1 \sin^{n-1}\theta_2 \ldots \sin \theta_n .$$

It is as simple as that.

References

1. D. J. Smith and Mavina K. Vamanamurthy, this MAGAZINE 62 (1989), 101-107.
2. H. Sagan, *Advanced Calculus*, Houghton-Mifflin Company, Boston 1974, p. 466.

Hans Sagan
North Carolina State University
Raleigh, NC 27695

CORRECTION

In Richard K. Guy's recent article, The Second Strong Law of Small Numbers, this MAGAZINE 63 (1990), 3-20, Example 38, the column heads should have been moved one column to the right. Row 8 of the table should have been moved one column to the right as well.

—Editor

Random Walks and Electric Networks,

by J. Laurie Snell and Peter Doyle

xiii + 159 pages. Hardbound

List: $25.00 MAA Member: $19.00

In this newest addition to the Carus Mathematical Monographs, the authors examine the relationship between elementary electric network theory and random walks, at a level which can be appreciated by the able college student. We are indebted to them for presenting this interplay between probability theory and physics in so readable and concise a fashion.

Central to the book is Polya's beautiful theorem that a random walker on an infinite street network in d-dimensional space is bound to return to the starting point when d = 2, but has a positive probability of escaping to infinity without returning to the starting point when d = 3. The authors interpret this theorem as a statement about electric networks, and then prove the theorem using techniques from classical electrical theory. The techniques referred to go back to Lord Rayleigh who introduced them in connection with an investigation of musical instruments.

In Part I the authors restrict themselves to the study of random walks on finite networks, establishing the connection between the electrical concepts of current and voltage and corresponding descriptive quantities of random walks regarded as finite state Markov chains. Part II deals with the idea of random walks on infinite networks.

Table of Contents

Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Honoring Great Mathematicians

Ramanujan Revisited

Proceedings of the Centenary Conference

edited by
George K. Andrews, Richard A. Askey, Bruce C. Berndt, K.G. Ramanathan, and Robert A. Rankin

"The excellent papers in this volume display the surprising extent to which Ramanujan's ideas are contributing to modern developments."
—*MATHEMATIKA*

1988, 632 pages, $59.00/ISBN: 0-12-058560-X

Analysis, et cetera

Research Papers Published in Honor of Jürgen Moser's 60th Birthday

edited by
Paul H. Rabinowitz and Eduard Zehnder

This book is intended for an introductory course in numerical analysis at the junior or senior undergraduate level. In addition to exceptional coverage of the normal range of topics in an introductory course, the book includes topics that have become particularly interesting in recent years. Important applications of the material are discussed throughout the book, and elucidated by many well-motivated exercises.

January 1990, 712 pages, $69.95
ISBN: 0-12-574249-5

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
Book Marketing Department #29040
1250 Sixth Avenue, San Diego, CA 92101

CALL TOLL FREE: 1-800-321-5068
FAX: 1-314-528-5001

Probability, Statistics, and Mathematics

Papers in Honor of Samuel Karlin

edited by
T.W. Anderson, Krishna B. Athreya, and Donald L. Iglehart

This volume consists of 24 papers written by prominent probabilists, statisticians, mathematicians, and economists on topics of current interest in their respective fields. Most of the authors have been collaborators or Ph.D. students of Samuel Karlin and the contributions reflect the broad spectrum of Professor Karlin's work.

1989, 424 pages, $39.95/ISBN: 0-12-058470-0

For Historians and Mathematicians

The History of Modern Mathematics

Volumes I and II

edited by
David Rowe and John McCleary

The History of Modern Mathematics brings together 24 leading authorities on 19th- and early 20th-century mathematics from eight countries. Particular attention has been given to various tensions between pure and applied mathematics as these arose in three different historical settings: early 19th-century France, Wilhelmian-era Germany, and the United States during World War II.

Volume I
December 1989, 472 pages, $39.50
ISBN: 0-12-599661-6

Volume II
December 1989, 344 pages, $37.50
ISBN: 0-12-599622-4

Set Price, Volumes I and II: $65.00

Quote this reference number for *free postage and handling* on your prepaid order ➡ **29040**

GEOMETRY & TOPOLOGY

FIRST CONCEPTS OF TOPOLOGY: The Geometry of Mappings of Segments, Curves, Circles, and Disks

W.G. Chinn and N.E. Steenrod

This clear and winning little book, for readers willing to come to genuine grips with the idea of a mathematical proof, presents topology . . . as mathematicians see it. The parlor tricks are gracefully alluded to here and there, but they are distinctly for after hours. The center of interest is the stuff itself: powerful notions of set theory are . . . exploited to define open sets and their coverings, and from them [to prove] the key theorems. .. One cannot any longer doubt that a single stroke of a knife exists that divides any irregular ham sandwich so that ham and both bread slices can be shared with perfect fairness by two consumers.

Philip Morrison in *Scientific American*

Contains over one hundred and fifty problems and solutions.

160 pp., 1966, ISBN 0-88385-618-2

List: $10.00 MAA Member: $8.00

Catalog Number NML-18

FROM PYTHAGORAS TO EINSTEIN

K.O. Friedrichs

Starting with area and dissection proofs of the Pythagorean theorem, Friedrichs proceeds gently, step by step, to cover vectors, coordinates, elastic and inelastic impacts, and relativistic space-time—ending with a derivation of a contemporary formula rivaling the Pythagorean theorem in fame, $E = mc^2$.

88 pp., 1965, ISBN 0-88385-616-6

List: $9.50 MAA Member: $7.60

Catalog Number NML-16

GEOMETRY REVISITED

H.S.M. Coxeter and S.L. Greitzer

After seeing a proof that a segment has a unique midpoint, students may wonder whether there are any interesting theorems in geometry. If either they or their teachers get a hold of this book, they will learn that there are many beautiful and nontrivial theorems in geometry. Among those found here are the theorems of Ceva, Menelaus, Pappus, Desargues, Pascal, and Brianchon. A nice proof is given of Morley's remarkable theorem on angle trisectors. The transformational point of view is emphasized: reflections, rotations, translations, similarities, inversions, and affine and projective transformations. Many fascinating properties of circles, triangles, quadrilaterals, and conics are developed.

193 pp., 1967, ISBN 0-88385-619-0

List: $11.75 MAA Member: $9.50

Catalog Number NML-19

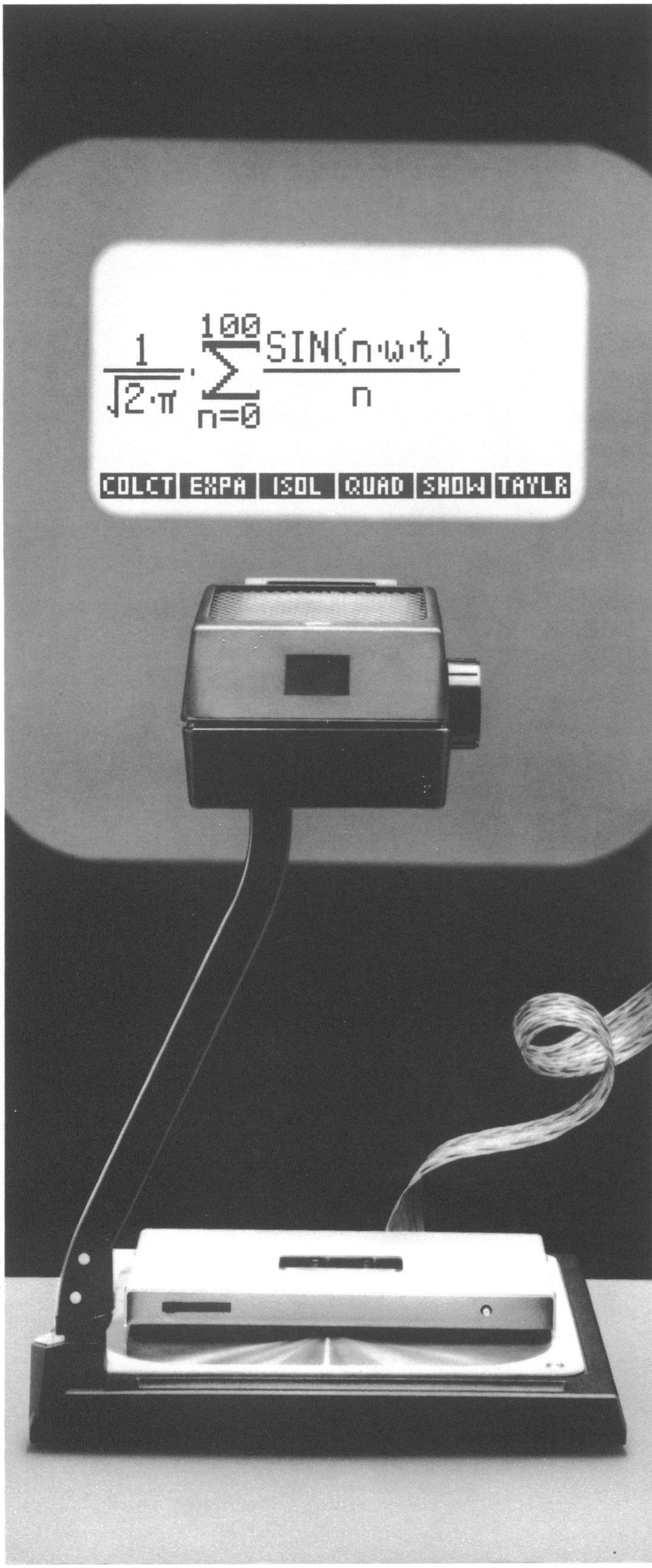
The very lucid presentation takes the reader from the elementary problems of plane Euclidean geometry to the fundamental concepts of non-Euclidean geometry . . . The book is rich in remarkable facts and thereby is very effective in promoting the significance and the value of geometry in mathematical teaching, a promotion which is very necessary . . . The always original developments use very simple tools . . . and soon proceed to higher configurations.

K. Strubecker in *Mathematical Reviews*

Order from:

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

Help your students discover more meaningful relationships.



Free classroom display device with purchase of 30 calculators.

Showing is much more powerful than telling. So we've developed special classroom displays for our two most advanced calculators.

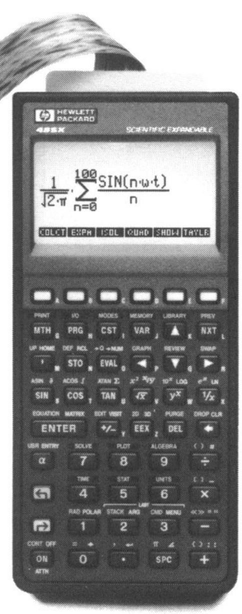
The new scientific expandable HP 48SX calculator and the cost-effective HP 28S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculator.

If your department or students purchase 30 HP 48SX or HP 28S calculators (or a mix of both), we'll give you a free HP 48SX and plug-in classroom display (an $850 retail value). Or a classroom display version of the HP 28S (a $500 retail value). And free teaching materials in the bargain.

Call us at **(503) 757-2004** between 8am and 3pm PDT for more details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends October 31, 1990, and applies to instructors at college and high school levels only.

There is a better way.

HEWLETT PACKARD



CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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