MATHEMATICS MAGAZINE

Vo. 59 No. 1
February 1986

A JULES VERNE CRYPTOGRAM • ADMISSIONS TEST
CONTINUED ROOTS • A SURPRISE FROM GEOMETRY

SOME INFORMATIVE AND USEFUL BOOKS FROM THE MAA . . .

Browse through this list and see if your library is missing some of these important books published by the Association

APPLICATIONS OF UNDERGRADUATE MATHEMATICS IN ENGINEERING—written and edited by Ben Noble. A collection of articles prepared by engineers for the Committee on the Undergraduate Program in Mathematics. 364 pp. Hardbound.
List: $19.00. MAA Member: $13.00.

ANNOTATED BIBLIOGRAPHY OF EXPOSITORY WRITING IN THE MATHEMATICAL SCIENCES, prepared by M. P. Gaffney and L. A. Steen. An invaluable reference source of expository articles in mathematics. 282 pp. Paperbound.
List: $12.00. MAA Member: $8.00.

AN ANNOTATED BIBLIOGRAPHY OF FILMS AND VIDEOTAPES FOR COLLEGE MATHEMATICS, by David I. Schneider. An up-to-date listing of films and videotapes available for classroom use. 107 pp. Paperbround.
List: $9.00. MAA Members: $6.00.

A BASIC LIBRARY LIST FOR TWO-YEAR COLLEGES, prepared by the Committee on Basic Library Lists. A recommended library nucleus for two-year colleges. 66 pp. Paperbound.
List: $8.00. MAA Member: $6.00.

A BASIC LIBRARY LIST FOR FOUR-YEAR COLLEGES, prepared by CUPM. Presents listings of books and journals that should be in every college library. 106 pp. Paperbound.
List: $9.00. MAA Member: $6.50.

THE CHAUVENET PAPERS. ***A Collection of Prize Winning Expository Papers in Mathematics,*** edited by James C. Abbott. Two-volumes of the collected prize winning Chauvenet Papers. Vol. 1—312 pp. Hardbound. Vol. 2—282 pp. Hardbound.
List: $21.00 each. MAA Member $16.00 each.
Two volume sets
List: $36.00. MAA Member: $27.00.

CRITICAL VARIABLES IN MATHEMATICS EDUCATION: ***Findings from a Survey of the Empirical Literature,*** by E. G. Begle. A joint publication of the MAA and the National Council of Teachers of Mathematics. List: $8.00. MAA Member: $6.40.

A COMPENDIUM OF CUPM RECOMMENDATIONS. Volumes I and II. A collection of the major recommendations of the Committee on the Undergraduate Program in Mathematics. Two volumes. 756 pp. Hardbound.
List: $16.50. MAA Member: $12.00.

THE WILLIAM LOWELL, PUTNAM MATHEMATICAL COMPETITION: PROBLEMS AND SOLUTIONS—1938-1964. Compiled by R. E. Greenwood, A. M. Gleason, and L. M. Kelly. Contains problems and solutions to the first 25 Putnam Exams. 652 pp. Hardbound.
List: $35.00. MAA Member: $26.00.

THE MATHEMATICAL ASSOCIATION OF AMERICA: ***Its First Fifty Years.*** An historical perspective of the Association. 170 pp. Hardbound.
List: $10.00. MAA Member: $5.00.

FIFTY YEAR INDEX OF THE MATHEMATICS MAGAZINE, edited by Lynn A. Steen, and J. Arthur Seebach. Cumulative index of volumes 1-50. 163 pp. Paperbound.
List: $10.00. MAA Member: $6.50.

INDEX OF THE AMERICAN MATHEMATICAL MONTHLY. Contains the tables of contents for each issue of volumes 1-80 as well as subject and author indices. 269 pp. Hardbound.
List:$19.00. MAA Member: $13.00.

PROFESSIONAL OPPORTUNITIES IN THE MATHEMATICAL SCIENCES, Tenth Edition. 1978. Designed for the student interested in a career in mathematics. 35 pp. Paperbound. $1.50 each. 95¢ for orders of five or more.

RECOMMENDATIONS ON A GENERAL MATHEMATICAL SCIENCES PROGRAM. Prepared by the Committee on the Undergraduate Program in Mathematics. (CUPM) 102 pp. Paperbound. $3.50 each.

MATHEMATICAL TIME EXPOSURES, by Issaac J. Schoenberg. Brings together topics from geometry, number theory, algebra and analysis. 288 pp. Paperbound.
List: $18.00. MAA Member: $13.50.
288 pp. Hardbound
List: $30.00. MAA Member: $22.50.

Order From:

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

MATHEMATICS MAGAZINE

THE MATHEMATICAL ASSOCIATION OF AMERICA

Volume 59, Number 1, February 1986
(ISSN— 0025-570X)

COVER: Deciphering decoded message in *La Jangada*, see pp. 2–3.

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

EDITORIAL POLICY

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

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

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

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

AUTHORS

Frederick Gass attended Phillips Academy, DePauw University, and Dartmouth College, receiving a Ph.D. in mathematics in 1968. He has been on the faculty of Miami University since 1968 except for one year spent as Visiting Scholar at Talladega College. He became interested in cryptology in 1974 after reading David Kahn's *The Codebreakers* and Sinkov's *Elementary Cryptanalysis*. As a pastime, he enjoys trying to write brief APL programs to aid the amateur cryptanalyst.

Edward Barbeau took his Ph.D. in functional analysis under F. F. Bonsall at the University of Newcastle-upon-Tyne in 1964, and has been at the University of Toronto since 1967. In 1982, he was publicity director for the joint AMS-MAA Summer Meeting. A session on the Analytic Hierarchy Process organized by T. L. Saaty attracted interest within and outside the mathematical community, and a story was run in a local newspaper. Professor Saaty suggested that it might be useful to publish an expository paper on this method of decision making of interest to politicians and management consultants; thus began a collaboration which resulted in the paper published here.

ILLUSTRATION

The cover illustration is from *La Jangada*, by Julio Verne, Gaspar, Madrid, 1882. "With his special alphabet in one hand and the document in the other."

The MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, five times a year: January, March, May, September, and November.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $11 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The non-member/library subscription price is $28 per year. Bulk subscriptions (5 or more copies) are available to colleges and universities for classroom distribution to undergraduate students at a 41% discount ($6.50 per copy—minimum order $32.50).

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Company, Middletown, CT 06457. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

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

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

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Solving a Jules Verne Cryptogram

To save the life of an innocent man, a secret message must be deciphered; Jules Verne provides an original technique

Frederick Gass
Miami University
Oxford, OH 45056

"... *KSPPSUVJHD*" is the end of a secret message that opens one of Jules Verne's lesser-known stories, *La Jangada*, known in English as *Eight Hundred Leagues on the Amazon*. Set in Brazil, the story is about Joam Dacosta, who stands wrongly accused of a heinous murder and diamond theft. The plot of this two-part story is long and involved, with Book One ("The Giant Raft") providing most of the adventure, and Book Two ("The Cryptogram") most of the suspense. Near the end, as gallows are being erected outside Joam's prison cell, his friends strive frantically to discover the message in the cryptogram, for by now it is clear that therein lies Joam's only hope. Even Judge Jarriquez attacks the problem. The final paragraph of the cryptogram, the only paragraph that is actually spelled out in the story, is as follows:

P H Y J S L Y D D Q F D Z X G A S G Z Z Q Q E H X G K F N D R X U
J U G I O C Y T D X V K S B X H H U Y P O H D V Y R Y M H U H P U
Y D K J O X P H E T O Z S L E T N P M V F F O V P D P A J X H Y Y
N O J Y G G A Y M E Q Y N F U Q L N M V L Y F G S U Z M Q I Z T L
B Q G Y U G S Q E U B V N R C R E D G R U Z B L R M X Y U H O H P (1)
Z D R R G C R O H E P Q X U F I V V R P L P H O N T H V D D Q F H
Q S N T Z H H H N F E P M Q K Y U U E X K T O G Z G K Y U U M F V
I J D Q D P Z J Q S Y K R P L X H X Q R Y M V K L O H H H O T O Z
V D K S P P S U V J H D

In (1), which I will regard as a cryptogram in itself, there are 276 letters and also several features that make it an interesting source of illustrations. My aim is for brevity and variety, using the Jules Verne cryptogram as motivation to discuss several interesting aspects of cryptanalysis. I encourage interested readers to consult [**1**], [**5**] and [**12**] for more information about the subject. References [**6**], [**8**], [**9**], and [**11**] are good for historical perspective, and [**2**] is the leading journal in the field. As you will see, some mathematical ideas begin to appear after we discuss a few interesting preliminaries. Reference [**10**] contains statistical details.

Let's go to work on Jules Verne's cryptogram. We approach it as scientific detectives, systematically forming hypotheses and checking them out. The first thing to consider is the language of the original message, the most obvious choices for us being English, Portuguese (because of the Brazilian setting), and French (the original language of Verne's tales). French would seem to be the most likely choice, but that particular detail of the problem will not be critical until later on in our investigation.

At any level of cryptanalysis it is important to identify other reasonable initial assumptions about the origin of a cryptogram. In the present case, we will assume that the writer of the message had essentially a one-time-only need for secrecy, and that the cryptogram was devised by him/her alone or in collusion with a very few confidants. Now let's move on to the central question: What method was used to transform the original message into a cryptogram?

One familiar way to transform a message is to replace some or all of the words and phrases by code words and phrases given in a special book that resembles a dictionary. In fact, the word "code" refers precisely to such a system, even though the general public often uses "secret code" and "code-breaking" to embrace all aspects of cryptic writing. Let's tentatively rule out the possibility of a code in the present case, because one is unlikely to go to the trouble of preparing a code book when there is evidently only one message at stake.

One standard alternative to code is **transposition**, whereby the original message is rearranged so as to be unintelligible. For example, the message could be divided into five-letter groups, and then each group rewritten in reverse order or subjected to some other fixed permutation. A more complicated version of this scheme plays a crucial role in Verne's *Journey to the Center of the Earth*, and a still more complicated one appears early in his *Mathias Sandorf*. (For an interesting and thorough discussion of Jules Verne as cryptographer, consult reference [4].)

To rule out transposition in the present case, we refer to the most fundamental piece of information at the cryptanalyst's disposal, the frequency distribution. For our cryptogram, this information is shown in TABLE 1.

i:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	X	Y	Z
f_i:	3	4	3	16	9	10	13	23	4	8	9	9	9	9	12	16	16	12	10	8	17	13	12	19	12

TABLE 1. **Frequency distribution of letters in the cryptogram (1).**

It is well known that certain letters—*e* and *t*, for instance—tend to appear with the highest frequency in standard English, and similar results are found in other languages. If this cryptogram were the result of a transposition, then all 23 of those *H*'s would have been present in the original message, and likewise all 16 *Q*'s, 12 *Z*'s and so forth—an unlikely possibility for any modern language.

A simple alternative to code and transposition is the scheme known as "monoalphabetic substitution," whereby each letter of the original message is replaced by a particular "cipher letter" substitute according to some correspondence such as the one pictured in (2).

Monoalphabetic Substitution

$$\begin{aligned} \text{Plain: } & a\ b\ c\ d\ e\ f\ g\ h\ i\ j\ k\ l\ m\ n\ o\ p\ q\ r\ s\ t\ u\ v\ w\ x\ y\ z \\ \text{Cipher: } & P\ Q\ S\ U\ V\ W\ X\ Y\ Z\ I\ N\ T\ E\ G\ R\ A\ L\ B\ C\ D\ F\ H\ J\ K\ M\ O \\ \text{Sample message: } & \text{f o u r s c o r e a n d s e v e n y e a r s . . .} \\ \text{Cryptogram: } & W\ R\ F\ B\ C\ S\ R\ B\ V\ P\ G\ U\ C\ V\ H\ V\ G\ M\ V\ P\ B\ C\ .\ .\ . \end{aligned} \tag{2}$$

The example in (2) shows how this substitution transforms a plain message into a cryptogram. If one were to draw up the frequency distribution of such a cryptogram, it is likely that *V* would be the most frequently used letter, possibly followed by *D*, since those two are the cipher replacements for *e* and *t*, respectively. (As a rule, I will use capitals for cipher text letters and lower case for plain text in this article.)

A careful reading of Verne's cryptogram reveals the occurrence of *HHH* at two locations, prompting one to question the likelihood of a monoalphabetic substitution, since those three *H*'s would have to result from three of whatever is the plain counterpart of *H*. Three-in-a-row is possible in a plain message, however. Consider phrases like "three eggs" and "small legs," for instance. Still, those two occurrences of "*HHH*" look suspicious, and after a fruitless search for

clues based on the assumption of a monoalphabetic substitution, Judge Jarriquez is prepared to consider an alternative hypothesis, as we will, later.

Although we won't stop to consider techniques used on monoalphabetic substitution cryptograms, there is a very nice mathematical scheme that the judge could have used to cast further doubt on the likelihood of a monoalphabetic. The idea—discovered by William F. Friedman in 1920 and published in [3]—is to calculate a statistic that measures the variation in the frequency distribution, and then compare that statistic with the value one would expect in a monoalphabetic case. This statistic, the **index of coincidence** (*I.C.*), is closely related to a formula used by geneticists to measure the diversity of a species.

Let $f_A, f_B, \ldots, f_Z$ be the frequency of letters $A, B, \ldots, Z$, respectively, in a given cryptogram that contains N letters. Then

$$I.C. = \sum_{i=A}^{Z} \frac{f_i}{N} \frac{f_i - 1}{N - 1} = \frac{1}{N(N-1)} \sum_{i=A}^{Z} f_i(f_i - 1). \tag{3}$$

For Verne's cryptogram we have $N = 276$, and from TABLE 1 we calculate $I.C. = 0.044$. Using either of the formulas in (3), one can make the following interpretation of Friedman's index: If two letters were chosen at random (without replacement) from the cryptogram, then the *I.C.* is the probability that those two letters would be alike. The probability of getting a particular letter as our first choice is f_i/N, and $(f_i - 1)/(N - 1)$ is the probability that our second letter will be the same.

What value would one expect the *I.C.* of a cryptogram to have, approximately? If the cryptogram were actually a plain, unenciphered message, say, in standard English, then it could be considered a random sample of N letters from an extremely large population. The relative frequencies of letters in that population are shown in TABLE 2. (Please note that any proposed frequency distribution for a modern language must be taken with a grain of salt. The one in TABLE 2 was generated by the author of [12] from a sample of 1000 letters.)

Letter:	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*
% Frequency:	7.3	0.9	3.0	4.4	13.0	2.8	1.6	3.5	7.4	0.2	0.3	3.5	2.5
Letter:	*n*	*o*	*p*	*q*	*r*	*s*	*t*	*u*	*v*	*w*	*x*	*y*	*z*
% Frequency:	7.8	7.4	2.7	0.3	7.7	6.3	9.3	2.7	1.3	1.6	0.5	1.9	0.1

TABLE 2. **Relative frequencies in standard english.**

If two letters are chosen at random from that larger population, then the probability of a matched pair is called the "kappa value" for that particular language. Therefore, we have $\kappa = \Sigma(p_i)^2$, where i varies among the letters of the alphabet, and p_i is a letter's probability according to the relative frequency distribution. So from TABLE 2 we have $p_a = .073$, and so on up to $p_z = .001$. The κ values for several languages are shown in (4).

Language	κ
English	0.066
French	0.076
German	0.076
Portuguese	0.079
Russian	0.053
Spanish	0.078

(4)

The fact that links the two previous paragraphs is this: The *I.C.* for a monoalphabetic substitution cryptogram is equal to that of the original plain text message, because in both cases

the summation involves the same numbers f_i (although, following the substitution, the numbers are associated with different letters). Because of this observation, we would expect the *I.C.* of the cryptogram to be an approximation of κ for the language of the original message.

In fact, it turns out that the *I.C.* is an unbiased estimator for κ. (Students of statistics may have noted already that the *I.C.* and κ bear some resemblance to sample variance and population variance formulas, respectively.) Comparing the *I.C.* of Verne's cryptogram with the information in (4), we might conclude that Verne's *I.C.* is a rather poor approximation of κ for English, French or Portuguese, and so the cryptogram is not likely to be a monoalphabetic. But wait. We cannot judge the goodness of an approximation without some sort of error estimate or (in this case) an alternative to monoalphabeticity that is suggested by $I.C. = 0.044$.

Polyalphabetic substitution

"Polyalphabetic" means that each plain letter has more than one possible cipher equivalent, so that *e*, for instance, might be represented by *J*, *H*, *G*, or even *E*, at the various locations where *e* occurs in the original message. One way to accomplish such a substitution is shown in (5).

Gronsfeld Polyalphabetic Substitution

Keyword: 5203

Key sequence:	5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 3 5 2 0 ...	(5)
Plain message:	f o u r s c o r e a n d s e v e n y e a r s a g o o u r f a t h e r s ...	
Cipher message:	*KQUUXEOUJCNGXGVHSA EDWUAJTQUUKCTKJTS* ...	

It is a simple scheme attributed to the 17th century Count of Gronsfeld, and it employs the digits of a "keyword" to determine cipher letters. If key digit 5, say, lies above a certain letter of the plain message, then the cipher letter that corresponds to it is the letter 5 positions later in the alphabet (with *A* being the letter that follows *Z*). If the key digit for a certain letter is 0, then the cipher letter is the same as the plain.

The effect of most polyalphabetic substitutions on a frequency distribution is to flatten it out. The distribution of letters in an original message is apt to exhibit the variety predicted by tables like the one in TABLE 2. But after polyalphabetic substitution, one finds that high frequency letters have spread their wealth of occurrences among several possible cipher equivalents, and likewise the poverty of a low frequency is suffered by more than just one cipher equivalent. The ultimate in flat distributions is the one wherein all letters are equally represented. For instance, if all 26 letters of standard English were equally likely, then we would have the flat distribution and κ value shown in (6). Of course, smaller alphabets yield larger κ(flat). If we assume French without a *w*, say, then $\kappa(\text{flat}) = 0.040$.

Kappa For a Flat Distribution

$$\begin{aligned} &i = A \text{ through } Z \text{ (26 letters)} \\ &p_i = 1/26 \approx 0.038 \\ &\kappa(\text{flat}) = \Sigma p_i^2 = \Sigma(1/26)^2 = 26(1/26)^2 = 1/26 \approx 0.038. \end{aligned} \tag{6}$$

Since the *I.C.* of Verne's cryptogram is closer to κ(flat) than to κ for English, French or Portuguese, we are led to entertain the hypothesis of a polyalphabetic substitution.

Judge Jarriquez has already arrived at that hypothesis, and he has focused upon the Gronsfeld as the most reasonable polyalphabetic scheme to consider first. But now the good judge has reached an impasse, for he sees no way to discover the keyword that unlocks the message. For him and for author Verne, logical analysis has run its course and must now be supplemented by a final burst of desperate searching and good fortune.

We, on the other hand, have several means of attacking a suspected Gronsfeld cryptogram. My plan is to complete this article by looking at several interesting and to some extent duplicating techniques, rather than simply pursue one line of reasoning toward the goal of solution.

Key length

Let's use (5) to collect some ideas about the structure of the Gronsfeld system. Since the keyword in that example has four digits, we say that 4 is the "key length," and that the plain and cipher messages may accordingly be partitioned into four "components." For example, the key digit 2 governs the component *ocaeysoar*... in the plain message and the component *QECGAUQCT*... in the cipher message (check the letters below the 2's). Notice that the cipher component is a very simple monoalphabetic variation of its corresponding plain component: in the example above, each cipher letter is simply two spaces beyond its plain equivalent in the alphabet. (This simple, shift-type monoalphabetic is called a Caesar Cipher, after its most famous user.) So, if one can divide a Gronsfeld cryptogram into its separate cipher components, the rest of the solution should be relatively easy. And the way to identify the components is to determine the key length.

In 1863, a Prussian military officer named F. W. Kasiski published a simple number-theoretic means of searching for the key length. Like so many techniques of cryptanalysis, it deals with pattern repetitions that may exist in the cryptogram. Also, it is the basis of the solution discussed in [7].

Look again at (5) and look for repeated sequences of letters in the cryptogram: there are two *QUU*'s and two *JT*'s. With the plain message and the key sequence before us, we see that the *JT* repetition is simply a fluke that represents no special interplay among the key sequence, the plain and the cipher; but the *QUU* is a different story. Generally speaking, the longer repetitions (*QUU* as opposed to *JT*, here) tend to be more significant because we feel intuitively that long repetitions probably happen by design rather than by chance. In the case of *QUU*, the cipher repetition happens precisely because both occurrences of "our" in the plain message coincided with the repetition of 203 in the key sequence. Let's call this a "special repetition."

What is special about a special repetition is the way the keyword repeats itself in the interval between the two occurrences. In (5), beginning with the first *QUU*, we find exactly six repetitions of the keyword 5203 before the next occurrence of *QUU*. This observation suggests that we look for long repetitions in any suspected Gronsfeld cryptogram, determine the lengths of the intervening intervals (figured as above for *QUU*), and proceed on the assumption that whole repetitions of the keyword fit exactly into those intervals. In other words, *the key length is a divisor of those interval lengths*. That is Kasiski's approach.

In TABLE 3 is the data for a Kasiski analysis of Verne's cryptogram. If we assume that these repetitions are all special, then the key length must be a common divisor of 186, 192, 60, 54, and 12, and so it must be 2, 3, or 6.

We could use Friedman's *I.C.* to help us identify the most likely key length among 2, 3, and 6. The idea behind it is to recall that each cipher component is really just a simple monoalphabetic substitution (specifically, a Caesar Cipher) of its plain counterpart. To test the hypothesis that the key length is n, one divides the cryptogram into n components and calculates the *I.C.* for each

DDQF at interval of length	$186 = 2 \cdot 3 \cdot 31$
RYM	$192 = 2^6 \cdot 3$
TOZ	$186 = 2 \cdot 3 \cdot 31$
RPL	$60 = 2^2 \cdot 3 \cdot 5$
HHH	$54 = 2 \cdot 3^3$
KYUU	$12 = 2^2 \cdot 3$

TABLE 3. **Kasiski analysis of Verne's cryptogram.**

Assumed keylength n	*I.C.*'s of the n components						
2	0.045	0.057					
3	0.055	0.052	0.054				
4	0.058	0.053	0.040	0.055			
5	0.041	0.040	0.042	0.047	0.050		
6	0.061	0.083	0.071	0.065	0.074	0.071	
7	0.036	0.044	0.042	0.039	0.042	0.047	0.046

TABLE 4

one, looking to see whether those values are reasonably close to κ for a spoken language—somewhere between .06 and .08. The results for $n = 2$, 3, and 6, among others, are shown in TABLE 4.

Another idea, one that bypasses the Kasiski approach, is simply to test *every* hypothetical key length from $n = 2$ on up, using the method described above. This approach works well when one is aided by a computer, and for Verne's cryptogram the results are shown in TABLE 4. Doesn't $n = 6$ suggest itself nicely? Before leaving the index of coincidence, I should mention that it was devised for use with ciphers much more complicated than the Gronsfeld. In that context its use here might be considered an example of mathematical overkill.

Finding the plain components

Let's assume now that Verne's cryptogram is a Gronsfeld with a key length of 6. Then it has six components, both in plain and in cipher, and the first of the cipher components begins *PYZZXRIX*.... (Just take the first letter of the cryptogram and every sixth one thereafter. Using modular arithmetic, we could say that the ith letter is in the 1st component if and only if $i \equiv 1 \bmod 6$.) I want to show you an interesting way to search for the corresponding plain component. Since we are getting close to the particulars of the original message, the time has come to think in terms of some particular language—in this case, French. Some frequency data for standard French are given in TABLE 5.

Our problem, of course, is that we do not know the first digit of the keyword, the digit that governs this first component. Call that digit d for the moment. If d happens to be 0, then the cipher and plain components are the same. If d happens to be 1, then the cipher component is letter-for-letter one position beyond its plain component in the alphabet. In (7) we see all the possibilities for this first component of the Verne cryptogram.

Digit d	Corresponding Plain Component
0	*pyzzxrix*...
1	*oxyyvqhv*...
2	*nvxxupgu*...
3	*muvvtoft*...
4	*ltuusnes*...
5	*ksttrmdr*...
6	*jrssqlcq*...
7	*iqrrpkbp*...
8	*hpqqojao*...
9	*goppnizn*...

(7)

The question is, which of the possible plain components is most likely the correct one? Another way to put the question is to ask which line contains the letters that are most likely to be part of

Letter:	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*
% Frequency:	9.4	1.0	2.6	3.4	15.9	1.0	1.0	0.8	8.4	0.9	0.0	5.3	3.2
Letter:	*n*	*o*	*p*	*q*	*r*	*s*	*t*	*u*	*v*	*x*	*y*	*z*	
% Frequency:	7.2	5.1	2.9	1.1	6.5	7.9	7.3	6.2	2.2	0.3	0.2	0.3	

TABLE 5. **Relative frequencies in standard French (without *w*).**

the original plain text message, and therein lies a clue to our next step: using the data in TABLE 5, we assign a probability to each line of component in (7). More specifically, we treat each line as a statistical experiment, with the letters being the results of independent trials. In (7), if $d = 1$, then the plain component is *oxyy*..., and the probability associated with the component would be the product of the individual letter possibilities, $p_o \cdot p_x \cdot p_y \cdot p_y \cdot \cdots$, an exceedingly small value. For the sake of convenience, one can replace each probability p by the corresponding percentage $100p$, which would yield an exceedingly large value, and then scale back a bit by finding instead the logarithm of that value:

$$\log(100p_o) + \log(100p_x) + \log(100p_y) + \log(100p_y) + \cdots .$$

(The percentage may be taken directly from the frequency distribution in TABLE 5.) In (8) we see the complete computer-generated results and the fact that $d = 4$ yields the plain component of highest probability. On this basis it appears that the first component of the plain message is *ltuusnes*... .

d	Sum of log-percentages for corresponding plain component
0	58.0
1	59.7
2	54.4
3	68.6
4	79.7
5	60.7
6	53.3
7	58.2
8	58.7
9	58.8

(8)

By using the above procedure on all six cipher components of the Verne cryptogram, we can discover the plain components and hence the original message. Before proceeding to the message, however, I want to show you my favorite means of solving a suspected Gronsfeld.

The probable word method

This method of solving Gronsfelds is interesting for at least two reasons. First, we try to imagine what thoughts might have occupied the writer of the cryptogram, and we use intuition or psychology or whatever to choose words that might have been used to express those thoughts. Second, with a well-chosen "probable word," we can discover the keyword and the original message in a very straightforward way.

Take the Verne cryptogram, for instance. If it really does concern the crime with which Joam Dacosta is charged, then "Dacosta," "diamant," and other (French) words dealing with particulars of the crime are probable word candidates. Let's stick with "Dacosta" for the sake of illustration. If "Dacosta" is mentioned in the original message, then somewhere in the cryptogram is a sequence of seven letters that forms a cipher equivalent of that word. Furthermore, the

Gronsfeld scheme guarantees that that sequence satisfies what I shall call the "Nines Condition": none of the cipher letters is more than nine positions later in the alphabet than its plain counterpart in "Dacosta."

Here, then, is what we do, with the aid of a computer if possible: check each sequence of seven consecutive cipher letters in the cryptogram to see if it satisfies the Nines Condition. (As a related problem, you might try to estimate the probability that a sequence of seven randomly-chosen letters will satisfy that condition. The probability is quite small.) In other words, we search for locations in the cryptogram where the cipher equivalent of "Dacosta" might be found.

Display (9) shows how this search begins. In each line of (9) a different sequence of cipher letters is checked as a possible location for "Dacosta"; and an "x" signifies that the cipher letter at that location is too far down the alphabet beyond its counterpart in "Dacosta." For instance when the first seven letters *PHYJSLY* of the cryptogram are checked, the computer prints "x7xx0xx" to show that only the second and fifth cipher letters are within the prescribed distance (a maximum of 9) of their counterparts. Thus *H* and *S* are, respectively, 7 and 0 positions beyond the "a" and "s" of "Dacosta."

Checking Possible Locations of "Dacosta" in the Gronsfeld

$$\begin{array}{rc} \text{Beginning at 1st cipher letter:} & x7xx0xx \\ \text{2nd} & 4x74x43 \\ \text{3rd} & x9xx593 \\ \text{4th} & 6x99x9x \\ \text{5th} & xxxxxx5 \\ \text{6th} & 8x1xxx3 \\ \vdots & \vdots \end{array} \tag{9}$$

A complete computer search reveals that only one of the 269 possible cipher sequences satisfies the Nines Condition. It is near the middle of the cryptogram, and it yields the printout

$$1343251. \tag{10}$$

Now we know a sequence of key digits for the cryptogram, and we could begin deciphering letters even without knowledge of the keyword. However, since we know the key length to be 6, it is evident that the keyword is 134325, or else 343251, or 432513, or one of the other three cycled versions of (10) (without the repeated 1). We go back to the start of the cryptogram and try out these sequences on the first six cipher letters, finding very quickly that 432513 produces French. With the key sequence 432513, we then recover the original message. Displayed below, with punctuation and spaces between words, it is a dramatic confession that even nonreaders of French can fairly well interpret. Its remorseful author claims sole responsibility for the crime in question, and he clears the name of Joam Dacosta.

Le véritable auteur du vol des diamants et de l'assassinat des soldats qui escortaient le convoi commis dans la nuit du vingt deux janvier mil huit cent vingt six n'est donc pas Joam Dacosta injustement condamné à mort c'est moi le misérable employé de l'administration du district diamantin, oui moi seul qui signe de mon vrai nom, Ortega.

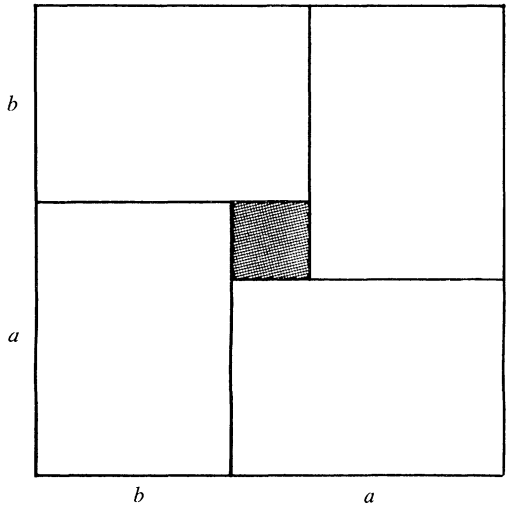
In *La Jangada*, Judge Jarriquez very nearly discovers the probable word method during his initial bout with the cryptogram, but after several hours he gives up in frustration. Next morning, when he learns that the writer might have been named Ortega, it finally dawns on him (!) He feverishly examines the end of the message (where any sincere declaration would bear a signature), derives the keyword, and just barely saves the life of Joam Dacosta. By virtue of this solution, Jules Verne is credited with the first published exposition of the probable word method for Gronsfeld ciphers.

I am grateful to the editors and a referee for their help in the preparation of this article and to Brian Winkel for valuable information and advice. I offer much-belated thanks to Chuck Oravec for introducing me to Kahn's *The Codebreakers*.

References

[1] H. Beker and F. Piper, Cipher Systems: The Protection of Communications, John Wiley and Sons, 1982.
[2] Cryptologia, Rose-Hulman Institute of Technology, Terre Haute, Indiana.
[3] W. F. Friedman, The Index of Coincidence and Its Application in Cryptography, Riverbank Laboratories, Publication No. 22, Geneva, Illinois, 1922.
[4] _____, Jules Verne as cryptographer, Signal Corps Bull., (1940) 70–107. This article is reprinted in Cryptography and Cryptanalysis Articles, v. 2, Aegean Park Press, Laguna Hills, California, 1976.
[5] H. F. Gaines, Cryptanalysis, Dover, New York, 1956.
[6] Martin Gardner, A new kind of cipher that would take millions of years to break, Scientific American, 237 (1977) 120–124.
[7] C. W. R. Hooker, The Jules Verne cipher, The Police Journal, London, 4 (1931) 107–119.
[8] David Kahn, The Codebreakers, Macmillan, New York, 1967.
[9] _____, Kahn on Codes, Macmillan, New York, 1983.
[10] S. Kullback, Statistical Methods in Cryptanalysis, Aegean Park Press, Laguna Hills, California, 1976.
[11] A. Lempel, Cryptology in transition, ACM Computing Surveys, II (1979) 280–303.
[12] A. Sinkov, Elementary Cryptanalysis—A Mathematical Approach, The New Mathematical Library no. 22, Mathematical Association of America, Washington, D.C., 1968.
[13] Jules Verne, Voyage au Centre de la Terre, Hetzel, Paris, 1864. Journey to the Center of the Earth, Dodd, New York, 1984.
[14] _____, La Jangada, Hetzel, Paris, 1881. Eight Hundred Leagues on the Amazon, Didier, New York, 1952.
[15] _____, Mathias Sandorf, Hetzel, Paris, 1885. Mathias Sandorf, Hachette, Paris, 1979.

Proof without words: The arithmetic mean–geometric mean inequality



$$(a+b)^2-(a-b)^2=4ab$$

$$\frac{a+b}{2}\geqq\sqrt{ab}$$

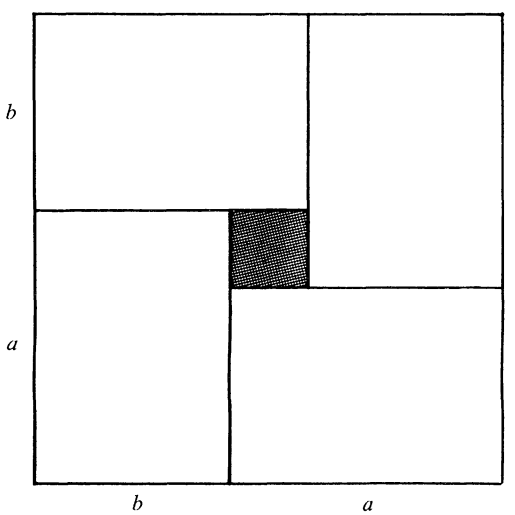
—Doris Schattschneider
Moravian College

I am grateful to the editors and a referee for their help in the preparation of this article and to Brian Winkel for valuable information and advice. I offer much-belated thanks to Chuck Oravec for introducing me to Kahn's *The Codebreakers*.

References

[1] H. Beker and F. Piper, Cipher Systems: The Protection of Communications, John Wiley and Sons, 1982.
[2] Cryptologia, Rose-Hulman Institute of Technology, Terre Haute, Indiana.
[3] W. F. Friedman, The Index of Coincidence and Its Application in Cryptography, Riverbank Laboratories, Publication No. 22, Geneva, Illinois, 1922.
[4] _____, Jules Verne as cryptographer, Signal Corps Bull., (1940) 70–107. This article is reprinted in Cryptography and Cryptanalysis Articles, v. 2, Aegean Park Press, Laguna Hills, California, 1976.
[5] H. F. Gaines, Cryptanalysis, Dover, New York, 1956.
[6] Martin Gardner, A new kind of cipher that would take millions of years to break, Scientific American, 237 (1977) 120–124.
[7] C. W. R. Hooker, The Jules Verne cipher, The Police Journal, London, 4 (1931) 107–119.
[8] David Kahn, The Codebreakers, Macmillan, New York, 1967.
[9] _____, Kahn on Codes, Macmillan, New York, 1983.
[10] S. Kullback, Statistical Methods in Cryptanalysis, Aegean Park Press, Laguna Hills, California, 1976.
[11] A. Lempel, Cryptology in transition, ACM Computing Surveys, II (1979) 280–303.
[12] A. Sinkov, Elementary Cryptanalysis—A Mathematical Approach, The New Mathematical Library no. 22, Mathematical Association of America, Washington, D.C., 1968.
[13] Jules Verne, Voyage au Centre de la Terre, Hetzel, Paris, 1864. Journey to the Center of the Earth, Dodd, New York, 1984.
[14] _____, La Jangada, Hetzel, Paris, 1881. Eight Hundred Leagues on the Amazon, Didier, New York, 1952.
[15] _____, Mathias Sandorf, Hetzel, Paris, 1885. Mathias Sandorf, Hachette, Paris, 1979.

Proof without words: The arithmetic mean–geometric mean inequality



$$(a+b)^2-(a-b)^2=4ab$$

$$\frac{a+b}{2}\geqq\sqrt{ab}$$

—DORIS SCHATTSCHNEIDER
Moravian College

Perron's Result and a Decision on Admissions Tests

Matrix theory is used to rank several options

Ed Barbeau
University of Toronto
Toronto, Ontario, Canada M5S 1A1

> *Her decision had been indicated in an instant, but it had been made after days and nights of anguished deliberation. She had known she would be asked, she had decided what she would answer, and, without the slightest hesitation, she had moved her hand to the right.*
>
> Frank R. Stockton
> **The lady, or the tiger?**

A choice between two options can be the result of "anguished deliberation." Still worse can be having to choose one of many courses of action. Factors to be considered are often contradictory in the options they indicate. Should one flip a coin or draw a straw? Generally, it will not do to consign the matter to a random device, which ignores whatever information and judgments that should be brought to bear. Rather, one would prefer to have at hand a technique which combines objectivity with an ability to cut through the confusion and uncertainty of ranking and weighting the relevant factors.

This paper treats the **analytic hierarchy process**, developed by T.L. Saaty and described by him in a number of publications (for example, [7], [8], [9]). We will not go into the difficulties of ranking two possibilities, but will suggest a way in which pairwise rankings can be synthesized into an ordering of more than two options. Our example will be a problem faced by a typical university in the Province of Ontario: what is the best method on which to base the admission of students?

Until the mid-60s, the province's education ministry operated the "departmentals," a universal system of high school graduation examinations whose results were virtually the sole criteria for an admission decision. When these were abandoned, the universities turned to grades assigned by the school and scores from aptitude tests. As you can imagine, when the aptitude tests in their turn were dropped, suspicions arose that marking standards in the schools were being relaxed. One indication was that government scholarships for high school graduates with at least an 80% average were awarded to 8% of the students in 1965 and 27% in 1981. Worse, there was evidence that the perceived inflation of grades was not uniform throughout the school system. Some students with high grades performed badly at the university level; perhaps they were preventing the admission of highly qualified students from more rigorous schools. As the government was deaf to entreaties to restore a universal and objective system of examinations, the universities were impelled to consider action on their own.

Three suggestions were put forward:

G: Continue to accept grades assigned by the high schools, but review individually those students with averages within, say, 3% of the cutoff point for admission, informally introducing whatever extra knowledge of the student or the school that might be available.

C: Calibrate the marks submitted by the high schools, adjusting upwards the marks from schools whose students have in the past performed better than average at university and downwards the marks from schools whose students earned university grades excessively lower than their school grades.

E: Institute a system of admission examinations which all candidates are required to take.

The question of choosing one of these options is governed by a number of considerations, not all of which lead to the same alternative.

F: Fairness. The method should be equitable. Acceptance of raw high school scores seems to be most unfair. Even with hand review of individual cases, the quality of the information available is likely to vary widely. Calibration seems to be fairer, but it, too, is flawed. Some schools might have sent too few students to university to produce a reliable calibration factor. In any case, the calibration of a student's mark is based on the past and may not give due weight to changes, such as a new teacher or principal, which might rapidly alter standards. Admission examinations, which all candidates take on an equal footing, seem fairest of all.

P: Predictability. One should have a fair indication of future success at the university. In this respect, it emerged from a 1977 study that the old provincial examinations were an indifferent indicator of future success and that, in fact, grades provided by the school were slightly better. Thus, it appears that we are doing about as well as possible, although some fine tuning might be possible through calibration or a cleverly designed objective examination.

L: Low cost. The method used should be economical and convenient to administer. Accepting high school grades involves a slight extra cost in considering borderline cases. There would be an initial expense in setting up a calibration formula, but once the system is in place, the calibration of marks can be closely linked with the entering of other admissions information and the continuing cost would be small. The cost of examinations is a much more serious factor. They must be prepared, administered, and marked, and the results would probably have to be entered into the data base at a time different than other admissions information. At least some of this might be offset by a candidate's fee, but this brings us to the fourth criterion.

A: Acceptability. The method should be politically acceptable. Calibration is resented by some schools and teachers' organizations who see it as a rating of schools; despite assurances of confidentiality, they are concerned about the use made of a comparison of schools. While teachers might prefer, but not be enthusiastic about, admissions tests, their superiors are worried about the possible demands on time and resources to administer such examinations. Furthermore, universities which forego admission tests might attract students from institutions which require them.

It is not clear which of the options *G*, *C*, or *E* should be adopted. Criterion *F* favours the third, but *E* is clearly inferior to the other two under criterion *L*. Option *C* is probably slightly favoured by *P*. As for criterion *A*, it points to *G*. Much depends on how seriously each criterion is taken. An eminent or wealthy institution could ignore *A* or *L*, while these criteria are significant to smaller and more impoverished colleges. Formulating a ranking of the four criteria and using this ranking to decide among the three options is a demanding task, for which some tools would be desirable.

One such tool is the **analytic hierarchy process** (AHP). It is by no means the only seriation technique available (see, for example, [**2**], [**3**], [**5**], [**11**]); in fact, Perron's Theorem lies at the bottom of one method for ranking competitors in a round-robin tournament [**6**, p. 44]. However, this process is capable of allowing for a decision problem to be decomposed into several levels. At each level, there is a pairwise comparison of the options according to higher-level criteria which are melded mathematically into an overall ranking. In the present example, there are three levels. At the third level are the three choices to be ranked. However, because of the difficulty of arriving at an *a priori* ranking of these causes, one introduces a second level—the four criteria. The three

choices are assessed separately with respect to the four criteria; the four criteria are ranked with respect to the overall goal which constitutes the first level of the hierarchy; this leads to a blended ranking of the choices which takes into account all the criteria.

The mathematical requirements for the AHP at this basic level are indeed modest. The matrix theory involved is accessible to a student in a first linear algebra course; the theorem that powers the engine is a result of Perron concerning eigenvectors of a matrix with positive entries, although in practice the necessary mathematical apparatus can be set up without a direct appeal to Perron's Theorem. All that is required of the user is the ability to make a comparison between two items; AHP then produces for him not only a ranking of the options but a measure of its reasonableness. Refinements of the process will endow it with the ability to respond to judgments changing over time and to criteria which not only are applicable to those of another level, but which feed into each other. However, the admissions problem will illustrate only the main line of the approach without these complications.

How to weight several options

Suppose we are given n options which have to be ranked in order of importance or significance: $Q_1, Q_2, \ldots, Q_n$. In the absence of some objectively determined property of the options, such as cost, it is difficult to come up with a ranking with confidence. We may be beset by second thoughts: should this pair really be ranked so far apart, or should that triple really be in that order? Our instinct is to look at the options a few at a time; but then it is difficult to get a synthesis. If it is a committee, rather than an individual, which is doing the ranking, the results can be anomalous. The members of a three-man committee might give individual rankings of three options as Q_1, Q_2, Q_3; Q_2, Q_3, Q_1; and Q_3, Q_1, Q_2, respectively. On a majority vote for each pair, the committee as a whole would rank Q_1 ahead of Q_2, Q_2 ahead of Q_3, and Q_3 ahead of Q_1. To cope with this intransitivity, one should try to get some sense of whether the decision maker regards one option as being slightly, or significantly, better than another; in other words, it might be desirable to have some numerical measure of the superiority of one option over another. Even if the committee, with the help of a numerical scale and some negotiation, manages to avoid the snare of intransitivity, it may not be preserved from a milder form of inconsistency. While it might agree that Q_1 is twice as important as Q_2 and that Q_2 is three times as important as Q_3, it may well shrink from assigning Q_1 six times the importance of Q_3.

To arrive at a procedure, let us work backwards. Suppose that we have actually succeeded in attaching to each option Q_i a positive real number w_i which measures its importance. Then it would be easy to deduce from this a measure of the relative importance of two of the options: Q_i, with weight w_i, can be regarded as being more important than Q_j, with weight w_j, by the factor $a_{ij} = w_i/w_j$ (we make use of a convention here: if $a_{ij} < 1$, then Q_i is actually *less* important than Q_j; alternatively, we can say Q_j is more important than Q_i by the factor $a_{ji} = w_j/w_i$). It is reasonable to let $a_{ii} = 1$. We now form an $n \times n$ matrix $\mathbf{A} = (a_{ij})$ which has the following properties:

(i) $a_{ii} = 1$, $a_{ij} > 0$, and $a_{ji} = a_{ij}^{-1}$ for all i, j.
(ii) $a_{ij}a_{jk} = a_{ik}$ for all i, j, k.
(iii) The matrix $\mathbf{A}$ has rank 1, with each column proportional to the vector $C = (w_1, w_2, \ldots, w_n)^T$ and each row proportional to the vector $R = (w_1^{-1}, w_2^{-1}, \ldots, w_n^{-1})$;
(iv) 0 is an eigenvalue of $\mathbf{A}$ with multiplicity $n-1$, and the trace of $\mathbf{A}$ is n; it follows from this that there is a remaining eigenvalue which is simple and equal to n;
(v) C is a column eigenvector and R is a row eigenvector of $\mathbf{A}$ corresponding to the eigenvalue n. Thus, in this special case, our relative weighting of the options Q_i appears in the form of an eigenvector corresponding to the largest positive eigenvalue of a matrix with positive entries.

Imagine that we now perturb the entries a_{ij} of $\mathbf{A}$. Its eigenvectors and eigenvalues will be correspondingly perturbed. However, if the perturbation is small, there will be an eigenvalue close

to n whose column eigenvector can be regarded as a pretty good approximation to the relative weighting of the Q_i. This suggests that we can look at the ranking problem in this fashion: For the n options, there is an ideal but unknown weighting of their significance by an n-vector of positive real numbers. In order to discover this vector, we assign to each pair (i, j) a positive real number a_{ij}, which measures the relative importance of Q_i and Q_j. The two are equally important when $a_{ij} = 1$; Q_i is more important than Q_j exactly when $a_{ij} > 1$. The only condition we impose on the assignment of the a_{ij} is the property (i) mentioned earlier. This is already a strong assumption, human psychology being what it is, as questions eliciting the relative importance of two options may draw quite different responses depending on how they are asked ([**12**]). A matrix **A** with entries a_{ij} satisfying (i) is called a **reciprocal** matrix.

Suppose it turned out that, with brilliant insight, we managed to pick the a_{ij} to achieve condition (ii). (In this case, we say that the matrix **A** is **consistent**.) Then, the kth column is equal to a_{jk} times the jth column, so that the rank of **A** is 1 and **A** satisfies (iv). Indeed, if $(c_1, c_2, \ldots, c_n)^T$ is any column and $(r_1, r_2, \ldots, r_n)$ any row eigenvector with eigenvalue n, we have $r_j/r_i = c_i/c_j = a_{ij}$. Thus, the eigenvectors can be used to weight the options in a way which is consistent with our pairwise comparisons.

In the case that **A** is not consistent, the situation is pleasantly satisfactory; **A** is subject to the following theorem.

PERRON'S THEOREM. *If* **A** *is a matrix with strictly positive entries, then* **A** *has a simple positive eigenvalue* $\lambda_{\max}$ *which is not exceeded in absolute value by any of its (complex) eigenvalues. Every (row or column) eigenvector corresponding to* $\lambda_{\max}$ *is a constant multiple of an eigenvector with strictly positive entries.*

This important result is treated in [**1**], [**4**] and [**10**] and is widely applicable in such areas as probability, numerical analysis, economics [**1**, p. 242], and demography. To get a rough idea of why this theorem holds, let $\mathcal{O}$ be the "positive" orthant consisting of all vectors in $\mathbf{R}^n$ which have nonnegative coordinates. Then **A**, considered as linear operator on $\mathbf{R}^n$, maps $\mathcal{O}$ into a proper convex subset of itself; indeed, the positive halves of the axes of $\mathbf{R}^n$ get mapped by **A** to rays in the interior of $\mathcal{O}$, and $\mathbf{A}(\mathcal{O})$, the image of $\mathcal{O}$ under **A**, is the convex hull of these rays. The sequence $\{\mathbf{A}^n(\mathcal{O}) \equiv \mathbf{A}(\mathbf{A}^{n-1}(\mathcal{O}))\}$ of successive images of $\mathcal{O}$ under A is a nested sequence of subcones of $\mathcal{O}$, each strictly smaller than its predecessor, which collapses down to a single direction as n increases. This direction determines an eigenvector of **A** with positive coordinates. Another way of looking at the result is to apply the Brouwer Fixed Point Theorem to the mapping ϕ defined on the simplex of vectors in $\mathcal{O}$ the sum of whose coordinates is 1, where $\phi(X)$ is that multiple of $\mathbf{A}(X)$ which lies on the simplex.

If **A** is any reciprocal matrix containing our numerical judgments concerning all the pairs of the options Q_i, let C be the positive column eigenvector for $\lambda_{\max}$ the sum of whose entries is 1; we can take the entries as measures of the relative importance of the Q_i. While this is reasonable for consistent matrices, how much confidence can we have in the process for nonconsistent matrices? First, there is a ready alternative. Instead of using the column vector, why not take the row eigenvector? We could form the reciprocals of its entries, normalize the vector obtained to make its entries add to 1, and use these to measure the relative importance of the Q_i. In general, the result will be different. Secondly, if our matrix is far from being consistent, then this points to some unreasonableness in our original judgments and we can hardly expect to get from them a reliable weighting. The best answer to the first point is: yes, there are alternatives which we could use, but the method stands up to the rigours of field testing quite well. In fact, with a proper choice of scale, one can recapture from subjective judgments the relative illuminations (as measured by the inverse square law) of chairs at various distances from a light source, or the relative distances of cities [**2**, pp. 38, 41]; another model which actually places cities roughly in their proper places on a map is discussed in [**5**]. The recapturing is not exact, however, and one might expect to do as well using row rather than column eigenvectors. As for the second point, we shall see later that we can actually provide a numerical measure for the acceptability of our

judgment matrix. Thus, the method is self-correcting in the sense that we can know when we ought to subject our judgments to closer scrutiny and revision.

How to balance different criteria

As we saw in the discussion of the opening section, there may be several criteria with which to rank our options and these may point in different directions. To handle this situation, we first attach weights to the criteria, $C_1, C_2, \ldots, C_s$, using the method of the last section: suppose these are entries of the column vector $Z = (z_1, z_2, \ldots, z_s)^T$, normalized so that its entries add up to 1.

Now evaluate the n options according to each criterion in isolation. For the jth criterion, suppose the weights are the entries of the column vector $(y_{1j}, y_{2j}, \ldots, y_{nj})^T$. An overall weighting of the options is found by taking a weighted average, using Z, of the weightings for the several criteria. In order that we actually do achieve the proper mix of the different criteria, the total weights for the column vectors for the various criteria should be the same; accordingly, we suppose that the column vectors $(y_{ij})^T$ are normalized so their coordinates add up to 1. We then get a blended weighting $(w_1, w_2, \ldots, w_n)^T$ of the n options with

$$w_i = y_{i1}z_1 + y_{i2}z_2 + \cdots + y_{in}z_n \qquad (1 \leqslant i \leqslant n).$$

Observe that $w_1 + w_2 + \cdots + w_n = 1$. More briefly, we can write $W = YZ$, where Y is the $n \times s$ matrix with entries y_{ij}.

It is straightforward to generalize this to a more complex decomposition of the decision process, in which there are several levels of criteria, those at one level being judged according to the criteria at the next higher level. At the lowest level are the options to be considered; at the highest, the most general overriding criteria. For each pair of adjacent levels, we can form an $r \times s$ matrix whose s columns represent the weightings of the r lower-level criteria with respect to the s higher-level criteria. If $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m$ are the matrices, with normalized columns, for each pair of adjacent levels from lowest to highest respectively, and Z is the weighting of the highest level criteria, the overall weighting of the options is given by a matrix product $\mathbf{Y}_1\mathbf{Y}_2 \cdots \mathbf{Y}_m Z$.

Now let us turn to the university admissions problem.

Applying the analytic hierarchy process

The first task is to come up with a numerical measure of the various pairwise comparisons. This requires a scale sensitive enough to classify the importance of one choice over another as mild, moderate, strong, or overwhelming, but not so fine as to lead to spurious or uncertain determinations. After much experimenting, the most credible weighting of possibilities is found to be achieved by a nine-point scale of relative importance, described in TABLE 1 [7, p. 53].

The three options we have to decide among are G(accepting high school grades), C (calibrating grades) and E (requiring admission tests). The criteria to be applied are F (fairness), P (predictability), L (low cost), and A (acceptability). In order to rank the criteria, we make pairwise comparisons. Suppose we rate L as slightly more significant than P; then we can assign to the pair (L, P) the number 2, and to the pair (P, L) the reciprocal $1/2$. On the other hand, assigning to the pair (F, P) the number 5 is an indication of our sense that fairness is much to be desired over predictability if we had to choose between them. A possible table of values, in which every entry measures the relative importance of the row variable over the column variable might be

	F	P	L	A
F	1	5	3	4
P	$\frac{1}{5}$	1	$\frac{1}{2}$	$\frac{1}{3}$
L	$\frac{1}{3}$	2	1	2
A	$\frac{1}{4}$	3	$\frac{1}{2}$	1.

These entries constitute a 4×4 matrix with a positive eigenvalue which exceeds the magnitudes of all the other eigenvalues. We take the normalized column eigenvector for this eigenvalue for a relative weighting of the criteria. In this example, it is $(0.543, 0.085, 0.213, 0.159)^T$ with eigenvalue 4.14. Thus fairness is the most important criterion, followed by economy, political acceptability and predictability, in that order.

Intensity of importance	Definition	Explanation
1	Equal importance	Two options contribute equally to the objective
3	One moderately more important than the other	Experience and judgment slightly favour one option over the other
5	One essential or strongly more important than the other	Experience and judgment strongly favour one option over the other
7	One has very strong or demonstrated importance relative to the other	One option is favoured very strongly over the other; its dominance is demonstrated in practice
9	Extreme importance	The evidence favouring one option over the other is conclusive
2, 4, 6, 8	Intermediate between adjacent scale values	Useful when compromise is needed
Reciprocals of above	If option i has one of above integers assigned to it when compared with j, then j has the reciprocal value when compared with i	

TABLE 1. **A nine-point scale of relative importance.**

A rating of the three options, G, C, E, with respect to the four criteria, F, P, L, A, might give the four arrays shown in TABLE 2.

F

How much fairer is the row choice than the column choice?

	G	C	E
G	1	1/3	1/5
C	3	1	1/3
E	5	3	1

P

What is the gain in predictability by taking the row choice rather than the column choice?

	G	C	E
G	1	1/2	2
C	2	1	3
E	1/2	1/3	1

L

How much more economical is the row choice than the column choice?

	G	C	E
G	1	3	6
C	1/3	1	4
E	1/6	1/4	1

A

How much more acceptable is the row choice than the column choice?

	G	C	E
G	1	5	4
C	1/5	1	1/3
E	1/4	3	1

TABLE 2. **Ratings of the three options G, C, E with respect to the four criteria F, P, L, A.**

The eigenvalues and eigenvectors of the matrices in TABLE 2 corresponding to the four criteria are the following:

eigenvalues:		F 3.03	P 3.01	L 3.05	A 3.09
eigenvectors:	G	0.105	0.297	0.644	0.674
	C	0.258	0.540	0.271	0.101
	E	0.637	0.163	0.085	0.226

Thus, two criteria favour accepting raw grades, one criterion favours calibration and the remaining criterion favours examinations. However, the criterion favouring examinations has the most importance. A matrix multiplication gives the overall weighting of the three options:

$$\begin{pmatrix} 0.105 & 0.297 & 0.644 & 0.674 \\ 0.258 & 0.540 & 0.271 & 0.101 \\ 0.637 & 0.163 & 0.085 & 0.226 \end{pmatrix} \begin{pmatrix} 0.543 \\ 0.085 \\ 0.213 \\ 0.159 \end{pmatrix} = \begin{pmatrix} 0.327 \\ 0.260 \\ 0.414 \end{pmatrix}.$$

The relative weights attached to grades, calibration and examinations are, respectively, 0.327, 0.260, and 0.414, so that having examinations is the preferred option. Despite their high cost, the perception that they enable the fairest admission process is conclusive. Now, of course, if our judgments change in any respect, then there will be a corresponding change in at least some of the matrices, resulting in a different weighting.

While the figures obtained in a subjective situation such as this are not as compelling as figures arising from a physical or engineering application of mathematics, nevertheless there is considerable value in going through the process. We have to isolate the important ingredients in the situation and then systematically assess their relative importance. Secondly, the figures themselves encourage us to review our analysis. We have gained a sense of the mechanisms which lead to our weighting. If the weights produced by the process are not in accord with our preconceived notions, we are encouraged to check the validity of our assessments along the way and rethink the whole situation. Either there is some factor which we did not take into account, or else we were a little extreme in some of our judgments.

Row vector versus column vector

Our earlier discussion indicates that it would be equally reasonable to take either the column or the row eigenvector of the reciprocal matrix in determining the relative weights of several options. In this section, we pursue a modest exploration of this question and find that, indeed, the two methods do not always yield the same ranking. To establish notation, let λ denote the Perron eigenvalue of the reciprocal matrix $\mathbf{A}$, let C denote one of its positive column eigenvectors, and R one of its positive row eigenvectors. Let R' be that column vector whose entries are the reciprocals of the corresponding entries of R. The row and column eigenvectors will yield exactly the same weighting if and only if C and R' are proportional. For any positive vector X, let $\overline{X}$ denote the normalization obtained by dividing each entry of X by the sum of all the entries; thus, the entries of $\overline{X}$ add up to 1.

A 2×2 reciprocal matrix is trivially consistent and the row and column eigenvectors give the same weighting. Let us suppose that $\mathbf{A}$ is the 3×3 matrix

$$\begin{pmatrix} 1 & u & v \\ u^{-1} & 1 & w \\ v^{-1} & w^{-1} & 1 \end{pmatrix}$$

with u, v, w all positive. Each column of $\mathbf{A}$ can be interpreted as measuring the relative significance of three options with respect to a fixed one. One might then expect that the overall weighting of the three options would be some kind of mean of the three columns. Keeping in mind the possibility that the row eigenvector and the column eigenvector give the same weighting when R' is proportional to C, we are led to examining two geometric mean vectors:

$$(u^{1/3}v^{1/3}, u^{-1/3}w^{1/3}, v^{-1/3}w^{-1/3})^T,$$

a column vector whose entries are the geometric means of the entries of the corresponding rows of the matrix, and

$$(u^{-1/3}v^{-1/3}, u^{1/3}w^{-1/3}, v^{1/3}w^{1/3}),$$

a row vector whose entries are the geometric means of the entries of the corresponding columns of the matrix. Both of these turn out to be eigenvectors with eigenvalue $1+y+y^{-1}$, where $y = u^{1/3}v^{-1/3}w^{1/3}$. That this is the Perron eigenvalue can be seen from the fact that the characteristic polynomial of **A**, which is

$$(1-x)^3 - 3(1-x) + (y^3 + y^{-3}) = (1-x+z)\left[(1-x)^2 - z(1-x) + (z^2-3)\right]$$

with $z = y + y^{-1}$, has one real root $1+z$ and two imaginary roots (except when $y=1$ and $x=1$ is a double root). Thus, when **A** is a 3×3 matrix, C and R can be determined by taking geometric means and thus give the same weighting.

We turn to the 4×4 case, and let G denote the column vector whose entries are the geometric means of the rows of **A** and H denote the row vector whose entries are the geometric means of the columns of **A**. From the reciprocal property of **A**, it follows that G and H' are proportional. The situation is now more interesting. For example, let **A** have the special form

$$\begin{pmatrix} 1 & u & v & w \\ u^{-1} & 1 & t & v \\ v^{-1} & t^{-1} & 1 & u \\ w^{-1} & v^{-1} & u^{-1} & 1 \end{pmatrix}.$$

with all entries positive. For such matrices, the row and column eigenvectors are simply related; if $C = (a, b, c, d)^T$, then it is easy to see that $R = (d, c, b, a)$ is an eigenvector. In fact, for suitable λ,

$$\begin{aligned} \lambda a &= a + bu + cv + dw \\ \lambda b &= au^{-1} + b + ct + dv \\ \lambda c &= av^{-1} + bt^{-1} + c + du \\ \lambda d &= aw^{-1} + bv^{-1} + cu^{-1} + d. \end{aligned}$$

Suppose that R' is proportional to C. Then $ad = bc$. Setting $p = ub/a = ud/c$, $q = wd/a$, $r = tc/b$, and $s = vc/a$, and equating four different expressions for λ, we obtain

$$\begin{aligned} 1+p+s+q &= p^{-1} + 1 + r + s = s^{-1} + r^{-1} + 1 + p \\ &= q^{-1} + s^{-1} + p^{-1} + 1, \end{aligned}$$

which is equivalent to

$$p - p^{-1} = r - q = \frac{r-q}{rq} \quad \text{and} \quad s - s^{-1} = \frac{1-qr}{r} = \frac{1-qr}{q}.$$

Either $r = q$, in which case it follows that $w = tu^2$, or else $rq = 1$, in which case it follows that $v^2 = wt$.

Conversely, suppose that in the matrix **A**, it turns out that $w = tu^2$. Then, G is a column eigenvector with eigenvalue $\lambda = 2 + \sqrt{v/ut} + \sqrt{ut/v}$, so that R' and C are proportional. On the other hand, if $v^2 = wt$, then G is an eigenvector with eigenvalue

$$\lambda = 2 + (u^{1/2}t^{1/4}/w^{1/4}) + (w^{1/4}/u^{1/2}t^{1/4}),$$

so that again R' and C are proportional.

To see what happens when both the conditions $w = tu^2$ and $v^2 = wt$ fail, consider the special case $u = v = t = 1$. Then λ satisfies the equation

$$\lambda^3 - 4\lambda^2 - 2(w + w^{-1} - 2) = 0.$$

For R', G, and C, we can take, respectively,

$$\begin{pmatrix} \dfrac{\lambda^2 - 2\lambda}{2(\lambda - 1 + w^{-1})} \\ 1 \\ 1 \\ \dfrac{\lambda^2 - 2\lambda}{2(\lambda - 1 + w)} \end{pmatrix} \quad \begin{pmatrix} w^{1/4} \\ 1 \\ 1 \\ w^{-1/4} \end{pmatrix} \quad \begin{pmatrix} \dfrac{2(\lambda - 1 + w)}{\lambda^2 - 2\lambda} \\ 1 \\ 1 \\ \dfrac{2(\lambda - 1 + w^{-1})}{\lambda^2 - 2\lambda} \end{pmatrix}.$$

These will, in general, differ. It might be suggested that, although the actual relative weights differ, the rankings induced by R' and C are the same. However, even this is not so. Consider the following example, designed to give $G = (1,1,1,1)^T$:

$$\mathbf{A} = \begin{pmatrix} 1 & 5 & 2 & 0.1 \\ 0.2 & 1 & 2.5 & 2 \\ 0.5 & 0.4 & 1 & 5 \\ 10 & 0.5 & 0.2 & 1 \end{pmatrix}.$$

In this case, $\lambda_{\max} = 8.02$, $\overline{C} = (0.22, 0.19, 0.26, 0.33)^T$, $\overline{R} = (0.33, 0.26, 0.19, 0.22)$, and $\overline{R'} = (0.18, 0.23, 0.31, 0.28)^T$. The column eigenvector ranks the four options in the order 4, 3, 1, 2, while the row eigenvector ranks them in the order 3, 4, 2, 1. Here, one has to be careful in coming to a conclusion. However, in this example, the fact that the Perron eigenvalue far exceeds the trace of the matrix gives us pause, as we shall see in the next section.

One has the sense that, in the general case, the vector G has an interesting role to play. In particular, it would be of interest to compare G with R' and C in the case when R' and C are proportional.

Consistency

A critic of the AHP method of ranking might complain that the pairwise comparisons that went into the matrix **A** could be wildly out of line. To take an extreme case, the option pairs (P, Q), (Q, R), (P, R) might be rated 3, 5, and 1/2 respectively, so that P is somewhat more important than Q, Q considerably more than R, while, in a direct comparison, P is less significant than R. While a thoughtful assessment is not likely to produce such an extreme anomaly, nevertheless, some inconsistency is bound to occur. Fortunately, there is a mathematical way of getting a handle on the situation.

If we review the case that **A** is a 3×3 matrix, we see that the Perron eigenvalue $1 + y + y^{-1}$ is always at least 3, with equality exactly when $y = 1$. But this condition is equivalent to $uw = v$, which in turn characterizes the consistency of the matrix **A**. Thus, the largest eigenvalue of **A** exceeds 3 if and only if the matrix is not consistent, and equals 3 otherwise.

More generally, suppose that $\mathbf{A} = (a_{ij})$ is an $n \times n$ matrix with positive entries satisfying $a_{ij} = 1/a_{ji}$. Then, if $\lambda_{\max}$ is its eigenvalue of maximum absolute value, $\lambda_{\max} \geqslant n$, *and* **A** *is consistent if and only if* $\lambda_{\max} = n$. The proof of this is pleasantly straightforward, and worth including here. Let $(w_1, w_2, \ldots, w_n)^T$ be any positive column eigenvector with eigenvalue $\lambda = \lambda_{\max}$. Then, for $1 \leqslant i \leqslant n$,

$$\lambda = \sum_{j=1}^{n} a_{ij} w_j w_i^{-1}.$$

Summing these equations over i yields

$$n\lambda = \sum_{i,j} a_{ij} w_j w_i^{-1} = \sum_{i \neq j} a_{ij} w_j w_i^{-1} + n, \tag{1}$$

taking account of $a_{ii} = 1$. With $y_{ij} = a_{ij} w_j w_i^{-1}$, we have $y_{ji} = y_{ij}^{-1}$, so that equation (1) can be written

$$n\lambda = \sum_{1 \leqslant i < j \leqslant n} \left(y_{ij} + y_{ij}^{-1} \right) + n. \tag{2}$$

The sum in (2) is taken over $\binom{n}{2} = (1/2)n(n-1)$ terms, so that the right-hand side is at least equal to $2((1/2)n(n-1)) + n = n^2$, with equality if and only if each $y_{ij} = 1$, in other words, if $a_{ij} = w_i/w_j$. But this characterizes the consistency of **A**, and the result follows.

Thus, the difference $\lambda_{\max} - n$ can be regarded as a measure of consistency. Since the sum of all the eigenvalues of **A** is n (the trace of **A**), $\lambda - n$ is the negative of the sum of the remaining eigenvalues of **A**. The average of these eigenvalues is $-\mu$, where

$$\mu = \frac{\lambda - n}{n-1}.$$

This is the *consistency index* of **A**. If μ is too large, then the process is likely to be defective and the judgments made should be reviewed. In practice, one judges μ to be satisfactory if it is no more than about 10% of the mean consistency index for a sample of 500 randomly generated matrices satisfying $a_{ij} = a_{ji}^{-1}$ with entries drawn from the set $\{1/9, 1/8, \ldots, 1/2, 1, 2, \ldots, 9\}$. In the table below, the first row gives the order of the matrix and the second row the random mean consistency index:

n:	1	2	3	4	5	6	7	8	9	10
μ:	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

In the admissions example, the 4×4 matrix has consistency index $0.14/3 = 0.05$, less than 10% of the random consistency index, which is 0.90. The consistency indices of all the 3×3 matrices in that example are also well within the acceptable range. However, the 4×4 matrix of the last section which produced different rankings from the row and column eigenvectors has consistency index $4.02/3 = 1.34$, which is poor. It would be interesting to investigate how strong is the connection between the consistency index and the coherence of the row and column rankings.

Conclusion

We pursued the admissions procedure example on the assumption that the criteria and options could be organized into a hierarchy in a cut and dried way, and that at each level, we could take each criterion in isolation from the others. However, life is usually more complicated than this. Often, the criteria used might be interrelated. For example, the fairness of an admissions procedure would to some extent be governed by its predictability and cost (especially if the candidates were charged a fee); political acceptability might also hinge on other factors—if the procedure is perceived as fair, it would be much easier to swallow by all concerned. One way to handle this would be to introduce into the hierarchy an extra level in which the four criteria F, P, L, A are weighted with respect to each of the same four criteria. These could be combined (by means of a matrix multiplication as indicated earlier) with the preliminary weightings of G, C, and E in terms of the criteria. More complex situations will involve a nonlinear sort of hierarchy in which components of the criteria will affect each other in either direction. These complications are taken up, with examples, in [**7**] and [**8**] (see [**7**, chapter 8, pp. 206–222] for a detailed discussion of how to handle a system with feedback).

AHP is capable of considerable refinement to cope with the complexity of a situation for which a decision is required. Built into it is the capacity to adjust conveniently one's ranking of options to new judgments and new pieces of information, whether slight or significant, as well as a warning bell in the form of the consistency index. AHP has been used in many practical situations of industrial or government policy, and has been used to second-guess Britain's going to war over the Falkland Islands, [**9**]; it has also been the focus of experiments in which the conclusions are subject to independent checking [**7**, pp. 38–42].

As for the question of admissions policy, what was decided in the end? Actually, at the time this was written, the end had not yet come. The government is winding up a substantial revision of the curriculum, and the Minister of Education has opened the door to the possibility of the province restoring some form of universal testing of high school graduates. A government commission looking into the universities has also recommended tests. All this has had the effect of bringing to the fore another criterion, "sensitivity to political instability." On this basis, it was perceived that the universities would be foolish to embark on an elaborate new venture while matters are still quite unsettled. For the time being, nothing will change.

The author warmly thanks T. L. Saaty for introducing him to the analytic hierarchy process and suggesting an expository paper. He would also like to thank Prof. Saaty and the referees for their helpful advice and a number of the references.

References

[1] A. Berman and R. J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, 1979.
[2] H. D. Brunk, Mathematical models for ranking from paired comparisons, J. Amer. Stat. Assoc., 55 (1960) 503–520.
[3] H. A. David, The Method of Paired Comparisons, Griffin's Statistical Monographs & Courses #12, 1963.
[4] F. R. Gantmacher, Applications of the Theory of Matrices, Interscience, New York, 1960.
[5] D. G. Kendall, A mathematical approach to seriation, Phil. Trans. Roy. Soc. Lond., A 269 (1970) 125–135.
[6] J. W. Moon, Topics on Tournaments, Holt, Rinehart & Winston, 1968.
[7] T. L. Saaty, The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill, 1980.
[8] _____, A scaling method for priorities in hierarchical structures, J. Math. Psychology, 15 (1977) 234–280.
[9] _____, Conflict resolution and the Falkland Islands invasions, Interfaces, 13 (1983) 68–83 (#6).
[10] E. Senata, Non-negative Matrices and Markov Chains, 2nd ed., Springer, New York, 1973, 1981.
[11] A. Shuchat, Matrix and network models in archaeology, this MAGAZINE, 57 (1984) 3–14.
[12] A. Tversky and D. Kahneman, The framing of decisions and the psychology of choice, Science, 211 (1981) 453–458.

Continued Roots

WALTER S. SIZER
Moorhead State University
Moorhead, MN 56560

Examples of infinitely nested roots appear from time to time in the literature and as problems to be considered. One instance was the problem presented by Ramanujan in the *India Journal of Mathematics* (see [4]) and later converted to problem A6 of the 1966 Putnam Examination, where the contestant was asked to "Justify the statement that

$$3 = \sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + 4\sqrt{1 + 5\sqrt{1 + \cdots}}}}}\ ,"$$

(see [16]). Another similar Putnam Examination problem was rephrased as a problem in this MAGAZINE in May, 1983 (for the problem and solution, see [12]). Other problems involving nested radicals are given in [10], [11], and [17]. There seems to be little general theory for nested radicals, however, and even no uniformity as to form or notation. This situation contrasts sharply with the case of other infinitely repeated operations (infinite series, products, and continued fractions), where the abstract theory is well documented in the literature. In this note we will consider certain infinitely nested roots and derive a few general results.

Our goal will be to consider what we shall call a **continued (square) root**, by which we shall mean an expression of the form

$$a_0 + \sqrt{}(a_1 + \sqrt{}(a_2 + \sqrt{}(a_3 + \sqrt{}(a_4 + \cdots)))), \tag{1}$$

where the a_i's are numbers. Familiarity suggests we limit ourselves to the case where the a_i's are real numbers, with $a_i \geq 0$ for $i \geq 1$, or even be more restrictive and require the a_i's to be integers.

Several questions about continued roots present themselves naturally at this point—in fact, they are precise analogues of familiar questions posed for series, products, and continued fractions. Some of these are:

(1) *What does it mean for a continued root to converge to a number L*?
(2) *What conditions on the a_i's guarantee convergence of the root*?
(3) *What numbers can be represented by continued roots*?
(4) *Is there any uniqueness to such representations*?
(5) *What numbers are represented by "terminating" or "repeating" continued roots*?

In our exploration of continued roots in this paper we shall give at least partial answers to the above five questions.

Definition and Convergence

For ease in notation we will denote the continued root in (1) by $\sqrt{}(a_0, a_1, \ldots)$, and we write $L_n = \sqrt{}(a_0, \ldots, a_n)$ for the continued root truncated after a_n. Patterning our definition after what is done with infinite sums and products and with continued fractions, we define

$$\sqrt{}(a_0, a_1, \ldots) = \lim_{n \to \infty} L_n,$$

provided the indicated limit exists.

One would like at this stage to identify circumstances under which the above limit exists. In the

event that the a_i's are all nonnegative real numbers for $i \geq 1$, the truncations L_i will form a nondecreasing sequence of real numbers; hence the sequence $\{L_i\}$ has a limit if and only if the L_i's are bounded (see, for example, [3], p. 16).

One case where the L_i's are bounded is the case in which the a_i's themselves are bounded. Suppose, for example, that $a_i \leq B$ for all natural numbers i. We may safely take B to be greater than or equal to 2, and so $a_i \leq B \leq B(B-1)$. Then it is easy to see that

$$L_i \leq \sqrt{}(a_0, B(B-1), \ldots, B(B-1)),$$

where there are i terms $B(B-1)$. But

$$L_1 \leq a_0 + \sqrt{}B(B-1) \leq a_0 + \sqrt{}B^2 = a_0 + B;$$

$$L_2 \leq a_0 + \sqrt{}(B(B-1) + \sqrt{}(B(B-1)) \leq a_0 + \sqrt{}(B(B-1) + \sqrt{}B^2)$$

$$= a_0 + \sqrt{}(B(B-1) + B) = a_0 + \sqrt{}B^2 = a_0 + B.$$

Continuing in this manner (or, rather, using a proof by mathematical induction) we get that $L_i \leq a_0 + B$, regardless of the index i. We have thus established the following

PROPOSITION. *If $\{a_i\}$ is a bounded sequence of real numbers and if a_i is nonnegative for $i \geq 1$, then $\sqrt{}(a_0, a_1, \ldots)$ converges.*

This first result is far from the best result one can get for convergence of continued roots, but it is a start. Naively, the next step might be to proceed as follows: suppose $a_1, a_2, \ldots$ are nonnegative real numbers with $\sqrt{}(a_0, a_1, \ldots) = L$, and suppose $M > 0$. Then

$$ML = M(a_0 + \sqrt{}(a_1 + \sqrt{}(a_2 + \cdots))) = Ma_0 + M\sqrt{}(a_1 + \sqrt{}(a_2 + \cdots))$$

$$= Ma_0 + \sqrt{}(M^2a_1 + M^2\sqrt{}(a_2 + \cdots)) = Ma_0 + \sqrt{}(M^2a_1 + \sqrt{}(M^4a_2 + M^4\sqrt{}(a_3 + \cdots)))$$

$$= Ma_0 + \sqrt{}(M^2a_1 + \sqrt{}(M^4a_2 + \sqrt{}(M^8a_3 + \cdots))) = \sqrt{}(Ma_0, M^2a_1, \ldots, M^{2^i}a_i, \ldots).$$

A careful examination in terms of limits shows that this procedure is indeed legitimate! Using the sequence $a_i = 1$ and $M = 2$ gives a convergent continued root for which the terms of the root are unbounded.

One might speculate that all continued roots with $a_i \geq 0$ for $i \geq 1$ converge (such a result does hold, after all, for continued fractions—see, for example, [9], p. 67). However, such is not the case, and in fact one gets the following.

THEOREM. *Suppose a_i is real for all $i \geq 0$ and that $a_i \geq 0$ for $i \geq 1$. Then $\sqrt{}(a_0, a_1, \ldots)$ converges if and only if the set*

$$S = \left\{\sqrt[2^i]{}a_i : i \geq 1\right\}$$

is bounded.

Proof. If: Suppose S is bounded by a number B. By the observation before the statement of the theorem, then, $\sqrt{}(a_0, a_1, \ldots) = B\sqrt{}(a'_0, a'_1, \ldots)$, where $0 \leq a'_i \leq 1$ for $i \geq 1$. The Proposition tells us that $\sqrt{}(a'_0, a'_1, \ldots)$ converges, and hence $\sqrt{}(a_0, a_1, \ldots)$ converges.

Only if: Suppose now that S is not bounded. Thus for any real number B there exists an N such that $\sqrt[2^N]{}(a_N) > B$. We wish to show that $\{L_i\}$ diverges, that is, that it increases without bound. Thus we want to show that for any B there exists a number $N > 0$ such that $L_n > B$ for $n > N$. Since the L_i's form a nondecreasing sequence it suffices to find an N such that $L_N > B$. Let B be given, and let N be such that $\sqrt[2^N]{}(a_N) > B$. A simple check then shows that $L_N \geq \sqrt[2^N]{}(a_N) > B$.

Representing numbers as continued roots

Once the question of convergence is settled one can "solve" some continued roots—particularly those with certain repeating patterns. We illustrate the technique with two examples. Note that it was important to settle the convergence question first, in order that our assumption that we can write $L = \sqrt{}(a_0, a_1, \ldots)$ be justified!

EXAMPLE. Let $L = \sqrt{}(0,1,1,1,\ldots)$. Then $L = \sqrt{}(1 + \sqrt{}(1 + \sqrt{}(1 + \cdots)))$, so $L^2 = 1 + \sqrt{}(1 + \sqrt{}(1 + \cdots))$, or $L^2 = 1 + L$. Solving, we get $L^2 - L - 1 = 0$, so $L = (1 \pm \sqrt{}(1+4))/2$; since $L > 0$ we get $L = 1/2 + \sqrt{5}/2$. This number is the "golden ratio" of the Greeks (see, for example, [**9**], p. 82). In fact, this example is just a special case of the more general formula

$$n = \sqrt{}(0, n(n-1), n(n-1), \ldots),$$

which can be shown to hold for any number $n > 1$.

EXAMPLE. Let $L = \sqrt{}(1 + \sqrt{}(7 + \sqrt{}(1 + \sqrt{}(7 + \cdots))))$. Here $L^2 = 1 + \sqrt{}(7 + \sqrt{}(1 + \sqrt{}(7 + \cdots)))$, so $L^2 - 1 = \sqrt{}(7 + \sqrt{}(1 + \sqrt{}(7 + \cdots)))$, $(L^2 - 1)^2 = 7 + \sqrt{}(1 + \sqrt{}(7 + \cdots)) = 7 + L$, or $L^4 - 2L^2 - L - 6 = 0$. By Descartes' rule of signs (see [**14**], vol. II, p. 471) this equation has one and only one positive root, which is L. By inspection, this root is 2, so $L = 2$.

These examples point to answers to two more questions about continued roots. The technique used to solve for the roots in the examples can be used to show that any terminating or repeating continued root represents a root of a monic polynomial of degree 2^i for $i \geqslant 1$; in case the entries a_i in the root are all integers, the root represents a root of a monic polynomial over the integers. The second observation to be made based on the examples is that there may be several distinct ways to represent the same number as a continued root. Thus, 2 can be represented $\sqrt{}(2,0,0,0,\ldots)$; $\sqrt{}(0,4,0,0,\ldots)$; $\sqrt{}(0,2,2,2,\ldots)$; or $\sqrt{}(0,1,7,1,7,\ldots)$.

The question remains of which real numbers can be represented by continued roots; for example, can π or $\sqrt[3]{9}$ be so represented? If we use real numbers in our continued roots we can of course represent any real number. For example, if x is nonnegative, $x = \sqrt{}(0, x^2, 0, 0, \ldots)$. Thus this question really concerns just the case where the a_i's are integers.

Let x be any number, and suppose we want to write x as $\sqrt{}(a_0, a_1, \ldots)$. If any of the numbers a_i, $i \geqslant 1$, are nonzero, necessarily nonnegative integers, then we will have $x = a_0 + r$, where $r \geqslant 1$. Thus if we are choosing a_0, we should choose it to be less than or equal to $x - 1$. Let us choose a_0 to be the integer so that x is in the interval $(a_0 + 1, a_0 + 2]$. A similar analysis shows that if $a_i > 0$ for any $i \geqslant 2$, then $x = a_0 + \sqrt{}(a_1 + r)$, where $r \geqslant 1$. Thus we might choose a_1 so that x is in the interval $(a_0 + \sqrt{}(a_1 + 1), a_0 + \sqrt{}(a_1 + 2)]$. Note that since $\sqrt{}(a_1 + 2) \leqslant 2$, a_1 is 0, 1, or 2. We can continue this pattern: assuming x is in the interval $(\sqrt{}(a_0, a_1, \ldots, a_i + 1), \sqrt{}(a_0, a_1, \ldots, a_i + 2)]$, choose a_{i+1} so that x lies in $(\sqrt{}(a_0, a_1, \ldots, a_{i+1} + 1), \sqrt{}(a_0, a_1, \ldots, a_{i+1} + 2)]$. Again, note that a_{i+1} is 0, 1, or 2 for $i > 0$.

The continued root $\sqrt{}(a_0, a_1, \ldots)$ we get in this manner converges by our first proposition. We would like to know that, for any choice of x, the value L of our continued root is equal to x. Since for each truncated root L_i we know that $x > L_i$, we conclude that $x \geqslant L$. To prove that $x = L$, we need to show that x cannot exceed L. We do this by assuming that $x - L = \varepsilon > 0$, and obtain a contradiction. Assume $x - L = \varepsilon > 0$. Since the L_i's converge to L, we can choose an i so that $L - L_i < \varepsilon/2$. Also we know that x is in the interval $(\sqrt{}(a_0, a_1, \ldots, a_i + 1), \sqrt{}(a_0, a_1, \ldots, a_i + 2]$, which we shall denote by $(A, B]$. We express what we know graphically in FIGURE 1. (Note that we use two number lines in FIGURE 1, not necessarily intended to have uniform scale, as we do not yet know how A and L compare.)

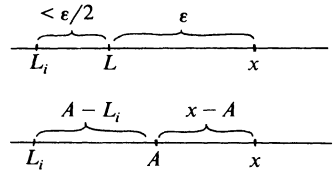


FIGURE 1

We complete the proof in essentially two steps. First, we show (i) that $A - L_i \geqslant B - A$. From this it will follow, since $B - A \geqslant x - A$ and $x - L_i \geqslant \varepsilon$, that $a - L_i \geqslant \varepsilon/2$, so $A > L$. We then prove (ii) that there is an n with $L_n \geqslant A$; it then follows that $L_n > L$, which will contradict the fact that L is the limit of the nondecreasing sequence $\{L_i\}$.

To show (i), that $A - L_i \geqslant B - A$, consider the function $f(x) = \sqrt{}(a_0, a_1, \ldots, a_{i-1}, x+1) - \sqrt{}(a_0, a_1, \ldots, a_{i-1}, x)$. Straightforward computation shows that $f'(x) < 0$ for $x > 0$, and thus that $f(x)$ is decreasing on $[0, \infty)$. Thus $f(a_0) \geqslant f(a_0 + 1)$; but this is the statement that $A - L_i \geqslant B - A$.

Now we show that (ii) there is an n with $L_n \geqslant A$. By inspection, if for any $k > i$, $a_k \neq 0$, then $L_k \geqslant A = \sqrt{}(a_0, a_1, \ldots, a_i + 1)$, so the only way we can have no $L_n \geqslant A$ is to have $a_m = 0$ for all $m > i$. But by construction, $x > A$, that is, $A = \sqrt{}(a_0, a_1, \ldots, a_i + 1) < x \leqslant \sqrt{}(a_0, a_1, \ldots, a_i + 2)$. Since $\lim_{j \to \infty} \sqrt[2^j]{2} = 1$, there is a k such that $\sqrt{}(a_0, a_1, \ldots, a_i + \sqrt[2^k]{2}) < x$. But in this case, even if a_j were 0 for $j = i+1, \ldots, i+k-1$, a_{i+k} would be at least 1 by construction and we would have (in any case) $L_{i+k} \geqslant A$. This was what we needed to complete the proof of the following.

THEOREM. *Any real number can be represented as a continued root* $\sqrt{}(a_0, a_1, \ldots,)$, *where the* a_i*'s are integers and for* $i \geqslant 1$, a_i *is* 0, 1, *or* 2.

EXAMPLE. We will illustrate the construction outlined above to get the first several entries in a continued root expansion of $\sqrt[3]{}9$. Recall that our rule is to choose each a_i so that

$$\sqrt{}(a_0, a_1, \ldots, a_i + 1) < x \leqslant \sqrt{}(a_0, a_1, \ldots, a_i + 2).$$

The computations will be done using a calculator; thus we get $x = \sqrt[3]{}9 \approx 2.0800838$. Since $2 < x \leqslant 3$, a_0 must equal 1. To determine a_1, note that

$$1 + \sqrt{}1 = 2,$$
$$1 + \sqrt{}2 \approx 2.4142136,$$

so $a_1 = 0$. Continuing,

$$1 + \sqrt{}(0 + \sqrt{}1) = 2,$$
$$1 + \sqrt{}(0 + \sqrt{}2) \approx 2.1892071,$$

so $a_2 = 0$;

$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}1)) = 2,$$
$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}2)) \approx 2.0905077,$$

so $a_3 = 0$;

$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}1))) = 2,$$
$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}2))) \approx 2.0442738,$$
$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}3))) \approx 2.0710755,$$
$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}4))) \approx 2.0905077,$$

so $a_4 = 2$;

$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(2 + \sqrt{}1)))) \approx 2.0710755,$$
$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(2 + \sqrt{}2)))) \approx 2.0797685,$$
$$1 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(0 + \sqrt{}(2 + \sqrt{}3)))) \approx 2.0857922;$$

so $a_5 = 1$.

Evidently our calculator would allow us to go a few more steps, but we would be limited by our seven-place accuracy. At any rate, there will be a continued root representation of $\sqrt[3]{}9$ which begins $\sqrt{}(1, 0, 0, 0, 2, 1, \ldots)$.

Notes on references

One other, older reference to continued roots in the literature is in François Viète's formula

$$2/\pi = \sqrt{\tfrac{1}{2}} \cdot \sqrt{\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{1}{2}}\right)} \cdot \sqrt{\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{1}{2}}\right)}\right)} \ldots$$

published in 1593 ([**2**]; also [**1**] and [**14**], vol. I, p. 312; see [**15**] for a proof of the formula). While the formula is an infinite product of finite continued roots, its convergence does imply that $\sqrt{(0, 1/2, 1/8, 1/128, \ldots)} = 1$. Many introductory calculus texts treat infinite series, as do several special books devoted to just that topic; see, for example, the books by J. A. Green ([**3**]) and James M. Hyslop ([**6**]). Such treatments of infinite products are harder to find, but the interested reader is referred to the appropriate sections of books by Hirschmann ([**5**]) and Knopp ([**8**]). Continued fractions are discussed in books by Khinchin ([**7**]) and Olds ([**9**]), and in the article by Richards ([**13**]). The development of continued roots presented in this note and the questions addressed parallel the usual approaches to series, infinite products, and continued fractions, and reinforce concepts of convergence seen in these more conventional areas.

The author is indebted to Edward J. Allen of the University of North Carolina at Asheville for providing several references.

References

[1] L. Baxter, Are π, e, and $\sqrt{2}$ equally difficult to compute?, Amer. Math. Monthly, 88 (1981) 50–51.
[2] Petr Beckmann, A History of π, The Golem Press, Boulder, Colorado, 1971, p. 94.
[3] J. A. Green, Sequences and Series, The Free Press, Glencoe, Illinois, 1958.
[4] G. H. Hardy, P. V. Seshu Aigar, and B. M. Wilson, Collected Papers of Srinivasa Ramanujan, Chelsea Pub. Co., New York, 1962, p. 323.
[5] I. I. Hirschmann, Infinite Series, Holt, Rinehart & Winston, New York, 1962.
[6] James M. Hyslop, Infinite Series, Interscience Publishers, Inc., New York, 1959.
[7] A. Ya Khinchin, Continued Fractions, University of Chicago Press (Phoenix Books), Chicago, 1964.
[8] Konrad Knopp, Infinite Series and Sequences, Dover Publishers, Inc., New York, 1956.
[9] C. D. Olds, Continued Fractions, New Math. Library, Math. Ass'n. of America, Washington, D.C., 1963.
[10] Problems 75 and 78, National Mathematics Magazine, 9 (1934–35) 208–210, 252.
[11] Problem 460, Amer. Math. Monthly, 24 (1917) 32–33.
[12] Problem 1174, this MAGAZINE, 57 (1984) 299–300.
[13] Ian Richards, Continued fractions without tears, this MAGAZINE, 54 (1981) 163–171.
[14] David Eugene Smith, History of Mathematics, vols. I, II, Ginn & Co., Boston, 1923, I, p. 312; II, p. 471.
[15] F. Rudio, Zeitschrift für Math. und Phys., 36 (1891) 139–140.
[16] William Lowell Putnam Mathematical Contest, A6, Amer. Math. Monthly, 74 (1967) 775.
[17] Leo Zippin, Uses of Infinity, New Math. Library, Math. Ass'n. of America, Washington, D.C., 1962, pp. 15, 51, 83.

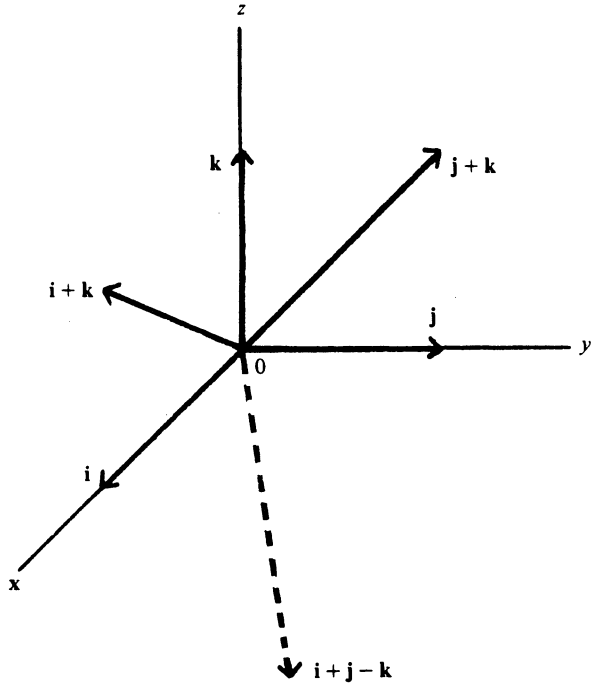
A Surprise from Geometry

Ross A. Honsberger
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

It is patently obvious that two vectors in the plane (all vectors are considered to issue from the origin), which meet at an angle that does not exceed a right angle, can be spun around the origin so that both vectors lie in the nonnegative quadrant (that is, the endpoint (x, y) of each vector has coordinates which are both nonnegative). It is not quite so obvious that a set of 3 vectors in 3-space, which in pairs meet at angles not exceeding a right angle, can always be spun around the origin to lie in the nonnegative octant. It is not at all obvious, but is also true, that any set of 4 vectors in 4-space, no 2 of which meet at an angle greater than a right angle, can be arranged to lie in the nonnegative orthant (orthant is the general term for quadrant and octant).

At this point, who can resist the conjecture that any set of n vectors in n-space, no 2 of which meet at an angle exceeding a right angle, can be rotated to a position so that all the vectors in the set are contained in the nonnegative orthant? Isn't it surprising that this is false for every $n > 4$?

We shall see that the set S, consisting of the five 3-dimensional vectors $\mathbf{i} = (1,0,0)$, $\mathbf{j} = (0,1,0)$, $\mathbf{i} + \mathbf{k} = (1,0,1)$, $\mathbf{j} + \mathbf{k} = (0,1,1)$, and $\mathbf{i} + \mathbf{j} - \mathbf{k} = (1,1,-1)$, cannot all lie in the nonnegative orthant of a space of any dimension.



First of all, it is easy to check that each pair of vectors in S meets at an angle not exceeding a right angle, and a glance at the figure shows that they fan out too far to fit into an octant of 3-space. We shall establish our general claim by arguing to a contradiction. Suppose, then, that in some n-space, our set S can be accommodated completely within the nonnegative orthant. Then the coordinates of each vector in S are all nonnegative. Since S does not contain the zero vector, none of these n-tuples of coordinates will consist entirely of 0's; in each case, at least one coordinate must actually be a positive number.

Now, the crux of our argument consists in showing that the positioning of S in the nonnegative orthant necessarily also brings into this orthant the companion vector $\mathbf{k} = (0, 0, 1)$, even though it does not belong to S.

While each coordinate of a vector in S is either positive or zero, a coordinate in the description of $\mathbf{k}$'s position may presumably be positive, zero, or negative. Let us investigate the feasibility of a negative coordinate in $\mathbf{k}$. If $\mathbf{k}$ were to have a negative coordinate in a component in which the vector $\mathbf{i}$ has a zero, then that component in their sum $\mathbf{i} + \mathbf{k}$ would have a negative value. But, since $\mathbf{i} + \mathbf{k}$ belongs to S, no component of $\mathbf{i} + \mathbf{k}$ is negative. Consequently, $\mathbf{k}$ can have a negative coordinate only in a position in which $\mathbf{i}$ has a positive coordinate (recall that the coordinates of $\mathbf{i}$ are either positive or zero). Similarly for the vector $\mathbf{j}$: a negative component in $\mathbf{k}$, opposite a zero in $\mathbf{j}$, would yield a contradictory negative component in the vector $\mathbf{j} + \mathbf{k}$ of S. As a result, $\mathbf{k}$ can have a negative coordinate only in a place in which both $\mathbf{i}$ and $\mathbf{j}$ have a positive coordinate.

But there are no such places! If there were, such a pair of positive coordinates would contribute a positive amount t to the dot product $\mathbf{i} \cdot \mathbf{j}$. However, since $\mathbf{i}$ and $\mathbf{j}$ are orthogonal, we have $\mathbf{i} \cdot \mathbf{j} = 0$ in every coordinate system, yet there would be no way to nullify the above contribution t because there are no negative coordinates in any vector of S (in particular, in $\mathbf{i}$ and $\mathbf{j}$). It follows, then, that $\mathbf{k}$ possesses no negative coordinates and must also reside in the nonnegative orthant.

Now we can conclude easily. Since both the vectors $\mathbf{k}$ and $\mathbf{i} + \mathbf{j} - \mathbf{k}$ have no negative coordinates, their dot product $\mathbf{k} \cdot (\mathbf{i} + \mathbf{j} - \mathbf{k})$ cannot be a negative number. However, obvious orthogonalities yield

$$\begin{aligned} \mathbf{k} \cdot (\mathbf{i} + \mathbf{j} - \mathbf{k}) &= \mathbf{k} \cdot \mathbf{i} + \mathbf{k} \cdot \mathbf{j} - \mathbf{k} \cdot \mathbf{k} \\ &= 0 + 0 - |\mathbf{k}|^2, \end{aligned}$$

which is negative, since $\mathbf{k}$ is not the zero vector, and the argument is complete.

This argument is due to the Israeli mathematician Moshe Roitman of the University of Haifa; it was most kindly communicated to me by his colleague Joe Zaks during his recent visit to Waterloo (summer, 1984). He also noted that L. M. Kelly and Shreedharan of Michigan State University in East Lansing had a similar example of a set of 5 vectors which lent itself to an easy argument based on inner products.

Proofs of the cases of vectors in 3-space and 4-space can be found in [**1**]. It is interesting that the concluding remark in this paper is an example of 8 vectors that need a space of at least 9 dimensions for their accommodation in the nonnegative orthant.

References

[1] L. J. Gray and D. G. Wilson, Nonnegative factorization of positive semidefinite nonnegative matrices, Linear Algebra and its Applications, 31 (1980) 119–127.

A Transfer Device for Matrix Theorems

William P. Wardlaw
U. S. Naval Academy
Annapolis, MD 21402

Our title refers to a method for obtaining a number of results for matrices over arbitrary commutative rings by "transferring" the corresponding results for matrices over the real numbers. The technique was suggested by a proof [5] in a calculus text which showed $\det(AB) = (\det A)(\det B)$ for A and B nonsingular, and then extended the result to singular A or B by continuity. More or less, the technique described in this note is an algebraic substitute for the use of continuity which can serve as a rigorous replacement for waving the hands and stating "For commutative rings, everything goes through as for fields." The existence of the transfer device obviates the need to do undergraduate linear algebra over commutative rings and suggests that a restriction to the field $\mathbf{R}$ of real numbers (or perhaps the field $\mathbf{C}$ of complex numbers) will suffice, since many results can be "transferred" to more general settings in a graduate course.

Throughout this note, R is an arbitrary commutative ring, $R^{m\times n}$ is the collection of all $m \times n$ matrices over R, $R_n = R^{n\times n}$, and $R[t]$ is the ring of polynomials over R. Here, $R[t]$ is considered to be the ring formally generated by t and R, containing R as the constant polynomials and all of the powers t^k for positive k, even if R does *not* have an identity 1. Finally,

$$\mathbf{M}(R,t) = \bigcup_{m,n\in\mathbf{N}} R[t]^{m\times n}$$

is the partial algebra of all elements of $R = R_1$, all polynomials in $R[t] = R[t]_1$, and all matrices with entries in $R[t]$. The two operations $+$ and $\cdot$ in $\mathbf{M}(R,t)$ are ordinary matrix addition (with $A + B$ defined when A and B are the same size) and matrix (or scalar) multiplication (with $A \cdot B$ defined when A is $m \times n$ and B is $n \times p$ *or* when either A or B is 1×1). The phrase partial algebra refers to the fact that the operations are not always defined.

Recall that the determinant of a square matrix $A = (a_{ij})$ in $\mathbf{M}(R,t)$ is

$$\det A = \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma)\, a_{1\sigma_1} a_{2\sigma_2} \cdots a_{n\sigma_n}.$$

The characteristic polynomial of A in R_n is $f_A = f_A(t) = \det(tI - A)$, and the (classical) adjoint or adjugate of A is $\operatorname{Adj} A = C^T$, the transpose of the cofactor matrix $C = (c_{ij})$, where $c_{ij} = (-1)^{i+j}\det A(i|j)$ and $A(i|j)$ is the matrix resulting from A by deleting the ith row and the jth column.

The transfer device and applications

Transfer Theorem. *Let R and R' be commutative rings and $\theta\colon R \to R'$ be a ring homomorphism. Then θ induces a homomorphism*

$$\phi\colon \mathbf{M}(R,t) \to \mathbf{M}(R',t)$$

satisfying:

(1) $\phi(a) = \theta(a)$ *for every* $a \in R$,
(2) $\phi(\sum_{i=0}^{n} a_i t^i) = \sum_{i=0}^{n} \theta(a_i) t^i$ *for* $a_0, a_1, \ldots, a_n \in R$,
(3) $A = (a_{ij}) \in R[t]^{m\times n}$ *implies* $\phi(A) = (\phi(a_{ij}))$,
(4) $\phi(A + B) = \phi(A) + \phi(B)$ *when $A + B$ is defined*,
(5) $\phi(A \cdot B) = \phi(A) \cdot \phi(B)$ *when $A \cdot B$ is defined*,
(6) $\phi(\det A) = \det(\phi A)$ *when A is square*,
(7) $\phi(f_A) = f_{\phi(A)}$ *when A is square, and*
(8) $\phi(\operatorname{Adj} A) = \operatorname{Adj}(\phi A)$ *when A is square.*

The proof of the Transfer Theorem is not difficult: we *define* ϕ by properties (1)–(3) and then prove properties (4)–(8). It is clear that ϕ is well defined by properties (1)–(3). The proof of properties (4) and (5), which show that ϕ is a homomorphism, is straightforward but tedious, and hence is omitted. Properties (6)–(8) then follow because $\det A$, the coefficients of f_A, and the entries in $\operatorname{Adj} A$ are all polynomials in the entries of A.

Throughout the remainder of this note applications of the Transfer Theorem will be demonstrated by using it to prove, first, some very well known theorems about determinants over commutative rings and, later, some less well known theorems.

THEOREM 1. *If A and B are square matrices over a commutative ring R, then*

$$\det(AB) = (\det A)(\det B).$$

Proof. Let $A = (a_{ij})$ and $B = (b_{ij})$ be $n \times n$ matrices over the ring R. Then let $\mathbf{A} = \langle A, B \rangle \equiv \langle a_{11}, \ldots, a_{nn}, b_{11}, \ldots, b_{nn} \rangle$ be the subring of R generated by the entries of A and B. Let $X = (x_{ij})$ and $Y = (y_{ij})$ be $n \times n$ matrices over the field $\mathbf{R}$ of real numbers with $2n^2$ independent transcendental entries x_{ij} and y_{ij} in $\mathbf{R} \setminus \mathbf{Q}$. Then let $K = \mathbf{Q}(X, Y) \equiv \mathbf{Q}(x_{11}, \ldots, x_{nn}, y_{11}, \ldots, y_{nn})$ be the $2n^2$-fold transcendental extension of $\mathbf{Q}$, and let $\mathbf{X} = \langle X, Y \rangle \equiv \langle x_{11}, \ldots, x_{nn}, y_{11}, \ldots, y_{nn} \rangle$ be the subring of K generated by the entries of X and Y. Since the transcendentals x_{ij} and y_{ij} are algebraically independent over $\mathbf{Q}$, they generate a free commutative ring (that is, the free algebra over the class of all commutative rings, with the free generating family $\{x_{ij}, y_{ij}\}$, as defined in [**1**]), which is actually just the set of all nonconstant polynomials in the polynomial ring $\mathbf{Z}[X, Y] \equiv \mathbf{Z}[x_{11}, \ldots, x_{nn}, y_{11}, \ldots, y_{nn}]$. Hence, there is a homomorphism $\theta\colon \mathbf{X} \to \mathbf{A}$ such that $\theta(x_{ij}) = a_{ij}$, $\theta(y_{ij}) = b_{ij}$ for each i and j. Defining ϕ as in the Transfer Theorem and using the fact that $\det(XY) = (\det X)(\det Y)$ for the matrices X and Y over $\mathbf{R}$, we obtain

$$\begin{aligned}\det(AB) &= \det(\phi X \cdot \phi Y) = \det \phi(XY) = \phi \det(XY) = \phi((\det X)(\det Y)) \\ &= (\phi(\det X))(\phi(\det Y)) = (\det(\phi X))(\det(\phi Y)) = (\det A)(\det B)\end{aligned}$$

after several applications of the Transfer Theorem.

The techniques of the above proof will be repeated with minor modifications to prove the theorems which follow. To save space and to relieve tedium, many of the details given above will be omitted.

THEOREM 2 (Cayley-Hamilton). *$A \in R_n$ implies $f_A(A) = 0$.*

Proof. For any $A \in R_n$, let $\mathbf{A} = \langle A \rangle$ be the subring of R generated by the entries a_{ij} of A and let $\mathbf{X} = \langle X \rangle$ be the free subring of $K = \mathbf{Q}(X)$ generated by the n^2 transcendental entries x_{ij} of X. Let ϕ be the canonical homomorphism from $\mathbf{M}(\mathbf{X}, t)$ to $\mathbf{M}(\mathbf{A}, t)$ given by the Transfer Theorem satisfying $\phi(x_{ij}) = a_{ij}$ for all i and j. Then

$$f_A(A) = f_{\phi X}(\phi X) = (\phi f_X)(\phi X) = \phi(f_X(X)) = \phi(0) = 0$$

follows from the Transfer Theorem and the Cayley-Hamilton Theorem $f_X(X) = 0$ for matrices over $\mathbf{R}$.

THEOREM 3. *If A is a square matrix over the commutative ring R, then*

$$A(\operatorname{Adj} A) = (\det A)\, I = (\operatorname{Adj} A)\, A.$$

Proof. Let A, X, and ϕ be as in the proof of Theorem 2. Then

$$X(\operatorname{Adj} X) = (\det X)\, I = (\operatorname{Adj} X)\, X$$

holds for the matrix X over $\mathbf{R}$, and so the images under ϕ of the above three expressions must also be equal, i.e.,

$$A(\operatorname{Adj} A) = (\det A)\, I = (\operatorname{Adj} A)\, A.$$

So far, we have used the Transfer Theorem only to "transfer" a theorem that is well known for

matrices over the real numbers to obtain the corresponding theorem for matrices over an arbitrary commutative ring R. The next three theorems are interesting not only because they are less well known than the preceding three, but also because even their proof for arbitrary matrices over the field **R** of real numbers makes use of the transfer theorem. The first two of these will be proved by obtaining the result for invertible matrices over **R** and then applying the transfer theorem to obtain the corresponding results for any matrices over an arbitrary commutative ring R. In particular, this establishes the results for singular matrices over the real numbers. (Note that this approach could also have been taken for Theorem 1; indeed, such a proof would be the algebraic equivalent of the continuity proof [**5**] that motivated this paper.)

THEOREM 4. *Let A and B be $n \times n$ matrices over a commutative ring R. Then*

$$\operatorname{Adj}(AB) = (\operatorname{Adj} B)(\operatorname{Adj} A).$$

Proof. Case 1. Assume A and B are invertible over **R**. Then

$$\begin{aligned}\operatorname{Adj}(AB) &= \det(AB)\cdot(AB)^{-1} = (\det A)(\det B)\,B^{-1}A^{-1}\\ &= (\det B)\,B^{-1}\cdot(\det A)\,A^{-1} = (\operatorname{Adj} B)(\operatorname{Adj} A).\end{aligned}$$

Case 2. $A, B \in R_n$. Let X, Y, and ϕ be chosen as in the proof of Theorem 1. Since the elements x_{ij} of X are algebraically independent, $\det X$ can be considered as a polynomial in the x_{ij} with rational coefficients, and X is singular if and only if this polynomial is identically 0. However, the substitution $x_{ij} = \delta_{ij}$ (where δ_{ij} is the Kronecker delta defined by $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$) gives $X = I$ and $\det X = 1$, so the polynomial $\det X$ cannot be identically 0. Thus X is an invertible matrix over **R**, and so is Y. Hence $\operatorname{Adj}(XY) = (\operatorname{Adj} Y)(\operatorname{Adj} X)$ by Case 1, and the Transfer Theorem gives

$$\operatorname{Adj}(AB) = \phi(\operatorname{Adj}(XY)) = \phi((\operatorname{Adj} Y)(\operatorname{Adj} X)) = (\operatorname{Adj} B)(\operatorname{Adj} A).$$

THEOREM 5. *If $A \in R_n$, then $\operatorname{Adj} A = p_A(A)$ is a polynomial p_A evaluated at A, where*

$$p_A(t) = (-1)^{n+1}[f_A(t) - f_A(0)]/t.$$

Proof. Case 1. Assume A is invertible over **R**. Then

$$Ap_A(A) = (-1)^{n+1}[f_A(A) - f_A(0)\cdot I] = (-1)^n f_A(0)\cdot I = (\det A)\cdot I = A(\operatorname{Adj} A)$$

implies that

$$p_A(A) = \operatorname{Adj} A$$

upon left multiplication by A^{-1}.

Case 2. Let A, X, and ϕ be as in the proof of Theorem 2. Then X is invertible over **R**, so $p_X(X) = \operatorname{Adj} X$. Recalling the definition of p_A, it is clear from the Transfer Theorem that the image under ϕ is $p_A(A) = \operatorname{Adj} A$.

The last two theorems were first proved for invertible matrices over the field **R** of real numbers and then "transferred" to matrices over arbitrary commutative rings. To carry out the transfers we needed to know that the "transcendental matrices" X and Y are invertible. To prove our last theorem, we will need a more subtle property of the matrices X and Y, namely, that their product XY has distinct eigenvalues in the field **C** of complex numbers.

THEOREM 6. *Let R be a commutative ring, let A be an $m \times n$ matrix over R, and let B be an $n \times m$ matrix over R, where $m \leqslant n$. Then*

$$f_{BA}(t) = t^{n-m} f_{AB}(t).$$

Proof. Case 1. Let A and B be $m \times n$ and $n \times m$ matrices, respectively, over the field **R** of real numbers such that AB has distinct nonzero eigenvalues $\lambda_1, \ldots, \lambda_m$ in the field **C** of complex

numbers. If λ is a nonzero eigenvalue of AB with eigenvector $\mathbf{v}$, then $AB\mathbf{v} = \lambda\mathbf{v}$ implies $BA(B\mathbf{v}) = \lambda(B\mathbf{v})$ and λ is an eigenvalue of BA, also. Thus, $\lambda_1, \ldots, \lambda_m$ are distinct nonzero eigenvalues of BA. Hence the $n \times n$ matrix BA has rank m and nullity $n - m$, so 0 is an $(n-m)$-fold eigenvalue of BA. Therefore,

$$f_{BA}(t) = t^{n-m}(t - \lambda_1) \cdots (t - \lambda_m) = t^{n-m} f_{AB}(t).$$

Case 2. Let A and B be $m \times n$ and $n \times m$ matrices, respectively, over the commutative ring R, and let $X = (x_{ij})$ and $Y = (y_{ij})$ be $m \times n$ and $n \times m$ matrices, respectively, with $2mn$ algebraically independent entries x_{ij}, y_{ij} in $\mathbf{R} \setminus \mathbf{Q}$. The $m \times m$ matrix XY is invertible over the reals, since its determinant is nonzero. This can be seen by considering $\det(XY)$ as a polynomial in x_{ij}, y_{ij}. The substitutions $x_{ij} = \delta_{ij}$ and $y_{ij} = \delta_{ij}$ give $\det(XY) = \det I = 1$, showing that $\det(XY)$ cannot be 0.

Moreover, XY has m distinct eigenvalues in $\mathbf{C}$. This can be seen as follows: The discriminant $D(f_{XY})$ of f_{XY} is a polynomial in the coefficients of f_{XY}, which are in turn polynomials in the entries x_{ij} and y_{ij} of X and Y. (See [**2**] or [**6**] for a description of the discriminant of a polynomial and its properties.) The matrix XY has a repeated eigenvalue if and only if the discriminant $D(f_{XY})$ is zero. However, considering $D(f_{XY})$ as a polynomial in x_{ij}, y_{ij}, the substitutions $x_{ij} = \delta_{ij}$ and $y_{ij} = i \cdot \delta_{ij}$ give $XY = \operatorname{diag}(1, 2, \ldots, m)$ with m distinct eigenvalues, showing that the polynomial $D(f_{XY})$ cannot be identically 0.

Defining the transfer homomorphism ϕ more or less as in the proof of Theorem 1, we use Case 1 to obtain

$$f_{YX}(t) = t^{n-m} f_{XY}(t),$$

and hence, upon taking images under ϕ,

$$f_{BA}(t) = t^{n-m} f_{AB}(t).$$

As an algebraist, I was unhappy when years ago I first encountered the "continuity" proof that $\det(AB) = (\det A)(\det B)$ given in [**5**], especially because the extension to singular A or B was so easy to carry out algebraically. However, it did motivate me to look for an algebraic equivalent of the continuity argument. My solution to this problem was improved by my exposure to the notion of a "generic element" in [**4**] while taking a graduate seminar in Lie algebras. Prior to obtaining these proofs, I had not seen Theorems 4–6 in the literature, but it was later pointed out to me that Theorem 6 for matrices over a field can be found in [**3**]. All of these results are so easy and natural that they probably appear somewhere in the literature. However, my interest was more in the technique than in the specific results. Perhaps some teachers and students of linear algebra may find some pleasure and utility in these ideas, just as I have.

References

[1] George Grätzer, Universal Algebra, Van Nostrand, Princeton, 1968, p. 162.

[2] Emil Grosswald, Topics from the Theory of Numbers, Macmillan, New York, 1966, p. 59.

[3] Nathan Jacobson, Lectures in Abstract Algebra, vol. II, Van Nostrand, New York, 1953, p. 106, Th. 16.

[4] _____, Lie Algebras, Interscience, New York, 1962, pp. 60–61.

[5] Murray H. Protter and Charles B. Morrey, Jr., Modern Mathematical Analysis, Addison-Wesley, Palo Alto, 1964, pp. 327–328.

[6] B. L. van der Waerden, Algebra, vol. 1, Ungar, New York, 1970, pp. 101–107.

Tiling Deficient Boards with Trominoes

I-Ping Chu
Richard Johnsonbaugh
DePaul University
Chicago, IL 60604

Suppose that we remove one square from an $n \times n$ board. A 7×7 board with a missing square is shown in FIGURE 1. Can we tile the remaining squares with right trominoes? (A **right tromino**,

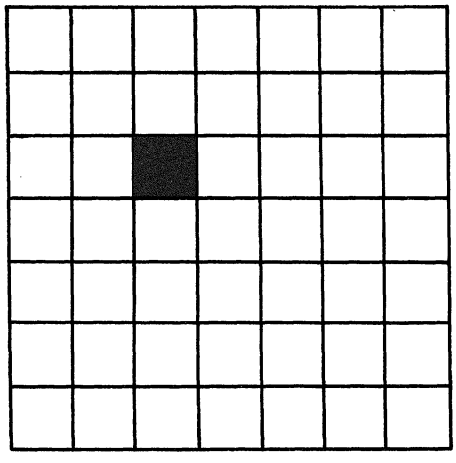


FIGURE 1

hereafter called simply a tromino, is an object made up of three squares as shown in FIGURE 2.) In this paper, by a **tiling** of a figure, we mean an exact covering of the figure by trominoes without

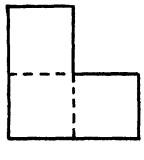


FIGURE 2

having any of the trominoes overlap each other or extend outside the figure. A tiling of the 7×7 board of FIGURE 1 is shown in FIGURE 3.

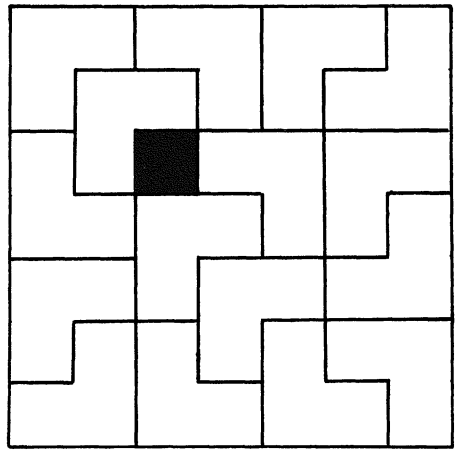


FIGURE 3

A tromino is a type of polyomino. Since polyominoes were introduced by Solomon W. Golomb [2] in 1954, they have been a favorite topic in recreational mathematics. A **polyomino of order k** consists of k squares joined at the edges. A tromino is a polyomino of order 3. Three squares in a row form the only other type of polyomino of order 3. (No one has yet found a simple formula for the number of polyominoes of order k.) Numerous combinatorial problems using polyominoes have been devised (see [3]).

We will call a board with one square missing a **deficient board**. In order for a deficient $n \times n$ board to be tiled by trominoes, 3 must divide $n^2 - 1$ or, equivalently, 3 must not divide n. It is a surprising fact that, except for the case $n = 5$, the condition $3 \nmid n$ is necessary and sufficient for a deficient board to have a tiling. Our proof gives an algorithm for constructing the tilings.

Before continuing to the next section, we invite the reader to find a tiling of a 7×7 board with a different square removed than in FIGURE 1, and also to find a deficient 5×5 board which cannot be tiled. (Some deficient 5×5 boards have tilings while others do not.)

Special cases

Golomb [2] gave a proof by induction that every deficient $n \times n$ board, where n is a power of two, can be tiled. We reproduce this proof since we will need the specific cases $n = 2$, 4, and 8. (This proof also appears in Golomb [3] and Liu [4].) Later (Theorem 2) we will give another proof of this result for $n > 8$.

PROPOSITION 1. *Every deficient $2^k \times 2^k$ board, $k \geqslant 1$, can be tiled.*

Proof. The proof is by induction on k. The case $k = 1$ is obvious.

Suppose we can tile a deficient $2^k \times 2^k$ board. Consider a deficient $2^{k+1} \times 2^{k+1}$ board. Divide the board into four $2^k \times 2^k$ boards as shown in FIGURE 4. Rotate the board so that the missing square is in the upper left quadrant. By the inductive assumption, the upper left $2^k \times 2^k$ board can

be tiled. Place one tromino T in the center, as shown in FIGURE 4, so that each square of T is in each of the other quadrants. These quadrants can now be considered deficient $2^k \times 2^k$ boards. Again, by the inductive assumption, these boards can be tiled. We now have a tiling of the $2^{k+1} \times 2^{k+1}$ board.

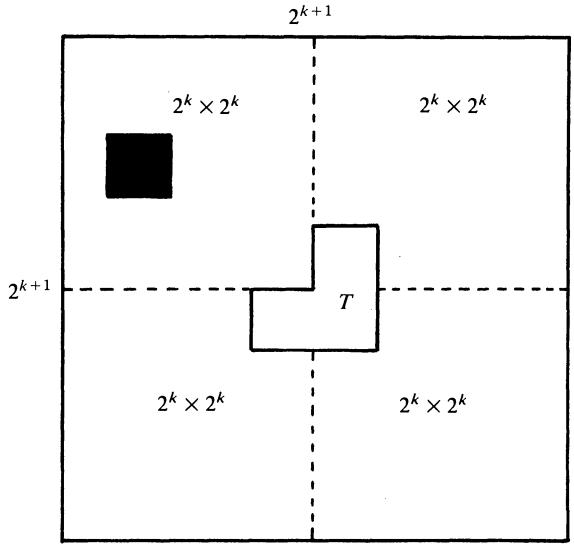


FIGURE 4

Our next proposition deals with the 5×5 board.

PROPOSITION 2. *A 5×5 board with one corner square removed can be tiled.*

Proof. If we eliminate the top two rows and the two columns at the extreme left of FIGURE 3, we obtain a tiling of the 5×5 board with one corner square removed.

An interesting fact, which we leave to the reader, is that if a square next to a corner square is removed from a 5×5 board, the resulting board cannot be tiled.

A trivial but useful fact is our next proposition.

PROPOSITION 3. *A $(2i) \times (3j)$ board, $i, j \geq 1$, can be tiled.*

Proof. A $(2i) \times (3j)$ board can be tiled with the 2×3 configurations shown in FIGURE 5.

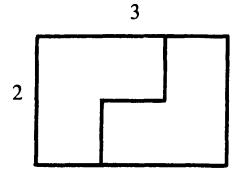


FIGURE 5

We also need to construct tilings for deficient 7×7 boards.

PROPOSITION 4. *Every deficient 7×7 board can be tiled.*

Proof. Let us denote the square in row i, column j by (i, j). Then, by symmetry, we need only consider 7×7 boards with squares (i, j) removed where $i \leqslant j \leqslant 4$. The solution when square $(1,1)$ is removed is shown in FIGURE 6. Not all trominoes of the tiling are shown. The 3×2 and 2×3

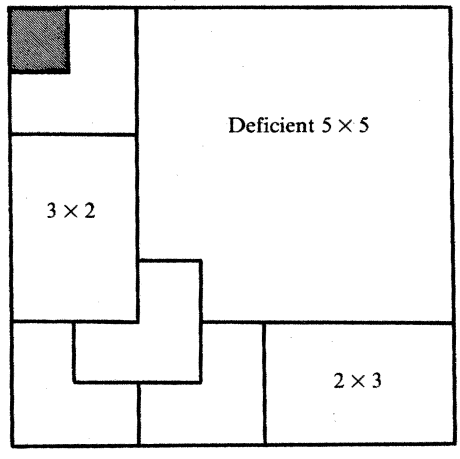


FIGURE 6

subboards have tilings by Proposition 3. The 5×5 subboard with the corner square removed has a tiling by Proposition 2. Essentially the same figure gives tilings in case square $(1,2)$ or $(2,2)$ is deleted.

FIGURE 7 gives a tiling in case square $(1,3)$ is deleted. Essentially the same figure gives tilings in case square $(1,4)$, $(2,3)$, $(2,4)$, or $(4,4)$ is deleted.

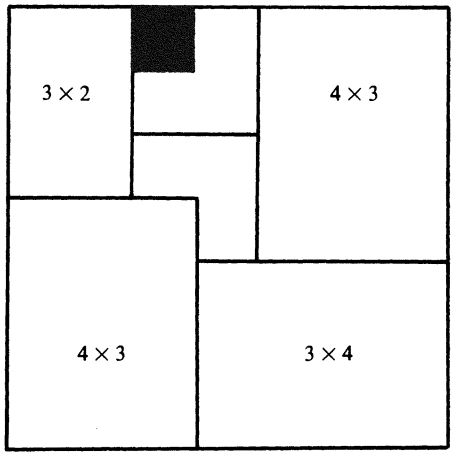


FIGURE 7

FIGURE 3 gives a tiling in case square (3,3) is deleted.
We leave the remaining case, where square (3,4) is deleted, to the reader.

General results

We are now ready to establish a general result for deficient square boards with odd side.

THEOREM 1. *We can tile any deficient $n \times n$ board if n is odd, $n > 5$, and $3 \nmid n$.*

Proof. The case $n = 7$ is given by Proposition 4.
The solution for $n = 11$ is shown in FIGURE 8. We first rotate the board so that the missing

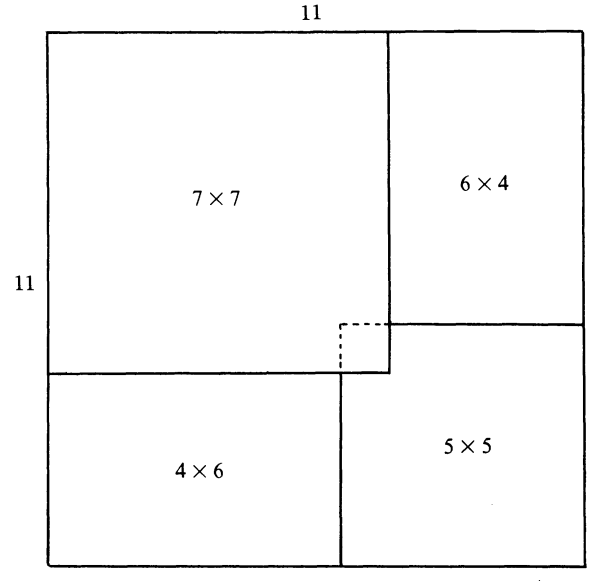


FIGURE 8

square is located in the 7×7 subboard. By Proposition 4, this deficient 7×7 subboard can be tiled. The 6×4 and 4×6 subboards can be tiled by Proposition 3. The 5×5 subboard with a corner square missing can be tiled by Proposition 2.

We can now proceed by induction. Suppose that n is odd, $n > 11$, $3 \nmid n$, and that deficient $k \times k$ boards where k is odd, $n > k > 5$, and $3 \nmid k$ can be tiled. FIGURE 9 shows a tiling of the deficient $n \times n$ board. We first rotate the board so that the missing square is located in the $(n-6) \times (n-6)$ subboard. Now $n-6$ is odd, $n-6 > 5$, and $3 \nmid n-6$; so, by the inductive assumption, this deficient $(n-6) \times (n-6)$ subboard can be tiled. Since n is odd, $n-7$ is even; thus, by Proposition 3, the $6 \times (n-7)$ and $(n-7) \times 6$ subboards can be tiled. By Proposition 4, the deficient 7×7 subboard can be tiled. We have tiled the $n \times n$ board.

Our final result deals with deficient square boards with even side. The proof is similar to the proof for deficient square boards with odd side.

THEOREM 2. *We can tile any deficient $n \times n$ board if n is even, $n > 1$, and $3 \nmid n$.*

Proof. The cases $n = 2$, 4, and 8 are given by Proposition 1.
FIGURE 10 shows a tiling of the deficient $n \times n$ board where n is even, $n > 8$, and $3 \nmid n$. We first rotate the board so that the missing square is located in the $(n-3) \times (n-3)$ subboard. Since $n-3$ is odd, $n-3 > 5$, and $3 \nmid n-3$, we may use Theorem 1 to conclude that the deficient $(n-3) \times (n-3)$ subboard can be tiled. By Proposition 3, the $3 \times (n-4)$ and $(n-4) \times 3$ subboards can be tiled. By Proposition 1, the deficient 4×4 subboard can be tiled. We have tiled the $n \times n$ board.

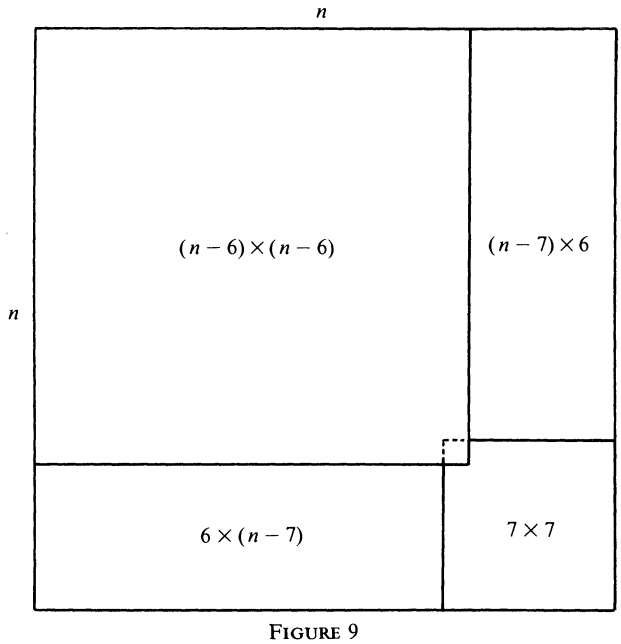


FIGURE 9

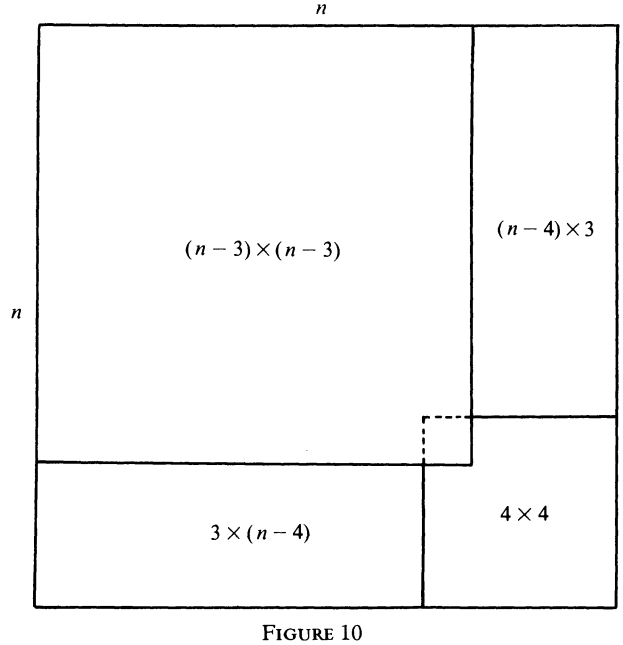


FIGURE 10

We conclude:

THEOREM 3. *If $n \neq 5$, then a deficient $n \times n$ board can be tiled with trominoes if and only if $3 \nmid n$.*

Related problems

Having classified the deficient square boards that can be tiled with trominoes, a number of other questions can be raised. For example:

1. Which deficient rectangular boards can be tiled with trominoes?

2. Which rectangular boards with k squares removed can be tiled with trominoes?

Proposition 3 showed that a $(2i) \times (3j)$ (nondeficient) board can be tiled with trominoes. We can ask:

3. Which (nondeficient) rectangular boards can be tiled with trominoes?

A new set of problems results if we ask the preceding questions about some other kind of polyomino. In this connection, de Bruijn [1] proved that if an $n \times m$ board is tiled by $a \times b$ rectangular polyominoes, then either a divides n or a divides m. Actually, de Bruijn's result was valid in an arbitrary number of dimensions; we have stated only the two-dimensional case. Of course, all of the above questions can be posed in an arbitrary number of dimensions.

Finally, once we have a tiling of a board, we can ask:

4. How many tilings of a particular type are there?

References

[1] N. G. de Bruijn, Filling boxes with bricks, Amer. Math. Monthly, 76 (1969) 37–40.

[2] S. W. Golomb, Checker boards and polyominoes, Amer. Math. Monthly, 61 (1954) 675–682.

[3] _____, Polyominoes, Scribner's, New York, 1965.

[4] C. L. Liu, Elements of Discrete Mathematics, McGraw-Hill, 2nd ed., New York, 1985.

Three Aspects of Fubini's Theorem

J. Chris Fisher
University of Regina
Regina, Canada S4S 0A2

J. Shilleto
6 Locksley Avenue, #5B
San Francisco, CA 94122

Which of the three propositions in the box—(1), (2) or (3)—would you consider to be the most palpably true? Our first choice is (1), while (3) is second, and (2) a close third. This is because

> Let $f(x, y)$, $\frac{\partial}{\partial x} g(x, y)$, and $\frac{\partial^2}{\partial y \partial x} h(x, y)$ be continuous real-valued functions in the rectangle $\{(x, y)\colon a \leqslant x \leqslant b, c \leqslant y \leqslant d\}$. Then:
>
> (1) $\displaystyle\int_a^x \int_c^y f(u, v)\, dv\, du = \int_c^y \int_a^x f(u, v)\, du\, dv,$
>
> (2) $\displaystyle\frac{\partial}{\partial x} \int_c^y g(x, v)\, dv = \int_c^y \frac{\partial}{\partial x} g(x, v)\, dv,$
>
> (3) $\displaystyle\frac{\partial^2}{\partial y \partial x} h(x, y) = \frac{\partial^2}{\partial x \partial y} h(x, y).$

the geometrical evidence for (1) provides a more compelling argument than the naturalness and sense of order of (2) and (3). In fact, (3)'s interpretation using velocities actually *detracts* from its believability (as we shall see)!

These statements are surprising in light of the fact that *using only the fundamental theorem of*

2. Which rectangular boards with k squares removed can be tiled with trominoes?

Proposition 3 showed that a $(2i) \times (3j)$ (nondeficient) board can be tiled with trominoes. We can ask:

3. Which (nondeficient) rectangular boards can be tiled with trominoes?

A new set of problems results if we ask the preceding questions about some other kind of polyomino. In this connection, de Bruijn [1] proved that if an $n \times m$ board is tiled by $a \times b$ rectangular polyominoes, then either a divides n or a divides m. Actually, de Bruijn's result was valid in an arbitrary number of dimensions; we have stated only the two-dimensional case. Of course, all of the above questions can be posed in an arbitrary number of dimensions.

Finally, once we have a tiling of a board, we can ask:

4. How many tilings of a particular type are there?

References

[1] N. G. de Bruijn, Filling boxes with bricks, Amer. Math. Monthly, 76 (1969) 37–40.

[2] S. W. Golomb, Checker boards and polyominoes, Amer. Math. Monthly, 61 (1954) 675–682.

[3] _____, Polyominoes, Scribner's, New York, 1965.

[4] C. L. Liu, Elements of Discrete Mathematics, McGraw-Hill, 2nd ed., New York, 1985.

Three Aspects of Fubini's Theorem

J. Chris Fisher
University of Regina
Regina, Canada S4S 0A2

J. Shilleto
6 Locksley Avenue, #5B
San Francisco, CA 94122

Which of the three propositions in the box—(1), (2) or (3)—would you consider to be the most palpably true? Our first choice is (1), while (3) is second, and (2) a close third. This is because

> Let $f(x, y)$, $\frac{\partial}{\partial x} g(x, y)$, and $\frac{\partial^2}{\partial y \partial x} h(x, y)$ be continuous real-valued functions in the rectangle $\{(x, y)\colon a \leqslant x \leqslant b, c \leqslant y \leqslant d\}$. Then:
>
> (1) $\displaystyle\int_a^x \int_c^y f(u, v)\, dv\, du = \int_c^y \int_a^x f(u, v)\, du\, dv,$
>
> (2) $\displaystyle\frac{\partial}{\partial x} \int_c^y g(x, v)\, dv = \int_c^y \frac{\partial}{\partial x} g(x, v)\, dv,$
>
> (3) $\displaystyle\frac{\partial^2}{\partial y \partial x} h(x, y) = \frac{\partial^2}{\partial x \partial y} h(x, y).$

the geometrical evidence for (1) provides a more compelling argument than the naturalness and sense of order of (2) and (3). In fact, (3)'s interpretation using velocities actually *detracts* from its believability (as we shall see)!

These statements are surprising in light of the fact that *using only the fundamental theorem of*

calculus and some routine manipulations, any one of these propositions can be derived from any other.

Many of our observations can be found in [**2**], and some of the ideas are suggested by exercises in [**1**, p. 793], [**3**, p. 61], and [**4**, pp. 464–465]. Nevertheless, they are missing from contemporary calculus texts and deserve occasional airings. In addition to bringing [**2**] back to light, our goal here is to emphasize the intuitive content of this circle of ideas.

Statement (1), a special case of Fubini's theorem, can be interpreted as follows:

> One gets just as much tomato to eat if he slices it from left to right or from back to front.

Compare this with the mental gymnastics required to untangle the interpretation of (3):

> A person walks on a hillside and points a flashlight along a tangent to the hill; then the rate at which the beam's direction changes when walking south and pointing east equals its rate of change when walking east and pointing south.

We leave the interpretation of (2) to the reader. (Hint: The left side of (2) is the rate of change of the cross-sectional area of the tomato slices mentioned above. Does your interpretation of (2) convince you of its validity?)

Proofs that the statement (i) implies (i $\pm$ 1) are readily found in textbooks (or see [**2**]). As a typical example, here is the standard proof that (1) implies (2). We assume (1) and define $f(x, y) = \frac{\partial}{\partial x} g(x, y)$. That is,

$$\int_a^x f(u, y)\, du = g(x, y) - g(a, y).$$

Then

$$\begin{aligned}\frac{\partial}{\partial x}\int_c^y g(x, v)\, dv &= \frac{\partial}{\partial x}\int_c^y \left(\int_a^x f(u, v)\, du + g(a, v)\right) dv \\ &= \frac{\partial}{\partial x}\int_c^y \int_a^x f(u, v)\, du\, dv + \frac{\partial}{\partial x}\int_c^y g(a, v)\, dv.\end{aligned}$$

Since $\int_c^y g(a, v)\, dv$ is a function of y only, its partial derivative with respect to x is zero, and (having assumed (1))

$$\begin{aligned}\frac{\partial}{\partial x}\int_c^y g(x, v)\, dv &= \frac{\partial}{\partial x}\int_a^x \int_c^y f(u, v)\, dv\, du \\ &= \int_c^y f(x, v)\, dv \\ &= \int_c^y \frac{\partial}{\partial x} g(x, v)\, dv.\end{aligned}$$

The proof's only nontrivial steps use the fundamental theorem of calculus. Indeed, one rather undesirable feature of this proof is that the details make it seem as if something more is involved. Let us therefore change our notation to one of operators to bring out the essence of the above argument. Define

$$D_x f := \frac{\partial f}{\partial x}, \qquad D_x^{-1} f := \int_a^x f(u, y)\, du,$$

$$D_y f := \frac{\partial f}{\partial y}, \quad \text{and} \quad D_y^{-1} f := \int_c^y f(x, v)\, dv.$$

In this notation, statements (1), (2), (3) become

(1) $$D_x^{-1}D_y^{-1} = D_y^{-1}D_x^{-1},$$

(2) $$D_x D_y^{-1} = D_y^{-1}D_x,$$

and

(3) $$D_x D_y = D_y D_x.$$

The fundamental theorem of calculus for $f = f(z)$ is *essentially* $D_z D_z^{-1} f = D_z^{-1} D_z f = f$, where "essentially" means that $D_z^{-1} D_z f$ should have a constant of integration. Of course, in the present context that constant eventually disappears (much as it did in the detailed proof), a fact that can conveniently be left as an exercise. With this warning, the proof that (1) implies (2) now reads

$$D_x D_y^{-1} \underset{\text{F.T.}}{=} D_x D_y^{-1}\left(D_x^{-1}D_x\right) \underset{(1)}{=} D_x\left(D_x^{-1}D_y^{-1}\right)D_x \underset{\text{F.T.}}{=} D_y^{-1}D_x$$

Here is (2) implies (3):

$$D_y D_x \underset{\text{F.T.}}{=} D_y D_x D_y^{-1} D_y \underset{(2)}{=} D_y D_y^{-1} D_x D_y \underset{\text{F.T.}}{=} D_x D_y.$$

The proofs that (3) implies (2) and (2) implies (1) can be obtained by interchanging D with D^{-1} in the lines above.

We should emphasize that because $D^{-1}Df$ differs from f by a constant, the above argument does not constitute a rigorous proof that (i) implies (i − 1). It is, however, an amusing exercise to decode such a symbolic argument to check that each constant of integration really does disappear. Here, for example, is a proof that (3) implies (2) (by decoding $D_y^{-1}D_x = D_y^{-1}D_x D_y D_y^{-1} = D_y^{-1}D_y D_x D_y^{-1} = D_x D_y^{-1}$):

$$\begin{aligned}
\int_c^y \frac{\partial}{\partial x} g(x,v)\,dv &\underset{\text{F.T.}}{=} \int_c^y \frac{\partial}{\partial x}\left(\frac{\partial}{\partial v}\int_c^v g(x,t)\,dt\right)dv \\
&\underset{(3)}{=} \int_c^y \frac{\partial}{\partial v}\frac{\partial}{\partial x}\int_c^v g(x,t)\,dt\,dv \\
&\underset{\text{F.T.}}{=} \frac{\partial}{\partial x}\int_c^y g(x,t)\,dt - \frac{\partial}{\partial x}\int_c^c g(x,t)\,dt \\
&= \frac{\partial}{\partial x}\int_c^y g(x,v)\,dv.
\end{aligned}$$

The ideas touched upon in this note seem to be appropriate for any calculus course, rigorous or not. At one level they provide an attractive way of proving (3): merely explain how it follows quickly from (1). At any level they provide the opportunity to stress normally unseen connections while providing one more chance to show (and show off) the power of the fundamental theorem of calculus.

Note finally that one can easily avoid the intermediate proposition (2), since (3) follows *directly* from (1):

$$D_y D_x = D_y D_x\left(D_y^{-1}D_x^{-1}\right)D_x D_y = D_y D_x\left(D_x^{-1}D_y^{-1}\right)D_x D_y = D_x D_y.$$

We would like to thank Jerry Marsden and John Wilker for their helpful comments and references.

References

[1] Jerrold Marsden and Alan Weinstein, Calculus, Benjamin/Cummings, 1980.
[2] R. T. Seeley, Fubini implies Leibniz implies $F_{yx} = F_{xy}$, Amer. Math. Monthly, 68 (1961) 56–57.
[3] Michael Spivak, Calculus on Manifolds, W. A. Benjamin, New York, 1965.
[4] R. E. Williamson, R. H. Crowell and H. Trotter, Calculus of Vector Functions, 3rd ed., Prentice-Hall, 1972.

PROBLEMS

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

LEROY F. MEYERS, Past Editor
The Ohio State University

Proposals

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

1231. *Proposed by Martin Feuerman, New Jersey Medical College, Newark.*

Let A be a $t \times t$ real symmetric matrix of rank $t-1$ such that $A\mathbf{1} = 0$, where $\mathbf{1}$ is the $t \times 1$ vector with each element equal to 1, and let

$$A^* = \begin{bmatrix} A & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix}.$$

(The prime denotes transpose.) Prove that A^* is nonsingular.

1232. *Proposed by J. T. Groenman, Arnhem, and D. J. Smeenk, Zaltbommel, The Netherlands.*

Let l be the Euler line of the nonisosceles triangle ABC (with sides a, b, c and angles α, β, γ), and let d be the internal angle bisector of γ. (The Euler line of a triangle contains the centroid, circumcenter, and orthocenter.) Prove that:

(a) l is perpendicular to d if and only if $\gamma = \pi/3$; and

(b) l is parallel to d if and only if $\gamma = 2\pi/3$.

1233. *Proposed by Robert E. Shafer, Berkeley, California.*

Prove that if $x > -1$ and $x \neq 0$, then

$$\frac{x^2}{1 + x + \frac{x^2}{12} - \frac{\frac{x^4}{240}}{1 + x + \frac{31}{252}x^2}} < \log^2(1+x) < \frac{x^2}{1 + x + \frac{x^2}{12} - \frac{\frac{x^4}{240}}{1 + x + \frac{x^2}{20}}}.$$

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

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

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

1234. *Proposed by the Computer Science Problem Seminar, Stanford University.*

A positive integer is said to be "sorted" if the digits in its decimal notation are nondecreasing from left to right.

(a) Let x be any integer whose decimal notation consists of an arbitrary number of 3's followed by an arbitrary number of 6's followed by a single 7. Prove that x^2 is sorted. For example, $33366667^2 = 1113334466688889$.

(b)* Which positive integers x are such that both x and x^2 are sorted?

1235. *Proposed by Ira Rosenholtz, The University of Wyoming.*

The book *Calculus in Vector Spaces* by Lawrence J. Corwin and Robert H. Szczarba contains the following in its discussion of local extrema for functions of several variables.

"Suppose f has local maxima at v_1 and v_2. Then f must have another critical point, v_3, because it is impossible to have two mountains without some sort of valley in between. The other critical point can be a saddle point (a pass between the mountains) or a local minimum (a true valley)."

(a) Show that the impossible is possible.

(b)* Is the impossible possible for polynomials?

[For related material see three articles in the May, 1985, issue of this MAGAZINE, pp. 146–150, as well as the article by Calvert and Vamanamurthy in *J. Austral. Math. Soc.*, ser. A, v. 29 (1980) 362–368.]

1236. *Proposed by Mihály Bencze, Săcele, Romania.*

Let the functions f and g be defined by

$$f(x) = \frac{\pi^2 x}{2\pi^2 + 8x^2} \quad \text{and} \quad g(x) = \frac{8x}{4\pi + \pi x^2} \quad \text{for all real } x.$$

(a) Prove that if A, B, and C are the angles of an acute-angled triangle, and R is its circumradius, then

$$f(A) + f(B) + f(C) < \frac{a+b+c}{4R} < g(A) + g(B) + g(C). \tag{1}$$

(b)* Determine functions f and g, where $f(x)$ and $g(x)$ have the form $x/(u + vx^2)$, with u and v real constants, for which the inequalities in (1) are best possible.

Quickies

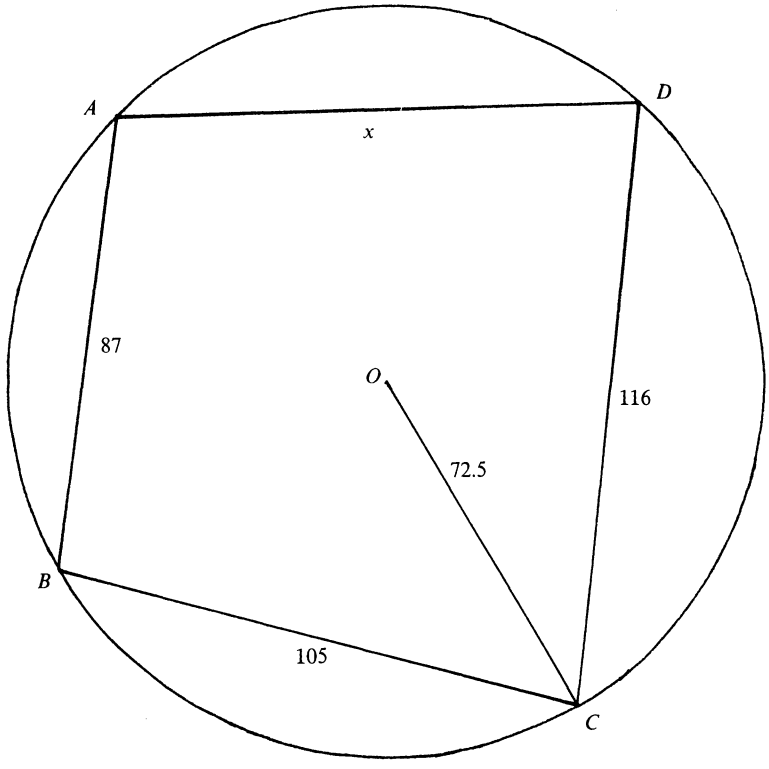
Answers to the Quickies are on pages 53–54.

Q704. *Submitted by M. S. Klamkin, University of Alberta.*

Determine the maximum value of

$$\cos^2 \angle POA + \cos^2 \angle POB + \cos^2 \angle POC + \cos^2 \angle POD,$$

where $ABCD$ is a face of a cube inscribed in a sphere with center O, and P is any point on the sphere.



Q705. *Submitted by John P. Hoyt, Lancaster, Pennsylvania.*

In the accompanying figure, $AB = 87$, $BC = 105$, $CD = 116$, and radius $OC = 72.5$. Find AD.

Q706. *Submitted by Bill Olk, student, Carroll College.*

Suppose that the function f is continuous on the interval $[a, b]$, is differentiable on (a, b), and vanishes at a and b. Show that for every real number r, there is a point c in (a, b) such that $f'(c) = r(f(c))^2$.

Q707. *Submitted by Zhang Zai-ming, Yuxi Teachers' College, Yuxi, Yunan, China.*

Let the perpendicular bisectors of the sides BC, CA, and AB of triangle ABC intersect the circumcircle of ABC in the points A', B', and C', respectively, so that A' is on the arc BC not containing A, and similarly for B' and C'. Continue the process by constructing triangle $A''B''C''$ from $A'B'C'$ in the same way, and so on. Show that the angles of triangle $A^{(n)}B^{(n)}C^{(n)}$ approach $\pi/3$ as $n \to \infty$.

Solutions

Sum of Inradii of a Dissected Triangle **January 1985**

1206. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

Let ABC be a triangle with sides a, b, and c and semiperimeter s. Let the side BC be subdivided using the points $B = P_0, P_1, \ldots, P_{n-1}, P_n = C$ in order. If r_i is the inradius of triangle

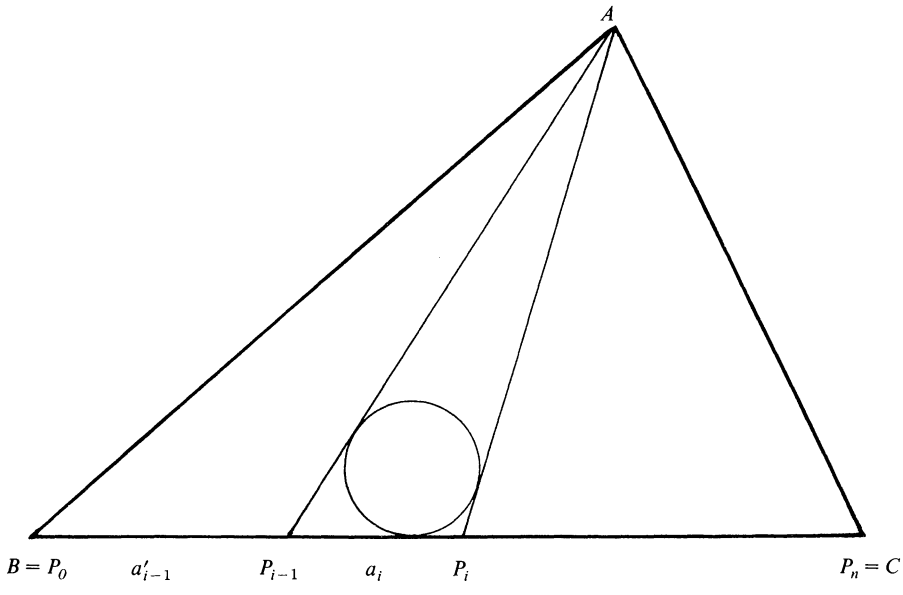


FIGURE 1

$AP_{i-1}P_i$ for $i = 1, \ldots, n$, prove that

$$r_1 + \cdots + r_n < \frac{1}{2} h_a \ln \frac{s}{s-a},$$

where h_a is the length of the altitude from vertex A.

Solution by Vania D. Mascioni, student, ETH Zürich, Switzerland.

For $i = 1, 2, \ldots, n$ let a_i be the base $P_{i-1}P_i$ and s_i the semiperimeter of triangle $AP_{i-1}P_i$, and let a_i' and s_i' be the corresponding quantities for triangle ABP_i. We show below that

$$\frac{s_{i-1}' - a_{i-1}'}{s_{i-1}'} \cdot \frac{s_i - a_i}{s_i} = \frac{s_i' - a_i'}{s_i'} \quad \text{for } 2 \leqq i \leqq n. \tag{1}$$

An easy induction yields

$$\frac{s-a}{s} = \prod_{i=1}^{n} \frac{s_i - a_i}{s_i}.$$

From the arithmetic-geometric mean inequality and the fact that $r_i s_i = \frac{1}{2} a_i h_a$ we obtain

$$\left(\frac{s-a}{s}\right)^{1/n} \leqq \frac{1}{n} \sum_{i=1}^{n} \frac{s_i - a_i}{s_i} = \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{a_i}{s_i}\right) = 1 - \frac{2}{nh_a} \sum_{i=1}^{n} r_i,$$

so that

$$\sum_{i=1}^{n} r_i \leqq \frac{nh_a}{2}\left(1-\left(\frac{s-a}{s}\right)^{1/n}\right),$$

which is stronger than the proposed inequality, which follows if we use $1-1/x<\ln x$ for $x>1$ with $x:=(s/(s-a))^{1/n}$.

Proof of (1). To simplify notation, the sides of triangles ABP_{i-1} and $AP_{i-1}P_i$ are relabeled as shown in FIGURE 2. Then (1) becomes

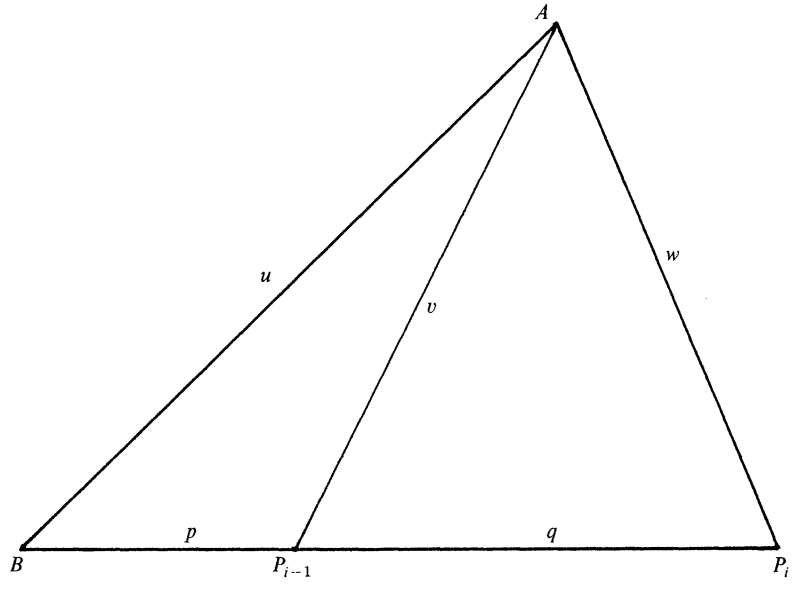


FIGURE 2. **Stewart's theorem.**

$$\frac{u+v-p}{u+v+p}\cdot\frac{v+w-q}{v+w+q}=\frac{u+w-p-q}{u+w+p+q},$$

and an easy (though boring) algebraic manipulation shows this is equivalent to

$$\left(v^2+p^2-u^2\right)q+\left(v^2+q^2-w^2\right)p=0.$$

Now by the law of cosines, this is equivalent to

$$2pqv\left(\cos\angle AP_{i-1}B+\cos\angle AP_{i-1}P_i\right)=0,$$

which is obvious, since $\angle AP_{i-1}B+\angle AP_{i-1}P_i=\pi$. Cf. also Stewart's theorem, in Coxeter and Greitzer, *Geometry Revisited*, p. 6.

Also solved by Jordi Dou (Spain), Václav Konečný & Ronald Shepler, L. Kuipers (Switzerland), Syrous Marivani, William A. Newcomb, Bjorn Poonen (student), J. M. Stark, Paul J. Zwier, and the proposer.

Most solvers used an estimate like

$$\sum_{i=1}^{n} r_i<\sum_{j=1}^{m} r'_j=\sum_{j=1}^{m}\frac{h_a\left(x'_j-x'_{j-1}\right)}{x'_j-x'_{j-1}+\sqrt{\left(x'_{j-1}\right)^2+\left(h_a\right)^2}+\sqrt{\left(x'_j\right)^2+\left(h_a\right)^2}}\approx\int_z^{z+a}\frac{h_a\,dx}{2\sqrt{x^2+h_a^2}},$$

where $A=(0,h_a)$, $B=(z,0)$, $C=(z+a,0)$, $P'_j=(x'_j,0)$, $[P'_0,\ldots,P'_m]$ is a strict refinement of the partition $[P_0,\ldots,P_n]$ of BC (i.e., each P_i is a P'_j, and $m>n$), and r'_j is the inradius of triangle $AP'_{j-1}P'_j$.

1207. *Proposed by Barry Powell, Kirkland, Washington.*

Prove that for each positive integer K there exist infinitely many even positive integers which can be written in more than K ways as the sum of two odd primes.

I. *Solution by Michael V. Finn, Annandale, Virginia.*

Let P be the set of odd primes, and let a_i be the number of ways in which $2i$ can be written as the sum of two elements of P. Suppose that the sequence $(a_i)_{i=1}^{\infty}$ has an upper bound M. Then for every $x \in (0,1)$ we have

$$\left(\sum_{p\in P} x^p\right)^2 = \sum_{i=2}^{\infty} a_i x^{2i} \leqq M \sum_{i=2}^{\infty} x^{2i} = \frac{Mx^4}{1-x^2}.$$

Hence

$$\sum_{p\in P} x^{p-1} = \frac{1}{x}\sum_{p\in P} x^p \leqq \sqrt{M}\,\frac{x}{\sqrt{1-x^2}}.$$

Then, since a power series can be integrated term by term within its interval of convergence, we have

$$\sum_{p\in P}\frac{1}{p} = \sum_{p\in P}\int_0^1 x^{p-1}\,dx = \int_0^1 \sum_{p\in P} x^{p-1}\,dx \leqq \sqrt{M}\int_0^1 \frac{x}{\sqrt{1-x^2}}\,dx = \sqrt{M}.$$

But it is known that $\Sigma_{p\in P}(1/p)$ is unbounded, so we arrive at a contradiction. Hence for every M, some a_i exceeds M.

II. *Solution by John A. Frohliger, St. Norbert College.*

Since every odd integer can be written as the sum of two primes in at most two ways, the problem is equivalent to the following:

For every positive integer K there exist infinitely many positive integers which can be written in more than K ways as the sum of two primes, where the sums $a+b$ and $b+a$ are considered distinct if $a \neq b$.

Proof. Suppose that only finitely many integers can be written as the sum of two primes in more than K ways, and that N is the largest of these integers. Since N can be written as the sum of two positive integers in $N-1$ ways, no integer can be written as the sum of two primes in N or more ways. Let n be a positive integer and $\pi(n)$ the number of primes not exceeding n. Then $(\pi(n))^2$ is the number of sums of two primes, neither exceeding n. Since no such sum exceeds $2n$ and no integer not exceeding $2n$ can be written as such a sum in N or more ways, we see that

$$(\pi(n))^2 < 2nN.$$

Hence

$$\left(\pi(n)\frac{\log n}{n}\right)^2 < 2N\frac{\log^2 n}{n}.$$

Now let $n \to \infty$. By the prime number theorem, the left side approaches 1, while the right side clearly approaches 0. Hence

$$1 \leqq 0,$$

which provides the desired contradiction.

Also solved by Andreas Müller (student, Switzerland), William A. Newcomb, Bjorn Poonen (student), Daniel A. Rawsthorne, William Staton, and the proposer.

Most solvers used the prime number theorem, although the weaker estimate $\pi(x) > ax/\log x$ with, say, $a = .1$,

due to Chebyshev and used by the proposer, is sufficient. *Rawsthorne* proved a generalization: let $R(n)$ be the number of representations of n as the sum of two primes. If $0 < \varepsilon < 1/4$, then there are infinitely many even integers n with $R(n) > (1/4 - \varepsilon)n/\log^2(n/2)$.

A Two-Term Product Inequality

January 1985

1208. *Proposed by Mihály Bencze, Săcele, Romania.*

Prove that if a and b are positive, then

$$\prod_{k=1}^{n} (a^k + b^k)^2 \geqq (a^{n+1} + b^{n+1})^n.$$

Composite of nearly identical, independent solutions by: Víctor Hernández, Universidad Autónoma de Madrid, Spain; Padmini T. Joshi, Ball State University; Michael M. Parmenter, Memorial University of Newfoundland, Canada; Richard E. Pfiefer, San Jose State University; Bjorn Poonen, student, Winchester, Massachusetts; Jan Söderkvist, student, Stockholm, Sweden; and Carl Wagner, University of Tennessee.

$$\begin{aligned}\prod_{k=1}^{n} (a^k + b^k)^2 &= \prod_{k=1}^{n} (a^k + b^k) \prod_{k=1}^{n} (a^{n+1-k} + b^{n+1-k}) \\ &= \prod_{k=1}^{n} (a^{n+1} + a^k b^{n+1-k} + a^{n+1-k} b^k + b^{n+1}) \\ &> \prod_{k=1}^{n} (a^{n+1} + b^{n+1}) = (a^{n+1} + b^{n+1})^n.\end{aligned}$$

Note that inequality is strict.

Also solved by Beno Arbel (Israel), David Boduch (student), Pedro Celis (Canada), Crist Dixon, Sheldon Degenhardt (student), Michael V. Finn, David C. Flaspohler, Riad Ghibril (student, Lebanon), Chico Problem Group, Gymnasium Bern-Kirchenfeld Problem Solving Group (12 students, Switzerland), Hans Kappus (Switzerland), M. S. Klamkin (Canada), Václav Konečný & Ronald Shepler, L. Kuipers (Switzerland), Eugene Levine, J. C. Linders (The Netherlands), Peter W. Lindstrom, Beatriz Margolis (France), Syrous Marivani, Vania Mascioni (student, Switzerland), Mike Molloy (student, Canada), Andreas Müller (student, Switzerland), Roger B. Nelsen, William A. Newcomb, Richard Orr, David Paget (Australia), Richard Parris, Kostas A. Petrakos, Daniel A. Rawsthorne, Joseph Sardinha, Jr., Volkhard Schindler (East Germany), Shannon Schumann (student), Michiel Smid (student, The Netherlands), J. M. Stark, B. Viswanathan (Canada), Michael Vowe (Switzerland), J. G. Wendel, Wong Ngai Ying (Hong Kong), Yan-Loi Wong (student), and the proposer. There was one incorrect solution. Late solution by Erhard Braune (Austria).

A Definite Integral

January 1985

1209. *Proposed by Themistocles M. Rassias, Athens, Greece.*

Evaluate

$$\int_0^\infty \frac{\sqrt{x} \log x}{(1+x)^2}\, dx.$$

I. *Solution by Víctor Hernández, Universidad Autónoma de Madrid, Spain.*

Use integration by parts, with $u = \sqrt{x} \log x$ and $dv = (1+x)^{-2}\, dx$, so that

$$\int_0^\infty \frac{\sqrt{x} \log x}{(1+x)^2}\, dx = \frac{1}{2} \int_0^\infty \frac{\log x}{\sqrt{x}(1+x)}\, dx + \int_0^\infty \frac{dx}{\sqrt{x}(1+x)}.$$

Now, letting $y = 1/x$, we have

$$\int_0^1 \frac{\log x}{\sqrt{x}(1+x)}\,dx = -\int_1^\infty \frac{\log y}{\sqrt{y}(1+y)}\,dy$$

and both improper integrals exist. Hence

$$\int_0^\infty \frac{\log x}{\sqrt{x}(1+x)}\,dx = 0,$$

from which it follows that

$$\int_0^\infty \frac{\sqrt{x}\log x}{(1+x)^2}\,dx = \int_0^\infty \frac{dx}{\sqrt{x}(1+x)} = [2\operatorname{Arctan}\sqrt{x}\,]_{x=0}^\infty = \pi.$$

II. *Solution by William A. Newcomb, Lawrence Livermore National Laboratory.*

A generalization is proved. Draw a cut in the complex plane from 0 to ∞ along the positive real axis, and define the range of θ to be from 0 to 2π in the formulas $z = re^{i\theta}$ (with $r > 0$), $\log z = \log r + i\theta$, and $\sqrt{z} = \sqrt{r}\,e^{i\theta/2}$. Let F be any rational function having no poles on the positive real axis and satisfying the further conditions

$$F(z) \quad \text{is real for positive real } z,$$
$$F(z) = O(r^{-2}) \text{ as } r \to \infty, \quad \text{and}$$
$$F(z) = O(r^{-1}) \text{ as } r \to 0.$$

Let $G(z) = F(z)\sqrt{z}\log z$. We apply the residue theorem to $\int_C G(z)\,dz$ around the closed contour C consisting of: the segment C_1 from ε to R (where $0 < \varepsilon < R$) along the upper edge of the cut; the circle C_2 of radius R centered at the origin and traversed in the positive or counterclockwise sense; the segment C_3 from R to ε along the lower edge of the cut; and the circle C_4 of radius ε centered at the origin and traversed in the negative sense. Let the poles z_k of G have the respective residues ρ_k. Now

$$\int_{C_1} G(z)\,dz + \int_{C_3} G(z)\,dz = \int_\varepsilon^R F(x)\sqrt{x}(\log x)\,dx + \int_R^\varepsilon F(x)(-\sqrt{x})(\log x + 2\pi i)\,dx$$
$$= 2\int_\varepsilon^R F(x)\sqrt{x}(\log x)\,dx + 2\pi i\int_\varepsilon^R F(x)\sqrt{x}\,dx,$$

$$\int_{C_2} G(z)\,dz = O\big(2\pi R\cdot R^{-2}\cdot\sqrt{R}(\log R + 2\pi)\big) = O\big(R^{-1/2}\log R\big) \text{ as } R \to \infty,$$

and

$$\int_{C_4} G(z)\,dz = O\big(2\pi\varepsilon\cdot\varepsilon^{-1}\sqrt{\varepsilon}(|\log\varepsilon| + 2\pi)\big) = O\big(\varepsilon^{1/2}|\log\varepsilon|\big) \text{ as } \varepsilon \to 0.$$

Hence by passage to the limit as $R \to \infty$ and $\varepsilon \to 0$ and use of the residue theorem we obtain

$$2\int_0^\infty F(x)\sqrt{x}(\log x)\,dx + 2\pi i\int_0^\infty F(x)\sqrt{x}\,dx = 2\pi i\sum_k \rho_k.$$

Hence

$$\int_0^\infty F(x)\sqrt{x}\,dx = \sum_k \operatorname{Re}\rho_k$$

and

$$\int_0^\infty F(x)\sqrt{x}(\log x)\,dx = -\pi\sum_k \operatorname{Im}\rho_k.$$

In particular, for $F(z) = (1+z)^{-2}$ we have

$$\operatorname{Res}(G,-1)=\left[\frac{d}{dz}(\sqrt{z}\log z)\right]_{z=-1}=\frac{\pi}{2}-i,$$

and so

$$\int_0^\infty \frac{\sqrt{x}\log x}{(1+x)^2}=\pi \quad \text{and} \quad \int_0^\infty \frac{\sqrt{x}}{(1+x)^2}\,dx=\frac{\pi}{2}.$$

Also solved by Nicolas Artemiadis (Greece), David Boduch (student), W. M. Causey, L. Matthew Christophe, Jr. (two solutions), John M. Coker, Roger Cuculière (France), Sheldon Degenhardt (student), Peter F. Ehlers (Canada), Irwin K. Feinstein, Edward Gade, 3rd, Ralph Garfield, Raymond Greenwell, Chico Problem Group, Hans Kappus (Switzerland), Panos Karambelas (student), M. S. Klamkin (Canada), L. Kuipers (Switzerland), Kee-wai Lau (Hong Kong), Randall Leigh, Robert Leslie, Peter Lindstrom, Beatriz Margolis (France), Syrous Marivani, Fran Masat, Vania Mascioni (student, Switzerland), Roger B. Nelsen (three solutions), Richard Parris, Kostas A. Petrakos, Bjorn Poonen (student), Wulf D. Rehder (three solutions), Volkhard Schindler (East Germany), Robert E. Shafer, Michiel Smid (student, The Netherlands), M. R. Spiegel (two solutions), J. M. Stark, John S. Sumner, Michael Vowe (Switzerland), Edward T. H. Wang (Canada, two solutions), Harry Weingarten, M. G. Wurtele, Paul J. Zwier, and the proposer.

The solutions submitted were of five main types: elementary evaluation using various substitutions; use of the gamma or beta function; use of contour integrals; use of infinite series; and table look-up, principally in Gradshteyn & Ryzhik, formula 4.252.4. The problem occurs, or can be reduced to one occurring, in several well-known textbooks. For example, *Nelsen* found it in Churchill et al., *Complex Variables and Applications*, fourth edition, p. 183, problem 9 (set $x=t^{-2}$). Evaluations of several related or more general integrals were submitted, among them the following:

$$\int_0^\infty \frac{\sqrt{x}(\log x)^2}{(1+x)^2}\,dx=\frac{\pi^3}{2} \qquad \text{(R. E. Shafer);}$$

$$\int_0^\infty \frac{x^p\log x}{(1+x)^n}\,dx=\pi\binom{n-2-p}{n-1}\left(\pi\cot(p\pi)+\sum_{j=0}^{n-2}\frac{1}{j-p}\right)\csc(p\pi)$$

if $0<p<1$, $n\geqq 2$, and n is an integer (M. R. Spiegel).

Rational Polynomials and Roots of Unity — January 1985

1210. *Proposed by J. Rosenblatt, The Ohio State University.*

For a fixed integer $n\geqq 3$, consider the polynomials $f(x)$ with rational coefficients and degree less than n such that $|f(\omega)|=1$ whenever ω is an nth root of unity. Must there be infinitely many such polynomials $f(x)$?

Solution by Daniel B. Shapiro, The Ohio State University.

The answer is YES. Let $\mathbf{Q}$ be the field of rational numbers and ζ any primitive nth root of unity, such as $\exp(2\pi i/n)$. The result is a consequence of the following Claim.

CLAIM. Let n be 4 or an odd prime. Then there are infinitely many $f(x)\in\mathbf{Q}[x]$ with $\deg f<n$ such that $|f(\zeta)|=1$ and $f(1)=1$ (and $f(-1)=1$ if $n=4$).

The Claim will settle the question in these special cases. For suppose that ω is an nth root of unity and that f is one of the functions whose existence is guaranteed by the Claim. If $\omega=1$ (or $\omega=-1$ when $n=4$), then clearly $|f(\omega)|=1$. If $\omega\neq\pm 1$, then Galois theory implies that $\omega=\zeta^\sigma$ for some automorphism σ of $\mathbf{Q}(\zeta)$. Since complex conjugation commutes with every automorphism of $\mathbf{Q}(\zeta)$, it follows that $|\alpha^\sigma|=|\alpha|$ for every $\alpha\in\mathbf{Q}(\zeta)$. Therefore, $|f(\omega)|=|f(\zeta^\sigma)|=|(f(\zeta))^\sigma|=|f(\zeta)|=1$.

Now any given $n\geqq 3$ has a divisor d which is either 4 or an odd prime. From the preceding argument we know that there are infinitely many polynomials $g(x)\in\mathbf{Q}[x]$ with $\deg g<d$ and with $|g(\mu)|=1$ whenever μ is a dth root of unity. For any such $g(x)$ we define $f(x)$ to be $g(x^{n/d})$. Then $\deg f<n$ and $|f(\omega)|=1$ whenever ω is an nth root of unity. Furthermore,

distinct g's provide distinct f's. Hence the problem is solved once we have established the Claim.

To prove the Claim, we first note that there are infinitely many $\alpha \in \mathbf{Q}(\zeta)$ with $\alpha\bar{\alpha} = 1$, e.g., $\alpha = (r+\zeta)/(r+\bar{\zeta}) = (r+\zeta)/(r+1/\zeta)$ with $r \in \mathbf{Q}$. It is easy to see that distinct r's provide distinct α's. We now construct for each α a polynomial $f(x) \in \mathbf{Q}[x]$ satisfying the conclusions of the Claim, with distinct α's providing distinct polynomials. If n is an odd prime, it suffices to take $f(x) = c_0 + c_1 x + \cdots + c_{n-2}x^{n-2} + a(1 + x + \cdots + x^{n-1})$, where $c_0 + c_1\zeta + \cdots + c_{n-2}\zeta^{n-2}$ with $c_j \in \mathbf{Q}$ is the unique representation of α in the $\mathbf{Q}$-basis $\{1, \zeta, \ldots, \zeta^{n-2}\}$ of $\mathbf{Q}(\zeta)$, and $a \in \mathbf{Q}$ is chosen suitably. Since $f(\zeta) = \alpha$, the condition $f(1) = 1$ forces $a = (1 - c_0 - c_1 - \cdots - c_{n-2})/n \in \mathbf{Q}$. If $n = 4$, it suffices to take $f(x) = c_0 + c_1 x + a(1 + x + x^2 + x^3) + b(1 + x^2)$, where $\alpha = c_0 + c_1 i$ with unique $c_j \in \mathbf{Q}$ and suitable a and b in $\mathbf{Q}$. Since $f(i) = \alpha$, the conditions $f(1) = f(-1) = 1$ force $a = -c_1/2 \in \mathbf{Q}$ and $b = (1 - c_0 + c_1)/2 \in \mathbf{Q}$. Thus every one of the infinitely many choices of α determines a suitable polynomial, which proves the Claim.

It may be noted that the result is false if $n \leqq 2$, since there are only two constant polynomials $f(x)$ with $f(1) = \pm 1$, and only four linear polynomials $f(x)$ with $f(\pm 1) = \pm 1$.

There was one incorrect solution.

Comments

966 (proposed January 1976; partial solution May 1977).

No solution was published for part (iii), which was to determine if it is possible to find a square and an interior point such that the distances from the interior point to the vertices and to the sides are all integers. *John P. Robertson* (Berkeley, California) has proved that there is no such square if it is required that two of the distances from the point to the sides be equal.

1094 (proposed March 1980; solution May 1981).

Late solution by Lee A. Hagglund (lost in editor's files).

1154 (proposed November 1982; solution January 1984).

The late *Henry E. Fettis* (Mountain View, California) provided a generalization, replacing the positive integer n by an arbitrary positive real number p. Let

$$F_p(x) = \sum_{k=0}^{\infty} (-1)^k \binom{p-1}{k} \frac{x^{k+1}}{(k+1)^2}.$$

Then $d(xF'(x))/dx = (1-x)^{p-1}$, and integration and substitution yield

$$F_p(1) = \frac{1}{p}\int_0^1 \frac{1-(1-t)^p}{t}\,dt = \frac{1}{p}(\psi(1+p) + C),$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ and C is Euler's constant.

Q677 (November 1982).

Benny N. Cheng (student, University of California, Berkeley) gives a direct proof. Let P be a polynomial of degree $n \geqq 2$ with real coefficients: $P(x) = ax^n + bx^{n-1} + cx^{n-2} + \cdots$. If $(n-1)b^2 < 2nac$, then P has at most $n-2$ real zeros. For suppose that P has n real zeros (it cannot have $n-1$). We may assume without loss of generality that $a = 1$. Let $\alpha_1, \ldots, \alpha_n$ denote the (real) zeros of P. Then $b = -\Sigma_i \alpha_i$ and $c = \Sigma_{i<j}\alpha_i\alpha_j$. Hence

$$(n-1)b^2 \geqq 2nc \Leftrightarrow (n-1)\left(\sum_i \alpha_i\right)^2 \geqq 2n\sum_{i<j}\alpha_i\alpha_j \Leftrightarrow n\left(\left(\sum_i \alpha_i\right)^2 - 2\sum_{i<j}\alpha_i\alpha_j\right) \geqq \left(\sum_i \alpha_i\right)^2 \Leftrightarrow n\sum_i \alpha_i^2 \geqq \left(\sum_i \alpha_i\right)^2,$$

which is nothing but the Cauchy-Schwarz inequality. Hence if P has n real zeros, then $(n-1)b^2 \geqq 2nac$, and the contrapositive follows.

Q677 is a generalization of Q626 (September 1975).

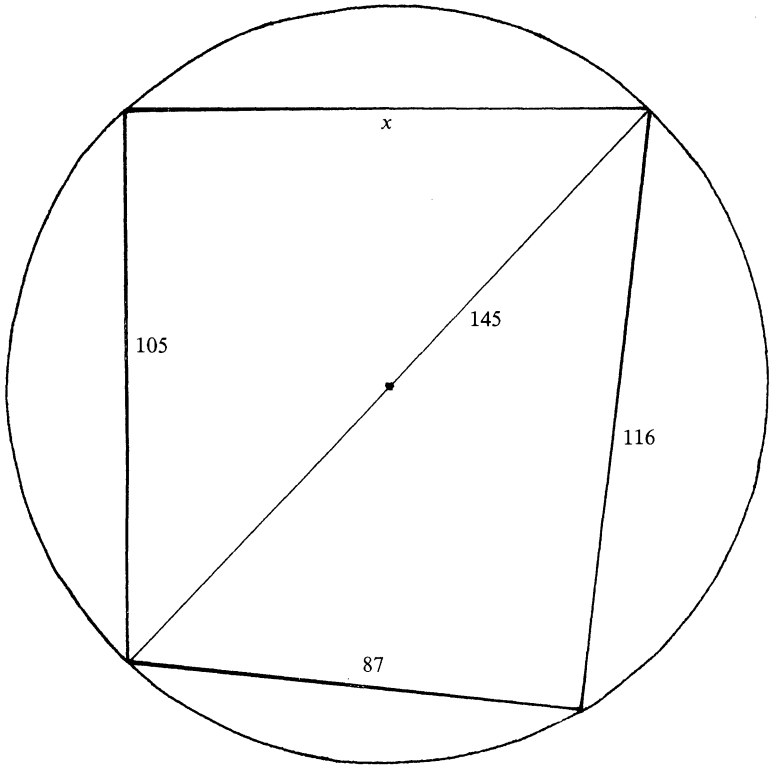
Answers

Solutions to the Quickies on pages 44–45.

A704. We choose a rectangular coordinate system so that the direction cosines of OA, OB, OC, and OD are $(\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}})$. Let the direction cosines of OP be (u, v, w). Then

$$\sum \cos^2 \angle POA = \sum\left(\frac{u}{\sqrt{3}} \pm \frac{v}{\sqrt{3}} \pm \frac{w}{\sqrt{3}}\right)^2 = \frac{4}{3} \quad \text{(constant)}.$$

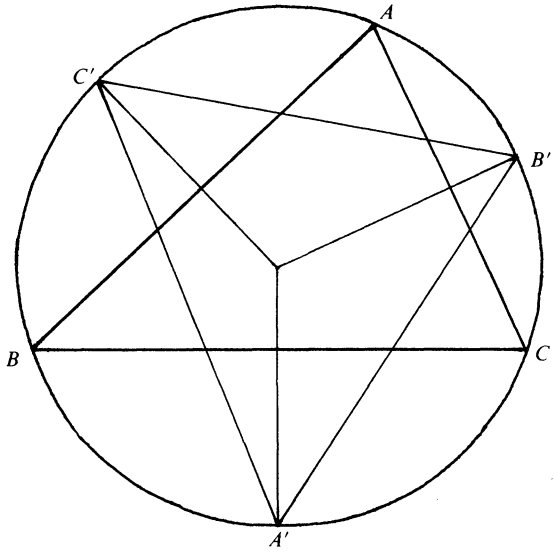
A705. Rearrange the sides of the quadrilateral as shown in the accompanying figure. The triangle whose sides are 87, 116, and 145, i.e., 3×29, 4×29, and 5×29, is a right triangle. Therefore the triangle whose sides are x, 105, and 145, i.e., x, 21×5, and 29×5, is a right triangle, and $x = 20 \times 5 = 100$.



Note. In the given quadrilateral, the diagonals had lengths of 143 and 144, the diagonals were perpendicular, and the area of the quadrilateral was 10296. In the rearranged quadrilateral, the diagonals have lengths of 143 and 145, the diagonals are not perpendicular, and the area is 10296.

A706. Let g be an antiderivative of f and set $h(x) = e^{-rg(x)}f(x)$ for all x in $[a, b]$. Then h satisfies the hypotheses of Rolle's theorem on $[a, b]$. Hence there is a c in (a, b) such that $h'(c) = -rg'(c)e^{-rg(c)}f(c) + e^{-rg(c)}f'(c) = 0$. Dividing out the nonzero exponentials and noting that $g'(c) = f(c)$ yields the desired result.

A707. From the figure we see that $A' = \frac{1}{2}(B + C) = \frac{1}{2}(\pi - A)$. By induction we obtain $A^{(n)} = \pi(\frac{1}{2} - \frac{1}{4} + \cdots - (-\frac{1}{2})^n) + (-\frac{1}{2})^n A$, which approaches $\pi/3$ as $n \to \infty$, and similarly for $B^{(n)}$ and $C^{(n)}$.



Ed. note. Several similar problems have appeared in the literature. In this MAGAZINE, problem 913 (v. 48 (1975) 246–247), A' is the intersection of the circumcircle with the median from A; in the MONTHLY, problem E2906 (v. 90 (1983) 338), A' is the intersection of the circumcircle with the angle bisector from A; in the MONTHLY, problem E1223 (v. 64 (1957) 274–275), and in *Crux Mathematicorum*, problem 554 (v. 7 (1981) 184–185 and v. 10 (1984) 197–198), A' is the point of tangency of the incircle with BC.

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.

Albers, Donald J., *et al.* (eds.), *New Directions in Two-Year College Mathematics: Proceedings of the Sloan Foundation Conference on Two-Year College Mathematics*, Springer-Verlag, 1985; xx + 491 pp, $24.

Report on the *first* national conference on mathematics education in two-year colleges. Most of the excellent essays and discussions are of new curricula and new tools and are of interest also to faculty at four-year colleges and universities.

Weeks, Jeffrey R., *The Shape of Space: How to Visualize Surfaces and Three-Dimensional Manifolds*, Dekker, 1985; x + 324 pp, $49.75.

What is the shape of space? This splendid book, which has no prerequisites except curiosity, treats the connections between topology, geometry, and cosmology. A wealth of illustrations, a minimum of notation, a little fantasy (à la Flatland), and lots of thought-provoking exercises make the book superbly stimulating. Even among professional mathematicians, few will fail to learn something new about geometries on 3-manifolds. Too bad about the price; authors of books like this one would serve potential readers better by striving for more inexpensive publication (e.g., in an MAA series).

Brancazio, Peter J., *Sport Science: Physical Laws and Optimum Performance*, Simon & Schuster, 1985; 400 pp, $9.95 (P).

Why should you string your tennis racket at 50 rather than 70 lbs.? or launch a basketball at the minimum-effort angle? There's less mathematics here than physics--in fact, the mathematics of the optimization is kept hidden, with only the results displayed in tables. Still, it's enjoyable to see how far a few simple mathematical models can go.

Honsberger, Ross, *Mathematical Gems III*, MAA, 1985; 250 pp, $27 ($21 to members).

This additional collection of capsules--whose technical demands seldom go beyond college freshman mathematics--is mostly dedicated to problems from discrete mathematics. The sole remaining problem is to get a book like this into the hands of each freshman interested in mathematics; once read, it will work its charm.

Stewart, Ian, The power of positive thinking, *Nature* 315 (13 June 1985) 539.

Relates recent progress by C. N. Delzell on Hilbert's 17th problem, on the representation of positive functions as sums of squares.

McMahon, Thomas A., and Bonner, John Tyler, *On Size and Life*, Scientific American, 1983; xiii + 255 pp.

An engineer and a biologist team to write for the general reader one of the most enjoyable natural history books of the decade. With hundreds of photographs, figures, and graphs, they illustrate--in both qualitative and quantitative terms--the consequences of different sizes, for the physiology, embryology, support structure, locomotion, and evolution of organisms. The fundamental concepts of similarity and allometric growth are investigated using log-log plots; even high-school students can enjoy this book.

Halmos, Paul R., *I Want to Be a Mathematician: An Automathography*, Springer-Verlag, 1985; xv + 421 pp, $41.50.

What's it like to become and be a research mathematician? Not bad, Paul Halmos might conclude. Would-be mathematicians will get some idea of what it's like; current practitioners are bound to find an anecdote about someone they've heard of or met. This congeries wanders in enjoyable fashion, alternately opinionated, interesting, judgmental, and inspiring. True to the neologism of the title, Halmos sticks to the mathematical side of his life; for example, one learns of his marriages only through accidental references. Non-mathematicians may get the wrong impression, that mathematicians' lives are as narrow as popularly suspected; in any case, Halmos is writing less for them than for his colleagues and successors.

Mackiw, George, *Applications of Abstract Algebra*, Wiley, 1985; v + 184 pp, $11.95 (P).

Provides a supplement on applications for a class studying groups, rings, and fields. Included are exact computing, error-correcting codes, block designs, crystallography, integer programming, cryptography, and combinatorics.

Moore, David S., *Statistics: Concepts and Controversies*, 2nd ed., Freeman, 1985; xvii + 350 pp, $19.95, $12.95 (P).

Second edition of an outstanding statistics book for readers interested in ideas rather than technique. Changes include updating data and topical examples, adding fresh non-numerical exercises, and providing some added material.

Hofstadter, Douglas R., *Metamagical Themas: Questing for the Essence of Mind and Pattern; An Interlocked Collection of Literary, Scientific, and Artistic Studies*, Basic Books, 1985; xxviii + 852 pp, $24.95.

Admirers of Hofstadter's former column in *Scientific American* will be overjoyed at this volume, which contains all of those 25-1/2 columns, plus further comments and eight additional essays. Those who know him only from *Gödel, Escher, Bach: An Eternal Golden Braid*--or worse yet, not at all!--should prepare for an extended treat by the master of self-reference, pattern, and perception. His cleverness wanders "all over the intellectual map--from sexism to music to art to nonsense, from game theory to artificial intelligence to molecular biology to the Cube."

Abraham, Ralph H., and Shaw, Christopher D., *Dynamics--The Geometry of Behavior: Part 3: Global Behavior*, Aerial Pr, 1985; xi + 123 pp, $26 (P).

Continues the authors' Visual Mathematics Library, in which mathematical concepts are presented without algebra or equations. This volume treats generic properties of dynamical systems, structural stability, heteroclinic and homoclinic tangles, and nontrivial recurrence.

Day, Lucille, The higher math, *California Monthly* (June-July 1985) 15-17 + cover.

Story on the mathematics research center in Berkeley, California.

Cleveland, William S., *The Elements of Graphing Data*, Wadsworth, 1985; xii + 323 pp, $27.95, $18.95 (P).

Practical hints on how to graph data for best effect, a how-to manual to accompany E. R. Tufte's *The Visual Display of Quantitative Information* (1983). Cleveland draws most of his examples from illustrations in *Science*, where 30% of the graphs give cause for discussion and improvement.

Brancazio, Peter J., The physics of kicking a football, *The Physics Teacher* 23:7 (October 1985) 403-407.

Constructs a model "of the trajectory of a football kick, using the laws of projectile motion and basic aerodynamics. This model is able to determine within a fairly narrow range the launching angles used for kickoffs and punts." The editors note: "The author gave up his Sunday afternoons and Monday nights for several months in order to obtain the data for this article."

Stewart, Ian, The duellist and the monster, *Nature* 317 (5 September 1985) 12-13.

Emmy Noether first asked which groups can occur as Galois groups of equations. Now it is known that the "monster" group--which rose to fame as the largest of the sporadic simple groups--is a Galois group. The proof by J. G. Thompson makes heavy use of the function theory of fuchsian groups.

Stewart, Ian, The Bieberbach gambit, *Nature* 316 (18 July 1985) 213-214.

An account of de Branges's proof of the Bieberbach conjecture, with more details of the mathematics than one finds in other popular versions. The reasons? The author is a mathematician, and the editors were willing to tolerate a little notation and some terminology.

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

"Our inability to cope with probabilities, says [Amos] Tversky, makes certainty appealing.... The result is that low probabilities seem greater than they are and high probabilities seem less.... Most people overestimated the numbers of deaths from causes that were sensational and underestimated more common causes of death that were less dramatic." Data are given on all kinds of risks.

MacHale, Desmond, *George Boole: His Life and Work*, Boole Press Ltd., 1985; xiii + 304 pp.

First full-length biography of George Boole (1815-1864). More than just a mathematical genius, he was a "child prodigy, self-taught linguist, turbulent academic, social reformer, poet, psychologist, humanitarian and lover of animals--truly a nineteenth-century polymath." Still a mystery, though, is why Boole at Cork and Hamilton at Dublin had almost nothing to do with each other.

Golub, Gene H., *Studies in Numerical Analysis*, MAA, 1984; x + 415 pp, $42 ($31 to members).

The ten contributions range over current areas of research in numerical analysis, including Newton's method, sparse matrices, conjugate gradient methods, and multigrid methods. J. H. Wilkinson's "The perfidious polynomial," in which he demonstrates that backwards error analysis *should* have been discovered in connection with root-finding on polynomials (instead of matrix eigenvalue problems), will become a classic.

Day, Lucille, The world's greatest living geometer, *California Monthly* (June-July 1985) 16-17.

Thumbnail sketch of S. S. Chern.

NEWS & LETTERS

26th INTERNATIONAL MATHEMATICAL OLYMPIAD

The following are excerpts from a report on the 26th International Mathematical Olympiad by M. S. Klamkin. The complete report, with details on the U.S. and Canadian teams, appears in Crux Mathematicorum.

The Twenty-Sixth International Mathematical Olympiad (IMO) was held this year in Finland from June 29 to July 9. Teams from 38 countries took part in the competition. This was again a record number of participating countries, up from last year's record of 34 countries. The team size was 6 students (maximum number) from each country, the same as for the last two years. However, if the number of participating countries continues to increase, the team size will probably be reduced to 4 students (as occurred in Hungary in 1982). Having a smaller team size should make it easier for countries with relatively small populations to field better teams. Additionally, the expenses will be reduced and the logistics made easier. The total number of students was also a record one of 208, up from last year's record of 192. The countries participating for the first time were China, Iran, Iceland, and Turkey.

The 1986, 1987, and 1988 IMO's are to be held in Poland, Cuba, and Australia, respectively. I fully expect to see a new record number of participating countries for the 1988 Australian IMO.

The six problems of the competition were assigned equal weights of 7 points each (the same as the last 4 IMO's) for a maximum possible score of 42. I believe that this year's competition was harder than the previous one, as evidenced by only ten students having scores of at least 35 (last year there were 24 such students), and only two perfect scores, 6 less than last year.

The first prize winners were:

Geza Kos	Hungary	42
Daniel Taturu	Romania	42
Gabor Megyesi	Hungary	38
Nikolai I. Chavdarov	Bulgaria	37
Philippe Alphonse	Belgium	36
Olga Leonteva	Soviet Union	36
Andrew Hassell	Australia	35
Vasil B. Daskalov	Bulgaria	35
Waldemar Horwat	U.S.A.	35
Nguyen T. Dung	Vietnam	35
Hagen V. Eitzen	West Germany	34
Radu Negulescu	Romania	34
Gelca Razvan	Romania	34
Jeremy Kahn	U.S.A.	34

As the IMO Competition is an individual event, the results are announced officially only for individual team members. However, team standings are usually compiled unofficially by adding up the scores of individual team members. Since there were quite a few teams with less than six students, these will be noted in the subsequent table. Congratulations to Romania, the originator of the IMO in 1959, which was first. A list of the top fifteen teams follows:

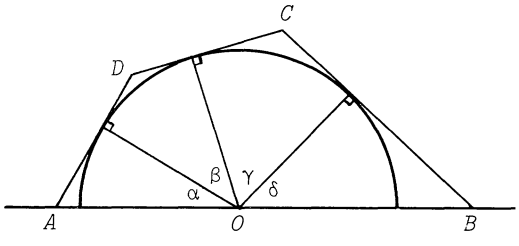
Rank	Country	Score
1	Romania	201
2	U.S.A.	180
3	Hungary	168
4	Bulgaria	165
5	Vietnam	144
6	U.S.S.R.	140
7	West Germany	139
8	East Germany	136
9	France	125
10	Great Britain	121
11	Australia	117
12-13	Canada	105
12-13	Czechoslovakia	105
14	Poland	101
15	Brazil	83

26TH INTERNATIONAL MATH OLYMPIAD SOLUTIONS

The solutions that follow have been especially prepared for publication in this MAGAZINE by Loren C. Larson, St. Olaf College.

1. A circle has center on the side AB of the cyclic quadrilateral $ABCD$. The other three sides are tangent to the circle. Prove that $AD + BC = AB$.

Sol. Label the figure as shown.



We may assume that the radius of the circle at O is unity. Thus, $AD + BC = (\tan\alpha + \tan\frac{\beta}{2}) + (\tan\frac{\gamma}{2} + \tan\delta)$.

We know that angle β equals the angle at B because they are both supplementary to the angle at D, and therefore β and δ are complementary. Similarly, α and γ are complementary. Using this, together with the half-angle formula for tangent (easily obtained from the double-angle formula), the last expression becomes

$$AD + BC = (\cot\gamma + \frac{\sec\beta - 1}{\tan\beta}) + (\frac{\sec\gamma - 1}{\tan\gamma} + \cot\beta) = \csc\gamma + \csc\beta = \sec\alpha + \sec\delta = AB.$$

2. Let n and k be given relatively prime natural numbers, $0 < k < n$. Each number in the set $M = \{1, 2, \ldots, n-1\}$ is colored either blue or white. It is given that

(i) for each $i \in M$, both i and $n-i$ have the same color, and

(ii) for each $i \in M$, $i \neq k$, both i and $|i - k|$ have the same color.

Prove that all numbers in M must have the same color.

Sol. Let $[x]$ denote the unique integer between 1 and k such that $[x] \equiv x \pmod{k}$.

Alternate applications of (i) and (ii) (see proof that follows) lead one to consider the sequence $x_0, x_1, \ldots, x_{k-1}$ defined recursively by $x_0 = k$ and

$$x_{i+1} = \begin{cases} [n-x_i] & \text{if } i \text{ is even,} \\ [-x_i] & \text{if } i \text{ is odd.} \end{cases}$$

We will show that x_i and x_{i+1} have the same color.

Suppose i is even. Repeated use of (ii) implies that x_i, x_i+k, x_i+2k, ..., x_i+qk all have the same color, where q is such that $x_i+qk < n \leq x_i+(q+1)k$. By (i), $n - (x_i+qk)$ also has the same color, and $n - (x_i+qk) = [n-x_i] = x_{i+1}$.

Suppose i is odd. By (ii), $|k-x_i|$ and x_i have the same color, and $|k-x_i| = k-x_i = [-x_i] = x_{i+1}$.

It follows that x_0, x_1, ..., x_{k-1} all have the same color.

An easy induction shows that $x_{2i-1} = [in]$ and $x_{2i} = [(k-i)n]$ for $i = 1,2,\ldots,\lfloor k/2 \rfloor$. Thus, $x_0, x_1, \ldots, x_{k-1}$ is a permutation of $[n]$, $[2n]$, ..., $[kn]$, and because k is relatively prime to n, the latter is a permutation of $1,2,\ldots,k$. The result now follows from repeated use of (ii).

3. For any polynomial $P(x) = a_0 + a_1x + \ldots + a_kx^k$ with integer coefficients, the number of coefficients which are odd is denoted by $w(P)$. For $i = 0,1,2,\ldots$ let $Q_i(x) = (1+x)^i$. Prove that if $i_1, i_2, \ldots, i_n$ are integers such that $0 \leq i_1 < i_2 < \ldots < i_n$, then

$$w(Q_{i_1} + Q_{i_2} + \ldots + Q_{i_n}) \geq w(Q_{i_1}).$$

Sol. We will induct on i_n. The inequality holds when $i_n = 0$ or 1. Suppose the result holds whenever $i_n < 2^s$, and now suppose that $2^s \leq i_n < 2^{s+1}$.

Case 1. Suppose that $2^s \le i_1$. Let $Q = Q_{i_1} + \ldots + Q_{i_n}$. For $k = 1,2,\ldots,n$, let $j_k = i_k - 2^s$ and let $\overline{Q} = Q_{j_1} + \ldots + Q_{j_n}$. Then $Q = Q + x^{2^s}\overline{Q} \equiv (1 + x^{2^s})\overline{Q} \pmod 2$. Since $\deg \overline{Q} < 2^s$ the preceding implies that $w(Q) = 2w(\overline{Q}) \ge 2w(Q_{j_1})$ (by induction) $= w((1 + x^{2^s})Q_{j_1}) = w(Q_{i_1})$.

Case 2. Suppose there is an integer t, $0 < t < n$ such that $i_t < 2^s \le i_{t+1}$. Let $P = Q_{i_1} + \ldots + Q_{i_t}$ and $Q = Q_{i_{t+1}} + \ldots + Q_{i_n}$. Let $j_k = i_k - 2^s$ for $k = t+1,\ldots,n$, and let $\overline{Q} = Q_{j_{t+1}} + \ldots + Q_{j_n}$. A case analysis shows that $w(P+\overline{Q}) + w(\overline{Q}) \ge w(P)$. Also, $P+Q = P+Q + x^{2^s}\overline{Q} \equiv P + (1+x^{2^s})\overline{Q} \pmod 2$ $\equiv P + \overline{Q} + x^{2^s}\overline{Q}$. Since $\deg(P+\overline{Q}) < 2^s$, the preceding shows that $w(P+\overline{Q}) + w(\overline{Q}) \ge w(P) \ge w(Q_{i_1})$ (inductive assumption).

This completes the induction.

4. Given a set M of 1985 distinct positive integers, none of which has a prime divisor greater than 26, prove that M contains at least one subset of four distinct elements whose product is the fourth power of an integer.

Sol. We will make use of the following.

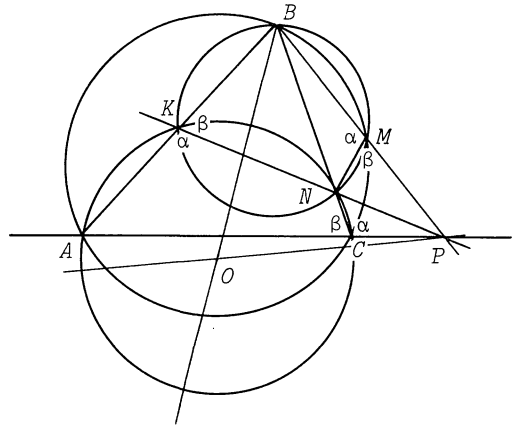
Lemma. Any subset S of M with more than 512 elements contains two elements whose product is a perfect square.

Proof of Lemma. The elements of S have the form $p_1^{n_1}p_2^{n_2}\ldots p_9^{n_9}$, where $p_1 = 2 < p_2 < \ldots < p_9 = 23$ are the nine prime numbers less than 26. The 9-tuple of exponents, $(n_1,n_2,\ldots,n_9)$, has one of 2^9 (=512) possible parity patterns. By the pigeonhole principle, two elements of S will have exponents with the same parity pattern. Their product is a perfect square.

By continued use of the Lemma, we can find distinct elements $a_1,b_1,a_2,b_2,\ldots,a_{513},b_{513}$ in M such that a_ib_i is a perfect square (apply the Lemma to the sets M, $M-\{a_1,b_1\}$, $M-\{a_1,b_1,a_2,b_2\}$, ...). Let $c_i = a_ib_i$. By the proof of the Lemma, there are distinct integers i and j such that c_ic_j is a perfect square, or equivalently, $c_i^2c_j^2 = a_ib_ia_jb_j$ is a perfect fourth power.

5. A circle with center O passes through the vertices A and C of triangle ABC, and intersects the segments AB and BC again at distinct points K and N, respectively. The circumscribed circles of the triangles ABC and KBN intersect at exactly two distinct points B and M. Prove that angle OMB is a right angle.

Sol. The common chords of the three pairs of circles are concurrent at their radical center P. Let α denote angle AKN and let $\beta = 180^\circ - \alpha$. We find (see figure) that $\angle NMP$ and $\angle NCP$ are supplementary, so that $MNCP$ is a cyclic quadrilateral. Therefore, $BM\cdot BP = BN\cdot BC = BO^2 - r^2$, and $PM\cdot PB = PN\cdot PK = PO^2 - r^2$, where r is the radius of the circle through A,C,N,K. Hence $PO^2 - BO^2 = BP(PM-BM) = PM^2 - BM^2$, or



equivalently, $PO^2 - PM^2 = BO^2 - BM^2$. From this it follows that OM is perpendicular to BM.

6. For every real number x_1, construct the sequence $x_1, x_2, \ldots$ by setting

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

for each $n \geq 1$. Prove that there exists exactly one value of x_1 for which $0 < x_n < x_{n+1} < 1$ for every n.

Sol. For each positive integer n, let $f_n(x) = x(x + \frac{1}{n})$ for $x > 0$. Set $a_1 = 0$ and $b_1 = 1$, and for $n \geq 2$, let $a_n = f_1^{-1} f_2^{-1} \cdots f_{n-1}^{-1}(1 - \frac{1}{n})$ and $b_n = f_1^{-1} f_2^{-1} \cdots f_{n-1}^{-1}(1)$. Define $F_n(x) = f_n f_{n-1} \cdots f_1(x)$ for $x > 0$. Then $F_n(a_n) = f_n(1 - \frac{1}{n}) = 1 - \frac{1}{n}$, $F_n(a_{n+1}) = 1 - \frac{1}{n+1}$, $F_n(b_{n+1}) = 1$, and $F_n(b_n) = f_n(1) = 1 + \frac{1}{n}$. Thus, $F_n(a_n) < F_n(a_{n+1}) < F_n(b_{n+1}) < F_n(b_n)$. Since F_n is an increasing function (each f_k is increasing), it must be the case that $a_n < a_{n+1} < b_{n+1} < b_n$.

Let $a = \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. The preceding work shows that $a \leq b$.

Let x_1 be any real number in the interval $(0,1)$. The condition that $x_{n+1} > x_n$ is equivalent to $x_n > 1 - \frac{1}{n}$, and this is equivalent to $x_1 > a_n$. The condition that $x_{n+1} < 1$ is equivalent to $x_1 < b_{n+1}$. Thus, $0 < x_n < x_{n+1} < 1$ holds for all n if and only if $a \leq x_1 \leq b$.

To prove uniqueness, it suffices to prove that $a = b$, and for this, it is sufficient to prove that $b_n - a_n < \frac{1}{n}$. The function $F_{n-1}(x)$ is convex (each f_k is convex), and therefore, because $F_{n-1}(0) = 0$ and $F_{n-1}(b_n) = 1$, it follows that $F_{n-1}(x) \leq \frac{x}{b_n}$ for $0 \leq x \leq b_n$. In particular, $1 - \frac{1}{n} = F_{n-1}(a_n) \leq a_n / b_n$. It follows that $b_n - a_n \leq \frac{1}{n} b_n < \frac{1}{n}$, and this completes the proof.

MAA AWARDS

At the annual Business Meeting of the Mathematical Association of America, held January 10, 1986, in New Orleans, Louisiana, three individuals received special recognition.

Arnold Ross of Ohio State University was awarded the Award for Distinguished Service to Mathematics. Professor Ross was chosen for this award for his "significant impact on mathematics on a national scale through his unique summer program for high school students. He has profoundly influenced many people early in their lives, among them, a great number of original, now eminent, colleagues in mathematics. Indeed, no major mathematics conference is without a few mathematicians who can tell of their experience in Professor Ross' summer programs."

George Miel of the University of Nevada, Las Vegas, was awarded the Chauvenet Prize "for a noteworthy expository or survey paper published in a North American journal in 1981-83." The article for which Professor Miel received the award was "Of calculations past and present: the Archimedean algorithm," which appeared in the *American Mathematical Monthly* 90 (1983), 17-35. The Committee on the Chauvenet Prize consisted of Peter J. Hilton (chair), Theodore W. Gamelin, and Lawrence A. Zalcman.

Edward W. Packel of Lake Forest College was awarded the MAA Book Prize "for a distinguished, innovative book published by the MAA." The book for which Professor Packel won the prize was *The Mathematics of Games and Gambling*, volume 28 in the New Mathematical Library series of the MAA. The Committee on the MAA Book Prize consisted of Doris Schattschneider (chair), J.A. Seebach, and Gary J. Sherman.

EDITOR OF MONTHLY NAMED

The Board of Governors of the MAA, at their meeting in Laramie, Wyoming, August 11, 1985, elected Herbert Wilf, of the University of Pennsylvania, Editor of the *American Mathematical Monthly* for a five year term beginning January, 1987. Professor Wilf will replace Paul Halmos.

GERHARD N. WOLLAN

Gerhard N. Wollan of Purdue University died on July 16, 1985. Professor Wollan was Editor of *MATHEMATICS MAGAZINE* from 1971 until 1976.

ANNOUNCEMENTS

The New York State Mathematics Association of Two-Year Colleges will hold its annual conference at Grossinger's Hotel in Grossinger, New York, April 18-20, 1986. For further information contact: Gerald M. Smith, NYSMATYC President-Elect, Cayuga Community College, Auburn, NY 13021 (Phone: (315) 255-1743).

Peter J. Hilton will be the principal speaker at the annual Pi Mu Epsilon Student Conference at St. John's University, Collegeville, MN 56321, March 14-15, 1986. Additional talks will be given by students who have been working on independent study or research projects. For more information contact Mike Gass at (612)363-3192 or Jerry Lenz at (612) 363-3193.

The Eugene Strens Memorial Conference on Intuitive and Recreational Mathematics and Its History will be held at the University of Calgary, July 27-August 2, 1986. Invited speakers include Elwyn Berlekamp, John Conway, H.S.M. Coxeter, Kee Dewdney, Aviezri Fraenkel, Martin Gardner, Ron Graham, Branko Grünbaum, Hendrik Lenstra, Willy Moser, Angela Newing, Roger Penrose, John Selfridge, Doris Schattschneider, and David Singmaster. For further information contact Richard Guy or Bill Sands, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4

INTERNATIONAL CONGRESS IN BERKELEY

For the first time since 1950 an International Congress of Mathematicians will be held in the United States. The last Congress in America was in 1950 in Cambridge, Massachusetts; the last on this continent in Vancouver in 1974. The highlight of the Congress for many will be the awarding of the Fields Medals. At each Congress since the Oslo Congress of 1936 these prizes have been given to the two (or in some years four) mathematicians under the age of 40 who have made important contributions to mathematics. Congresses are held only every four years. The Fields Medals are viewed as comparable to Nobel Prizes, though the criteria for selection are quite different.

The Congress in Berkeley will take place August 3-11, 1986. There will be 19 areas of mathematics covered.

For more information write ICM-86, Post Office Box 6887, Providence, RI 02940.

USCMI PRE-CONGRESS SERIES OF INVITED SURVEY TALKS

On the afternoon of August 2nd, 1986, The United States Commission on Mathematical Instruction will sponsor a series of invited survey talks aimed at enhancing understanding and appreciation of some of the major research-related work which will be discussed at ICM-86.

The USCMI invites recommendations of potential speakers and their areas of interest. Please send all suggestions to the session organizer: Warren Page, New York City Technical College, 300 Jay Street, Brooklyn, NY 11201.

Further details, including the names of speakers and titles of their survey talks will be announced in a forthcoming issue of *MATHEMATICS MAGAZINE*.

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

MATHEMATICS MAGAZINE VOL 59, NO. 1, February 1986