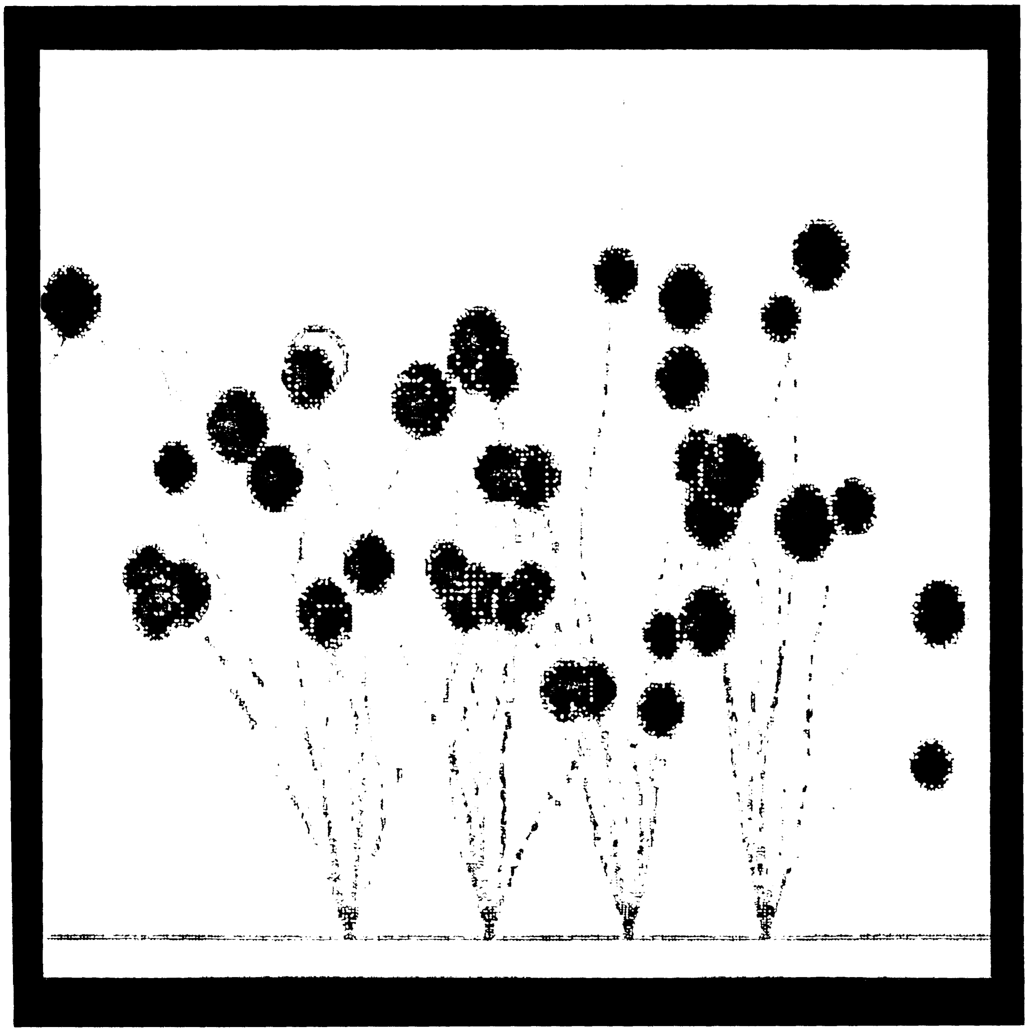
Vol. 67, No. 4 October 1994

MATHEMATICS MAGAZINE



- Quantitative Estimates for Polynomials in One or Several Variables
- Convolutions and Computer Graphics
- Wagering in Final *Jeopardy!*

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

The full statement of editorial policy appears in this *Magazine*, Vol. 64, pp. 71–72, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher.

Send new manuscripts to: Martha Siegel, Editor, Mathematics Magazine, Towson State University, Towson, MD 21204. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and two copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added. Do not use staples.

AUTHORS

Bernard Beauzamy was born in 1949. He received his Ph.D. in mathematics (thèse d'Etat) in 1976, under the supervision of Laurent Schwartz, at the University of Paris 7. He was one of the founders of the "Institut de Calcul Mathématique" in 1987, and has been executive director of ICM since that date.

After a misspent youth as a budding concert pianist, **Per Enflo** lost his way in the labyrinths of Banach space geometry. Disappointing the unabashed but naive optimists, he exhibited a space without the approximation property in 1972 and then, in 1976, he celebrated the bicentennial with an operator without an invariant space. Like the Constitution, his example took 11 years to gain acceptance.

Paul S. Wang, born in China and raised in Taiwan, received his Ph.D. from M.I.T. in 1971. His research in computer algebra and symbolic computation focuses on the mechanization of mathematics with advanced computer techniques. It is in this fertile ground of interactions between mathematics and computer science he met and continues to collaborate with Bernard Beauzamy.

Anne M. Burns received a Ph.D. in mathematics from the State University of New York at Stony Brook in 1976. In an earlier life she studied art at Pratt Institute. A conference on Computer Graphics for the Arts and Sciences at NYU in 1987 opened her eyes to the possibilities of using computer graphics as an educational tool.

Rhonda L. Hatcher received her Ph.D. from Harvard University in 1987. Although her area of specialty is number theory, she became interested in game theory while teaching a freshman seminar on the subject at St. Olaf College. The present article arose from observations she and her husband and co-author George Gilbert made about the sometimes peculiar betting strategies of *Jeopardy!* contestants.

After obtaining a B.A. in mathematics from Washington University, **George T. Gilbert** received a Ph.D. from Harvard University, specializing in number theory. His other mathematical interests include problem-solving and dabbling in mathematical game theory. Watching contestants win with very low scores sparked the author's interest in Final *Jeopardy!*

Vol. 67 No. 4 October 1994

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025–570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

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

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

Postmaster: Send address changes to Mathematics Magazine Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036-1385.
PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Quantitative Estimates for Polynomials in One or Several Variables

From Analysis and Number Theory to Symbolic and Massively Parallel Computation

BERNARD BEAUZAMY
Institut de Calcul Mathématique
75011 Paris, France

PER ENFLO
PAUL WANG
Kent State University
Kent, OH 44242

0. Introduction

Polynomials are basic objects in mathematics. The behavior of any system usually depends on several constraints (the variables) and so is given by one or several functions of one or several variables. These functions, in turn, provided they are sufficiently continuous (which is usually the case), can be approximated by polynomials, in some range and within some accuracy. So the study of any system, no matter how complicated it is, starts with the study of polynomials.

As an example, the position of a robot in a plane may be given by two polynomials (one for each coordinate), each of them depending on several variables: directions of wheels, current in each motor, and so on.

Polynomials appear also as technical tools. For instance, for a system depending linearly on the data (thus given by a matrix), the stability depends on the zeros of the characteristic polynomial of the matrix (cf. Marden [30]), and so locating zeros is a major problem in the theory of automata.

Polynomials play a central role in mathematics, for instance, in analysis (complex or real), number theory, approximation theory, and numerical analysis.

Classical results about polynomials fall into two types. The first one is purely qualitative: It says that something exists. An easy example is the Bezout identity: If P and Q are two polynomials, with no root in common, there exist two other polynomials, A, B such that $AP + BQ = 1$ (no information is provided on A and B; nobody even knows on what this information should depend). The second type depends on the degree. This is the case for Bernstein's inequality, a basic result in complex analysis: If P is a polynomial of degree n and P' its derivative, then

$$\|P'\|_\infty \leq n\|P\|_\infty,$$

where $\|P\|_\infty = \max_{|z|=1}|P(z)|$ (the value of $|P'|$ indicates the size of a slope, so, for

Supported by the National Science Foundation (U.S.A.), the C.N.R.S. (France), the Ministry of Defense (France) and Digital Equipment Corp.

instance, it shows how large the polynomial remains near one of its maxima). This is also the case for Gelfond's theorem in number theory (a result that is useful for the study of transcendental numbers): If P and Q are two polynomials with degree m and n respectively, then

$$|P \cdot Q|_\infty \geq 2^{-(m+n)} |P|_\infty \cdot |Q|_\infty,$$

where, this time, $|P|_\infty = \max|a_j|$, if $P = \sum_j a_j z^j$.

Both types of results are quite unsatisfactory for two reasons. First, sharp estimates are often needed, especially for computational purposes (and precisely when the degree is high), and second, some problems are not connected with the degree. For instance, how many zeros does a polynomial have in the disk, centered at 0, with radius 1/2? One may say: at most its degree, but this answer is trivial and useless. The correct—and useful—answer depends on the relative importance of the low degree terms inside the polynomial, as we shall see later.

We present here new directions for research and recent results, with a general purpose: to get new instruments of measure for polynomials, allowing us to obtain quantitative estimates, where only qualitative ones were known, or more precise ones, when some were already known. This article should be considered as an introduction to a research topic: No proofs are given (only references). But we try instead to present the motivations and the directions for further development. The field is entirely new (it was created in 1985), and we have not tried to present its historical roots. On the contrary, we emphasize the links with other branches of mathematics; some of these branches (such as number theory) are extremely old, but our new concept gives a fresh approach, a fresh point of view, which brings new questions very quickly. On such a new topic, opportunity for research comes almost immediately.

Interactions with computer science are also emphasized; as will be seen in Section 3, the need for fast factorization algorithms introduces very interesting questions about polynomials, on which our work casts a new light. Conversely, the strict requirements of massively parallel programming obliged us to invent a new spatial representation for polynomials in many variables, and this new representation finally turned out to be very fruitful, at the theoretical level.

We hope to convince the reader that research in mathematics is not what people usually think, and does not necessarily require many years of study before the first question can be met. But let us now be more specific.

Technically, two tools will be essential. The first one is a measure of the importance of the terms of low degree inside the whole polynomial; the second is a new norm, with weighted coefficients. Both of them are very simple.

In order to measure the importance of the low degree terms, we have to use a norm on the space of polynomials, and which norm we use will depend on the specific problem we investigate. For $P = \sum a_j z^j$, we have already met

$$\|P\|_\infty = \max_{|z|=1} |P(z)|, \qquad |P|_\infty = \max_j |a_j|.$$

We can also use

$$\|P\|_2 = \left(\int_0^{2\pi} |P(e^{i\theta})|^2 \frac{d\theta}{2\pi} \right)^{1/2}, \qquad |P|_2 = \left(\sum_j |a_j|^2 \right)^{1/2},$$

and

$$\|P\|_1 = \int_0^{2\pi} |P(e^{i\theta})| \frac{d\theta}{2\pi}, \qquad |P|_1 = \sum_j |a_j|.$$

These notations refer to the following convention: For the double-bar norms, $\|\cdot\|$, P is considered as a function on the unit circle; for the single-bar norms, $|\cdot|$, P is identified with the sequence of its coefficients, $(a_0, a_1, \dots, a_n)$.

These norms are comparable:

$$|P|_\infty \leq \|P\|_1 \leq |P|_2 = \|P\|_2 \leq \|P\|_\infty \leq |P|_1 \leq \sqrt{\deg(P)}\,|P|_\infty.$$

Since the $|\cdot|_1$-norm is simplest, we use it first, in order to define our concept.

1. Concentration at Low Degrees for Polynomials in One Variable, and Applications

The notion of *concentration at low degrees* for a polynomial was introduced by Bernard Beauzamy and Per Enflo in 1985. It gives, with just one new concept, quantitative results in several branches of mathematics, and governs seemingly unrelated phenomena, such as the location of the zeros and the size of the polynomial in a given interval. Moreover, the estimates obtained are *independent of the degree*.

A. Definition of the concentration. Let $P(z) = a_0 + a_1 z + \cdots + a_k z^k + \cdots + a_n z^n$ be a polynomial with complex coefficients. Assume $0 < d \leq 1$, and $0 \leq k \leq n$. We say that P has *concentration d at degree k* if

$$\sum_{j=0}^{k} |a_j| \geq d \sum_{j=0}^{n} |a_j|. \tag{1}$$

Of course, if the degree of the polynomial is precisely k, this polynomial has concentration $d = 1$ at this degree. Of course also, if P has concentration d at degree k, it will also have concentration d', for any d', $0 < d' \leq d$. So, in order to be more precise, we consider the quotient $\sum_{j \leq k}|a_j| / \sum_{j=0}^{n}|a_j|$ and we call it the *concentration factor* of the polynomial $P(z)$ at degree k.

For instance, the polynomial $1 + z^{100}$ has degree 100, but concentration $1/2$ at degree 0. For some applications, it may be more interesting, more accurate, and quicker (in terms of computing speed) to consider it as a polynomial with concentration $1/2$ at degree 0 than as a polynomial of degree 100. Also, all the polynomials

$$P_n(z) = 1 + \left(1 - \frac{1}{n}\right)z + \left(1 + \frac{1}{n}\right)z^2 + z^n \qquad (n \geq 3)$$

have one thing in common: They all have concentration $3/4$ at degree 2. They do not have the same number of zeros, and their zeros do not stay at the same place when n increases, but, as we will see, due to the concentration property, the number of their zeros in any given disk remains uniformly bounded, independently of n.

More generally, we wish to replace the actual degree of the polynomial by the concentration factor (at a given degree) in order to obtain estimates *independent* of the actual degree of the polynomial.

The first application of the concept is that of a *generalized Jensen's Inequality*, which governs the size of the subset of the unit circle on which a polynomial with concentration at low degree is large.

B. The generalized Jensen's Inequality. The classical Jensen's Inequality (see for instance Rudin [36]) asserts that, for a polynomial $P = \sum a_j z^j$, with $a_0 \neq 0$,

$$\int_0^{2\pi} \log|P(e^{i\theta})| \frac{d\theta}{2\pi} \geq \log|a_0|. \tag{2}$$

If $a_0 = 0$ but $a_1 \neq 0$, applying the above inequality to $P(z)/z$ gives $\log|a_1|$ on the right-hand side; if both a_0 and a_1 are 0 but a_2 is not, applying it to $P(z)/z^2$ gives $\log|a_2|$, and so on.

This inequality is quite important, for the following reason: If $|P(z)|$ is small at some z, then $\log|P|$ is extremely negative. Thus, to control from below the whole integral $I = \int_0^{2\pi} \log|P(e^{i\theta})|d\theta/2\pi$ (that is, to give a statement of the form $I \geq c$, for some c) allows one to control at once the size of any set of the type $\{\theta\colon |P(e^{i\theta})| < \varepsilon\}$ and make sure that none of these sets is too big.

The proof of Jensen's Inequality is not obvious, and the fact that the integral converges is not immediate either. To see this, one writes P as a product $P = \lambda\prod_{j=1}^{n}(z - z_j)$, where the z_j's are the zeros of P. Then a term $\int_0^{2\pi} \log|e^{i\theta} - z_j|d\theta/2\pi$, with $|z_j| \neq 1$, causes no difficulty, since $|e^{i\theta} - z_j|$ is bounded away from zero, $0 \leq \theta < 2\pi$. So the only difficulty comes from the terms $\int_0^{2\pi} \log|e^{i\theta} - z_j|d\theta/2\pi$, with $|z_j| = 1$. By a change of variables, they reduce to $\int_0^{2\pi} \log|e^{i\theta} - 1|d\theta/2\pi$. This integral is handled by a contour integration in the complex plane, and the value is 0 (see Rudin [36]).

Connected with Jensen's Inequality is Mahler's measure

$$M(P) = \exp\int_0^{2\pi}\log|P(e^{i\theta})|\frac{d\theta}{2\pi},$$

which plays an important role in number theory, since it is multiplicative: $M(PQ) = M(P)\cdot M(Q)$.

Now, inequality (2) has a serious drawback: It depends on just one coefficient, namely a_0, and therefore it is discontinuous. When a_0 is small, you apply it as it is, and when a_0 vanishes, you apply it to a_1. For instance, consider the family of polynomials

$$P_n(z) = \frac{1}{n} + z + z^2.$$

For each n, by (2),

$$\int_0^{2\pi}\log|P_n(e^{i\theta})|\frac{d\theta}{2\pi} \geq \log\frac{1}{n} \to -\infty \text{ when } n \to \infty,$$

whereas $P_n \to P$, with $P = z + z^2$, and

$$\int_0^{2\pi}\log|P(e^{i\theta})|\frac{d\theta}{2\pi} \geq 0.$$

Therefore, there is a need for a better inequality, not using just a_0. Such an inequality exists. If $P(z) \not\equiv 0$ is of exact degree k, a result due to Kurt Mahler [28] asserts that

$$\int_0^{2\pi}\log\left(\frac{|P(e^{i\theta})|}{\Sigma_0^k|a_j|}\right)\frac{d\theta}{2\pi} \geq -k\log 2.$$

But this result is not completely satisfactory either, because it involves the degree, and you get no uniform bound on the family of polynomials

$$P_n(z) = \frac{1}{n} + z + z^2 + \frac{1}{n}z^n.$$

A more satisfactory answer is a lower bound for $\int_0^{2\pi} \log|P(e^{i\theta})|d\theta/2\pi$ depending only on the importance of the low degree terms in the polynomial. This is given by a result of the first two authors [4], which asserts that, if $P(z)$ satisfies (1) (and thus does not need to be of degree k anymore), then

$$\int_0^{2\pi} \log\left(\frac{|P(e^{i\theta})|}{\sum_0^k |a_j|}\right)\frac{d\theta}{2\pi} \geq B(d,k), \tag{3}$$

where the constant $B(d,k)$ is defined as the maximum value of the function

$$f_{d,k}(t) = t\log\frac{2d}{(t-1)\left(\left(\frac{t+1}{t-1}\right)^{k+1}-1\right)} \qquad (t>1). \tag{4}$$

This gives the rough estimate

$$B(d,k) \geq 2\log(2d/3^k).$$

Let $C(d,k)$ be the *best* (that is the largest) constant satisfying (3). The precise value of $C(d,k)$ is unknown. However, we showed in [5] that, for $d = 1/2$,

$$C(1/2,k) \leq -2k\log 2, \tag{5}$$

and that, asymptotically, when $k \to +\infty$,

$$C(1/2,k) \geq -2k. \tag{6}$$

The precise value of $C(d,k)$ has been computed by A. K. Rigler, S. Y. Trimble, and R. S. Varga [35], for the class of *Hurwitz polynomials*. A Hurwitz polynomial has real, positive coefficients, and its roots are either real and negative, or pairwise conjugate with negative, real parts. In other words, a Hurwitz polynomial is a product of terms of the form $z + a$ (with a real positive) and $z^2 + pz + q$, with p, q real positive.

For this class, the best constant, denoted by $C^H(d,k)$, is, for $d = 1/2$, given by:

$$C^H(1/2,k) = -(2k+1)\log 2, \qquad k = 0,1,\ldots, \tag{7}$$

which agrees with the result of (5) and (6). The *extremal polynomial*, in the class of Hurwitz polynomials (that is, the one that gives equality in (3) with the constant $C^H(1/2,k)$), is explicitly given by $(z+1)^{2k+1}$, $k = 0,1,\ldots$. Indeed, since the binomial coefficients satisfy $\binom{2k+1}{j} = \binom{2k+1}{2k+1-j}$, these polynomials are symmetric, and so have concentration $1/2$ at degree k. A complete description of all the constants $C^H(d,k)$ is given in [35], for every $0 < d \leq 1$ and $k = 0,1,\ldots$.

Instead of (1), the concentration can be measured with other norms: l_p or L_p, $1 \leq p \leq \infty$. For instance, one may consider the problem: Find

$$\inf\left\{\int_0^{2\pi}\log\left(\frac{|P(e^{i\theta})|}{\|P\|_\infty}\right)\frac{d\theta}{2\pi} : \sum_{j=0}^{k}|a_j| \geq d\|P\|_\infty\right\}. \tag{8}$$

In the case $k = 1$, this problem was solved in [6]. The solution is the unique constant $c < 0$ that satisfies the equation:

$$e^c(1-2c) = d.$$

The problem is still open for other values of k. Problems related to (8), but involving other norms than $\|P\|_\infty$, were studied by L. Bonvalot [14].

Besides the above results that are deduced from a generalized Jensen's Inequality, another area where striking results already have been obtained is that of *products of polynomials*.

C. Products of polynomials in one variable, with complex coefficients. To perform the product of two polynomials is of course a basic operation, on which one naturally wants quantitative estimates. But, as we will see in Section 3, these results are the key to fast factorization algorithms in computer science.

The first result, due to the first two authors [4], is that if P and Q both have concentration at fixed degrees, then the norm of their product is bounded from below, with a constant depending only on the concentration data, and not on the actual degrees of the polynomials. If, $P, Q, P \cdot Q$ are written:

$$P(z) = \sum_{j=0}^{J} a_j z^j, \quad Q(z) = \sum_{m=0}^{M} b_m z^m, \quad (P \cdot Q)(z) = \sum_{n=0}^{J+M} c_n z^n,$$

and satisfy

$$\sum_{j \le k} |a_j| \ge d \sum_{j \ge 0} |a_j| \qquad \text{and} \qquad \sum_{j \le k'} |b_j| \ge d' \sum_{j \ge 0} |b_j|,$$

then

$$\sum |c_n| \ge \lambda(d, d'; k, k') \sum_{j \ge 0} |a_j| \sum_{m \ge 0} |b_m|, \tag{9}$$

where (as stated) $\lambda(d, d'; k, k')$ depends only on d, d', k, k' and *not* on the precise degrees of $P(z)$ and $Q(z)$.

In number theory, a result of special importance is Gelfond's Theorem (see for instance M. Waldschmidt [44]). Here the constant $\lambda(d, d'; k, k')$, for fixed d, d', is also exponential in k, k', so the order of magnitude is the same as in Gelfond's theorem, but the scope of application is larger, since the degrees do not appear.

There is a much deeper result [4] concerning products of polynomials: If one polynomial has a large coefficient and the other some concentration at a fixed degree, then the product has a large coefficient. More specifically:

If P and Q satisfy

$$\max_j |a_j| \ge d \sum_{j \ge 0} |a_j| \tag{10}$$

and

$$\left(\sum_{m=0}^{k} |b_m|^2 \right)^{1/2} \ge d' \left(\sum_{m \ge 0} |b_m|^2 \right)^{1/2}, \tag{11}$$

then

$$\max |c_n| \ge \lambda(d, d'; k) \left(\sum_{i \ge 0} |a_i|^2 \right)^{1/2} \left(\sum_{m \ge 0} |b_m|^2 \right)^{1/2}, \tag{12}$$

where (as stated) $\lambda(d, d'; k)$ depends only on d, d', k, and *not* on the precise degrees of $P(z)$ and $Q(z)$.

The proof, using a complicated induction procedure, gives (for fixed d, d') a constant depending on k as a triple exponential (that is $e^{-(e^{e^k})}$). Further progress in this direction was later made by P. Enflo [18]: A double exponential suffices (that is

$e^{-(e^k)}$). It is quite likely (but has not been proven yet) that the correct order of magnitude is of exponential type.

D. A challenge. Though products of polynomials look elementary, precise constants are not easy to obtain, even in particular situations. Let's take, as an example, a very special case of the theorem above.

Assume, in this theorem, that Q has concentration $1/2$ at degree 0, that is

$$Q(z) = 1 + b_1 z + \cdots + b_m z^m, \tag{13}$$

with $\sum_1^m |b_j| \leq 1$ (which is stronger than (11)).

Assume that in P the largest coefficient is the last one:

$$P(z) = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n,$$

with $\sum_0^{n-1} |a_j| \leq 1$ (thus P satisfies (10) with $d = 1/2$). Then a precise result on these lines is due to C. Fabre [19]: The product $P \cdot Q$ satisfies $|P \cdot Q|_\infty \geq 1/2$. Under these precise assumptions, the constant $1/2$ is the best possible.

But now, take the same Q as before, but P under the more general form

$$P = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n + a_{n+1} z^{n+1} + \cdots + a_N z^N,$$

with

$$\sum_0^{n-1} |a_j| + \sum_{n+1}^{N} |a_j| \leq 1$$

(so P also satisfies (10) with $d = 1/2$, but the largest coefficient is not necessarily the last). Then, what is the greatest lower bound on $|P \cdot Q|_\infty$? We know it is less than $1/2$, but we do not know its value. This problem is not simple, and we would like to offer a prize for its solution: a week in Paris or in Kent (at the winner's choice!), trip included.

More generally, as we pointed out, in *most* cases previously studied, the precise values of the best constants mentioned are *unknown*. For polynomials of a fixed degree, the corresponding estimates have led to considerable study and deep theorems. Let us mention for instance Kurt Mahler [28], [29], Arestov [1], Bell [12], Newman [33], Beller–Newman [13], Kahane [22], [23]. Study of problems such as (8), for large values of k, is clearly related to their work, and, as the solution of the case $k = 1$ already shows, involves deep considerations in harmonic analysis. The determination of the precise values in the estimates does not only involve computational accuracy but a better understanding of the problems themselves. They have been satisfactorily solved, so far, in only a very *limited* number of cases.

The techniques used in the existing proofs use either complex analysis or combinatorics. Probabilistic techniques might be used: for instance considering the coefficients of the polynomials as independent random variables, and trying to prove that, with some probability, the product is bounded from below. Such a probabilistic approach has been used by J.-P. Kahane in [23], in a closely related context.

E. Where the roots are. Finding the location of the zeros is one of the major problems. As is well known, if the degree is at least 5, no exact algebraic solution can be given, so the procedure has to be numerical. Many algorithms exist, either to find the complex zeros, or the real ones. Also, many results are known about the size of the smallest disk containing all zeros, or containing a given number of zeros (see Marden

[30]). For instance, a result due to Cauchy says that all the zeros of the polynomial

$$P(z) = a_0 + a_1 z + \cdots + a_n z^n$$

are contained in the disk $|z| < r$, where r is the positive root of the equation

$$|a_0| + |a_1|x + \cdots + |a_{n-1}|x^{n-1} = |a_n|x^n.$$

If we come back to polynomials with concentration d at degree k, we observe that not all $a_0, a_1, \ldots, a_k$ can be zero, so 0 cannot be a root of order $k+1$. But in fact, a much more precise statement can be given, and not too many roots can be too close to 0. The following result was obtained by Sylvia Chou [15].

There is a radius $r(d,k) > 0$ such that any polynomial P satisfying (1) has at most k roots in the open disk centered at 0, with radius $r(d,k)$. For $d \leq 1/2$, the value of this radius is

$$r(d,k) = \left(\frac{1}{1-d}\right)^{1/(k+1)} - 1.$$

Precise numerical estimates for this radius, when $d > 1/2$, were obtained by the same author in [16]; it was computed exactly for the class of Hurwitz polynomials.

B. Beauzamy and Sylvia Chou [10] proved that if the polynomial has concentration d at degree k, and if the zeros are written in increasing order of moduli,

$$0 \leq |z_1| \leq |z_2| \leq \cdots, \tag{14}$$

then the quantity $|\sum_{j>k} 1/z_j|$ is bounded from above by a number depending only on d and k, for which they gave numerical estimates.

There is an extension of the classical Bernstein's Inequality (cited in the introduction). This extension is valid for Hurwitz polynomials, and is independent of the degree. Indeed, if P is Hurwitz and has concentration d at degree k, then

$$\|P'\|_\infty \leq C(d,k)\|P\|_\infty.$$

Conversely, if P is a Hurwitz polynomial satisfying

$$\|P'\|_\infty \leq C\|P\|_\infty,$$

it has concentration $d(C)$ at a degree $k(C)$. (S. Chou [15], [16], B. Beauzamy and Sylvia Chou [10]).

F. From polynomials to analytic functions. Instead of considering polynomials and defining the concentration with the sum of moduli of coefficients, one can use the l_2-norm, and extend the definition to the H^2 functions. This is the class of functions of the form

$$f(z) = \sum_0^\infty a_j z^j,$$

which satisfy $(\sum_0^\infty |a_j|^2)^{1/2} < \infty$ (and so are analytic inside the unit disk). For such a function, we say it has concentration d at degree k (measured in the l_2-norm) if:

$$\left(\sum_0^k |a_j|^2\right)^{1/2} \geq d\left(\sum_0^\infty |a_j|^2\right)^{1/2}. \tag{15}$$

This definition looks slightly more complicated than the one we gave in (1), but it has a major advantage. We are now in a Hilbert space.

Previous results on polynomials about the number of zeros in a disk extend to H^2 functions with concentration d at degree k (but the proofs are very different, and use tools from harmonic analysis). The first author proved in [8] that in any disk $D(r)$ with $0 < r < 1$, the number of zeros of f is bounded by a number that depends only on d, k, r (for which precise numerical estimates are given). From this result it follows that if the zeros z_j of f are written in increasing order of modulus, as in (14), the speed of growth to 1 of the sequence $|z_j|$ is bounded from below by a function depending only on d and k. This was improved by Maria Girardi [21] who showed that for an H^2-function with concentration d at degree k, the quantity $\Sigma_{j \geq 1}(1 - |z_j|)$ is bounded from above by a number depending only on d and k. This is one more example of a new quantitative theory, since the classical result asserts only that this sum is finite.

For such a function, for every $\varepsilon > 0$, the set of points where $|f| < \varepsilon$ can be covered by a union of disks, with sum of radii depending only on d, k, and ε, and tending to 0 when $\varepsilon \to 0$ ([8]). The importance of such results in computer science will be described in Section 3.

2. Polynomials in Several Variables

The concept of concentration at low degrees also makes sense for a polynomial in several variables. Let

$$P(z_1, \dots, z_N) = \sum_{\alpha} a_\alpha z_1^{\alpha_1} \dots z_N^{\alpha_N}, \tag{16}$$

(where $\alpha = (\alpha_1, \dots, \alpha_N)$) be a polynomial, with complex coefficients and N variables. Let's define $|\alpha| = |\alpha_1| + \cdots + |\alpha_N|$. We say that P has concentration d at *total* degree k if

$$\sum_{|\alpha| \leq k} |a_\alpha| \geq d \sum_{\alpha} |a_\alpha|. \tag{17}$$

As before, we define the l_1-norm of P by:

$$|P|_1 = \sum_{\alpha} |a_\alpha|.$$

A theorem due to P. Enflo [17] asserts that if P, Q are polynomials with concentration d, d' at degree k, k' respectively, that is satisfy

$$\sum_{|\alpha| \leq k} |a_\alpha| \geq d \sum_{\alpha} |a_\alpha| \tag{18}$$

and

$$\sum_{|\beta| \leq k'} |b_\beta| \geq d' \sum_{\beta} |b_\beta|, \tag{19}$$

then if the product PQ is written as $\Sigma_\gamma c_\gamma z_1^{\gamma_1} \dots z_N^{\gamma_N}$, we have

$$\sum_{|\gamma| \leq k+k'} |c_\gamma| \geq \lambda(d, d'; k, k') \sum_{\alpha} |a_\alpha| \sum_{\beta} |b_\beta|. \tag{20}$$

The interesting point is that λ depends neither on the specific degrees of P and Q, nor on the number of variables. This theorem was the key tool in the second author's construction of an operator without invariant subspaces [17]. But it is only an existence statement; nothing is known about the size of λ.

A paper by B. Beauzamy, E. Bombieri, P. Enflo, and H. Montgomery [7] continues the research on these lines. If we define the l_p norm of P as $(\Sigma_\alpha |a_\alpha|^p)^{1/p}$, and the L_p norm of P as

$$\|P\|_p = \left(\int \cdots \int |P(e^{i\theta_1}, \dots, e^{i\theta_N})|^p \frac{d\theta_1}{2\pi} \cdots \frac{d\theta_N}{2\pi} \right)^{1/p},$$

then similar results hold for all these norms, with constants independent both of the degrees, and of the number of variables, but of unknown size. However, there is a norm for which the product result has a sharp statement.

If P is a homogeneous polynomial of degree m, we define

$$[P]_2 = \left(\sum_{|\alpha|=m} \frac{\alpha!}{m!} |a_\alpha|^2 \right)^2, \tag{21}$$

(where $\alpha! = \alpha_1! \cdots \alpha_N!$). If P and Q are homogeneous of degree m and n respectively, then Bombieri's inequality is

$$[P \cdot Q]_2 \geq \sqrt{\frac{m!n!}{(m+n)!}} [P]_2 [Q]_2, \tag{22}$$

and this constant is the best possible. The extremal pairs, that is the pairs of P's and Q's for which the smallest product is obtained, were characterized by J.-L. Frot, C. Millour [20] and by Bruce Reznick [34]. We will come back to this result when we speak about parallel computation, in Section 4.

Comparison between real and complex sup-norms for polynomials in many variables was made by Richard Aron, Bernard Beauzamy, and Per Enflo in [3], where sharp estimates were obtained. Previous results in related areas had been obtained by R. Aron and J. Globevnik [2], A. Tonge [40], and Y. Sarantopoulos [37], [38], [39]. Further estimates were recently given by Miguel Lacruz [24].

3. Symbolic Computation

This is one of the most striking applications of the concept "quantitative estimates" Two closely related problems are considered: upper bounds for coefficients in polynomial factorizations and fraction decompositions, and location of zeros.

Let $P = \Sigma_0^n a_j z^j$ be a polynomial in one complex variable, with integer coefficients ($a_j \in \mathbb{Z}$). If P is factored as $P = Q \cdot R$, where Q and R are themselves in $\mathbb{Z}[z]$, what is the maximum of the coefficients in Q and in R? Can it be estimated before the decomposition is written? This problem has received special attention, because the existence of such an a priori bound is an essential feature for the design of efficient factorization algorithms in symbolic computation (H. Zassenhaus [50], Paul Wang and B. M. Trager [48], Paul Wang [45], [46], [47], Vilmar Trevisan and Paul Wang [43]).

Indeed, the classical computer algorithms for factorization work as follows. Suppose we want to factor P in $\mathbb{Z}[z]$. We can assume that P is primitive (no factor divides all coefficients in P) and that P and its derivative P' are relatively prime.

We choose a prime q, not dividing the leading coefficient of P. Let P_0 be the image of P in $\mathbb{Z}_q[z]$. More precisely, we divide the coefficients of P by q, until all of them are in the range $[-(q-1)/2, (q-1)/2]$. For instance, the polynomial P = $16z^2 - 8z - 15$ becomes $P_0 = 2z^2 - z - 1$, for $q = 7$.

Then P_0 is factored in $\mathbb{Z}_q[z]$. This is fast, since all coefficients of P_0 are smaller than q. Let $P_0 = Q_0 \cdot R_0$ be the factorization in $\mathbb{Z}_q[z]$ (we write only two factors for simplicity). Then

$$P \equiv P_0 = Q_0 \cdot R_0, \qquad (\mathrm{mod}\, q).$$

A lemma due to Hensel (H. Zassenhaus [50]) says that we can lift the factors modulo q into a decomposition modulo q^2. More precisely, we can find Q_1, R_1 in $\mathbb{Z}_{q^2}[z]$ such that:

$$Q_1 \equiv Q_0 \pmod{q}, \qquad R_1 \equiv R_0 \pmod{q},$$

and if $P_1 = Q_1 \cdot R_1$, then $P \equiv P_1 \pmod{q^2}$.

Let B be the bound we are looking for, namely the maximum (in modulus) of all coefficients in any factor of P. We repeat the lifting procedure with $q^2, q^4, \dots, q^{2^k}$, until $q^{2^k} \geq 2B$, so we find $Q_2, Q_3, \dots, Q_k, R_2, R_3, \dots R_k$, with

$$Q_j \equiv Q_{j-1} \quad \left(\mathrm{mod}\, q^{2^{j-1}}\right), \qquad R_j \equiv R_{j-1} \quad \left(\mathrm{mod}\, q^{2^{j-1}}\right),$$

and

$$P \equiv Q_k \cdot R_k \quad \left(\mathrm{mod}\, q^{2^k}\right).$$

We stop the lifting process at this point, and the algorithm now becomes a trial process: In $\mathbb{Z}[z]$, we try to divide P by $Q_k, R_k, \dots,$ or by combinations of these factors. If P has true factors, they will appear this way. However, it may happen that P is irreducible, though each factor in $\mathbb{Z}_{q^j}[z]$ was non-trivial.

The existence of the a priori bound B is therefore essential to determine the stopping time for the lifting process. Since this process is costly, the bound should be as small as possible.

The connection with product results, given in Section 1.C, is easy to describe. Assume $P = Q \cdot R$, both with integer coefficients, and assume we have proved a product result, such as

$$|Q \cdot R|_1 \geq \lambda |Q|_1 \cdot |R|_1.$$

Since R has integer coefficients, $|R|_1 \geq 2$, and so

$$|Q|_1 \leq \frac{1}{2\lambda}|P|_1, \tag{23}$$

the required bound.

The first estimates were given by Hans Zassenhaus [50]. Later, M. Mignotte [32] made use of Mahler's measure, and obtained the estimate

$$|Q|_1 \leq 2^n |P|_2. \tag{24}$$

A result of the first author [9], using the tools of the previous section, strongly improves upon this estimate. Indeed, he showed that the coefficients b_j of any factor Q of P satisfy

$$\max_j |b_j| \leq \frac{3^{3/4}}{2\sqrt{\pi}} \frac{3^{n/2}}{\sqrt{n}} [P]_2, \tag{25}$$

where $[P]_2$ is the norm defined in the previous section, which, in the one-variable case, is just

$$[P]_2 = \left(\sum_0^n \frac{1}{\binom{n}{j}} |a_j|^2 \right)^{1/2}. \tag{26}$$

The proof of (25) is very easy from the product result (22). Assume $P = Q \cdot R$, both factors having integer coefficients, so $[R]_2 \geq 1$. Then

$$[P]_2 \geq \sqrt{\frac{m_1!m_2!}{n!}}\,[Q]_2,$$

with $m_1 = \deg(Q)$, $m_2 = \deg(R)$. Taking the minimum of $m_1!m_2!$ when $m_1 + m_2 = n$ and using Stirling's formula yields (25).

The improvement over (24) is twofold. First, 2^n is replaced by $3^{n/2}/\sqrt{n}$, which is smaller, and second the classical l_2-norm is replaced by the new norm $[P]_2$, which is usually much smaller, since it carries significant weights in the denominators.

A computer implementation was realized by Vilmar Trevisan [41], [42], and showed strong improvements due to the new method: The a priori bound may, for instance, become 10 times smaller, for low-degree polynomials.

Future development of symbolic computation must include counting zeros of a given polynomial in a given region, for (as we have seen at the beginning), this is highly related to stability properties of dynamical systems. The work of the first author [8], described in section 1.F above, shows that inside a disk, the number of zeros depends only on the concentration of the polynomial at low degrees (and not on the degree itself); the subsequent works of Sylvia Chou [15], [16], and of Maria Girardi [21], provide precise estimates for this number. So, though the effective algorithms have not been written yet, the theoretical results already exist.

Another problem would also be worth considering: a priori bounds for coefficients in fraction decompositions.

4. Massively Parallel Computation on Polynomials

The present development of parallel computing turns it into a very efficient tool, but a quite rigid one, especially when we deal with *Single Instructions Multiple Data* machines, meaning that all processors are executing the same instruction at the same time. So far, only a limited number of problems, of situations, have been handled by S.I.M.D. parallel computing: mostly matrix computations and partial differential equations, by discretization.

The results of Beauzamy-Bombieri-Enflo-Montgomery [7] provide a canonical way of writing a many-variable polynomial on a hypercube. We now describe it.

We start with the polynomial written the usual way:

$$P(x_1, \dots, x_N) = \sum_{|\alpha| = m} a_\alpha x_1^{\alpha_1} \cdots x_N^{\alpha_N}.$$

Then we use Taylor's formula in order to write it in *symmetric* form, that is

$$P(x_1, x_2, \dots, x_N) = \sum_{i_1, \dots, i_m = 1}^{N} c_{i_1, \dots, i_m} x_{i_1} \cdots x_{i_m} \tag{27}$$

with

$$c_{i_1, \dots, i_m} = \frac{\partial^m P}{\partial x_{i_1} \cdots \partial x_{i_m}}. \tag{28}$$

This just means that a term $x_1 x_2$ is written as $\frac{1}{2}(x_1 x_2 + x_2 x_1)$, a term $x_1 x_2^2$ becomes $\frac{1}{3}(x_1 x_2 x_2 + x_2 x_1 x_2 + x_2 x_2 x_1)$, and so on.

Now we construct the hypercube. For this, we divide the segment $[0,1]$ into N equal pieces, by introducing the points

$$0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N}{N}.$$

In the hypercube $[0,1]^m$, we consider the points

$$M_{i_1,\dots,i_m} = \left(\frac{i_1}{N}, \dots, \frac{i_m}{N}\right),$$

where $i_1, \dots, i_m$ take any values in $\{1, \dots, N\}$. So there are N^m such points.

We now fill the hypercube the following way: At each point $M_{i_1,\dots,i_m}$ we put the coefficient $c_{i_1,\dots,i_m}$ defined in (28). We have obtained a representation of the polynomial on the hypercube.

For example, the polynomial $P = 4x_1x_2 - x_3^2$ has degree 2 and 3 variables. It will be represented as a cube of dimension 2, that is in a plane, by the matrix

$$H(P) = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

The weighted norm $[P]_2$ described and used previously appears simply as the l_2-norm on the hypercube, that is $(\Sigma_{i_1,\dots,i_m} |c_{i_1,\dots,i_m}|^2)^{1/2}$.

B. Beauzamy, J.-L. Frot, and C. Millour [11] showed that this representation of a many-variable polynomial on a hypercube allows full use of the massively parallel structure of a S.I.M.D. machine, in order to perform basic operations on polynomials, such as sums, products, pointwise evaluations, partial derivatives, and so on. Computer implementation was realized on a *Connection Machine* (in $*$-LISP), and showed considerable improvements in computing time, compared with sequential processing, when the number of variables becomes high (see [11]).

Moreover, as an unexpected theoretical benefit, this new representation allowed an easy description of the extremal pairs in the product results, that is of the pairs (P, Q) such that $[P \cdot Q]_2$ is as small as possible. These extremal pairs are those for which every column of the cube associated to P is orthogonal (in the usual euclidean sense) to every column of the cube associated to Q.

One sees here a true interaction between mathematics and computer science: the problem of representing a polynomial so it could be understood by special computers originated from computer science and required effective mathematical tools, but in turn these tools benefitted from the question.

5. Conclusion

We have seen throughout this presentation how quantitative estimates could be used in various and important areas of mathematics and computer science. There is one feature, however, worth emphasizing. The estimates we obtain are independent of the number of variables. This is of course essential for parallel computing, but this feature has a special meaning in modelling control theory. Indeed, usually, one wants to act on a control variable u in order to maximize some data on some system. The control variable is taken in a Hilbert space, or at least in some reflexive space (Besov, Sobolev, ...), in order to make use of the classical tools of weak compactness.

But this is unrealistic in the case where the global constraint u is made up of a lot of small ones, $u=(u_1, u_2, \dots, u_N)$, where each of them has a *right of veto*, that is should not fall below a certain level, $u_j < \varepsilon$ (we take $0 \le u_j \le 1$ for instance). Indeed, if u is in a Hilbert space, $u = \Sigma u_j e_j$, where e_j is some Hilbertian basis, and then $u_j \to 0$ when $j \to \infty$, which means that all constraints except a finite number are assumed to be negligible. That's exactly contrary to our "right of veto" situation.

So, in order to model such phenomena, one must allow each elementary variable to vary freely between 0 and 1. This means that $u=(u_1, u_2, \dots)$ is to be taken in a space of *bounded* sequences, namely l_∞. However, this space is not reflexive, and the necessary tools for optimization will disappear. So, in order to solve the problem, first fix N, look at $(u_1, \dots, u_N)$ and solve the problem in $l_\infty^{(N)}$, and then let $N \to \infty$. But, for given N, one needs the solution to be independent of N, that is, precisely to be independent of the number of variables.

Acknowledgement. The authors wish to thank Bruce Reznick for his kind and patient help in the preparation of this article.

REFERENCES

1. Arestov, V. V., Inequalities for different metrics for trigonometric polynomials, *Mat. Zametki*, Vol. 27 (1980), 4.
2. Aron, Richard and Globevnik, Josip, Interpolation by analytic functions on c_0, *Math. Proc. Cambridge Phil. Soc.*, 1988, 104, pp. 295–302.
3. Aron, Richard, Beauzamy, Bernard, and Enflo, Per, Polynomials in many variables: real vs. complex norms, *J. of Approx. Theory* 74, 2 (1992), 181–198.
4. Beauzamy, Bernard and Enflo, Per, Estimations de produits de polynômes, *J. of Number Theory*, 21-3 (1985), pp. 390–412.
5. Beauzamy, Bernard, Jensen's Inequality for polynomials with concentration at low degrees, *Numerische Math.* 49 (1986), pp. 221–225.
6. Beauzamy, Bernard, A minimization problem connected with a generalized Jensen's Inequality, *J. Math. Anal. and Applications*, Vol. 145, 1, January 1990, pp. 137–144.
7. Beauzamy, Bernard, Bombieri, Enrico, Enflo, Per, and Montgomery, Hugh, Products of polynomials in many variables, *J. of Number Theory*, Vol. 36, 2, October 1990, pp. 219–245.
8. Beauzamy, Bernard, Estimates for H^2 functions with concentration at low degrees and applications to complex symbolic computation, *J. für die Reine und Angewandte Math*, Vol. 433 (1992), 1–44.
9. Beauzamy, Bernard, Products of polynomials and a priori estimates for coefficients in polynomial decompositions: A sharp result, *J. of Symbolic Computation*, 13, (1992), pp. 463–472.
10. Beauzamy, Bernard and Chou, Sylvia, On the zeros of polynomials with concentration at low degrees, II, in *J. of Math. Anal. and Applications*, Vol. 175, 2 (1993), 360–379.
11. Beauzamy, B., Frot, J.-L., and Millour, C., Massively parallel computations on many-variable polynomials: when seconds count, Special Volume, "Maths and Computer Science," Annals of Maths and IA, M. Nivat and S. Grigorieff, eds., 1994.
12. Bell, E. T., Exponential polynomials. *Annals of Maths*, Vol. 35 2 (1934), pp. 258–277.
13. Beller, E. and Newman, D. J., An extremal problem for the geometric mean of a polynomial, *Proc. Amer. Math. Soc.*, 19 (1973), pp. 313–317.
14. Bonvalot, L., Moyenne Géométrique des fonctions des espaces de Hardy et polynômes concentrés aux bas degrés, *Thèse de Troisième Cycle*, Université de Paris 7, 1986.
15. Chou, Sylvia, On the roots of polynomials with concentration at low degrees, *J. of Math. Anal. and Applications*, Vol. 149, 2, July 1, 1990, pp. 424–436.
16. Chou, Sylvia, Séries de Taylor et concentration aux bas degrés, *Thèse*, Université de Paris VI, 1990.
17. Enflo, Per, On the invariant subspace problem in Banach spaces, *Acta Math.*, Vol. 158 (1987), pp. 213–313.
18. Enflo, Per, The largest coefficient in products of polynomials, *Functions Spaces, Lecture Notes in Pure and Appl. Math*, **136, Edwardsville, II, pp. 97–105.**
19. Fabre, C., La meilleure constante dans un produit de polynômes, *Note Comptes Rendus*, Acad. Sci. Paris, 307, (1988) pp. 767–770.
20. Frot, J.-L., and Millour, C., Rank of a polynomial and extremality, *to appear*.

21. Girardi, Maria, Bounding zeros of H^2 functions by concentrations, *J. of Math. Anal. and Applic.*, Vol. 163, 3 (1994), 605–612.
22. Kahane, J.-P., *Séries de Fourier Absolument Convergentes*, Springer-Verlag, New York, NY, 1970.
23. Kahane, J.-P., Polynômes à coefficients unimodulaires sur le cercle unité, *Séminaire d'Analyse Fonctionnelle*, 1979–80, Exposé 9, Ecole polytechnique.
24. Lacruz, Miguel, Ph.D. Thesis, Kent State University, 1991.
25. Langevin, M., Estimations du module d'un polynôme dans le plan complexe (preprint).
26. Lenstra, A. K., Lenstra, H. W., and Lovasz, L., Factoring polynomials with integer coefficients, *Math Annalen* 261 (1982), *pp*, 515–531.
27. Levin, B., On the distribution of zeros of entire functions, *Amer. Mat. Soc.*, Vol. 5.
28. Mahler, K., An application of Jensen's formula to polynomials, *Mathematika* 7, (1960), pp. 98–100.
29. Mahler, K., On two extremum properties of polynomials, *Ill. J. of Math*, 7 (1963), pp. 681–701.
30. Marden, Morris, Geometry of Polynomials, *Amer. Math. Soc. Math. Surveys*, 3rd edition, 1985.
31. McGehee, O., Pigno, L., and Smith, B., Hardy's inequality and the L_1 norm of exponential sums, *Ann. of Math*, 113 (1981), pp. 613–618.
32. Mignotte, M., An inequality about factors of polynomials, *Mathematics of Computation*, Vol. 28, 128, (1974), pp. 1153–1157.
33. Newman, D. J., An L_1-extremal problem for polynomials, *Proc. Amer. Math. Soc.*, 16-2, (1965), pp. 1287–1290.
34. Reznick, Bruce, An inequality for products of polynomials, *Proc. Amer. Math. Soc.*, Vol. 117, 4 (1993), 1063–1073.
35. Rigler, A., Trimble, R. S., and Varga, R., Sharp lower bounds for a generalized Jensen's Inequality, *Rocky Mountain J. of Math*, 19, 1989, pp. 353–373.
36. Rudin, W., *Real and Complex Analysis*, 3rd edition, Tata McGraw Hill, 1983.
37. Sarantopoulos, Yannis, Estimates for polynomial norms on $L_p(\mu)$ spaces, *Math. Proc. Cambridge Phil. Soc.*, 1986, 99, pp. 263–271.
38. Sarantopoulos, Yannis, Extremal multilinear forms on Banach Spaces, *Proc. Amer. Math. Soc.*, Vol. 99, 2, 1987, pp. 340–346.
39. Sarantopoulos, Yannis, Polynomials on certain Banach Spaces, *Bull, Soc. Math. Greece*, 28 (1987) part 4, pp. 89–102.
40. Tonge, Andrew, The Von Neumann inequality for polynomials in several Hilbert-Schmidt operators, *J. of the London Math. Soc.* (2), 18, 1978, pp. 519–526.
41. Trevisan, V., Recognition of Hurwitz polynomials, *SIGSAM Bulletin*, Vol. 24, 4, October 1990.
42. Trevisan, V., Computing a sharp bound for the coefficients in polynomial factorizations, to appear.
43. Trevisan, V. and Wang, P., Practical factorization of univariate polynomials over finite fields, *Proceedings of the ISAAC'91*, Bonn, July 1991.
44. Waldschmidt, M., *Nombres Transcendants*, Lecture Notes, Springer Verlag, NY, 1978.
45. Wang, P., An improved multivariate polynomial factoring algorithm, *Math. Comp.*, Vol. 32, 144, October 1978, pp. 1215–1231.
46. Wang, P., Parallel p-adic construction in the univariate polynomial factoring algorithm, *Proc., Second Macsyma Users Conference*, Washington, DC, June 1979, pp. 310–317.
47. Wang, P., A p-adic algorithm for univariate partial fractions, *Proceedings of the 1981 ACM Symposium on Symbolic and Algebraic Computation*, pp. 212–217.
48. Wang, P and Trager, B. M., New algorithms for polynomial square-free decompositions over the integers, *SIAM J. of Computing*, Vol. 8, 3, August 1979, pp. 300–305.
49. Wang, P., Early detection of true factors in univariate polynomial factorizations, in *Computer Algebra, Springer Lecture Notes in Computer Science* 162 (1983), 225–235.
50. Zassenhaus, Hans, On Hensel factorization, *J. of Number Theory*, 1, 1969, pp. 291–301.

Convolutions and Computer Graphics

ANNE M. BURNS
Long Island University
Brookville, NY 11548

The convolution of two functions is central to the operation of filtering in signal processing and in image processing. Now the concept of convolution has found a major application in computer graphics. Conversely, computer graphics makes it possible to visualize the effect of convolving two functions and so can be used as a valuable educational tool; there is an endless variety of graphical effects that can be achieved using two-dimensional convolutions.

First let us review a few facts about convolutions. Let $f(x)$ and $k(x)$ be real-valued functions of a real variable. Recall that the convolution of f with k is a new function of a real variable

$$F(t)=\int_R f(x)k(t-x)\,dx.$$

For example, let f be a simple rectangular function

$$f(x)=\begin{cases}0 & \text{for} \quad x<-h\\ 1 & \text{for} \quad -h\leqslant x\leqslant h\\ 0 & \text{for} \quad h<x.\end{cases}$$

See FIGURE 1 for $h=1$.

Then a straightforward calculation shows that the convolution of f with itself is

$$F(t)=\begin{cases}0 & \text{for} \quad t<-2h\\ t+2h & \text{for} \quad -2h\leqslant t<0\\ 2h-t & \text{for} \quad 0\leqslant t\leqslant 2h\\ 0 & \text{for} \quad 2h<t.\end{cases}$$

See FIGURE 2 for $h=1$.

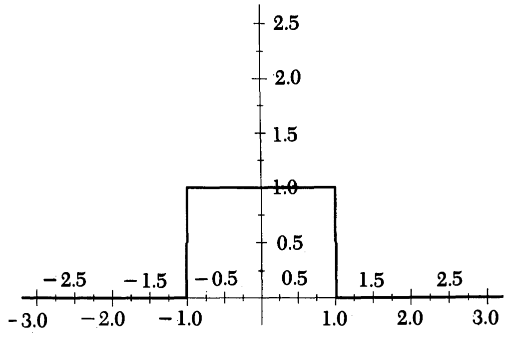


FIGURE 1

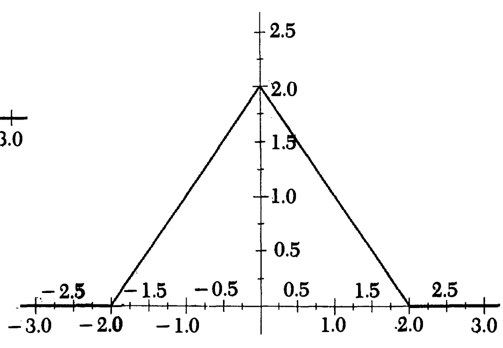


FIGURE 2

If we now take the resulting function $F(x)$ and convolve it with $f(x)$ to obtain

$$G(t)=\int_R F(x)f(t-x)\,dx,$$

another straightforward calculation yields

$$G(t)=\begin{cases}0 & \text{for} \quad t<-3h\\ (t+3h)^2/2 & \text{for} \quad -3h\leqslant t<-h\\ -t^2+3h^2 & \text{for} \quad -h\leqslant t<h\\ (t-3h)^2/2 & \text{for} \quad h\leqslant t\leqslant 3h\\ 0 & \text{for} \quad 3h<t.\end{cases}$$

See FIGURE 3.

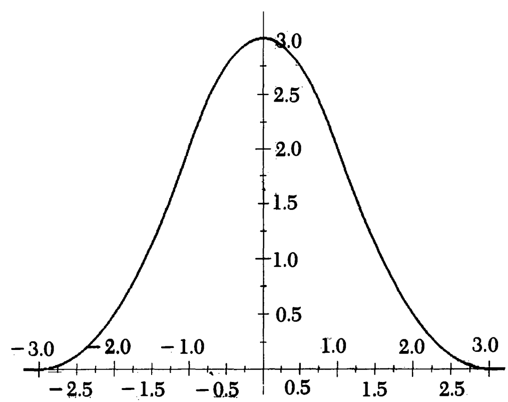


FIGURE 3

In this example we have let $k(x)$, called the *kernel*, be the rectangular function of height 1 and width $2h$. We see that convolving a function with $k(x)$ has the effect of "spreading" and "smoothing" the function f. An easy exercise shows that the operation of convolution is commutative, and thus the second convolution could also be thought of as convolving the rectangular function and the triangular function. Here we see that convolving the rectangle with the triangle results in more spreading and peaking than does convolving the rectangle with itself.

In the operation of filtering, a kernel or weighting function is convolved with the input signal and a weighted image is computed for each output signal. Two kernels of importance in this technique are:

(1) the sinc kernel, $k(x)=(w/\pi)\sin(wx)/(wx)$;

(2) the Gaussian kernel, $k(x)=\exp(-x^2/(2\sigma^2))/(\sqrt{2\pi}\,\sigma)$.

It is instructive to sketch the graphs of these functions.

The generalization to two dimensions is straightforward. The convolution of two real-valued functions of two real variables, $f(x,y)$ and $k(x,y)$, is a new function of two real variables

$$F(s,t)=\int_R\int_R f(x,y)k(s-x,t-y)\,dx\,dy.$$

If the kernel $k(x,y)$ has compact support, that is

$$k(x,y)=0 \quad \text{if} \quad x<-h \quad \text{or } x>h \quad \text{or} \quad y<-r \quad \text{or } y>r,$$

then making the substitution

$$\begin{aligned} u &= s-x, & du &= -dx \\ v &= t-y, & dv &= -dy, \end{aligned}$$

we obtain the integral

$$\begin{aligned} F(s,t) &= \int_R\int_R f(s-u,t-v)k(u,v)(-du)(-dv) \\ &= \int_R\int_R f(s-u,t-v)k(u,v)\,du\,dv \\ &= \int_{-r}^{r}\int_{-h}^{h} f(s-u,t-v)k(u,v)\,du\,dv. \end{aligned}$$

In computer graphics when an image is rendered using graphs of mathematical relations such as lines, circles, and polygons, the boundary between two regions may appear jagged rather than smooth, as we are actually plotting discrete points. This effect is known as *aliasing*. We saw that convolving a function with a rectangular or triangular kernel has the effect of spreading and smoothing the function. This suggests that the application of the discrete form of the convolution might be employed to smooth boundaries, a technique called *antialiasing*. We shall describe how this is achieved and also mention some other ways that convolutions may be used in computer graphics. We shall also describe how computer graphics may be used in the classroom to illustrate the results of various convolutions.

Since we are working on a computer screen, we will be concerned with a discrete subspace of two-dimensional Euclidean space. Let S be the discrete subspace of two-dimensional Euclidean space consisting of those points that have integer coordinates. In S, if we choose a kernel function that is zero outside of the rectangle $\{(x,y)\mid -h\leqslant x\leqslant h,\ -r\leqslant y\leqslant r\}$, then for all $(x,y)\in\mathbb{Z}\times\mathbb{Z}$ the convolution of a function $f(x,y)$ with $k(x,y)$ becomes

$$F(s,t)=\sum_{v=-r}^{r}\sum_{u=-h}^{h} f(s-u,t-v)k(u,v),$$

or

$$F(x,y)=\sum_{j=-r}^{r}\sum_{i=-h}^{h} f(x-i,y-j)k(i,j),$$

or

$$F(x,y)=\sum_{j=-r}^{r}\sum_{i=-h}^{h} f(x+i,y+j)k(i,j).$$

Now how does all this relate to a computer screen?

We shall consider the computer screen as a bounded subspace of S, where the x-coordinates are $0,1,2,3,\dots,X$ and the y-coordinates are $0,1,2,3,\dots,Y$, where $X+1$ and $Y+1$ are the maximum number of pixels in the horizontal and vertical directions. We will consider functions $f(x, y)$ that take on the values $0, 1, 2, \ldots, N-1$, where N is the number of colors available on our computer. This way we can "graph" our function by plotting the color numbered $f(x, y)$ at location (x, y) on the screen. One simple way of devising such a function is to take any real-valued function, round it to the nearest integer, take its absolute value and reduce the result modulo N. Another way is to take a real-valued function $g(x, y)$ with range $[s, t]$, then let $f(x, y) = (N-1)|g(x, y)|/\max(|s|, |t|)$ rounded to the nearest integer.

After seeing a few examples, the imaginative student will be able to invent any number of his or her own such functions. In the examples we will use the notation round (a) to denote the real number a rounded to the nearest integer.

Example 1. $f(x, y) = \text{round}(|(y - a\sin(x))(x - b\sin(y))|) \bmod N$, for $0 \leqslant x \leqslant 8$, $0 \leqslant y \leqslant 8$.

This function is shown in FIGURE 4 with $a = 5$, $b = 2$, and $N = 16$.

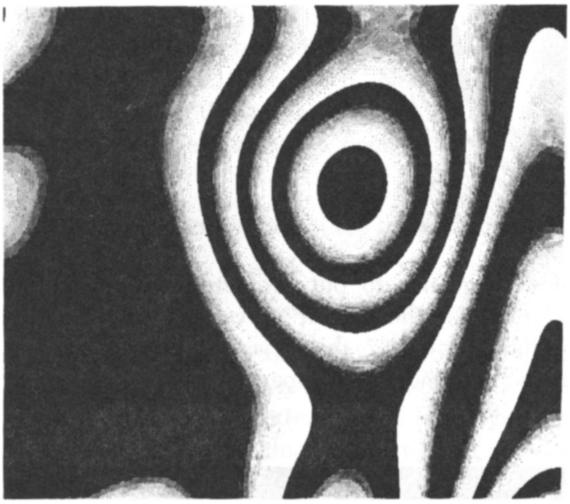


FIGURE 4

Example 2. $g(x, y) = \text{round}(x - y - (5\sin(x)\sin(y) + 3\sin(3x)\sin(3y) + 2\sin(5x)\sin(5y)))$, for $0 \leqslant x \leqslant 2\pi$, $0 \leqslant y \leqslant \pi$, and $f(x, y) = |g(x, y)| \bmod N - 1$. See FIGURE 5.

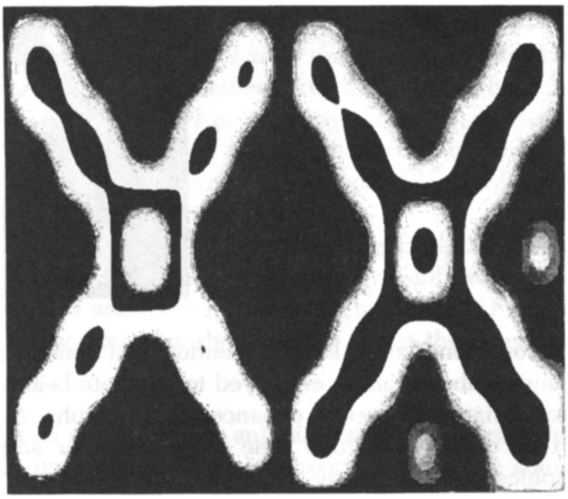


FIGURE 5

Notice that in Examples 1 and 2 we are graphing the level curves of the function.

Example 3. Begin with a complex-valued function $h(z)$, where $z = x + iy$. Let $f(z) = \text{round}(|h(z)|) \bmod N$. In FIGURE 6, we let $h(z) = 100(z-a)(z-b)(z-c)$ with $a = 1 + 0.2i$, $b = 0.2 + 0.8i$ and $c = 1.2 + 1.2i$, $-0.2 \leqslant \text{re}(z) \leqslant 1.65$, $-0.2 \leqslant \text{im}(z) \leqslant 1.65$.

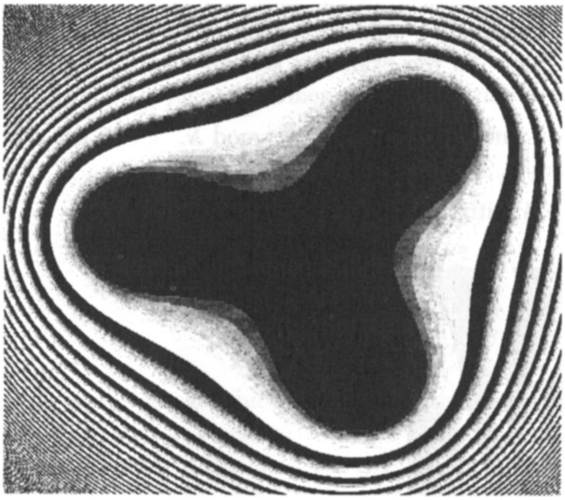
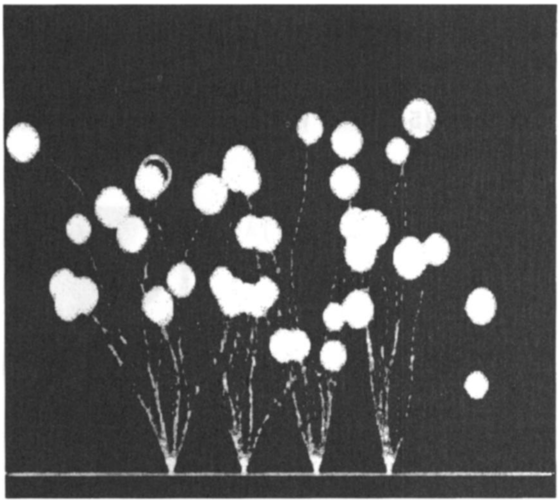


FIGURE 6

Example 4. Use available graphics commands to put geometric shapes (points, filled-in circles, filled-in rectangles, etc.) on the screen in various colors, either randomly or according to some formula. FIGURE 7 shows jagged lines of random lengths starting at the bottom of the screen and then topped with disks of random radii.

FIGURE 7



Before continuing, let me mention that functions such as those described in the examples above can be employed to generate beautiful pictures. A judicious choice of color assignment greatly enhances these graphs. Although the number of colors that can be displayed at one time on a computer screen is limited by the amount of graphics memory, many computers allow the user to use the pixel value from the graphics memory as an index into a color look-up table where the actual colors can be chosen from a large selection of colors. Thus the colors can be assigned to the look-up table in a way that achieves the effect of a continuous color ramp. For a thorough discussion on the assignment of colors see [8].

Before we continue our investigation into convolutions, a few pointers on constructing functions that yield interesting graphs should serve to get the imaginative student started.

Our function should map each pixel (x, y), where $0 \leqslant x \leqslant X$ and $0 \leqslant y \leqslant Y$ into the integers 0 through $N-1$ where N is the number of colors.

As we mentioned earlier, one idea that comes to mind is to graph the level curves of $|f(x, y)|$ properly scaled and rounded to the nearest integer. Or, to get interesting patterns, we can multiply $|f(x, y)|$ by a scale factor that makes the result greater than $N-1$ and then reduce that result modulo N.

One method for finding functions with interesting level curves is the following: Choose a real-valued function of one variable, $v = g(u)$, which has several zeros, or places where its graph crosses the u-axis. Now choose another such function and reverse the independent and dependent variables, writing $u = h(v)$. Then let $f(u, v) = |v - g(u)|\,|u - h(v)|$. Now f is a nonnegative function of two variables whose zeros are the graphs of g and h and whose level curves will loosely "follow" those curves. Next, choose a domain for f that includes the zeros of g and h:

$$a \leqslant u \leqslant b, \qquad c \leqslant v \leqslant d.$$

Then for each screen pixel (x, y), let

$$u = (b-a)x/X + a, \qquad v = (d-c)y/Y + c,$$
$$\text{and} \qquad \text{color index} = |f(u, v)| \bmod N.$$

Now, returning to convolutions, with one of these functions defined and mapped to the screen, we are ready to see the effect of convolving our function with a kernel defined on the set of ordered pairs of integers (x, y) with $-h \leqslant x \leqslant h$, $-r \leqslant y \leqslant r$.

Assume we have defined a function $f(x, y)$ for each pixel (x, y), $x = 0, 1, 2, \dots, X$, $y = 0, 1, 2, \dots, Y$, that ranges over the set of colors. We define a kernel $k(x, y)$, $x = -h, -h+1, \dots, h-1, h$, and $y = -r, -r+1, \dots, r-1, r$, where h and r are integers. We can store the kernel as a two-dimensional array $k[i, j]$. The convolution of $f(x, y)$ with $k(i, j)$, which we are denoting $F(x, y)$, is given by

$$F(x, y) = \sum_{j=-r}^{r} \sum_{i=-h}^{h} f(x+i, y+j) k(i, j).$$

We can visualize the value of $F(x, y)$ by picturing the array that represents the kernel superimposed over the screen, where the center element, $k(0, 0)$, is placed over pixel (x, y), and we sum over all the $(2h+1)(2r+1)$ pixels surrounding (x, y) that are covered by the kernel array, the present value (color) of the pixel times the corresponding element in the kernel. We might use periodic boundary conditions or we could simply assign zeros to locations outside the boundaries.

We want to show the effect of the convolution on our computer screen. Each pixel (x, y) has now been assigned a new number, but the number probably won't be an integer and it may not be between 0 and $N-1$. We must map $F(x, y)$ into the integers $0, 1, 2, \dots, N-1$. Again a few examples should serve to get the imaginative reader started.

Averaging To achieve the effect of smoothing or antialiasing that was mentioned earlier, we want a kernel roughly proportional to the Gaussian kernel. The color of neighboring pixels will influence the new color at each pixel, with those closest to it having the most influence.

We begin with such a kernel:

$$k(i,j)=\begin{cases}1/4 & \text{if } i=j=0\\ 1/8 & \text{if } i=0 \text{ and } j=\pm 1, \text{ or } i=\pm 1 \text{ and } j=0\\ 1/16 & \text{if } i=-1 \text{ and } j=\pm 1, \text{ or } i=1 \text{ and } j=\pm 1\\ 0 & \text{if } |i|>1, \text{ or } |j|>1.\end{cases}$$

As we mentioned earlier, k can be stored as a two-dimensional array, and we can visualize $k(i,j)$ as follows:

$$\begin{array}{ccc} 1/16 & 1/8 & 1/16\\ 1/8 & 1/4 & 1/8\\ 1/16 & 1/8 & 1/16. \end{array}$$

Henceforth we shall represent a kernel by its associated array. This kernel has been normalized, which means that the sum of the entries is 1. With a normalized kernel, $k(i,j)$, we can assign to the pixel with coordinates (x,y) the value of $F(x,y)$ rounded to the nearest integer. Alternatively, since integer arithmetic is much faster on the computer and an enormous number of computations are being performed, it would be more efficient to perform the convolutions using a kernel with integer values that are proportional to the entries in the normal kernel, such as

$$\begin{array}{ccc} 1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1. \end{array}$$

Then let s stand for the sum of the elements in the array representing the kernel, that is

$$s=\sum_{j=-r}^{r}\sum_{i=-h}^{h}k(i,j).$$

In our example $s=16$. We assign a new color to each pixel (x,y) using the formula $F(x,y)/s$ rounded off to the nearest integer. Notice that this assigns to each pixel a new color that is a weighted average of its present color and the present colors of neighboring pixels. If colors have been assigned to give the effect of a continuous color ramp or a gray scale ramp, the effect of this is to blur boundaries between two regions of different colors. For example, suppose colors 0 through 15 have been assigned as shades of gray from black to white, and we have drawn a white square on the screen illustrated by the following situation:

$$\begin{array}{ccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 15 & 15 & 15 & 0 & 0\\ 0 & 0 & 15 & 15 & 15 & 0 & 0\\ 0 & 0 & 15 & 15 & 15 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0. \end{array}$$

We apply a convolution to each point on the screen using the kernel from our example,

1	2	1
2	4	2
1	2	1.

After applying the convolution and averaging, the colors are assigned to the screen as follows:

0	0	0	0	0	0	0
0	1	3	4	3	1	0
0	3	8	11	8	3	0
0	4	11	15	11	4	0
0	3	8	11	8	3	0
0	1	3	4	3	1	0
0	0	0	0	0	0	0.

Notice the effect of smoothing and spreading that we observed in the continuous case. The reader should verify that this is indeed the result. An instructive exercise is to apply the convolution to the result repeatedly and find the limiting value. The reader should also experiment with a variety of kernels, including a 5-by-5 kernel and a nonsymmetric kernel. Repeated convolutions of a kernel such as

1	1	1
1	1	1
1	1	1

with itself should be compared with repeated convolutions of the rectangle with itself in the continuous case. The effect of shading on one side only can be accomplished by using a one-sided kernel, such as

1	1	0
1	1	0
0	1	0.

FIGURE 8 shows a scene created from simple geometric forms after which various convolutions have been performed. The sky is composed of random-sized, light-colored circles placed randomly on a dark background, then averaged several times using the kernel

1	2	1
2	4	2
1	2	1.

The sky is then reflected onto the bottom third of the screen, which is overlaid by randomly placed horizontal lines of random length that are several shades darker than the spot on which they are placed. The bottom third of the picture is convolved with the kernel

0	0	0
1	1	1
0	0	0.

The water in the background is convolved several times with this kernel, while the foreground water is convolved only once.

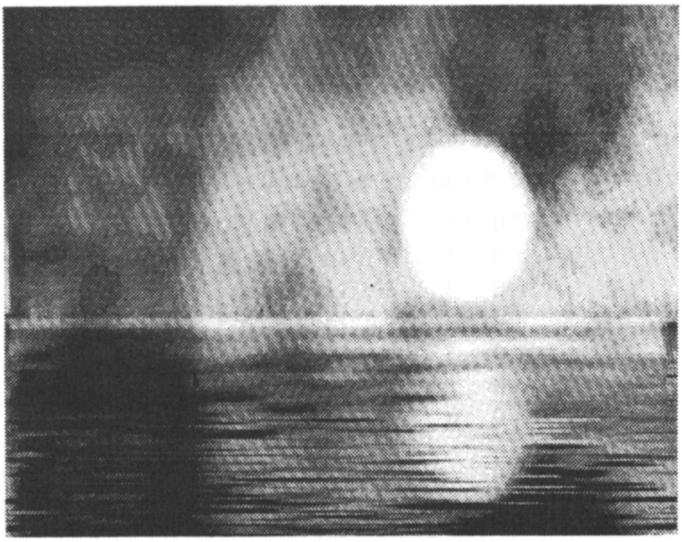


FIGURE 8

Thresholding This technique is used in processing images. For simplicity suppose we have just two colors, black = 0 and white = 1. We choose a threshold T, and color the pixel

$$\begin{array}{ll} 1, & \text{if } F(x,y) > T \\ 0, & \text{otherwise.} \end{array}$$

Depending on whether the threshold is low or high, an image can be expanded or contracted; if the kernel is one-sided an image can be shifted right or left, or up or down. This technique can be used to sharpen an image as opposed to smoothing it out. An excellent account of the theory of thresholding can be found in [6].

For endless variations an initial image can be created on the screen and convolved and thresholded any number of times, varying the kernel and/or the threshold at each stage.

Colorful patterns and pictures Convolving a function with a nonnormalized kernel may result in values outside the range of colors. In order to use colors to show the effect of the convolution we can reduce the result modulo the number of colors. A good exercise is the following: Start with a single pixel assigned the color 1 in the center of the screen. Apply repeatedly to the screen a convolution with a kernel such as

$$\begin{array}{ccc} 3 & 5 & 3 \\ 5 & 8 & 5 \\ 3 & 5 & 3 \end{array}$$

where the result is reduced modulo 16. Assign harmonious colors to the numbers 0 through 15. This exercise can be done in class on graph paper, changing the entries in the kernel and the modulus. My students are often surprised at the results.

It is hoped that the reader can use his or her imagination to expand on these few examples, and to invent many new, useful, educational and entertaining ways to exploit the theory of convolutions as applied to computer graphics.

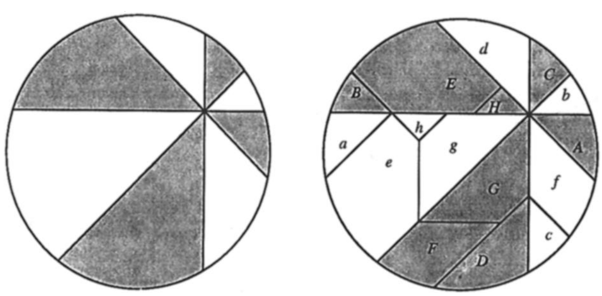
REFERENCES

1. Alain Fournier and Eugene Fiume, Constant-time filtering with space-variant kernels, *Computer Graphics* 22 (1988), 229–237.
2. Paul Heckbert, Filtering by repeated integration, *Computer Graphics* 20 (1986), 315–321.
3. R. C. Jennison, *Fourier Transforms and Convolutions for the Experimentalist*, Pergamon Press, Inc., NY, 1961.
4. Don P. Mitchell and Arun N. Netravali, Reconstruction filters in computer graphics, *Computer Graphics* 22 (1988), 221–228.
5. J. M. Ogden, E. H. Adelson, J. R. Bergen, and P. J. Burt, Pyramid-based computer graphics, *Course Notes #5*, ACM SIGGRAPH '88.
6. Kendall Preston, Jr. and Michael J. B. Duff, *Modern Cellular Automata Theory and Applications*, Plenum Press, New York, 1984.
7. Rod Salmon and Mel Slater, *Computer Graphics, Systems and Concepts*, Addison-Wesley Publishing Co., Reading, MA, 1987.
8. Kenneth Turkowski, Anti-aliasing in topological color spaces, *Computer Graphics* 20 (1986), 315–321.
9. H. Joseph Weaver, *Theory of Discrete and Continuous Fourier Analysis*, John Wiley & Sons, Inc., NY, 1989.

Proof without Words: Fair Allocation of a Pizza

Dedicated to the Late Professor Joseph D. E. Konhauser

THE PIZZA THEOREM: *If a pizza is divided into eight slices by making cuts at* 45° *angles from an arbitrary point in the pizza, then the sums of the areas of alternate slices are equal.*



NOTES: This result, discovered by L. J. Upton, is true when n, the number of pieces, is $8, 12, 16, \ldots$, but is false for $n = 2, 4, 6, 10, 14, 18, \ldots$. The positive results are in the references. For the negative, the case $n = 4$ is easily handled, while if $n \equiv 2 \pmod 4$ we have the following argument of Don Coppersmith (IBM). It suffices, by continuity, to take the special point on the boundary of the unit circle and one of the chords to be tangent at the point. Then the black and white areas can each be expressed in terms of π and algebraic numbers; their equality would contradict π's transcendence. In fact, the difference between black and white areas is $2\pi/n - \tan(2\pi/n)$. Also see Proposal 1457, this issue.

1. L. J. Upton, Problem 660, this MAGAZINE 41 (1968) 46.
2. Stanley Rabinowitz, Problem 1325, *Crux Mathematicorum* 15 (1989) 120–122.

—LARRY CARTER[1]
IBM THOMAS J. WATSON RESEARCH CENTER
YORKTOWN HEIGHTS, NY 10598

—STAN WAGON
MACALESTER COLLEGE
ST. PAUL, MN 55105

[1]Work done while author was visiting the Department of Computer Science at the University of Colorado.

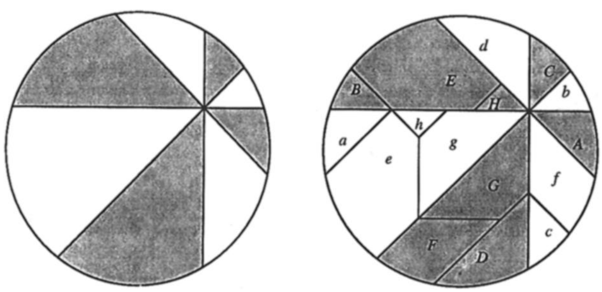
REFERENCES

1. Alain Fournier and Eugene Fiume, Constant-time filtering with space-variant kernels, *Computer Graphics* 22 (1988), 229–237.
2. Paul Heckbert, Filtering by repeated integration, *Computer Graphics* 20 (1986), 315–321.
3. R. C. Jennison, *Fourier Transforms and Convolutions for the Experimentalist*, Pergamon Press, Inc., NY, 1961.
4. Don P. Mitchell and Arun N. Netravali, Reconstruction filters in computer graphics, *Computer Graphics* 22 (1988), 221–228.
5. J. M. Ogden, E. H. Adelson, J. R. Bergen, and P. J. Burt, Pyramid-based computer graphics, *Course Notes #5*, ACM SIGGRAPH '88.
6. Kendall Preston, Jr. and Michael J. B. Duff, *Modern Cellular Automata Theory and Applications*, Plenum Press, New York, 1984.
7. Rod Salmon and Mel Slater, *Computer Graphics, Systems and Concepts*, Addison-Wesley Publishing Co., Reading, MA, 1987.
8. Kenneth Turkowski, Anti-aliasing in topological color spaces, *Computer Graphics* 20 (1986), 315–321.
9. H. Joseph Weaver, *Theory of Discrete and Continuous Fourier Analysis*, John Wiley & Sons, Inc., NY, 1989.

Proof without Words: Fair Allocation of a Pizza

Dedicated to the Late Professor Joseph D. E. Konhauser

THE PIZZA THEOREM: *If a pizza is divided into eight slices by making cuts at 45° angles from an arbitrary point in the pizza, then the sums of the areas of alternate slices are equal.*



NOTES: This result, discovered by L. J. Upton, is true when n, the number of pieces, is $8, 12, 16, \ldots$, but is false for $n = 2, 4, 6, 10, 14, 18, \ldots$. The positive results are in the references. For the negative, the case $n = 4$ is easily handled, while if $n \equiv 2 \pmod 4$ we have the following argument of Don Coppersmith (IBM). It suffices, by continuity, to take the special point on the boundary of the unit circle and one of the chords to be tangent at the point. Then the black and white areas can each be expressed in terms of π and algebraic numbers; their equality would contradict π's transcendence. In fact, the difference between black and white areas is $2\pi/n - \tan(2\pi/n)$. Also see Proposal 1457, this issue.

1. L. J. Upton, Problem 660, this MAGAZINE 41 (1968) 46.
2. Stanley Rabinowitz, Problem 1325, *Crux Mathematicorum* 15 (1989) 120–122.

—LARRY CARTER[1]
IBM THOMAS J. WATSON RESEARCH CENTER
YORKTOWN HEIGHTS, NY 10598

—STAN WAGON
MACALESTER COLLEGE
ST. PAUL, MN 55105

[1]Work done while author was visiting the Department of Computer Science at the University of Colorado.

Wagering in Final *Jeopardy!*

GEORGE T. GILBERT
RHONDA L. HATCHER
Texas Christian University
Fort Worth, TX 76129

1. Introduction

In the popular television game show *Jeopardy!* three contestants accumulate points by giving the correct questions corresponding to the answers provided by the emcee. While the contestants show a great deal of self-confidence in wagering in Final *Jeopardy!*, the last stage of the game, they appear to give little thought to betting strategy. In this article, we look for a reasonable strategy for wagering in Final *Jeopardy!* We consider betting strategies from two perspectives. First, we describe a practical strategy to use against two contestants who behave in a way predicted by actual game data. After this, we determine how the game might be played if all of the contestants were knowledgeable in game theory.

2. Description of Final *Jeopardy!*

Before the start of Final *Jeopardy!*, the three contestants have earned scores based on their performances in the first two stages of *Jeopardy!* Any contestant with a positive score at this point goes on to play Final *Jeopardy!*, a segment of the game consisting of a single answer and question.

At the beginning of Final *Jeopardy!*, a question category is announced and the contestants are asked to secretly wager an integer between zero and their current scores. The emcee announces the answer and the contestants are given 30 seconds to write down the corresponding question. The contestants who give the correct question have their wagers added to their scores; the contestants giving incorrect questions have their wagers subtracted from their scores. At the end of Final *Jeopardy!*, the player with the highest score is declared *Jeopardy!* Champion and wins the amount of his or her score in dollars and the right to return to the next show to play again. The second- and third-place players do not win any money and instead are sent home with consolation prizes. In the case of a tie for first place, those players win their scores and return to play in the next show. However, if all three contestants have a final score of zero, there is no winner and none of the contestants returns to play again.

To illustrate, let's look at an actual game. We will call the contestants who were in first, second, and third place at the beginning of Final *Jeopardy!* players A, B, and C, respectively. In this game, player C gave the correct question, and players A and B gave incorrect questions. The contestants' starting scores, wagers, and final scores were as follows:

	A	B	C
Starting Score	10000	8100	3200
Wager	6201	8100	3100
Final Score	3799	0	6300

Player C came from far back to win!

3. A practical strategy

By a practical strategy in Final *Jeopardy!*, we mean the betting strategy a particular contestant should use against two contestants with playing ability and betting strategy consistent with empirical data. For example, in the game described in the previous section, if player B had known player A's wager, a passive bet would allow player B to win whenever player A missed. We will see that player B could have anticipated player A's wager.

We examined in detail 170 games where Final *Jeopardy!* had three players with different scores. We provide a brief summary of our findings and describe an *optimal strategy* for each player based on the data. An optimal strategy maximizes the probability of a win or a tie and minimizes the likelihood of a tie if it does not significantly increase the risk of losing.

In discussing the betting behavior of the players, we use the following notation:

$$A = \text{player A's score at the beginning of Final } \textit{Jeopardy!},$$

$$A_R = A + (\text{amount of player A's wager}),$$

$$A_W = A - (\text{amount of player A's wager}).$$

We define B, B_R, B_W, C, C_R, and C_W similarly.

In 41 of the 170 games we analyzed, player A had a score at the beginning of Final *Jeopardy!* that was more than double player B's score. In other words, $A > 2B$. In every such game player A's wager was small enough to assure a win.

The remaining 129 games all fell into the category $A < 2B$, so player A had not clinched a win. Among these, there were 8 games with special relationships between the scores, which we omitted from consideration.

Let's first look at the betting behavior of player A in the remaining 121 games with $A < 2B$. In 117 of these games, player A bet so that $A_R \geq 2B$, thus assuring a win or tie with a correct answer. Of these 117 games, player A bet to tie in only 6 games, and in 60 games player A bet to win by one dollar. In only 8 games did player A bet to beat $2B$ by such a large margin that player A took on any significant strategic risk.

To examine player B and player C's strategies when $A < 2B$, consider the two cases $\lceil \frac{3}{2}B \rceil < A < 2B$ and $A < \lceil \frac{3}{2}B \rceil$. Note that if $\lceil \frac{3}{2}B \rceil < A < 2B$, then player A can bet $2B - A + 1$ so that $A_R = 2B + 1$. In order for player A to lose, player A must get the question wrong and one of the other players must get the question right. There is no such bet when $A < \lceil \frac{3}{2}B \rceil$.

In 22 of the 36 games where $\lceil \frac{3}{2}B \rceil < A < 2B$, player B bet within \$100 of all he or she had. In only 5 games was $B_R < A$, and in 3 of those $B_R > A_*$, where $A_* = 2A - 2B - 1$, player A's score with a wrong answer when wagering $2B - A + 1$. Player C bet within \$100 of everything in 21 of the 36 games. In only 4 games with $2C > A$ did player C bet so that $C_R < A$ and in only one of these was $C_R < A_*$.

The 85 games where $A < \lceil \frac{3}{2}B \rceil$ were potentially more interesting because the players had several strategies available. Yet in 43 of these games, player B bet within \$100 of everything, and in only 22 of the games was $B_W > A_*$. Player C wagered just as aggressively. When $C < A_*$, player C (justifiably) wagered within \$100 of everything in 13 of the 18 games, and only one wager could be called passive. In the 67 games with $C > A_*$, player C wagered within \$100 of everything in 30, while in 25 bet so that $C_W > A_*$.

The key fact pervading the data is that most players wager aggressively.

In order to arrive at a practical strategy for a contestant in Final *Jeopardy!*, we consider the distribution of right and wrong answers in Final *Jeopardy!* Each game falls into one of the eight categories listed in Table 1, where, for example, RRW is the category of games where the first- and second-place players give the correct answer and the third-place player gives the wrong answer. We tabulated the following empirical distribution.

TABLE 1. Right-Wrong Distribution

Category	Number	%
RRR	33	19.41
RRW	20	11.76
RWR	20	11.76
RWW	22	12.94
WRR	12	7.06
WRW	25	14.71
WWR	9	5.29
WWW	29	17.06

While one might suspect that the right-wrong distribution would be dependent upon the relative strengths of the players as measured by their scores upon starting Final *Jeopardy!*, we saw no indication that it differed appreciably.

Using the betting strategy data, the right-wrong distribution data, and quantitative reasoning, there is a fairly simple strategy for each player. In every case, player A should wager $2B - A + 1$, so that $A_R = 2B + 1$. Our suggested strategies for players B and C are summarized in Table 2.

TABLE 2. Recommended Wagers for Players B and C

Score Conditions	Player B	Player C
$\lceil\frac{3}{2}B\rceil < A < 2B$	B	C
$B + \lceil\frac{C}{2}\rceil < A < \lceil\frac{3}{2}B\rceil,\ A + C < 2B$	$\min\left\{\begin{matrix} 3B - 2A - 1 \\ \lvert B - 2C - 1\rvert \end{matrix}\right\}$	C
$B + \lceil\frac{C}{2}\rceil < A < \lceil\frac{3}{2}B\rceil,\ 2B < A + C$	$3B - 2A - 1$	C
$A < B + \lceil\frac{C}{2}\rceil,\ \lceil\frac{3C}{2}\rceil < B < 2C$	$2C - B + 1$	$2B + C - 2A - 1$
$A < B + \lceil\frac{C}{2}\rceil,\ 2C < B$	$B - 2C - 1$	0
$A < B + \lceil\frac{C}{2}\rceil < 2C,\ A + C < 2B$	$2C - B + 1$	0
$A < B + \lceil\frac{C}{2}\rceil < 2C,\ 2B < A + C$	$B - C - 1$	0

It would now be interesting to see how the outcomes of the observed games would change if exactly one player changed from the strategy he or she had used to our suggested strategy. We applied this test to each of the players A, B, and C in the 121 games in which $A < 2B$. As it turned out, no changes in wins occurred in games in the category $\lceil\frac{3}{2}B\rceil < A < 2B$ since, in this category, the players usually followed our strategy. On the other hand, for the 85 games in the category $A < \lceil\frac{3}{2}B\rceil$, the number of wins increased for each player, especially players B and C, who often bet in a way quite different from our practical strategy. The changes in wins for $A < \lceil\frac{3}{2}B\rceil$ are shown in Table 3.

It is important to note that we have not considered how the games should be played if all of the contestants are aware of the right-wrong distribution and knowledgeable in game theory. This will be addressed in the next two sections.

TABLE 3. Changes in Wins when $A < \lceil \frac{3}{2}B \rceil$

	Original Data		Using Practical Strategy	
	Wins	Ties	Wins	Ties
Player A	51	1	53	0
Player B	28	1	37	0
Player C	5	0	13	0

4. Two-player Final *Jeopardy!*

In this section, we examine Final *Jeopardy!* from a game theoretic point of view. For simplicity, we first consider the situation in which there are only two contestants. The two-player case occasionally arises because, if one of the contestants fails to have a positive score at the beginning of Final *Jeopardy!*, only the two remaining players go on to play Final *Jeopardy!*

As before, we call the contestants in first and second places player A and player B. Let P_{RW} be the probability that player A answers correctly and player B incorrectly, and similarly define P_{RR}, P_{WR}, and P_{WW}. We assume that both players know the probabilities and that none of the probabilities is zero.

Recall that we have assumed that the objectives of a player in Final *Jeopardy!* are to maximize the probability of a win or tie and to minimize the likelihood of a tie if it does not significantly increase the risk of losing. We now quantify these objectives. The outcome of a game from player A's perspective is assigned a value of 0 if player A loses, 1 if player A wins, and $1 - \varepsilon$, where ε is a small positive number, if player A ties. Similarly, we assign a value to the outcome of a game from player B's perspective.

The game of Final *Jeopardy!* for two players can be described by a matrix whose rows are indexed by player A's possible wagers (0 to A) and whose columns are indexed by player B's possible wagers (0 to B). The entry corresponding to player A wagering i and player B wagering j is denoted $a_{ij} \backslash b_{ij}$. Here, a_{ij} is the expectation of the outcome with respect to player A in the situation that player A wagers i and player B wagers j, in other words, the probability that player A wins plus $(1 - \varepsilon)$ times the probability that players A and B tie. The number b_{ij} is the corresponding value for player B. We remind the reader that neither player wins if both players have final scores of 0. In game theory, such a matrix is called a *payoff matrix*, and the values a_{ij} and b_{ij} are called the *payoffs* to player A and player B. According to the objectives of the game, both players want to maximize their payoffs. Note that, in our case, the values a_{ij} and b_{ij} do not involve the actual monetary payoff. This is consistent with our assumption that the main objective is to win or tie without regard to monetary payoff.

We note here that if $\varepsilon = \frac{1}{2}$, then for every i and j the values of a_{ij} and b_{ij} sum to 1. Such a game is called a *constant sum game*. Although such games are easier to analyze, we assume that $\varepsilon < \frac{1}{2}$, since a tie seems only marginally less desirable than winning in *Jeopardy!*

We first use the payoff matrix representation of the game to analyze the case $\lceil \frac{3}{2}B \rceil < A < 2B$. In presenting the payoff matrix, we abbreviate $(1 - \varepsilon)P_{RR}$ by P^*_{RR}, and similarly use P^*_{RW}, P^*_{WR}, and P^*_{WW}. The payoff matrix is presented in block form in FIGURE 1.

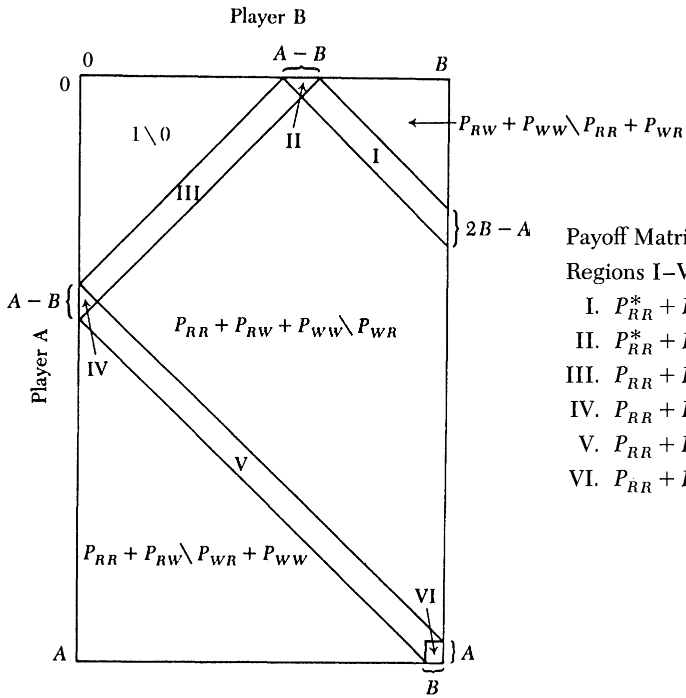


FIGURE 1
Payoff Matrix when $\lceil \frac{3}{2}B \rceil < A < 2B$.
Regions I–VI have the following entries:

I. $P_{RR}^* + P_{RW} + P_{WW} \backslash P_{RR}^* + P_{WR}$
II. $P_{RR}^* + P_{RW} + P_{WR}^* + P_{WW} \backslash P_{RR}^* + P_{WR}^*$
III. $P_{RR} + P_{RW} + P_{WR}^* + P_{WW} \backslash P_{WR}^*$
IV. $P_{RR} + P_{RW} + P_{WR}^* + P_{WW}^* \backslash P_{WR}^* + P_{WW}^*$
V. $P_{RR} + P_{RW} + P_{WW}^* \backslash P_{WR} + P_{WW}^*$
VI. $P_{RR} + P_{RW} \backslash P_{WR}$

Suppose that both players are rational players. First notice that for $i > 2B - A + 1$ and for all j,

$$a_{2B-A+1,j} \geq a_{ij},$$

and the inequality is strict for some j, including $j = 2A - 3B - 1$. We can say that the strategy of player A betting $2B - A + 1$ *dominates* any strategy of player A betting i with $i > 2B - A + 1$. Therefore, a rational player A would choose to bet $i \leq 2B - A + 1$. Now, player B is also a rational player and will anticipate player A betting $i \leq 2B - A + 1$. Hence, player B need only consider the *reduced* payoff matrix which has all of the rows with $i > 2B - A + 1$ deleted. For $j < B$ and $i \leq 2B - A + 1$,

$$b_{iB} \geq b_{ij},$$

and the inequality is strict for some i, including $i = 2B - A$. It follows that in the reduced payoff matrix, player B's strategy of betting B dominates any other strategy. Therefore, player B should bet exactly B. Player A, being rational, will know that player B will wager exactly B, and will therefore be left with the optimal strategy of betting exactly $2B - A + 1$. We have reached what is called an *equilibrium point*, where player A wagers $2B - A + 1$ and player B wagers B. An equilibrium point is a pair of strategies where neither player can improve his or her payoff by changing strategy unilaterally. Changing strategies may give the same payoff. In this sense, an equilibrium point is stable. A game may have one, several, or no equilibrium points. In fact, in the game of Final *Jeopardy!* with $\lceil \frac{3}{2}B \rceil < A < 2B$, there are many equilibrium points. For most values of P_{RR}, P_{RW}, P_{WR}, P_{WW}, and ε, the set of equilibrium points is any wager by player A between $2B - A + 1$ and $A - B - 1$, inclusive, paired with a wager by player B between $A - B + 1$ and B, inclusive. Although the values of a_{ij} and b_{ij} are then the same for all of the equilibrium points

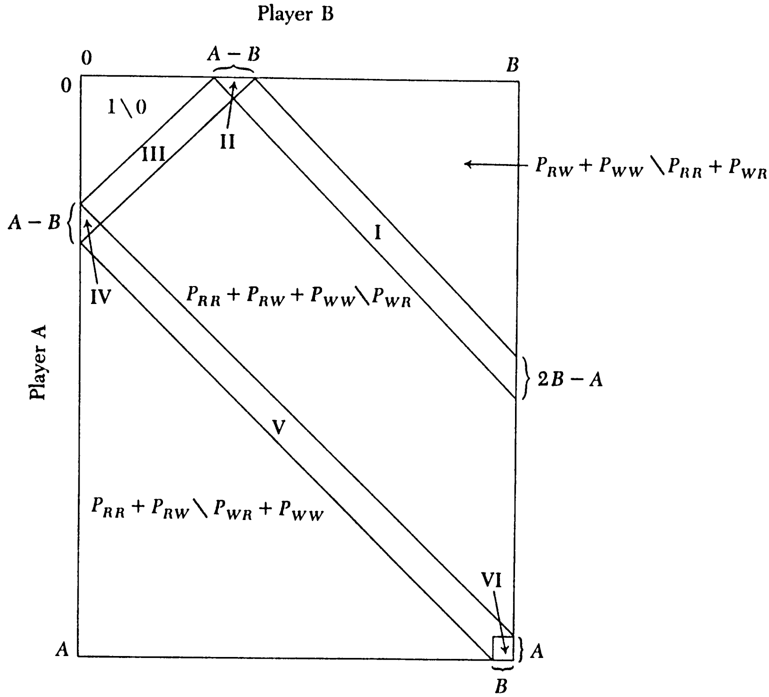


FIGURE 2

Payoff Matrix when $A < \lceil \frac{3}{2}B \rceil$.

Regions I–VI have the following entries:

I. $P^*_{RR} + P_{RW} + P_{WW} \backslash P^*_{RR} + P_{WR}$
II. $P^*_{RR} + P_{RW} + P^*_{WR} + P_{WW} \backslash P^*_{RR} + P^*_{WR}$
III. $P_{RR} + P_{RW} + P^*_{WR} + P_{WW} \backslash P^*_{WR}$
IV. $P_{RR} + P_{RW} + P^*_{WR} + P^*_{WW} \backslash P^*_{WR} + P^*_{WW}$
V. $P_{RR} + P_{RW} + P^*_{WW} \backslash P_{WR} + P^*_{WW}$
VI. $P_{RR} + P_{RW} \backslash P_{WR}$

of our game, there exist games with equilibrium points that do not have the same payoff values. It is also important to note that not all equilibrium points can be found by using dominating strategies. However, unless an equilibrium point can be found by use of dominating strategies, there is not necessarily any reason to view it as the natural outcome of the game.

We now turn to the case $A < \lceil \frac{3}{2}B \rceil$. The payoff matrix in this case is essentially the same as the matrix in FIGURE 1 with only the shape altered. In particular, the row $2B - A$ is now below the row $A - B$. Using the idea of dominance, player A wagers $i \leq 2B - A + 1$, but player B may or may not be able to eliminate some strategic possibilities, depending on the values of A, B and the right-wrong probabilities. We can show, however, that there are no equilibrium points in the case $A < \lceil \frac{3}{2}B \rceil$. To see this, we begin with the easily verified assertion that player A cannot win or tie a WR game at an equilibrium point. Therefore, since player A can win all RR, RW, and WW games by wagering exactly the same amount as player B, it follows that at any equilibrium point $a_{ij} \backslash b_{ij}$, player A must win all such games. Otherwise, player A could benefit by moving to the strategy of wagering j. On the other hand, suppose we are at a point $a_{ij} \backslash b_{ij}$ where player A wins all of the RR, RW, and WW games. Then, if $A_R \leq 2B$ at this point, player B can benefit by switching to the strategy of betting everything, since by doing so player B would win or tie all of the RR games and win all of the WR games. If $A_R > 2B$ at the point $a_{ij} \backslash b_{ij}$, then player B can benefit by switching to the strategy of betting nothing and therefore winning all of the WR and WW games. Hence the point $a_{ij} \backslash b_{ij}$ cannot be an equilibrium point. Therefore, we can conclude that in the case $A < \lceil \frac{3}{2}B \rceil$, there are no equilibrium points.

Since there are no equilibrium points when $A < \lceil \frac{3}{2}B \rceil$, we will take a somewhat different viewpoint in analyzing the game. If, instead of a single game, we view each game as one of a series, the betting strategies might be analogous to those used in poker, where players sometimes bluff in order to create uncertainty as to when they have good hands. If a player follows a single betting strategy repeatedly in a series of Final *Jeopardy!* games, then an opponent could react to that strategy. We will examine what happens if each player uses a *mixed strategy*, where each possible wager is made with a fixed probability. Each player will seek a mixed strategy that will result in the maximum expected payoff.

Let $\alpha = (\alpha_0, \dots, \alpha_A)$ and $\beta = (\beta_0, \dots, \beta_B)$ be the mixed strategies used by players A and B respectively, where α_i is the probability the player A wagers i and β_j is the probability that player B wagers j. When player A and player B use the mixed strategies α and β, the expected payoff to player A is given by

$$E_A(\alpha, \beta) = \sum_{i,j} \alpha_i \beta_j a_{ij},$$

and the expected payoff to player B is given by

$$E_B(\alpha, \beta) = \sum_{i,j} \alpha_i \beta_j b_{ij}.$$

A strategy pair (α, β) is an equilibrium if neither player can improve his or her expected payoff by a unilateral change in distribution; in other words, if

$$E_A(x, \beta) \leq E_A(\alpha, \beta) \quad \text{and} \quad E_B(\alpha, y) \leq E_B(\alpha, \beta)$$

for any probability distributions $x = (x_0, \dots, x_A)$ and $y = (y_0, \dots, y_B)$. This is equivalent to the condition that neither player can benefit by switching unilaterally to a single unmixed strategy. It follows from a result of J. F. Nash [4, Theorem 1] that an equilibrium exists for all games that can be described by a payoff matrix.

To illustrate the ideas given above, we look at a simple example. Suppose $A = 4$, $B = 3$, $P_{RR} = .4$, $P_{RW} = .2$, $P_{WR} = .1$, $P_{WW} = .3$, and $\varepsilon = .1$. In this case, there are three mixed-strategy equilibria, listed in Table 4.

TABLE 4. Mixed Strategy Equilibria and Expected Payoffs

	1	2	3
α_0	.1397	0	.0484
α_1	.2112	.5543	0
α_2	.2648	.1558	.7258
α_3	.3843	.2899	.2258
α_4	0	0	0
β_0	.4426	.3298	.4312
β_1	.0590	0	.0734
β_2	.0574	.3830	0
β_3	.4411	.2872	.4954
$E_A(\alpha, \beta)$	.7478	.7896	.7486
$E_B(\alpha, \beta)$	.3357	.3778	.3806

We see that both player A and player B would prefer the second or third equilibrium to the first since the values of $E_A(\alpha, \beta)$ and $E_B(\alpha, \beta)$ are higher for both players in the second and third equilibria. One might be tempted to conclude that the second and third equilibria are optimal in some sense, but we have to be careful here. As a matter of fact, there exist several other pairs of mixed strategy distributions that offer higher expected payoffs for both players. An example of such a pair of distributions is $\alpha = (1, 0, 0, 0, 0)$ and $\beta = (0, 1, 0, 0)$ (where player A always wagers zero and player B always wagers 1). In this case, $E_A(\alpha, \beta) = .95$ and $E_B(\alpha, \beta) = .45$, so we see that both players are better off in this situation than at any of the equilibria. In fact $\alpha = (1, 0, 0, 0, 0)$ and $\beta = (0, 1, 0, 0)$ is the pair of distributions that maximizes the sum $E_A(\alpha, \beta) + E_B(\alpha, \beta)$. It follows immediately that there is no other pair of distributions where one player can do better and the other player can do at least as well. A pair of distributions with this last property is said to be *Pareto-optimal*. This idea occurs frequently in economics [2, pp. 21–22]. It is important to note that although the pair (α, β) is optimal in this sense, it is not an equilibrium.

We are now faced with the dilemma of having three mixed-strategy equilibria, which are stable in the sense that neither of the players can benefit by changing his or her strategy unilaterally, but we know that there exist other mixed-strategy pairs that offer better expected payoffs for both players. This situation is similar to the famous *Prisoners' Dilemma* problem. (See, for instance, [5, pp. 99–102].) In Final *Jeopardy!* such a dilemma is more striking in the case $\lceil \frac{3}{2}B \rceil < A < 2B$, where we obtained one of the equilibrium points by use of dominating strategies. It turns out that for most values of P_{RR}, P_{RW}, P_{WR}, P_{WW}, and ε, these equilibrium points are not optimal, in the sense that there exist other unmixed-strategy pairs that have better payoffs for both player A and player B. However, since no negotiation is allowed, we should really think of the equilibrium point that we found using dominating strategies as the solution to our game.

We mention another use of the idea of mixed strategies at this point. Recall that even when dominating strategies cannot be used to arrive at a single equilibrium point, it is often possible to reduce the game somewhat. We can sometimes reduce the game still further if a mixed strategy dominates an unmixed strategy [3]. Because it does not aid our study, we will not pursue this idea.

Although we reserve our analysis of the three-player situation for the next section, we remark here that the ideas of this section could be applied to the particular three-player game where player A is assumed to wager exactly $2B - A + 1$ and miss. This game can be analyzed as a two-player game with players B and C. The payoff matrix is similar to the one we have studied in this section, except that since the players must have a final score of at least $A_* = 2A - 2B - 1$ in order to win or tie, some of the winning possibilities are eliminated from the matrix.

We conclude this section with a question for the reader to think about: Suppose you are one of two rational players tied with \$5000 in two-player Final *Jeopardy!* What would you wager?

5. Three-player Final *Jeopardy!*

The game of three-player Final *Jeopardy!* can be described by a three-dimensional payoff matrix whose entries are of the form $a_{ijk} \backslash b_{ijk} \backslash c_{ijk}$, where a_{ijk} is the expectation of the outcome with respect to player A in the situation that player A wagers i, player B wagers j, and player C wagers k. The numbers b_{ijk} and c_{ijk} are the corresponding values for player B and player C, respectively. Define the probabilities P_{RRR}, P_{RRW}, P_{RWR}, P_{RWW}, P_{WRR}, P_{WRW}, P_{WWR}, and P_{WWW} in a manner similar to the previous section. The empirical distribution from Table 1 has $P_{RRR} = .1941$, and so forth. As in the two-player case, we are interested in finding equilibrium points, especially those that arise through dominance, since they can be considered natural outcomes of games where all three players are rational.

For simplicity, we search only for equilibrium points that do not involve the possibility of a tie between two or more players. It is easy to show that if A, B, C, $2B$, and $2C$ are distinct, and if A is not equal to any of the values $\lceil \frac{B}{2} \rceil + C$, $B + \frac{C}{2}$, or $B + C$, then an equilibrium point cannot involve a tie.

In searching for equilibrium points, we could begin by using the idea of dominating strategies as in the two-player case. However, in the three-player case, we found that rather than starting our analysis by looking for dominating strategies, it was easier to find equilibrium points by determining necessary and sufficient conditions for an entry to be an equilibrium point.

Given a fixed right-wrong distribution, as in Table 1 for example, where each of the right-wrong categories is assumed to have nonzero probability, the values in the payoff matrix entry $a_{ijk} \backslash b_{ijk} \backslash c_{ijk}$ are determined by the winner or winners in each of the eight right-wrong categories of a game where players A, B, and C wager i, j, and k, respectively. We call the corresponding eight-tuple of winners the *right-wrong outcome* of the point.

There are 17 possible right-wrong outcomes that do not involve ties. It can be shown that only six of these outcomes, listed in Table 5, can occur at an equilibrium point. The argument is similar to the one used to show that there are no equilibrium points in the two-player game when $A < \lceil \frac{3}{2} B \rceil$.

TABLE 5. Right-Wrong Outcomes of Equilibrium Points

Outcome	RRR	RRW	RWR	RWW	WRR	WRW	WWR	WWW
1	A	A	A	A	A	A	A	A
2	A	A	A	A	B	B	A	A
3	A	A	A	A	B	B	C	A
4	B	B	A	A	B	B	A	A
5	A	A	A	A	C	B	C	B
6	B	B	A	A	B	B	C	A

Under certain conditions, each of the six outcomes in Table 5 may correspond to an equilibrium point. First, there are certain inequalities the values A, B, and C must satisfy, which we refer to as the score conditions. They ensure that no player can gain a win or tie in any of the eight right-wrong categories by unilaterally switching strategy without a collateral loss. In some cases, there are additional conditions on the right-wrong distribution or the value of ε, which we call *right-wrong conditions*. They ensure that no player can make a net gain in his or her payoff. Unfortunately, the right-wrong conditions are rather complex, and in one instance involve 25 subcases, so we will not give them here. In Table 6, we have listed the score conditions corresponding to each outcome and noted whether or not the outcome has right-wrong conditions.

Using the empirical right-wrong distributions given in Table 1, equilibrium points without potential ties exist if, and only if, $\lceil \frac{3}{2} B \rceil < A$ and $A \neq 2B$, and the equilibrium points have outcomes 1, 2, or 3. However, if the level of difficulty of the questions or the ability of the players is changed, then it is conceivable that some of the other right-wrong conditions would be satisfied and that we would have more games with equilibrium points. Only the conditions of outcome 6 are unimaginable in any real game situation. Only one type of outcome can occur at the equilibrium points of a

TABLE 6. Score Conditions

Outcome	Score Conditions	Right-Wrong Conditions
1	$2B < A$	no
2	$\lceil \frac{3}{2}B \rceil < A \leq 2B,\ B + C < A$	only if $A = 2B$
3	$A \leq B + C,\ B < \frac{A}{2} + \frac{C}{2},\ B + \frac{C}{2} < A$	sometimes
4	$2C < A \leq B + C$	yes
5	$A \leq \frac{B}{2} + C$	yes
6	$\frac{3}{2}C < A \leq \frac{2}{3}B + \frac{2}{3}C$	yes

given game, unless the unlikely condition $P_{RRR} + P_{RRW} = P_{WWR}$ is satisfied, in which case outcomes 3 and 4 may occur in the same game.

When the outcomes do turn out to correspond to equilibrium points, dominance can be used to reduce the payoff matrix to subgame where all remaining payoff entries are identical for outcome 1 and nearly always for outcomes 2, 3, and 4. In games where outcomes 5 or 6 correspond to equilibrium points, it seems unlikely that the payoff matrix can be reduced in this way, although we have only verified this for specific examples.

In the games without equilibrium points, we can look for mixed strategies resulting in an equilibrium as in the two-player case. Nash's Theorem guarantees the existence of an equilibrium point. However, while the two-player case involves only linear systems of equations, the equations in the three-player case are nonlinear. They are usually impossible to solve exactly and much more difficult to solve numerically.

6. Concluding remarks

Unless this article reaches a much wider audience than we expect, a *Jeopardy!* contestant would do significantly better than a typical player by following our suggested practical strategy. If the game is played by three rational players, then clear-cut optimal strategies do not always exist. However, even when optimal strategies cannot be found, game-theoretic analysis can be used to shed some light on the situation.

The reader who is interested in learning more about game theory may want to look at the references, which include the foundational work in game theory, **[6]**.

REFERENCES

1. S. J. Brams and D. M. Kilgour, *Game Theory and National Security*, Blackwell, New York, 1988.
2. J. W. Friedman, *Oligopoly and the Theory of Games*, North-Holland, Amsterdam, 1977.
3. A. J. Jones, *Game Theory: Mathematical Models of Conflict*, Ellis Horwood Limited, Chichester, England, 1980.
4. J. F. Nash, Non-cooperative games, *Ann. of Math.* (2) 54 (1951), 286–295.
5. E. Packel, *The Mathematics of Games and Gambling*, MAA, Washington, DC, 1981.
6. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 3rd edition, Princeton University Press, Princeton, NJ, 1953.
7. N. N. Vorob'ev, *Game Theory*, Springer-Verlag New York, 1977.
8. J. Wang, *The Theory of Games*, Oxford University Press, New York, 1988.

NOTES

People Who Know People

MICHAEL O. ALBERTSON
Smith College
Northampton, MA 01063

Two classic problems in a first combinatorics course [**2**, **4**] are:

A) *In every party with six or more people there are either three who are mutual acquaintances or three who are total strangers*; *and*
B) *In every party with two or more people there are two people who have the same number of acquaintances.*

It is entirely standard to rephrase these problems in the language of graphs. To that end construct a graph G whose vertex set $V(G)$ corresponds with the people at the party. Two vertices in G are joined by an edge if the corresponding people are acquainted. Three mutual acquaintances are represented by three vertices each pair of which is joined by an edge—in short a triangle, denoted by K_3. Three total strangers become three vertices, no pair of which is joined by an edge. It is convenient to introduce the complementary graph, denoted by G^c. G^c has the same vertex set as G, but two vertices are joined by an edge in G^c precisely if they are not so joined in G. Thus three total strangers form a triangle in G^c. The number of acquaintances of a given individual, say x, is counted by the number of edges in G incident with the vertex x. This is called the degree of x and is denoted by $\deg(x)$. These two classic problems then become:

A) *If G is any graph with $|V(G)| \geq 6$, then either G or G^c (or both) contains a triangle*; *and*
B) *In any graph G with $|V(G)| \geq 2$, some pair of vertices have the same degree.*

Both of these problems have elementary solutions that rely on the pigeonhole principle. Problem A is the first not entirely trivial instance of a Ramsey Theorem [**5**]. Phrased as an edge coloring problem and restricted to $|V(G)| = 6$, it appeared on the Putnam Exam in 1953 [**6**]. The party formulation was an elementary problem in the MONTHLY in 1958 [**3**].

The surprising idea that these two might be combined arose during a conversation with Jennifer Gyori, a Smith undergraduate.

THEOREM. *Let G be a graph with $|V(G)| \geq 6$. Then either G or G^c contains a K_3 in which two of the vertices have the same degree.*

In the vernacular, every party with at least six people contains three mutual acquaintances, some pair of which have the same number of acquaintances, or three total strangers, (again) some pair of which have the same number of acquaintances.

REMARKS. 1) This is a best possible result in two senses. First, the 5-cycle (or pentagon graph) contains no triangle and is its own complement. Thus the hypothesis that $|V(G)| \geq 6$ is necessary. Second, FIGURE 1 shows a graph with eight vertices in which only two vertices have the same degree. The vertices are labelled with their degrees. A similar construction works for any number of vertices.

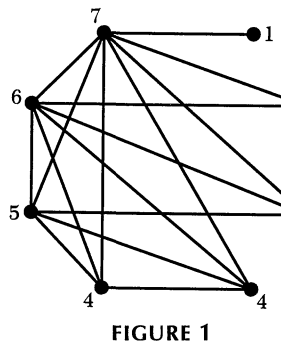


FIGURE 1

2) It is straightforward to check that the theorem is true if $|V(G)| = 6$. (In the classic Ramsey problem, once you know that for every graph on six vertices either it or its complement contains a triangle, then the same automatically holds for larger graphs.) This does not imply the entire result here, since the addition of vertices and edges will possibly alter the degrees.

3) The proof is elementary, but it illustrates the advantage of finding a good place to divide in a divide-and-conquer argument.

Proof. Case 1: G contains three vertices with the same degree. Let H denote the restriction of G to these three vertices. If H contains either no edges or three edges, we have found the triangle in G^c or G (respectively). If H contains two edges, then its complement contains just one edge, so we may assume that H consists of three vertices, say $\{x, y, z\}$ and one edge, say $\{(x, y)\}$.

If there is a vertex in G adjacent to both x and y, then we have found the required triangle in G. Thus a vertex, say w, that is adjacent to x cannot be adjacent to y. Then w must be adjacent to z or else $\{w, y, z\}$ forms a triangle with a repeated degree in G^c. Thus z is adjacent to all of the neighbors of y except x and all of the neighbors of x except y, and these two neighborhoods are disjoint. Consequently,

$$\deg(x) = \deg(z) \geq \deg(x) - 1 + \deg(y) - 1 = 2\deg(x) - 2$$

from which it follows that $\deg(x) \leq 2$. If $\deg(x) = \deg(y) = \deg(z) = 2$, then x, y, z, and their neighbors form a 5-cycle. If $\deg(x) = \deg(z) = 1$, then x, z, and their neighbors form two independent edges. Since we assumed that $|V(G)| \geq 6$, there must, in either event, be a vertex, say t, such that $\{t, x, z\}$ forms a triangle in G^c. What has been shown so far is that if G is any graph with six or more vertices, some three having the same degree, then two of these three are in a triangle in either G or G^c.

Case 2: We assume G is a graph with $V = |V(G)|$ (≥ 6) vertices such that neither G nor G^c contains a triangle in which two vertices have the same degree, and every vertex degree occurs no more than twice. We will show that such a G cannot exist, but for the moment pretend it does. Some of the vertices with repeated degrees might be adjacent in G. Let $\{u_1, v_1\}, \ldots, \{u_r, v_r\}$ denote all pairs of vertices such that $\deg(u_i) = \deg(v_i)$ and u_i is adjacent to v_i. Some of the vertices with repeated degrees might not be adjacent in G. Let $\{a_1, b_1\}, \ldots, \{a_s, b_s\}$ denote all pairs of vertices such that $\deg(a_j) = \deg(b_j)$ and a_j is not adjacent to b_j.

Finally let $w_1, \ldots, w_t$ be all those vertices whose degrees are unique. Since G contains no three vertices with the same degree, we know that there are $r+s+t$ different degrees and that $V = 2r + 2s + t$.

We will now bound the degrees of each of the three types of vertices. The vertex u_i must be adjacent to at least one of a_j or b_j for each j, or else $\{u_i, a_j, b_j\}$ will form a triangle in G^c. In addition u_i will be adjacent to v_i. Thus $\deg(u_i) \geq s + 1$. Since by assumption, u_i and v_i are not in a triangle in G, these two cannot have a common neighbor. Since they are adjacent, we have

$$s + 1 \leq \deg(u_i) \leq 1 + \frac{V-2}{2} = \frac{V}{2}. \tag{1}$$

The vertex a_j will be adjacent to at most one of u_i or v_i for each i, or else $\{a_j, u_i, v_i\}$ will form a triangle in G. Moreover, a_j is not adjacent to b_j. Thus $\deg(a_j) \leq r + 2(s-1) + t$. Since by assumption, a_j and b_j are not in a triangle in G^c, every other vertex in G must be adjacent to at least one of these two. Thus

$$\frac{V-2}{2} \leq \deg(a_j) \leq V - r - 2. \tag{2}$$

Finally consider the vertex w_k. It cannot be adjacent to both u_i and v_i for any i without creating a triangle in G, and it must be adjacent to at least one of a_j and b_j or create a triangle in G^c. Thus

$$s \leq \deg(w_k) \leq V - 1 - r. \tag{3}$$

Combining (1), (2), and (3), we conclude that if z is an arbitrary vertex in G, then

$$\min\left\{s, \frac{V-2}{2}\right\} \leq \deg(z) \leq \max\left\{V - r - 1, \frac{V}{2}\right\}.$$

Subcase A: Suppose $s < (V-2)/2$ and $(V/2) < V - r - 1$. Given these inequalities, we know $s \leq \deg(z) \leq V - r - 1$. The number of different possible degrees equals $(V - r - 1) - s + 1 = r + s + t$. Recall that this is the number of different degrees that G must have. Thus every possible degree must occur. Let w_1 denote the vertex of degree s: It must be one of the w's, since $\deg(u_i) \geq s + 1$ and $\deg(a_j) \geq (V-2)/2$.

Let w_t denote the vertex of degree $V - r - 1$: It must be one of the w's, since $\deg(u_i) \leq (V/2)$ and $\deg(a_j) \leq V - r - 2$. By construction, w_1 is adjacent to exactly one of a_j or b_j for each j and no other vertex. Similarly w_t is adjacent to every vertex in G except one of u_i or v_i for each i. From w_1's perspective w_1 is not adjacent to w_t, while from w_t's perspective w_t is adjacent to w_1. This contradiction shows that our imagined G cannot exist.

Subcase B: Suppose $s \geq (V-2)/2$ or $(V/2) \geq V - r - 1$. Since $V = 2r + 2s + t$, the first inequality simplifies to $2r + t \leq 2$ while the second simplifies to $2s + t \leq 2$. If both hold, then

$$V = 2r + 2s + t \leq (2r + t) + (2s + t) \leq 4,$$

contrary to our original hypotheses. Thus, just one of these inequalities holds. Moreover, these inequalities are complementary in the sense that one holds for a graph G if, and only if, the other holds for G^c. Thus we may assume that $s \geq (V-2)/2$. Since this is equivalent to $2r + t \leq 2$, either $r = 1$ and $t = 0$ or $r = 0$ and $t \leq 2$. We will dispose of these extreme possibilities one at a time.

If $r = 1$, then there is exactly one pair of adjacent vertices of the same degree. By

(2), the number of possible degrees for the vertices a_j equals $(V/2)-1$. Since $s \geq (V-2)/2$, every possible degree must actually occur. By (1), $\deg(u_1) = s+1$. Thus for some j, $\deg(a_j) = \deg(u_i)$, contradicting our original assumption that the degrees of the a_j's and u_i's were all distinct.

If $r = 0$, then by (2), the number of possible degrees for the a_j's equals $V - 2 - (V-2)/2 + 1 = s + (t/2)$. If $t = 0$, then we may assume that both a_1 and b_1 have degree $s-1$, while both a_s and b_s have degree $V-2$. This cannot occur because both a_s and b_s would have to be adjacent to both a_1 and b_1, while a_1 and b_1 would be adjacent to at most one of a_s and b_s. If $t = 1$, then we may assume that a_1 and b_1 have degree $s-1$, while $\deg(w_1) = V-1$. This cannot occur since if it did, w_1 would be adjacent to every vertex in G including both a_1 and b_1, but a_1 is assumed to be adjacent to just one of a_j and b_j for $j > 1$ and nothing else. Finally, if $t = 2$, we may assume that there exists z such that $\deg(z) = s-1$ and that $\deg(w_2) = V-1$. The vertex z cannot be w_1 by (3). If the vertex z were one of the a_j's, it could have degree $s-1$ only by not being adjacent to w_2, contradicting the assumption that $\deg(w_2) = V-1$.

Possible generalizations. The development of Ramsey graph theory suggests the possibility that if a graph is large enough either it or its complement ought to contain a clique of any fixed size with a repeated degree. In fact this does not occur. Specifically, there exist arbitrarily large graphs in which no two vertices of the same degree are adjacent and whose complements contain no K_4 [1].

REFERENCES

1. Michael O. Albertson and David M. Berman, Ramsey graphs without repeated degrees, *Congressus Numerantium*, 83 (1991) pp 91–96.
2. Michael O. Albertson and Joan P. Hutchinson, *Discrete Mathematics with Algorithms*, John Wiley & Sons, Inc., New York, 1988.
3. C. W. Bostwick, Problem E 1321, *Amer. Math. Monthly*, 65 (1958), p. 446: Solution, ibid. 66 (1959), pp. 141–142.
4. Richard A. Brualdi, *Introductory Combinatorics*, North Holland, New York, 1977, page 23, problems 8 & 10.
5. R. Rothschild, B. Graham, and J. Spencer, *Ramsey Theory, 2nd Edition*, Wiley Interscience, New York, 1990.
6. A. M. Gleason, R. E. Greenwood, and L. M. Kelly, *The William Lowell Putnam Mathematical Competition—Problems and Solutions*: 1938–1964. MAA, Washington, DC, 1980.

Proof by Game:

Flip a fair coin. Advance 2 for heads, 1 for tails.

Start		Go back 2 spaces	End

Summing the probabilities of reaching End in exactly n turns, $n \geqslant 2$, yields

$$\frac{1}{4} + \frac{1}{8} + \frac{2}{16} + \frac{3}{32} + \frac{5}{64} + \frac{8}{128} + \frac{13}{256} + \cdots = 1$$

—Kay P. Litchfield
GTE Government Systems Corp.
(student) University of Utah

(2), the number of possible degrees for the vertices a_j equals $(V/2)-1$. Since $s \geq (V-2)/2$, every possible degree must actually occur. By (1), $\deg(u_1)=s+1$. Thus for some j, $\deg(a_j)=\deg(u_i)$, contradicting our original assumption that the degrees of the a_j's and u_i's were all distinct.

If $r=0$, then by (2), the number of possible degrees for the a_j's equals $V-2-(V-2)/2+1=s+(t/2)$. If $t=0$, then we may assume that both a_1 and b_1 have degree $s-1$, while both a_s and b_s have degree $V-2$. This cannot occur because both a_s and b_s would have to be adjacent to both a_1 and b_1, while a_1 and b_1 would be adjacent to at most one of a_s and b_s. If $t=1$, then we may assume that a_1 and b_1 have degree $s-1$, while $\deg(w_1)=V-1$. This cannot occur since if it did, w_1 would be adjacent to every vertex in G including both a_1 and b_1, but a_1 is assumed to be adjacent to just one of a_j and b_j for $j>1$ and nothing else. Finally, if $t=2$, we may assume that there exists z such that $\deg(z)=s-1$ and that $\deg(w_2)=V-1$. The vertex z cannot be w_1 by (3). If the vertex z were one of the a_j's, it could have degree $s-1$ only by not being adjacent to w_2, contradicting the assumption that $\deg(w_2)=V-1$.

Possible generalizations. The development of Ramsey graph theory suggests the possibility that if a graph is large enough either it or its complement ought to contain a clique of any fixed size with a repeated degree. In fact this does not occur. Specifically, there exist arbitrarily large graphs in which no two vertices of the same degree are adjacent and whose complements contain no K_4 [1].

REFERENCES

1. Michael O. Albertson and David M. Berman, Ramsey graphs without repeated degrees, *Congressus Numerantium*, 83 (1991) pp 91–96.
2. Michael O. Albertson and Joan P. Hutchinson, *Discrete Mathematics with Algorithms*, John Wiley & Sons, Inc., New York, 1988.
3. C. W. Bostwick, Problem E 1321, *Amer. Math. Monthly*, 65 (1958), p. 446: Solution, ibid. 66 (1959), pp. 141–142.
4. Richard A. Brualdi, *Introductory Combinatorics*, North Holland, New York, 1977, page 23, problems 8 & 10.
5. R. Rothschild, B. Graham, and J. Spencer, *Ramsey Theory, 2nd Edition*, Wiley Interscience, New York, 1990.
6. A. M. Gleason, R. E. Greenwood, and L. M. Kelly, *The William Lowell Putnam Mathematical Competition—Problems and Solutions*: 1938–1964. MAA, Washington, DC, 1980.

Proof by Game:

Flip a fair coin. Advance 2 for heads, 1 for tails.

Start		Go back 2 spaces	End

Summing the probabilities of reaching End in exactly n turns, $n \geqslant 2$, yields

$$\frac{1}{4}+\frac{1}{8}+\frac{2}{16}+\frac{3}{32}+\frac{5}{64}+\frac{8}{128}+\frac{13}{256}+\cdots=1$$

—Kay P. Litchfield
GTE Government Systems Corp.
(student) University of Utah

The Importance of a Game

MARK F. SCHILLING
California State University
Northridge, CA 91330

Organized play for many of the major sports that are popular around the world (baseball, basketball, hockey, tennis, and volleyball, for example) involves a *contest* between two opponents in which the victor is the team or individual that is the first to win a prescribed number of subcontests. Often during these contests one will hear in conversation or on a broadcast a statement such as "This game is the most important game of the series". Why are some games perceived as being more important than others, when each game counts the same toward the total number of victories required to win the contest?

The purpose of this article to analyze the meaning and validity of such statements from a probabilistic perspective and to provide a means of quantifying importance, in order that the comparative significance of games can be measured. In this paper, the word *game* will refer to the subcontests that comprise a contest, and one participant in the contest will be arbitrarily denoted "the contestant", or "C". The key observation is that the importance of a game depends on context. Two distinct versions of importance, developed below, are therefore appropriate. It will be assumed that the outcome of each game in a contest is independent of all other outcomes. Participants in such contests frequently dispute the reasonableness of this supposition, citing the effects of momentum and other factors, primarily psychological; however, evidence from statistical analysis of contests [2] suggests that actual experience is quite compatible with the independence assumption.

Conditional and a priori importance To fix notation, let g be the number of a particular game in the sequence of games comprising a contest, and suppose that contestant C has won k of the previous $g-1$ games.

Definition 1. The *conditional importance* $I(g|k)$ of game g is the difference between (1) the conditional probability that C wins the contest if C *wins* the game and (2) the conditional probability that C wins the contest if C *loses* the game, given the status of the contest up to game g.

Note that $I(g|k)$ is always nonnegative and is the same for both contestants. Here is an example: Suppose two teams are playing a series in which the first team to win four games (commonly known as a "best of seven" series) emerges victorious. Assume that four games have been played so far, with each team having earned two wins. Then the conditional importance of the fifth game is

$$\begin{aligned} I(5|2) &= \mathrm{P}(\text{C wins series}|\text{C wins game 5}) - \mathrm{P}(\text{C wins series}|\text{C loses game 5}) \\ &= \mathrm{P}(\text{C wins series}|\text{C leads 3 games to 2}) \\ &\quad - \mathrm{P}(\text{C wins series}|\text{C trails 2 games to 3}) \\ &= \mathrm{P}(\text{C wins at least 1 of the 2 remaining games}) \\ &\quad - \mathrm{P}(\text{C wins both remaining games}). \end{aligned}$$

If each team is equally likely to win each game independently of the outcomes of any of the previous games, then the conditional importance of game 5 is

$$I(5|2) = 3/4 - 1/4 = 1/2.$$

That is, since the contestant's chances of winning the series will be 75% if game 5 is won and 25% if it is lost, the conditional importance of game 5 is 75% − 25% = 50%.

If the final game of a contest with a limited number of games (such as the seventh game of a "best of seven" series), is played, it clearly has the highest possible conditional importance, namely 1 − 0 = 100%. No other game can have a conditional importance as high as this, provided each team has a positive probability of winning each game.

Why then, are other games ever considered the "most important game"? The fifth game in a "best of seven" series—the example shown above—is often described this way, as is the seventh game in a set in tennis. And can it really be true that one game is any more or less important than any other?

The answers to these questions depend on the fact that a game will have no importance at all if it is not played. For example, the seventh game of a "best of seven" series is played only if the first six games are split evenly, with three wins for each contestant. Otherwise the conditional importance of game seven is zero—even if the game were played it could not change the winner of the contest. It can therefore be argued that importance should take into account the chance that the game will be necessary, and more generally, the possible states of the contest when that game is played (if at all), as well as the conditional importance of the game given these possible states.

Definition 2. The *a priori importance* of game g is $I(g) = EI(g|k)$ = the *expected* conditional importance of the game.

Conditional importance and a priori importance will clearly coincide for game $g = 1$. We shall see, however, that in general their forms are quite different.

Assume henceforth that to win a contest it is necessary to win a fixed number $n > 1$ of games, with at most $2n - 1$ games necessary in all, although the same ideas can easily be extended to other types of contests. Note that in the example above, the conditional importance of game 5 is simply P(C wins *exactly* 1 of the remaining 2 possible games). In general,

$$\begin{aligned} I(g|k) &= \text{P(C wins at least } n-k-1 \text{ of the remaining possible games)} \\ &\quad - \text{P(C wins at least } n-k \text{ of the remaining possible games)} \\ &= \text{P(C wins exactly } n-k-1 \text{ of the } 2n-1-g \text{ possible games} \\ &\qquad\qquad \text{remaining after game } g). \qquad (1) \end{aligned}$$

Now let $P_g(k)$ be the probability that C wins k of the first $g - 1$ games. Then from Definition 2,

$$\begin{aligned} I(g) &= \sum_{k=0}^{g-1} I(g|k) P_g(k) \\ &= \sum_{k=0}^{g-1} \text{P(C wins } n-k-1 \text{ games after game } g) \times \\ &\qquad \text{P(C wins } k \text{ games before game } g) \\ &= \text{P(C wins } n-1 \text{ of the games other than game } g). \qquad (2) \end{aligned}$$

Note that the event in (2) is that C wins exactly half of the $2n - 2$ games other than game g. This makes sense intuitively since, if the games other than game g result in an equal number of victories for each participant, then game g represents a "decisive game". With this interpretation, any contest will turn out to have either n

decisive games or *no* decisive games, depending on whether the contest does or does not extend to its last possible game. Furthermore, using (2) we can represent the a priori importance as

$$I(g) = \mathrm{P}(\mathrm{C} \text{ wins the contest} | \mathrm{C} \text{ wins game } g) - \mathrm{P}(\mathrm{C} \text{ wins the contest} | \mathrm{C} \text{ loses game } g),$$

which is the *unconditional* parallel to conditional importance.

Two simple models are appropriate for many contests of the "majority wins" type described above.

Model 1. The games follow a binomial model in which the win probability $0 < p < 1$ for contestant C is the same for each game.

Model 2. The contest consists of independent games that are of two types, say "type 1" and "type 2", each following a binomial model with win probabilities $0 < p_1$, $p_2 < 1$ respectively.

Model 2 reduces to Model 1 when $p_1 = p_2$. Model 1 is appropriate when the conditions are essentially the same for each game. Model 2 is representative of sporting contests in which the games are played at two sites, normally the home towns of the two contestants, with type 1 and type 2 being the games played at these two sites. This model would also be appropriate in tennis (volleyball) when a set represents a contest, since the win probability for each game (point) often depends heavily on who is serving.

First let us analyze the two importance quantities under Model 1. Note that although a contest will normally terminate as soon as the requisite number of games is won by one of the participants, all analyses can be carried out as if all $2n - 1$ games are actually played, since the winner is the same in either case. Define

$$\binom{u}{v} = \begin{cases} \dfrac{u!}{v!(u-v)!} & \text{for } 0 \le v \le u; \\ 0 & \text{otherwise.} \end{cases}$$

Since at most $2n - 1$ games can be played, equation (1) gives for the conditional importance

$$I(g|k) = \binom{2n-1-g}{n-k-1} p^{n-k-1} q^{n-g+k}, \tag{3}$$

where $q = 1 - p$.

We can compute the a priori importance by making use of the following combinatorial result (see [1], p. 64), which will be referred to as the *hypergeometric identity* because of its direct relationship to the hypergeometric probability distribution:

$$\sum_{k=0}^{m} \binom{a}{k}\binom{b}{m-k} = \binom{a+b}{m}.$$

This identity can be seen to count the number of ways to choose m items from a set of $a + b$ items of which a are of one kind and b are of another.

Now since $P_g(k) = \binom{g-1}{k} p^k q^{g-1-k}$, the a priori importance for Model 1 is

$$I(g) = \sum_{k=0}^{g-1} \binom{2n-1-g}{n-k-1}\binom{g-1}{k} p^{n-1} q^{n-1}$$

$$= \binom{2n-2}{n-1} p^{n-1} q^{n-1}, \tag{4}$$

upon using the hypergeometric identity. The a priori importance is *the same* for each game under Model 1. Note that its value is also the probability that the contest will extend to the last game.

Under Model 2, the conditional and a priori importances are obtained from convoluting the two distinct binomial distributions involved. Write $q_i = 1 - p_i$ and let n_i represent the number of games of type i remaining after game g, for $i = 1, 2$. Then the conditional importance of game g under Model 2 is

$$I(g|k) = \sum_{j=0}^{n_1} \binom{n_1}{j} p_1^j q_1^{n_1-j} \binom{n_2}{n-k-1-j} p_2^{n-k-1-j} q_2^{n_2-n+k+1+j}. \tag{5}$$

The a priori importance is *not* the same for each game under Model 2, since $I(g)$ depends on the total number of possible games t_i of each type, $i = 1, 2$, other than game g, as well as on the win probabilities. Decomposing over $j =$ the number of type 1 games other than game g that are won by C yields for (2)

$$I(g) = \sum_{j=0}^{t_1} \binom{t_1}{j} \binom{t_2}{n-1-j} p_1^j q_1^{t_1-j} p_2^{n-1-j} q_2^{t_2-n+1+j}. \tag{6}$$

This result can also be computed directly from the definition of a priori importance by repeated application of the hypergeometric identity. Looking at (6), we see that since game g is itself one of the two types of games, but is not counted by either t_1 or t_2, all games of type 1 must have equal a priori importance, as must all games of type 2; however these two values normally will not be equal to each other.

Analysis of the importance functions The degree of conditional importance attached to a game depends on the extent to which the ultimate outcome of the contest remains in doubt at that point. To gain insight into this phenomenon, let us examine $I(g|k)$ for Model 1 more closely. Let $N = 2n - 1 - g$, the number of possible games remaining after game g, and write $m = n - k - 1$ for the number of additional wins the contestant C must attain to win the contest if game g is also won. The conditional importance (3) then takes the simple binomial form

$$I(g|k) = \binom{N}{m} p^m q^{N-m}.$$

Using standard properties of the binomial distribution, we find that for given N and p, $I(g|k)$ is a unimodal function on $m = 0, 1, \dots, n$ reaching its maximum value at $m = [(N+1)p]$, where $[x]$ represents the greatest integer less than or equal to x. Similarly, as a function of p with N and m fixed, $I(g|k)$ is maximized at $p = m/N$ (by elementary calculus). From either perspective, conditional importance is greatest when the proportion of the possible remaining games that C must win is at or near to C's win probability for a single game. Comparable behavior occurs under Model 2 unless the values of p_1 and p_2 are very different from each other.

Natural questions to ask about a priori importance are:

(1) When is $I(g)$ small? large?
(2) When are games of one type (e.g., "home" games) more or less important than games of the other type ("away" games)?

The first question is easier to answer than the second: $I(g)$ can be made arbitrarily close to 0—for example, if p_1 and p_2 tend to 1. $I(g)$ can also become arbitrarily close to 1 if, for instance, $t_1 = t_2$ and p_1 tends to 1 while p_2 tends to 0.

With regard to the second question, if $p_1 = p_2$ then all games are equally important. In general, it turns out that the a priori importance of type 1 games can be greater than, lower than, or equal to that of type 2 games, regardless of which type comprises the majority of games in the contest.

Consider those contests in which one more game is scheduled of one type, say type 1, than of the other type. This is the usual setup in sporting contests; for example, the World Series in Major League Baseball has four games in one stadium and three in another, if the Series extends to the maximal number of games. If game g is of type 1, then $t_1 = t_2 = n - 1$, while if game g is of type 2, then $t_1 = n$ and $t_2 = n - 2$. Denote the a priori importances for type 1 and type 2 games by I_1 and I_2, respectively. Although a simple analytical result appears unobtainable, numerical studies indicate an interesting relationship, which is summarized in FIGURES 1, 2, and 3 below. These graphs indicate the sign of $\Delta I = I_1 - I_2$ for "best of three" ($n = 2$), "best of five" ($n = 3$) and "best of seven" ($n = 4$) contests, respectively, as a function of p_1 and p_2. In each case, both positive and negative values of ΔI can occur.

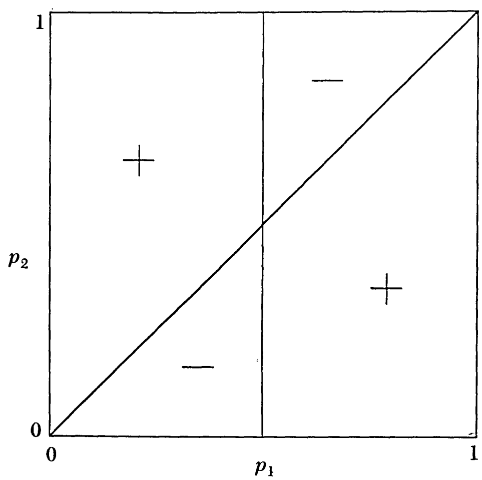


FIGURE 1
Sign of ΔI for "best of three" contests.

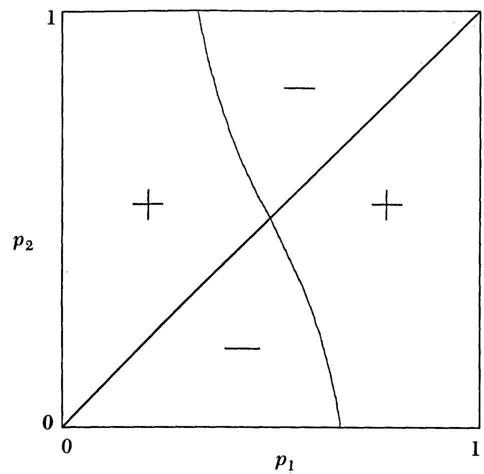


FIGURE 2
Sign of ΔI for "best of five" contests.

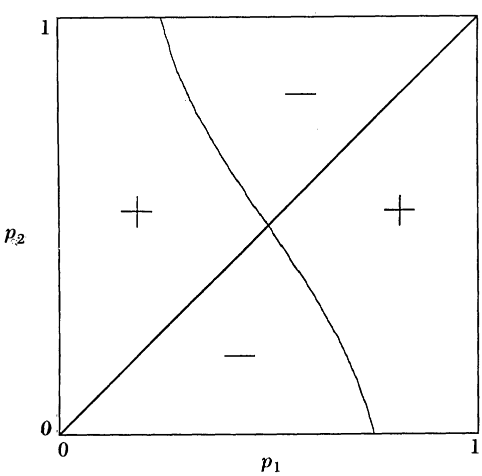


FIGURE 3
Sign of ΔI for "best of seven" contests.

Specifically, suppose that type 1 games are the home games for contestant C and type 2 games are the games played away from home. Invariably this means that p_1 will be greater than p_2, a fact generally known as the "home field (or court) advantage". In this setting, the home games will be *more* important than the away games if $p_1 > p_2 > .5$, that is, if C is favored to win each game in the contest, while the home games will be *less* important than the away games if C is the underdog at both sites. If each participant is favored at its own site, however, the relative a priori importance depends further on p_1 and p_2 as well as on n, with the home games of C being the more important ones if p_1 is sufficiently large compared to p_2.

The following example indicates the reasonableness of the result: Suppose a "best of three" contest is to be played and that each team is heavily favored to win at its home site, with two games to be played at the home site of the contestant C and one at the opponent's home site. According to FIGURE 1, C's home games are the more important ones. This makes sense intuitively because C should win the contest regardless of the outcome of the away game. Thus that game is relatively unimportant, while C's home games are quite important because losing a home game would decrease C's chances of winning the contest from near 1 to near 0.

Application: the 1988 NBA Championship Series To illustrate the concepts of conditional and a priori importance in a real world setting, the values of these quantities are estimated for the 1988 National Basketball Association (NBA) Championship Series, which was played between the Los Angeles Lakers and the Detroit Pistons. The series was a "best of seven" affair, with four games held in Los Angeles and three held in Detroit. Los Angeles will therefore be the contestant C in this application, and games in L.A. will be the type 1 games while type 2 games will represent those in Detroit. It is necessary to estimate p_1 and p_2 in order to estimate the importance values. The following procedure leads to reasonable results:

1. First estimate the *odds* $o(L.A.)$ in favor of Los Angeles defeating Detroit in a single game at a neutral site from the season records of each team as follows: L.A. won 62 games and lost 20 in 1987–88, Detroit won 54 and lost 28, so we estimate $o(L.A.)$ by

$$\hat{o}(L.A.) = \frac{62}{20} \times \frac{28}{54} = 1.607.$$

(This method has been suggested by several authors; see [3] and [4] for details and references.)

2. Data from NBA games indicate that the home team wins approximately twice as often as the away team, so we modify $\hat{o}(L.A.)$ to obtain home odds $\hat{o}_1(L.A.) = 2 \times \hat{o}(L.A.) = 3.215$ and away odds $\hat{o}_2(L.A.) = .5 \times \hat{o}(L.A.) = 0.804$. This leads to the following estimates for p_1 and p_2:

$$\hat{p}_1 = \frac{\hat{o}_1(L.A.)}{\hat{o}_1(L.A.)+1} = .763 \quad \text{and} \quad \hat{p}_2 = \frac{\hat{o}_2(L.A.)}{\hat{o}_2(L.A.)+1} = .446.$$

Table 1 provides the schedule of games, the results, and the values of $I(g)$ and $I(g|k)$ for the series, computed for Model 2 from these estimates of p_1 and p_2. For comparison, conditional importance values are also given for Model 1 with $p = .5$, which would apply if each team were equally likely to win each game.

Observe in Table 1 that the games in Los Angeles had greater a priori importance than those in Detroit, in agreement with FIGURE 3 for the estimates of p_2 and p_2 given above. This can be understood intuitively by noting that although L.A. was favored to win the series, a loss of just one of their home games would have been decisive if each other game were won by the home team. The games in Detroit were less important because, as long as Los Angeles won at home, the games in Detroit could not change the outcome of the series. The oft-stated "importance of holding onto the home-court advantage" is therefore seen to be well justified in this case.

TABLE 1. A Priori and Conditional Importances, 1988 NBA Championship Series

Game	Site	$I(g)$	$I(g\|k)$	$I(g\|k)$ (Model 1)	Winner
1	L.A.	.289	.289	.312	Detroit
2	L.A.	.289	.364	.312	L.A.
3	Detroit	.235	.369	.375	L.A.
4	Detroit	.235	.226	.375	Detroit
5	Detroit	.235	.362	.500	Detroit
6	L.A.	.289	.763	.500	L.A.
7	L.A.	.289	1.000	1.000	L.A.

Since this series went to the maximum number of games, conditional importance naturally increased dramatically toward the end. It is interesting that the fourth game turned out to be the game with the lowest conditional importance. The reason is that Los Angeles was in a very strong position at that time, ahead two games to one—they would have been heavily favored to win the series *regardless* of the outcome of game 4. Note also that the fifth game did not have critical importance (for similar reasons), despite the traditional dogma about such games mentioned earlier.

Concluding remarks There may be at least one practical application of the concepts of a priori and conditional importance. If the audiences for sporting contests tend to sense subjectively the approximate relative importances of the games, then advance calculations could provide a valuable means of predicting audience interest for television and radio networks, ticket sellers, and promoters. Rates for advertising time, for example, might be set in part according to the estimated a priori importance of a contest's games, or based on conditional importance (if deadline considerations permit) according to the progress of a series. Studies of past broadcast ratings and future surveys would be of interest in order to determine whether such connection exists.

The notions presented here have been illustrated only for contests of a specific kind, where the requirement is to win a predetermined number of games before the adversary does. However, the definitions and techniques can be adapted easily to any situation where the ultimate outcome depends probabilistically on the occurrence of a number of component events.

REFERENCES

1. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd. edition, John Wiley & Sons, Inc., New York, 1968.
2. T. Gilovich, R. Vallone and A. Tversky, The hot hand in basketball: On the misperception of random sequences, *Cogn. Psych.* 17 (1985), 295–314.
3. T. Jech, The ranking of incomplete tournaments: a mathematician's guide to popular sports, *Amer. Math. Monthly* 90 (1983), 246–266.
4. M. Stob, A supplement to "A mathematician's guide to popular sports", *Amer. Math. Monthly* 91 (1984), 277–282.

Matrix Representation of Finite Fields

WILLIAM P. WARDLAW*
U.S. Naval Academy
Annapolis, MD 21402

Most undergraduate texts in abstract algebra show how to represent a finite field $\mathbf{F}_q$ over its prime field $\mathbf{F}_p$ by clearly specifying its additive structure as a vector space or a quotient ring of polynomials over $\mathbf{F}_p$ while leaving the multiplicative structure hard to determine, or they explicitly illustrate the cyclic structure of its multiplicative group without clearly connecting it to the additive structure. In this note we suggest a matrix representation that naturally and simply displays both the multiplicative and the additive structures of the field $\mathbf{F}_q$ (with $q = p^d$) over its prime field $\mathbf{F}_p$. Although this representation is known (see [3] p. 65, for example), it does not appear to be widely used in abstract algebra texts.

To illustrate these ideas, let us first consider the field $\mathbf{F}_8$ of eight elements over its prime field $\mathbf{F}_2$. The additive structure of $\mathbf{F}_8$ is that of the three-dimensional vector space $\mathbf{V} = \{(0\ 0\ 0), (1\ 0\ 0), (0\ 1\ 0), (0\ 0\ 1), (1\ 1\ 0), (1\ 0\ 1), (0\ 1\ 1), (1\ 1\ 1)\}$ over $\mathbf{F}_2$. However, it is not at all clear how to define products of these vectors to get the multiplicative structure of $\mathbf{F}_8$! It can be shown that extending the multiplication table

	(1 0 0)	(0 1 0)	(0 0 1)
(1 0 0)	(1 0 0)	(0 1 0)	(0 0 1)
(0 1 0)	(0 1 0)	(0 0 1)	(1 1 0)
(0 0 1)	(0 0 1)	(1 1 0)	(0 1 1)

(1)

for the basis $\mathbf{B} = \{(1\ 0\ 0), (0\ 1\ 0), (0\ 0\ 1)\}$ of $\mathbf{V}$ by bilinearity gives the multiplicative structure of $\mathbf{F}_8$, although a direct proof would be tedious.

A more usual, as well as more useful, treatment (see [1], p. 171 or [3], p. 25, Thm. 1.6.1) is to represent

$$\mathbf{F}_8 \cong \mathbf{F}_2[x]/(x^3 + x + 1) \tag{2}$$

as the ring of all polynomials over $\mathbf{F}_2$ modulo the third-degree irreducible polynomial $x^3 + x + 1$. If we let $a \in \mathbf{F}_8$ denote the residue class of x modulo $x^3 + x + 1$, we have $a^3 + a + 1 = 0$. Recalling that the characteristic is 2, it is then easy to see that $a^3 = a + 1$, $a^4 = a^2 + a$, $a^5 = a^2 + a + 1$, $a^6 = a^2 + 1$, and $a^7 = 1$, so

$$\begin{aligned} \mathbf{F}_8 &= \{0, 1, a, a^2, a^3, a^4, a^5, a^6\} \\ &= \{0, 1, a, a^2, a+1, a^2+a, a^2+a+1, a^2+1\}. \end{aligned} \tag{3}$$

Thus, the multiplicative group $\mathbf{F}_8^* = \langle a \rangle$ of $\mathbf{F}_8$ is simply the cyclic group of order 7 generated by a. The second formulation in (3) makes the additive structure easy to see, although it obscures the multiplicative structure a little. One can use the

*Research supported in part by the U.S. Naval Academy Research Council and by the Naval Research Laboratory, Radar Division, Identification Branch.

abbreviated multiplication table

$$\begin{array}{c|ccc} & 1 & a & a^2 \\ \hline 1 & 1 & a & a^2 \\ a & a & a^2 & a+1 \\ a^2 & a^2 & a+1 & a^2+a \end{array} \tag{4}$$

along with the distributive law to multiply elements of $\mathbf{F}_8$. (Comparing tables (1) and (4) is one fairly easy way to prove that the multiplication given by table (1) satisfies the field axioms.) Alternatively, one can use the relation $a^3+a+1=0$ to multiply the elements given in the second formulation in (3). This is the standard representation of a finite field, and it is reasonably satisfactory. However, the transition from addition to multiplication still leaves something to be desired.

If we pick any element b of the field $\mathbf{F}_8$, left multiplication by b is a linear transformation L_b on the vector space $\mathbf{V}=\mathbf{F}_8$ over $\mathbf{F}_2$. If we choose any basis $\mathbf{B}'$ of $\mathbf{V}=\mathbf{F}_8$ over $\mathbf{F}_2$, we can find the matrix $[L_b]=[L_b]_{\mathbf{B}'}$ of L_b with respect to that basis. If we fix the basis $\mathbf{B}'$ and find the matrix of each element of $\mathbf{F}_8$ in this way, it is clear that the resulting set of matrices form a field isomorphic to $\mathbf{F}_8$! Thus, each choice of basis gives a different matrix representation of $\mathbf{F}_8$.

It appears at first glance that we must have a multiplication table for the field before we can get the matrix representation. But there is a way to get around this difficulty.

Let

$$A=\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

be the companion matrix (see [1], p. 264, [2], pp. 229–230, or [5], p. 201, Dfn. 5.2.16) of the irreducible third-degree polynomial $f(x)=x^3+x+1$ over the field $\mathbf{F}_2$. Then $f(A)=0$, so the powers of A satisfy the relations satisfied by a above; in particular, the matrix A generates the cyclic group $\langle A\rangle$ of order 7 isomorphic to $\mathbf{F}_8^*$, and the ring of matrices

$$\mathbf{F}_2[A]=\{0,I,A,A^2,A^3,A^4,A^5,A^6\}$$

is isomorphic to the field $\mathbf{F}_8$. That was easy, wasn't it?

Indeed, a bit too easy, as we shall see. Consider now the irreducible polynomial $g(x)=x^2+1$ over the three-element field $\mathbf{F}_3$. We see that its companion matrix B has multiplicative order 4:

$$B=\begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix},\ B^2=\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},\ B^3=\begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix},\ B^4=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}=I.$$

Not enough elements for $\mathbf{F}_9$! And the powers of B are not closed under addition. Fortunately, there is a fairly simple cure: Adjoin the matrices 0, $I+B$, $I+B^3$, $B+B^2$, and B^2+B^3 to the set of powers of B to obtain the ring $\mathbf{F}_3[B]$ of matrices generated by B. Since $g(B)=B^2+I=0$, it is clear that the ring $\mathbf{F}_3[B]$ is isomorphic to the field $\mathbf{F}_9$. Thus, B provides a matrix representation $\mathbf{F}_3[B]$ of the nine-element field, and we say that B is a *generator* of the field $\mathbf{F}_9$.

But we would like to have a *cyclic generator* of $\mathbf{F}_9$; that is, a matrix M such that the multiplicative group $\mathbf{F}_9^*$ of $\mathbf{F}_9$ is isomorphic to the cyclic group $\langle M\rangle$ generated by M. This, too, is not terribly difficult. An eight-element cyclic group has exactly $\varphi(8)=4$ generators, none of which is a power of an element of order 4. Thus, the multiplicative group $\mathbf{F}_3[B]^*\cong\mathbf{F}_9^*$ is cyclically generated by any of the four nonzero matrices in $\mathbf{F}_3[B]$ that are not powers of B. The reader can easily verify that the matrix $M = I + B = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$ is a cyclic generator of $\mathbf{F}_9$.

Note that the set $\mathbf{F}_3[B]$ is spanned (over $\mathbf{F}_3$) by the matrices I and B, and also by I and M. That is, $\mathbf{F}_3[B] = L(I, B) = L(I, M)$. If $\mathbf{B}$ and $\mathbf{M}$ are the ordered bases (I, B) and (I, M), respectively, we see that

$$L_B\colon \begin{array}{l} I \mapsto B \;= 0 \cdot I + 1 \cdot B \\ B \mapsto B^2 = 2 \cdot I + 0 \cdot B \end{array} \quad \text{so} \quad [L_B]_{\mathbf{B}} = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix} = B,$$

$$L_M\colon \begin{array}{l} I \mapsto M \;\;= 1 \cdot I + 1 \cdot B \\ B \mapsto MB = 2 \cdot I + 1 \cdot B \end{array} \quad \text{so} \quad [L_M]_{\mathbf{B}} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} = M,$$

and

$$L_M\colon \begin{array}{l} I \mapsto M \;\;= 0 \cdot I + 1 \cdot M \\ M \mapsto M^2 = 1 \cdot I + 2 \cdot M \end{array} \quad \text{so} \quad [L_M]_{\mathbf{M}} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} = A.$$

Since A is similar to M, it follows that A is another cyclic generator of $\mathbf{F}_9$. Moreover, A is the companion matrix of its characteristic polynomial $f_A(x) = x^2 + x + 2$. We call A a *canonical cyclic generator* of $\mathbf{F}_9$, and call the representation

$$\mathbf{F}_3[A] = \{0, I, A, A^2, A^3, A^4, A^5, A^6, A^7\}$$

a *canonical cyclic representation* of $\mathbf{F}_9$.

Of course, all of these ideas generalize for arbitrary finite fields. (Indeed, they generalize to finite extensions of *any* field, but we restrict the treatment here to finite extensions of fields $\mathbf{F}_p$ with p prime.) Let p be a prime number and let $q = p^e$ be the e^{th} power of p. Then $\mathbf{F}_q$ is a q element field containing $\mathbf{F}_p = \mathbf{Z}_p = \mathbf{Z}/(p)$ (the integers modulo p) as its prime field. Let $m(x)$ be any irreducible polynomial of degree e over $\mathbf{F}_p$, and let B be the companion matrix of $m(x)$. The ring $\mathbf{F}_p[B]$ of sums of powers of B is isomorphic to the field $\mathbf{F}_q$, and is thus a matrix representation of $\mathbf{F}_q$. Locate a matrix M in $\mathbf{F}_p[B]$ that has period (multiplicative order) $q - 1$. M is necessarily a cyclic generator of $\mathbf{F}_q$. The companion matrix A of the minimum polynomial $m_M(x) = m_A(x)$ is a canonical cyclic generator of

$$\mathbf{F}_p[A] = \{0, I, A, A^2, \dots, A^{q-2}\} \cong \mathbf{F}_q.$$

Note that if C is any $e \times e$ matrix over $\mathbf{F}_p$, then the ring $\mathbf{F}_p[C]$ generated by C is isomorphic to $\mathbf{F}_q$ if, and only if, the characteristic polynomial $f_C(x)$ of C is irreducible, only if the sequence $\mathbf{C} = (I, C, C^2, \dots, C^{e-1})$ of powers of C is independent. In this case, the matrix $[L_C]_{\mathbf{C}}$ of left multiplication by C, with respect to the basis $\mathbf{C}$, is the companion matrix of $f_C(x)$. C is a cyclic generator of $\mathbf{F}_q$ if, and only if, C is a primitive $(q-1)^{\text{st}}$ root of unity in $\mathbf{F}_p[C]$.

There is another, possibly easier, method of getting a canonical cyclic generator of $\mathbf{F}_q$. Recall that the n^{th} cyclotomic polynomial $\mathbf{c}_n(x)$ is defined to be the product

$$\mathbf{c}_n(x) = \prod (x - a) \tag{5}$$

taken over all $\varphi(n)$ primitive n^{th} roots a of unity. Since every root of $x^n - 1 = 0$ is a primitive d^{th} root of unity for some divisor d of n, it follows from (5) that

$$x^n - 1 = \prod_{d \mid n} \mathbf{c}_d(x). \tag{6}$$

One can use (6) to obtain the recursive formula

$$\mathbf{c}_n(x) = (x^n - 1) \Big/ \prod_{d \mid n,\, d < n} \mathbf{c}_d(x). \tag{7}$$

It follows inductively from (7) that $\mathbf{c}_n(x)$ is a monic polynomial with integer coefficients of degree (from (5)) $\varphi(n)$. The cyclotomic polynomials are all irreducible over the rational number field (see [3], p. 61, Thm. 2.4.7, [4], p. 162, or [5], p. 289, Thm. 6.3.13), but they usually factor over finite fields. It will be useful later to note that if $n = r^d$ is a power of a prime r, then it follows inductively from (7) that

$$\mathbf{c}_n(x) = (x^n - 1)/(x^{n/r} - 1), \ (n = r^d, \ r \text{ prime}). \tag{8}$$

Every element of $\mathbf{F}_q$ (p prime and $q = p^e$) is a root of

$$x^q - x = x(x^{q-1} - 1) = 0, \tag{9}$$

since $\mathbf{F}_q$ is the splitting field of $x^q - x$, and every nonzero element is a $(q-1)$st root of unity. If $m(x)$ is a monic irreducible factor of $\mathbf{c}_{q-1}(x)$, and a is a root of $m(x)$, then a is a primitive $(q-1)$st root of unity. (Note that $m(x)$ is necessarily of degree e.) It follows that if A is the $e \times e$ companion matrix of $m(x)$, then A is a canonical cyclic generator of $\mathbf{F}_q$.

Conversely, if A is a canonical cyclic generator of $\mathbf{F}_q$ over $\mathbf{F}_p$, then its minimum polynomial $m_A(x)$ is an irreducible factor of the cyclotomic polynomial $\mathbf{c}_{q-1}(x)$ in $\mathbf{F}_p[x]$. This observation can lead to a method of factoring cyclotomic polynomials. This is a related but different topic that we will not pursue here.

Let us conclude with two examples that use the method of factoring cyclotomic polynomials to obtain canonical cyclic representations of $\mathbf{F}_8$ over $\mathbf{F}_2$, and of $\mathbf{F}_9$ over $\mathbf{F}_3$. (We have treated these cases more naively above.)

For $\mathbf{F}_8$ over $\mathbf{F}_2$, $e = [\mathbf{F}_8 : \mathbf{F}_2] = 3$, so the factors of $\mathbf{c}_7(x)$ are cubic:

$$\begin{aligned} \mathbf{c}_7(x) &= (x^7 - 1)/(x - 1) \\ &= x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\ &= (x^3 + x + 1)(x^3 + x^2 + 1). \end{aligned}$$

(The factorization of $\mathbf{c}_7(x)$ was particularly easy, since its factors are the only irreducible polynomials of degree three over $\mathbf{F}_2$.) Since $x^3 + x + 1$ and $x^3 + x^2 + 1$ are irreducible factors of $\mathbf{c}_7(x)$, it follows that their companion matrices

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

are canonical cyclic generators of $\mathbf{F}_8$ over $\mathbf{F}_2$.

For $\mathbf{F}_9$ over $\mathbf{F}_3$, we would like to factor

$$\mathbf{c}_8(x) = (x^8 - 1)/(x^4 - 1) = x^4 + 1.$$

Since $e = [\mathbf{F}_9 : \mathbf{F}_3] = 2$, the factors are quadratic. It is not hard to see that the monic irreducible quadratics over $\mathbf{F}_3$ are $x^2 + 1$, $x^2 - x - 1$, and $x^2 + x - 1$. The desired factorization is

$$\mathbf{c}_8(x) = x^4 + 1 = (x^2 + x - 1)(x^2 - x - 1),$$

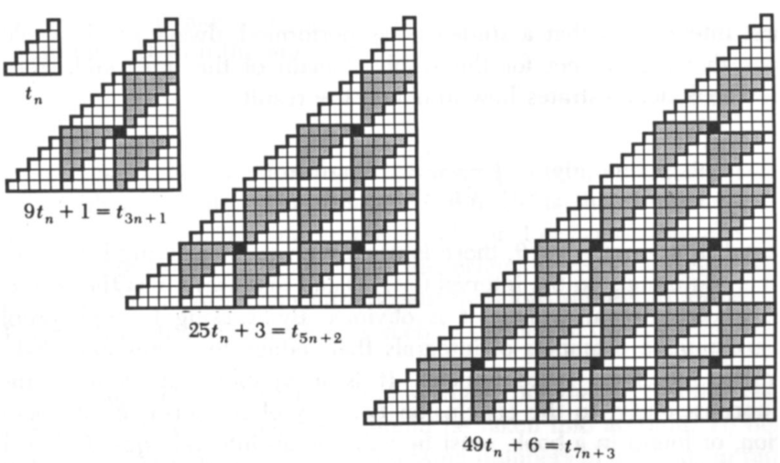
so the canonical cyclic generators of $\mathbf{F}_9$ over $\mathbf{F}_3$ are the corresponding companion matrices,

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

REFERENCES

1. I. N. Herstein, *Topics in Algebra*, Blaisdell, Waltham, MA, 1964.
2. K. Hoffman and R. Kunze, *Linear Algebra*, 2nd edition, Prentice Hall, Inc., Englewood Cliffs, NJ, 1971.
3. R. Lidl and H. Niederreiter, *Introduction to Finite Fields and their Applications*, Cambridge University Press, New York, 1986.
4. B. L. van der Waerden, *Modern Algebra*, Vol. 1, Ungar, New York, 1969.
5. E. A. Walker, *Introduction to Abstract Algebra*, Random House, Inc., New York, 1987.

Proof without Words: A Triangular Identity



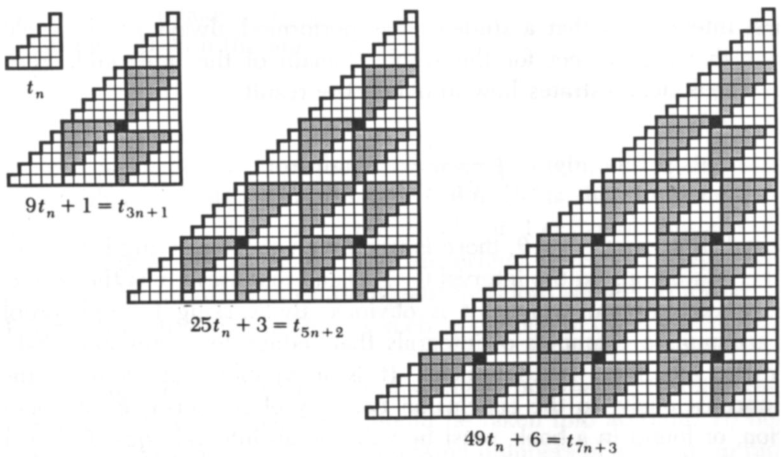
$$t_n = 1 + 2 + \cdots + n \Rightarrow (2k+1)^2 t_n + t_k = t_{(2k+1)n+k}$$

—Roger B. Nelsen
Lewis and Clark College
Portland, OR 97219

REFERENCES

1. I. N. Herstein, *Topics in Algebra*, Blaisdell, Waltham, MA, 1964.
2. K. Hoffman and R. Kunze, *Linear Algebra*, 2nd edition, Prentice Hall, Inc., Englewood Cliffs, NJ, 1971.
3. R. Lidl and H. Niederreiter, *Introduction to Finite Fields and their Applications*, Cambridge University Press, New York, 1986.
4. B. L. van der Waerden, *Modern Algebra*, Vol. 1, Ungar, New York, 1969.
5. E. A. Walker, *Introduction to Abstract Algebra*, Random House, Inc., New York, 1987.

Proof without Words: A Triangular Identity



$$t_n = 1 + 2 + \cdots + n \Rightarrow (2k+1)^2 t_n + t_k = t_{(2k+1)n+k}$$

—Roger B. Nelsen
Lewis and Clark College
Portland, OR 97219

The Importance of Being Continuous

D. J. JEFFREY
The University of Western Ontario
London, Ontario, Canada N6A 5B7

1. Introduction Every student of calculus learns its Fundamental Theorem, usually stated as follows [1]. *If f is continuous on $[a,b]$, then the function defined by*

$$g(x)=\int_a^x f(t)\,dt \tag{1}$$

is continuous on $[a,b]$ and differentiable on (a,b), and $g'(x)=f(x)$. Once students have learned this theorem, they proceed to learn a variety of ways to obtain closed-form expressions for the integrals of given integrands. These include memorizing a table of simple integrals, applying substitutions, learning to consult tables of integrals such as the CRC Handbook [2] or Gradshteyn and Ryzhik [3], and, increasingly, asking computer algebra systems, such as *Derive*, *Maple*, or *Mathematica*.

The first integrations that a student sees performed always result in closed-form expressions that are correct for the whole domain of the integrand. For example, when a textbook demonstrates how to obtain the result

$$\int\frac{dx}{\sqrt{1-x^2}}=\arcsin x$$

using the substitution $x=\sin\theta$, there is usually no accompanying statement that the integrand is continuous on the interval $(-1,1)$ and that $\arcsin x$ is the correct integral over that whole interval. Maybe it is obvious. By working through problem sets, students accumulate examples of integrals that reduce to expressions that are valid wherever the integrand is continuous. It is inevitable that students then jump, perhaps unconsciously, to the conclusion that any closed-form result obtained by a substitution, or found in a book, must be valid on an interval equal to the domain of the integrand. This paper considers some cases in which that conclusion is false.

Consider the function $f=\sqrt{1-\cos x}$. FIGURE 1 shows that f is continuous for all x, and therefore the fundamental theorem says there must exist a function g, also continuous for all x, which is its integral. Using the substitution $u=\cot\frac{1}{2}x$, we can derive the equation

$$\int\sqrt{1-\cos x}\,dx=-2\sqrt{1-\cos x}\,\cot\frac{1}{2}x. \tag{2}$$

This equation is returned by *Mathematica*, for example. The right-hand side of (2) is discontinuous at even multiples of π, as FIGURE 1 shows, and so although it was obtained by a standard procedure, it is not the integral g that the theorem says exists.

There are two attitudes we might take toward an equation such as (2). On the one hand, we could say that since (2) is correct only on the interval $(0,2\pi)$, the fault is one of notation. Technically, it is always incorrect to quote a formula without specifying the interval on which it applies, and here we have a case in which the common practice of leaving the interval unspecified is particularly misleading. With

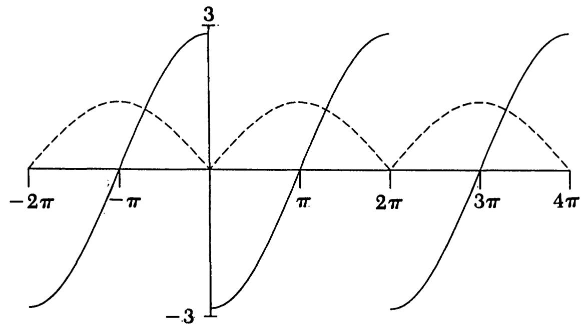


FIGURE 1
The integrand and integral in equation (2). - - -, the integrand. ——, the discontinuous integral.

this attitude, the problem becomes one of determining the interval of validity of a given closed-form expression and ensuring that the interval is properly displayed. On the other hand, we could say that since the fundamental theorem tells us that there does exist an integral of f that is continuous everywhere on the real line, our object should be to capture it in a closed-form expression. With this attitude, the problem becomes one of modifying expressions like (2) to enlarge their intervals of validity, and this is the view we take here.

We can contrast (2) with the equation

$$\int \frac{dx}{x^2} = -\frac{1}{x}. \tag{3}$$

This equation is not valid on intervals containing the origin, and again textbooks are unlikely to say this explicitly. The difference, though, is that we cannot do better, because the discontinuity in $1/x$ is the result of $1/x^2$ violating the conditions required by the Fundamental Theorem at the origin. Discontinuities such as those seen in (2) should be regarded as spurious, and therefore subject to further investigation, while discontinuities or singularities such as the one in (3) are genuine and must be accepted.

This is not a paper on computer algebra systems, but the perspective they add to the problem is an interesting one, and should be taken into account. To begin with, the spread of these systems is allowing increasing numbers of students to range much more widely in calculus than before, making it more likely that they will discover the difficulties discussed here for themselves. In addition, experienced mathematicians use these systems and demand the maximum in convenience from them, which means the designers must automate tasks that previously were left unaddressed, such as allowing for discontinuities. Finally, at present no algebra system has a good method for informing the user that a result it has obtained is valid only on a restricted interval, and so it is important to return results that are valid on as wide an interval as possible.

In the following sections we shall look at various integration problems in which expressions arise that contain spurious discontinuities, in the sense defined above. As sources of discontinuous expressions, we shall consider integral tables, substitutions and special types of integrands. It is important to realize that tables of integrals and calculus textbooks harbour an astonishingly large number of expressions with spurious discontinuities, and the fact that these entries have successfully maintained their population for over 100 years clearly indicates that they have no natural predators, even though they are one species that humans should be encouraged to endanger.

2. Tables of integrals We discuss this source of discontinuous integrals using a specific example, an entry appearing in two of the most popular tables of integrals, Gradshteyn and Ryshik [3] where it is formula 2.132.1, and the CRC handbook [2] where it is formula 77.

$$\int \frac{dx}{x^4+1} = \frac{1}{4\sqrt{2}} \ln \frac{x^2+\sqrt{2}x+1}{x^2-\sqrt{2}x+1} + \frac{1}{2\sqrt{2}} \arctan \frac{x\sqrt{2}}{1-x^2}. \tag{4}$$

The left-hand side of this equation is continuous for all x, but the right-hand side contains discontinuities at $x = \pm 1$. There is nothing printed in the tables to warn the reader of the discontinuities or, equivalently, to advise the reader of the intervals on which the formula is valid. To make matters worse, the editors of these tables have overlooked an alternative to (4), which is free of discontinuities.

$$\int \frac{dx}{x^4+1} = \frac{1}{4\sqrt{2}} \ln \frac{x^2+\sqrt{2}x+1}{x^2-\sqrt{2}x+1} + \frac{1}{2\sqrt{2}} \arctan\left(x\sqrt{2}+1\right) + \frac{1}{2\sqrt{2}} \arctan\left(x\sqrt{2}-1\right). \tag{5}$$

FIGURE 2 shows plots of (4) and (5) on the same axes.

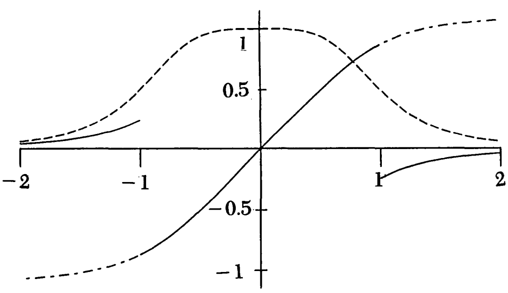


FIGURE 2

The functions in equations (4) and (5). - - -, the integrand. ——, the discontinuous integral (4). -·-·-, the continuous integral (5).

Equation (5) shows that the discontinuities in (4) are unnecessary, and even misleading, because they masquerade as properties of the integrand, whereas they are really artifacts of the process that derived the expression. It is likely that (4) was originally derived by obtaining (5) first and then (see [3]) making incorrect use of the formula

$$\arctan x + \arctan y = \begin{cases} \arctan \dfrac{x+y}{1-xy}, & \text{for } xy < 1, \\ \arctan \dfrac{x+y}{1-xy} + \pi, & \text{for } xy > 1 \text{ and } x > 0, \\ \arctan \dfrac{x+y}{1-xy} - \pi, & \text{for } xy > 1 \text{ and } x < 0. \end{cases}$$

Probably, the first line of this formula was taken to apply for all x and y, a misconception easily carried away from a casual reading of Abramowitz and Stegun [4]. By comparing (4) and (5), we can see that there is a temptation to prefer (4) because of its compactness, but the discontinuities it introduces negate this advantage.

A search through any extensive table of integrals will find entries in which there are terms of the form arctan(P/Q), where P and Q are polynomials. If Q has roots within the range of integration, then the expression will contain discontinuities. A form for an integral that is free of spurious discontinuities must be counted as superior to one that is not, and it is surely shoddy editing to print the inferior form. Also, there is the question of efficiency. If the user must check every entry in the tables for continuity, the use of the tables becomes less efficient.

3. Discontinuity from substitution The most important substitution that leads to spurious discontinuities is the Weierstrass, or $\tan \frac{1}{2}x$, substitution. For example, the function $3/(5-4\cos x)$ is continuous and positive for all real x, and so its integral should be continuous and monotonically increasing. By letting $u = \tan \frac{1}{2}x$, we obtain

$$\int \frac{3\,dx}{5-4\cos x} = \int \frac{6\,du}{1+9u^2} = 2\arctan\left(3\tan\frac{1}{2}x\right). \tag{6}$$

The final expression in (6) is discontinuous at odd multiples of π, as can be seen in FIGURE 3. The unsatisfactory nature of (6) has rarely been noted. The only published commentary on this class of integrals that acknowledges that special treatment is required is the introduction to the CRC tables [2], where it is stated that the 'correct branch' of the inverse tangent must be used when applying the formula.

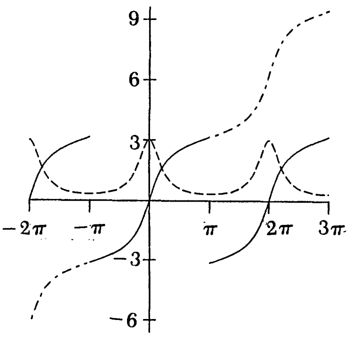


FIGURE 3

The functions in equations (6) and (7). - - -, the integrand. ——, the discontinuous integral (6). -·-·-, the continuous integral (7).

A new method for evaluating integrals such as (6) that always yields continuous expressions has been developed [5]. When applied to the integral above, it gives

$$\int \frac{3\,dx}{5-4\cos x} = 2\arctan\left(3\tan\frac{1}{2}x\right) + 2\pi\left\lfloor \frac{x+\pi}{2\pi} \right\rfloor, \tag{7}$$

where $\lfloor\ \rfloor$ is the floor function. An interesting facet of this result is that each term separately is discontinuous at odd multiples of π, but the algorithm arranges things so that together they make a continuous function. FIGURE 3 shows plots of the functions appearing in (6) and (7). We can express the floor function in terms of an inverse tangent as

$$2\pi\lfloor (x+\pi)/2\pi \rfloor = x - 2\arctan\left(\tan\frac{1}{2}x\right),$$

and substituting this into (7) and combining inverse tangents gives the following compact variation on (7):

$$\int \frac{3\,dx}{5-4\cos x} = x + 2\arctan\frac{\sin x}{2-\cos x}. \tag{8}$$

A generalization of this expression that is suitable for integral tables is as follows. If

$$p^2 > q^2 + r^2,$$

then

$$\int \frac{dx}{p+q\cos x + r\sin x} = \frac{x}{\Delta} + \frac{2}{\Delta}\arctan\frac{r\cos x - q\sin x}{p + q\cos x + r\sin x + \Delta}, \tag{9}$$

where $\Delta = \operatorname{sgn}(p)\sqrt{p^2-q^2-r^2}$. The right-hand side is continuous for all possible values of the parameters, and should be used in place of all similar formulae, wherever found.

Any substitution $x = u(s)$ can lead to an antiderivative with a spurious discontinuity if $u(s)$ contains a singularity. For example, since the substitution $x = 1/s$ is singular at the origin, results obtained using it should be checked for continuity there. The integrand in (10) is continuous everywhere, having a removable singularity at the origin, but the substitution $1/s$ gives a discontinuous integral.

$$\int \frac{e^{1/x}}{\left(1+e^{1/x}\right)^2}\frac{dx}{x^2} = \int \frac{-e^s}{\left(1+e^s\right)^2}\,ds = \frac{1}{1+e^{1/x}}. \tag{10}$$

The integrand and integral are plotted in FIGURE 4. We can remove the jump by using the signum function. The jump in the function at the origin is -1, and so a continuous expression for the integral is

$$\int \frac{e^{1/x}}{\left(1+e^{1/x}\right)^2}\frac{dx}{x^2} = \frac{1}{1+e^{1/x}} + \frac{1}{2}\operatorname{sgn} x.$$

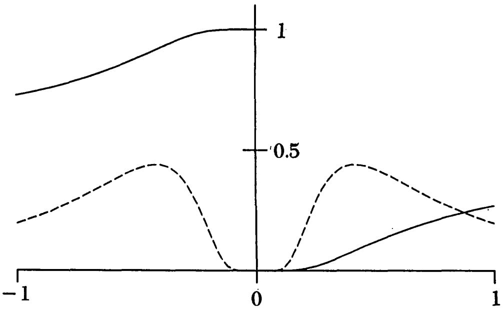


FIGURE 4
The functions in equation (10). - - -, the integrand. ——, the discontinuous integral.

4. Piecewise continuous functions Introductory calculus books and integral tables do not address the problems of integrating piecewise continuous functions, even though the step function, and similar functions, are very useful in applications and are used freely in physics and engineering books. Recently, Botsko [6] has stated and proved a generalization of the Fundamental Theorem that requires only that f be integrable, and not necessarily continuous, and hence covers piecewise continuous functions. The central topic of this paper, namely the importance of ensuring that integrals are continuous, is an important factor in the development of correct rules for integrating piecewise-defined functions.

The Heaviside step function is defined by

$$H(x)=\begin{cases}1, & \text{for } x\geq 0;\\ 0, & \text{for } x<0.\end{cases}$$

As an example of how we want to work with this function, consider the following differential equation coming from elementary beam theory. The bending moment $M(x)$ in a beam extending from $x=0$ to $x=l$ and supporting point loads P_a and P_b at $x=a$ and $x=b$ is given by the equation

$$\frac{dM}{dx}=\begin{cases}K, & \text{for } 0\leq x\leq a;\\ K+P_a, & \text{for } a\leq x\leq b;\\ K+P_a+P_b, & \text{for } b\leq x\leq l.\end{cases}$$

where K is an unknown constant. The equation is to be solved subject to boundary conditions representing free ends, to wit, $M(0)=M(l)=0$. Most students would solve this by integrating each line separately to obtain

$$M=\begin{cases}Kx+A_1, & 0\leq x\leq a;\\ (K+P_a)x+A_2, & a\leq x\leq b;\\ (K+P_a+P_b)x+A_3, & b\leq x\leq l,\end{cases}$$

Matching the solutions at $x=a$ and at $x=b$ gives

$$A_1=A_2+P_a a \quad\text{and}\quad A_2=A_3+P_b b.$$

The boundary conditions now give $A_1=0$ and $K=P_a(a-l)/l+P_b(b-l)/l$. The problem is not finished, however, because we must integrate twice more the same way to obtain the deflection.

A much quicker way to proceed is to write

$$\frac{dM}{dx}=K+P_aH(x-a)+P_bH(x-b),$$

and develop rules for integrating H. At first sight, it might seem that the rule is very simple, namely

$$\int f(x)H(x-a)\,dx=H(x-a)\int f(x)\,dx.$$

For most functions f, however, the right-hand side will violate our principle that the integral must be continuous. The better form is

$$\int f(x)H(x-a)\,dx=H(x-a)\int_a^x f(t)\,dt. \tag{11}$$

The right-hand side is now continuous everywhere that the integral of f is. Applying this to the differential equation above, we get

$$M=(x-a)P_aH(x-a)+(x-b)P_bH(x-b)+Kx+A,$$

and the boundary conditions give the same solution as before. The deflection can now be easily obtained by integrating twice more to find

$$y = \frac{1}{6}(x-a)^3 P_a H(x-a) + \frac{1}{6}(x-b)^3 P_b H(x-b) + \frac{1}{6}Kx^3 + Bx + C.$$

The boundary conditions $y(0) = y(l) = 0$ give us B and C. This method of solution is more convenient than the first one for people working by hand, and vastly more convenient for those using an algebra system.

5. Conclusions In calculus textbooks, it is popular to include a section on the use of integral tables. In view of the results in section 2, the textbooks should warn students that any expression extracted from a table might contain a spurious discontinuity. With equal force, we should require the editors of handbooks to check their tables thoroughly. In effect, an entry in a table should not be considered correct unless it is continuous on as wide an interval as possible. The alternative would be to note the interval upon which the integral is valid, without attempting to broaden it, but this would be less useful to the reader. Similar comments can be applied to computer algebra systems.

The Weierstrass substitution discussed in section 3 was once a standard topic in calculus textbooks, albeit an advanced topic. It appears less frequently now, but if it is treated, an analysis of the discontinuity and its correct handling should be included. The material of section 4 comes from my experience of watching students tackle problems such as the one described, and from implementing the solution on algebra systems.

Acknowledgement. This paper springs directly from discussions with David Stoutemyer and Al Rich, the developers of *Derive*, a computer algebra program. Almost all of the items discussed here have been implemented in *Derive*, and I am grateful to its developers for their interest. I am also grateful to the developers of *Maple* for stimulating discussions, and for adopting some of the ideas above.

REFERENCES

1. J. Stewart, *Calculus, 2nd edition*, Brooks/Cole Publishing Co., Belmont, CA, 1991.
2. W. H. Beyer, *Handbook of Mathematical Sciences, 6th edition*, CRC Press, Inc. Boca Raton, FL, 1987.
3. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, San Diego, CA, 1979.
4. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, Mineola, NY, 1965.
5. D. J. Jeffrey and A. D. Rich, The evaluation of trigonometric integrals avoiding spurious discontinuities, *ACM Trans. Math. Software*, submitted.
6. M. W. Botsko, A fundamental theorem of calculus that applies to all Riemann integrable functions, *this* MAGAZINE, 64 (1991), 347–348.

A New (?) Test for Convergence of Series

K. P. S. BHASKARA RAO
Indian Statistical Institute
Bangalore 560 059, India

While teaching a course on real analysis, I wanted to motivate, or, 'discover' the root test for the students. An attempt at this led to the 'discovery' of the following test. With this in mind, along with the proof, I shall include its motivation.

Let Σa_n be a series of strictly positive terms. Define

$$\alpha = \overline{\lim}\,(\log a_n/\log n) \text{ and } \beta = \underline{\lim}(\log a_n/\log n).$$

The test

$$\begin{aligned} &\text{If } \alpha < -1, && \text{then } \Sigma a_n \text{ converges.} \\ &\text{If } \beta > 1, && \text{then } \Sigma a_n \text{ diverges.} \end{aligned}$$

Proof. We start with a known result about the convergence or divergence of a family of series, depending on a parameter p:

$$\begin{aligned} &\Sigma\frac{1}{n^p} \text{ converges} && \text{if } p > 1 \\ &\text{and diverges} && \text{if } 0 \le p \le 1. \end{aligned}$$

Thus $p = 1$ is the critical value, or, turning point.

We consider applying the comparison test to test the convergence of Σa_n. Let us try to express this in a nice and natural way.

Firstly If there is a $p > 1$ such that $a_n \le (1/n^p)$ for all but finitely many n's, then Σa_n converges.

Rewriting: If there is a $p > 1$ such that $(\log a_n/\log n) \le -p$ for all but finitely many n's, then Σa_n converges.

Rewriting again: If $\overline{\lim}\,(\log a_n/\log n) < -1$, then Σa_n converges.

So, if $\alpha < -1$, then Σa_n converges.

Secondly If there is a $p \le 1$ such that $(\log a_n/\log n) \ge (1/n^p)$ for all but finitely many n's, then Σa_n diverges.

Rewriting: If there is a $p \le 1$ such that $(\log a_n/\log n) \ge -p$ for all but finitely many n's, then Σa_n diverges.

We cannot rewrite this in terms of β. But at least we can say:

If $\underline{\lim}(\log a_n/\log n) > -1$, then there is a $p \le 1$ such that $(\log a_n/\log n) \ge -p$ for all but finitely many n's.

So, if $\beta > -1$, then Σa_n diverges.

Thirdly The above arguments fail for the case $\beta \le -1 \le \alpha$.

Remark 1. A similar analysis with the family $\{\Sigma_n p^n\}_{p>1}$ in place of $\{\Sigma_n(1/n^p)\}_{p>0}$ would give the 'root test' in place of our test. A little care should be exercised in the second part, while proving Σa^n diverges if $\overline{\lim}\, a_n^{1/n} > 1$.

Remark 2. Any family of nonnegative series $\{\Sigma_n f_p(n) : p > 0\}$ for which the convergence or the divergence is known, should, by a similar analysis, lead to another test for convergence.

Remark 3. From the test above it follows that the Dirichlet series $\Sigma_{n \geq 1} a_n n^s$ with $a_n > 0$ for all n

$$\text{converges if} \quad s < -\overline{\lim} \frac{\log a_n}{\log n} - 1$$

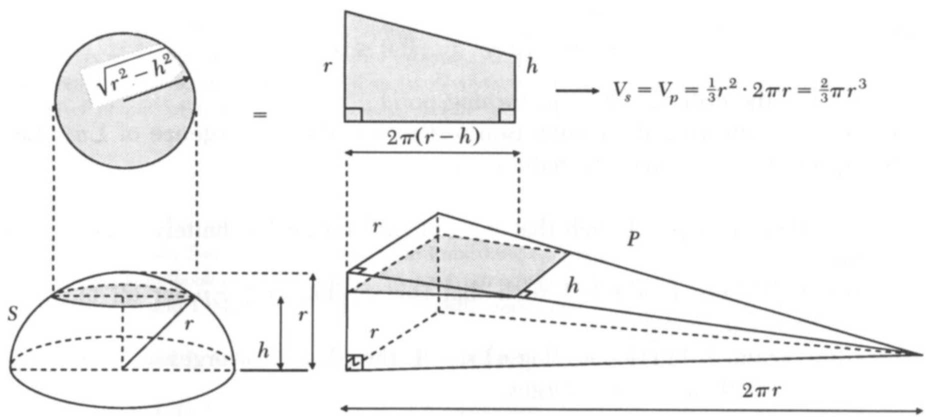
$$\text{and diverges if} \quad s > -\underline{\lim} \frac{\log a_n}{\log n} - 1.$$

Acknowledgement. Thanks to Adam Fieldsteel for pointing out an error in an earlier draft of this note.

REFERENCE

1. Ross, K. A., *Elementary Analysis: The Theory of Calculus*, Springer-Verlag New York, 1986.

Proof without Words: The Volume of a Hemisphere via Cavalieri's Principle*



*Tzu Geng, son of the most celebrated mathematician in ancient China, Tzu Chung Chih, was believed to be the first to develop the principle in the 5th century A.D.

—SIDNEY H. KUNG
JACKSONVILLE UNIVERSITY
JACKSONVILLE, FL 32211

Remark 2. Any family of nonnegative series $\{\Sigma_n f_p(n): p > 0\}$ for which the convergence or the divergence is known, should, by a similar analysis, lead to another test for convergence.

Remark 3. From the test above it follows that the Dirichlet series $\Sigma_{n\geq 1} a_n n^s$ with $a_n > 0$ for all n

$$\text{converges if} \quad s < -\overline{\lim}\,\frac{\log a_n}{\log n} - 1$$

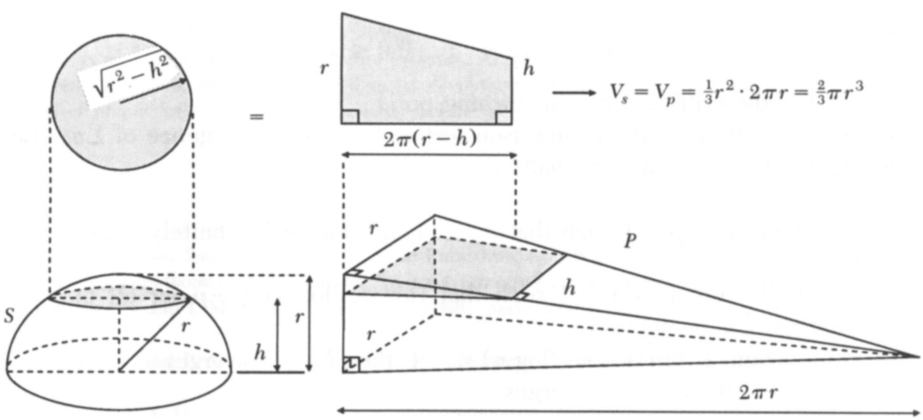
$$\text{and diverges if} \quad s > -\underline{\lim}\,\frac{\log a_n}{\log n} - 1.$$

Acknowledgement. Thanks to Adam Fieldsteel for pointing out an error in an earlier draft of this note.

REFERENCE

1. Ross, K. A., *Elementary Analysis: The Theory of Calculus*, Springer-Verlag New York, 1986.

Proof without Words: The Volume of a Hemisphere via Cavalieri's Principle*



*Tzu Geng, son of the most celebrated mathematician in ancient China, Tzu Chung Chih, was believed to be the first to develop the principle in the 5th century A.D.

—Sidney H. Kung
Jacksonville University
Jacksonville, FL 32211

PROBLEMS

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

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by March 1, 1995.

1454. *Proposed by Barry Cipra, Northfield, Minnesota.*

Suppose n people put their names in a hat (on slips of paper), and the names are redistributed at random (using the uniform probability distribution on the space of permutations). Those who receive their own name drop out, while the rest repeat the procedure. On average, how many rounds will it take until all have gotten their own name back?

1455. *Proposed by Jiro Fukuta, Gifu-ken, Japan.*

In a hexagon $A_1A_2A_3A_4A_5A_6$ inscribed in a circle with center O, let M_i, $i = 1, 2, \ldots, 6$, be the midpoints of the sides A_iA_{i+1}, where $A_7 = A_1$. Prove that if $\triangle\, M_1M_3M_5$ and $\triangle\, M_2M_4M_6$ are equilateral, $\triangle\, A_1A_3A_5$ and $\triangle\, A_2A_4A_6$ are also equilateral.

1456. *Proposed by Howard Morris, Chatsworth, California.*

Show that the only sequence of numbers (α_i) that satisfies the conditions

(i) $$\alpha_i > 0 \quad \text{for all } i \geq 1, \quad \text{and}$$

(ii) $$\alpha_{i-1} = \frac{i\alpha_i + 1}{\alpha_i + i} \quad \text{for all } i > 0,$$

is the sequence $\alpha_i = 1$ for all i.

ASSISTANT EDITORS: CLIFTON CORZAT, BRUCE HANSON, RICHARD KLEBER, KAY SMITH, *and* THEODORE VESSEY, *St. Olaf College and* MARK KRUSEMEYER, *Carleton 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.*

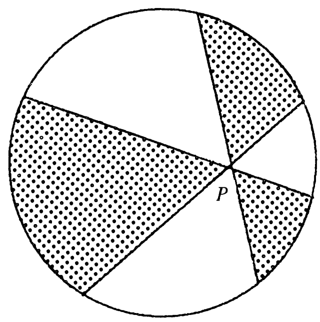
Solutions should be written in a style appropriate for Mathematics Magazine. Each solution should begin on a separate sheet containing the solver's name and full address.

Solutions and new proposals should be mailed in duplicate to Loren Larson, Department of Mathematics, St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057-1098 or mailed electronically via fax: (507) 663-3549 or e-mail: larson@stolaf.edu.

1457. *Proposed by Larry Carter, IBM Watson Research Center, Yorktown Heights, New York, John Duncan, University of Arkansas, Fayetteville, Arkansas, and Stan Wagon, Macalester College, St. Paul, Minnesota.*

a. For a point P inside a circle draw three chords through P making six 60° angles at P and form two regions by coloring the six "pizza slices" alternately black and white. Prove that the region containing the center has the larger area.

b.* Prove that if five chords make ten 36° angles at P, then the region containing the center has the lesser area.



1458. *Proposed by Dave Trautman, The Citadel, Charleston, South Carolina.*

Let $(a_i)_{i=1}^n$ and $(b_i)_{i=1}^n$ be sequences of positive integers with $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$.

a. Evaluate

$$\sup\left|\left(\sum_{i=1}^{n} \frac{a_i^2}{a_i^2+b_i^2}\right) - \frac{n}{2}\right|.$$

b. Same as (a) under the additional restriction that for all i, $A \le (a_i/b_i) \le B$, where A and B are fixed positive constants.

(Note: This problem has its genesis in the work of the baseball statistician Bill James. James constructs what he calls the "Pythagorean Model," which states that if a team scores a runs in a season and gives up b runs in the same season, then its winning percentage should be approximately $a^2/(a^2+b^2)$. The sum of the winning percentages of n teams in a league is $n/2$. This problem explores how far off this Pythagorean Model could be for the sum of the predicted winning percentages.)

Quickies

Answers to the Quickies are on page 309.

Q823. *Proposed by Norman Schaumberger, Hofstra University, Hempstead, New York.*

If n and k are positive integers, show that

$$n^{kn} \ge (n^k + n^{k+1} + \cdots + 1)^{n-1},$$

with strict inequality if $n > 1$. ($k = 1$ gives the familiar result, $n^n > (n+1)^{n-1}$.)

Q824. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Determine all positive rational solutions of $x^x = y^y$ with $x \geq y > 0$.

Q825. *Proposed by Paul Bracken, University of Waterloo, Waterloo, Ontario, Canada.*

Let $T^{(n+1)}(x) = T(T^{(n)}(x))$ for $n = 1, 2, 3, \dots$, where $T^{(1)}(x) = T(x)$ and the map $T:[0,1] \to [0,1]$ is defined as follows:

$$T(x) = \begin{cases} 2x & 0 \leq x \leq 1/2, \\ -2x+2 & 1/2 < x \leq 1. \end{cases}$$

Determine $\int_0^1 T^{(n)}(x)\,dx$.

Solutions

Remainder modulo a cyclotomic polynomial **October 1993**

1428. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Determine the remainder when $(x^2-1)(x^3-1)\cdots(x^{16}-1)(x^{18}-1)$ is divided by $1+x+x^2+\cdots+x^{16}$.

Solution by F. J. Flanigan, San Jose State University, San Jose, California.

The remainder is 17. More generally, for p an odd prime, the remainder when $F_p(x) = (x^2-1)(x^3-1)\cdots(x^{p-1}-1)(x^{p+1}-1)$ is divided by $\Phi_p(x) = 1+x+x^2+\cdots+x^{p-1}$ is the constant p.

To see this, write $F_p(x) = Q(x)\Phi_p(x) + R(x)$, where $R(x)$ is a polynomial of degree at most $p-2$. Let ζ be a root of $\Phi_p(x)$, that is, a primitive pth root of unity. Clearly $F_p(\zeta) = R(\zeta)$. But

$$\begin{aligned} F_p(\zeta) &= (\zeta^2-1)\cdots(\zeta^{p-1}-1)(\zeta^{p+1}-1) = (1-\zeta)(1-\zeta^2)\cdots(1-\zeta^{p-1}) \\ &= \Phi_p(1), \end{aligned}$$

where we have used $\zeta^{p+1} = \zeta$ and the standard factorization $\Phi_p(x) = (x-\zeta)(z-\zeta^2)\cdots(x-\zeta^{p-1})$, as well as the fact that $p-1$ is even (to reverse each factor).

From this we learn that $R(\zeta) = F_p(\zeta) = \Phi_p(1) = 1+1+\cdots+1 = p$, the given prime. But this is true for each of the $p-1$ primitive pth roots of unity. Since the polynomial $R(x)$ has degree no larger than $p-2$, $R(x) \equiv p$ for all x, proving the claim.

Note: For a positive integer n, recall that a complete positive reduced residue system modulo n is a set $\kappa = \{k_1, k_2, \dots, k_{\phi(n)}\}$ of positive integers, no two of which are congruent modulo n and each of which is coprime to n. For n and κ as above, define $E_\kappa(x) = \prod_{k \in \kappa}(x^k - 1)$. Then, for $n \geq 3$, the remainder when $E_\kappa(x)$ is divided by the nth cyclotomic polynomial $\Phi_n(x)$ is the integer $\Phi_n(1)$. The proof given above, *mutatis mutandis*, works here as well.

Also solved by the Anchorage Math Solutions Group, Armstrong State College Problem Group, Charles Ashbacher, Brian Beasley, J. C. Binz (Switzerland), Walter Blumberg, Stan Byrd and Terry Walters, D. K. Cohoon, Con Amore Problem Group (Denmark), Patrick Costello, Robert L. Doucette, E. S. Freidkin, Jiro Fukuta (Japan), David Hankin, N. S. Hekster (The Netherlands), Francis M. Henderson, Richard Holzsager, The Javelina Problem Solvers, H. K. Krishnapriyan, Robert L. Lamphere, Kee-Wai Lau (Hong Kong), Robert H. Lewis, O. P. Lossers (The Netherlands), Reiner Martin (student), Allan Pedersen (Denmark), Gábor Pete (student, Hungary), Iwan Praton, F. C. Rembis, Rockford College Problem Group, John S. Sumner and Kevin L. Dove, Michael Vowe (Switzerland), A. N.'t Woord (The Netherlands), and the proposer. There were three incorrect solutions.

Angles inside a convex n-gon **October 1993**

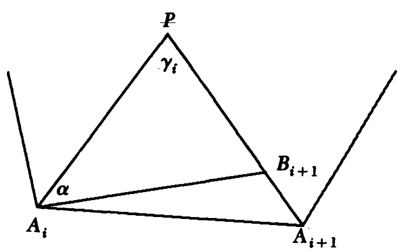
1429. *Proposed by Wee Liang Gan, student, Singapore.*

Let P be a point inside the convex n-gon $A_1A_2 \cdots A_n$. Prove that at least one of the angles $\angle PA_iA_{i+1}$, $i = 1, 2, \ldots, n$ is less than or equal to $(1/2 - 1/n)\pi$. (All subscripts are taken modulo n.)

Solution by The Javelina Problem Solvers, Texas A&M University, Kingsville, Texas.

Let α denote the smallest of the angle $\angle PA_iA_{i+1}$ and let $\gamma_i = \angle A_iPA_{i+1}$. For $i = 1, 2, \ldots, n$, pick B_{i+1} on PA_{i+1} so that $\angle PA_iB_{i+1} = \alpha$ (see figure).
Then, using the law of sines,

$$\frac{PA_i}{PA_{i+1}} \le \frac{PA_i}{PB_{i+1}} = \frac{\sin(\alpha + \gamma_i)}{\sin \alpha}.$$



It follows that

$$1 = \prod_{i=1}^{n} \frac{PA_i}{PA_{i+1}} \le \prod_{i=1}^{n} \frac{\sin(\alpha + \gamma_i)}{\sin \alpha} \le \left(\frac{\sin(\sum_{i=1}^{n}(\alpha + \gamma_i)/n)}{\sin \alpha} \right)^n = \left(\frac{\sin(\alpha + 2\pi/n)}{\sin \alpha} \right)^n$$

where, in the last inequality, we have used a special case of Jensen's inequality (see, for example, *Maxima and Minima Without Calculus*, Ivan Niven, MAA Dolciani Series, No. 6, 1981, pp. 92–95).

Thus, $\sin \alpha \le \sin(\alpha + 2\pi/n)$. We conclude that $\alpha \le \pi/2 - \pi/n$; if $\alpha > \pi/2 - \pi/n$, then $\sin(\alpha + 2\pi/n) < \sin(\pi/2 + \pi/n) < \sin \alpha$, a contradiction.

Also solved by Walter Blumberg, Con Amore Problem Group (Denmark), Jiro Fukuta (Japan), Richard Holzager, Pavlos B. Konstadinidis (student), O. P. Lossers (The Netherlands), Andreas Müller (Switzerland), Allan Pedersen (Denmark), Michael Vowe (Switzerland), and the proposer. There was one incorrect solution.

A Newton-Raphson Recurrence **October 1993**

1430. *Proposed by David Doster, Choate Rosemary Hall, Wallingford, Connecticut.*

Solve the recurrence

$$x_{n+1} = \frac{(x_n^4 + 18x_n^2 + 9)}{4x_n^3 + 12x_n}, \qquad n \ge 0, \quad x_0 = 2.$$

Solution by Stephen C. Locke, Florida Atlantic University, Boca Raton, Florida.

We will show that $x_n = p_n/q_n$, where p_n and q_n are integers such that $p_n + q_n\sqrt{3} = (2+\sqrt{3})^{4^n}$.

Let $x_n = p_n/q_n$, for integers p_n and q_n. Then

$$\frac{p_{n+1}}{q_{n+1}} = \frac{p_n^4 + 18p_n^2q_n^2 + 9q_n^4}{4p_n^3q_n + 12p_nq_n^3}.$$

It is natural to set $p_{n+1} = p_n^4 + 18p_n^2q_n^2 + 9q_n^4$, and $q_{n+1} = 4p_n^3q_n + 12p_nq_n^3$. Then

$$\begin{aligned} p_{n+1} \pm q_{n+1}\sqrt{3} &= p_n^4 + 6p_n^2\left(q_n\sqrt{3}\right)^2 + \left(q_n\sqrt{3}\right)^4 \pm 4p_n^3\left(q_n\sqrt{3}\right) \pm 4p_n\left(q_n\sqrt{3}\right)^3 \\ &= \left(p_n \pm q_n\sqrt{3}\right)^4. \end{aligned}$$

Since we may choose $p_0 = 2$ and $q_0 = 1$, this establishes

$$p_n + q_n\sqrt{3} = \left(2+\sqrt{3}\right)^{4^n}, \qquad p_n - q_n\sqrt{3} = \left(2-\sqrt{3}\right)^{4^n}.$$

Solving for p_n and q_n yields

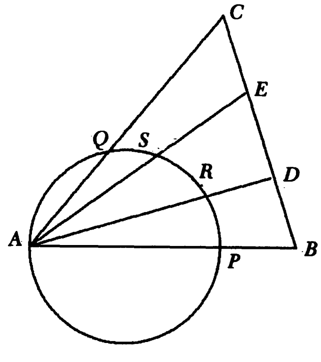
$$x_n = \frac{p_n}{q_n} = \sqrt{3}\,\frac{\left(2+\sqrt{3}\right)^{4^n} + \left(2-\sqrt{3}\right)^{4^n}}{\left(2+\sqrt{3}\right)^{4^n} - \left(2-\sqrt{3}\right)^{4^n}}.$$

Also solved by the Aardvork Problem Solving Group, Anchorage Math Solutions Group, David H. Arnold, Charles Ashbacher, J. C. Binz (Switzerland), Walter Blumberg, Paul Bracken (Canada), Con Amore Problem Group (Denmark), Jim Conklin, Paul Deiermann, Robert L. Doucette, David Graves, E. S. Freidkin, Richard Heeg, Richard Holzager, Joe Howard, The Javelina Problem Solvers, Hans Kappus (Switzerland), Pavlos B. Konstadinidis (student), Kee-Wai Lau (Hong Kong), Norman F. Lindquist, O. P. Lossers (The Netherlands), Beatriz Margolis (France), Reiner Martin (student) Patrick Dale McCray, Howard Morris, Jeremy Ottenstein (Israel), Phillip P. Ray, Herman Roelants (Belgium), Heinz-Jürgen Seiffert (Germany), Nicholas C. Singer, J. Sriskandarajah, John S. Sumner, Michael Vowe (Switzerland), WMC Problems Group, Lamarr Widmer, Kenneth S. Williams (Canada), Donald F. Winter, A. N.'t Woord (The Netherlands), Staffan Wrigge (Sweden), and the proposer. There was one unsigned solution. Several solvers interpreted the problem as asking for the limit of the sequence and found it to be $\sqrt{3}$.

Cevians and ratios **October 1993**

1431. *Proposed by Jiro Fukuta, Gifu-ken, Japan.*

In the given triangle *ABC*, let *AD*, *AE* be any cevians from *A* to *BC*. A circle drawn through *A* cuts *AB*, *AC*, *AD*, *AE*, or their extensions, at the points *P*, *Q*, *R*, *S*, respectively .



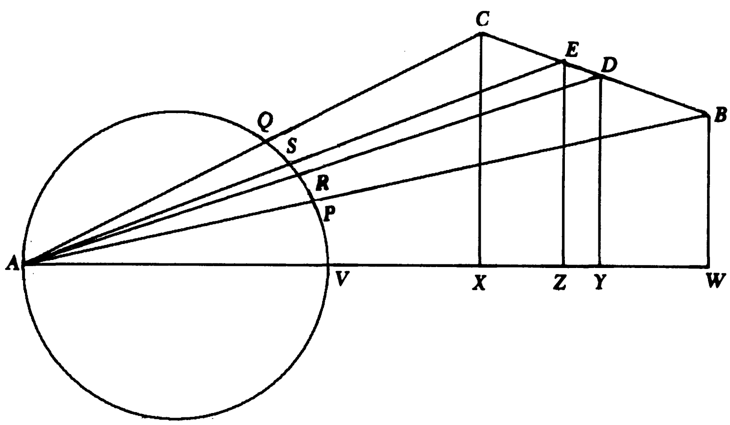
Prove that

$$\frac{AP \cdot AB - AR \cdot AD}{AS \cdot AE - AQ \cdot AC} = \frac{BD}{EC},$$

where $AP, AB, \ldots$ denote the lengths of the directed line segments $AP, AB, \ldots$.

Solution by David Jeremiason, student, Augustana College, Sioux Falls, South Dakota.

Let AV be a diameter of the circle, and let W, X, Y, Z be the feet of the perpendiculars from B, C, D, E, respectively, onto AV, or its extension (if V and P coincide, then so do B and W, etc.).



Then the triangles AVP, AVQ, AVR, AVS, are similar, respectively, to ABW, ACX, ADY, AEZ. Hence

$$\frac{AP}{AV} = \frac{AW}{AB}, \quad \frac{AQ}{AV} = \frac{AX}{AC}, \quad \frac{AR}{AV} = \frac{AY}{AD} \quad \frac{AS}{AV} = \frac{AZ}{AE}.$$

Therefore,

$$\frac{AP \cdot AB - AR \cdot AD}{AS \cdot AE - AQ \cdot AC} = \frac{AV \cdot AW - AV \cdot AY}{AV \cdot AZ - AV \cdot AX} = \frac{AW - AY}{AZ - AX} = \frac{YW}{XZ} = \frac{BD}{EC}.$$

Also solved by the Anchorage Math Solutions Group, Walter Blumberg, John F. Goehl, Jr., Richard Holzsager, The Javelina Problem Solvers, Emil F. Knapp, O. P. Lossers (The Netherlands), Gábor Pete (student, Hungary), Michael Vowe (Switzerland), Harry Weingarten, The Westmont College Problem Solving Group, and the proposer.

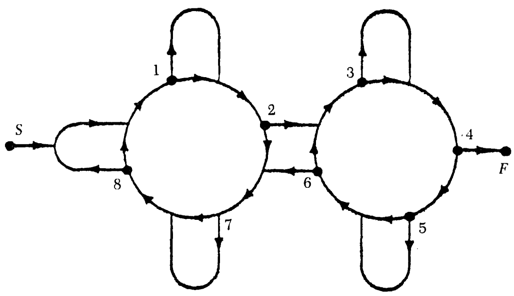
Wrong Turns in Roundabout Roadway **October 1993**

1432. *Proposed by Elliott A. Weinstein, Baltimore, Maryland.*

The map shown below is that of a notorious double-roundabout roadway in northern England. Consider a random "trip" on this roadway, which starts at S and finishes at F. Assume that the probability of choosing a particular continuation at any decision node is $1/2$. Define a *wrong turn* at any node as the selection which results in a trip to F longer than the minimal trip to F from that node. Let the word *route* mean the immediate sideroad that follows a wrong turn at node 1, 3, 5, 7, or 8.

a. Find the expected number of wrong turns.

b. Find the expected number of wrong turns, given that any particular route will have at most one wrong turn made onto it.



Solution by Richard Holzsager, The American University, Washington, D.C.

We will look at the expectations in which we are interested in terms of the contribution from the jth visit to the ith node, which is half the probability of making the visit. To calculate this, note that the probability of returning for another visit to a node after the jth visit is independent of j. Therefore, if we use p_i to denote the probability of the first visit to node i, and q_i to denote the probability of return, then the probability of the jth visit is $p_i q_i^{j-1}$.

For $i = 1$ or 2, $p_i = 1$, but q_i is slightly harder to calculate. To return after n trips around the second rotary, the probability is $1/2^{2n+1}$ (one factor for the turn at 2, n for the turns at 4 and n for the turns at 6). These form a geometric series with sum $q_i = 2/3$. At this point we can note that for $i = 7$ or 8, this same calculation shows that $p_i = q_i = 2/3$.

For $i = 3$ or 4, p_i is again 1, and the probability q_i of return is simply that of making a wrong turn at 4, that is $1/2$. Similarly, $p_i = q_i = 1/2$ for $i = 5$ or 6.

To solve (a), we need only multiply each $p_i q_i^{j-1}$ by $1/2$ and sum over i and j. For each i, the series is geometric, with respective sums $3/2$, $3/2$, 1, 1, $1/2$, $1/2$, 1, 1, giving the answer 8.

For (b), the contributions of the jth visit to vertex 1, 3, 5, 7, or 8 is $1/2j$ (half the probability that no wrong turns have yet been made at this vertex), so we have to sum $p_i q_i^{j-1} 2^{-j}$ at these vertices. Again, each series is geometric, with sums $3/4$ at 1, $2/3$ at 3, $1/3$ at 5, and $1/2$ at 7 and 9. Adding these to the previously calculated sums at nodes 2, 4, and 6 gives $23/4$.

Part (a) also solved by Michael Andreoli, Con Amore Problem Group (Denmark), Robert L. Doucette, Milton P. Eisner, Herbert Gintis, and Michael Vowe (Switzerland). Part (a) may be viewed as an absorbing Markov chain. Parts (a) and (b) also solved by David Callan, The Javelina Problem Solvers, O. P. Lossers (The Netherlands), WMC Problems Group, and the proposer.

Answers

Solutions to the Quickies on page 304.

A823. The AM-GM inequality gives

$$n^k = \frac{(n^{k+1}-1)+1}{n} = \frac{(n-1)(n^k+n^{k-1}+\cdots+1)+1}{n}$$
$$= \frac{(n^k+n^{k-1}+\cdots+1)+\cdots+(n^k+n^{k-1}+\cdots+1)+1}{n}$$
$$\geq (n^k+n^{k-1}+\cdots+1)^{(n-1)/n}.$$

A824. Obviously, one set of solutions is $x = y$. To obtain all the others, we let $x = 1/u$ and $y = 1/v$ to give $u^v = v^u$ where $v > u$. It is known that all the solutions of the latter equation are given by

$$v = (1 + 1/m)^{m+1}, \qquad u = (1 + 1/m)^m, \qquad m = 1, 2, \ldots .$$

(See W. Sierpinski, *Elementary Theory of Numbers*, Hafner, New York, 1964, 106–107.)

Hence

$$x = (1 + 1/m)^{-m}, \qquad y = (1 + 1/m)^{-1-m}.$$

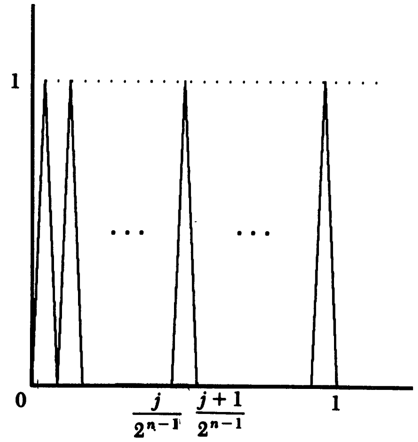
A825. After n iterations, the graph of $T^n(x)$ has 2^{n-1} spikes, as shown below.

Each spike can be represented analytically by

$$T_j^{(n)}(x) = \begin{cases} 2^n x - 2j & j/2^{n-1} < x \le (2j+1)/2^n, \\ -2^n x + 2(j+1) & (2j+1)/2^n < x \le (j+1)/2^{n-1}, \\ 0 & \text{otherwise}, \end{cases}$$

for $j = 0, 1, 2, \ldots, 2^{n-1} - 1$. Since there are 2^{n-1} spikes, the total area is

$$\int_0^1 T^{(n)}(x)\,dx = 2^{n-1}\left(\frac{1}{2} \cdot \frac{1}{2^{n-1}} \cdot 1\right) = \frac{1}{2}.$$



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.

Dillon, Sam, Perfect score for Americans in world math tourney, *New York Times* (20 July 1994) B7; National Edition, A16.

The six U.S. participants in the 1994 International Mathematical Olympiad in Hong Kong all received perfect scores! This was the first time in the 35-year history of the competition that an entire team received perfect scores. Although the U.S. team members were all boys—as has been true in the past—girls from Austria and the U.K. also achieved perfect scores. Sample question: "Show that there exists a set A of positive integers with the following property: For any infinite set S of primes there exist two positive integers m in A and n not in A, each of which is a product of k distinct elements of S for some $k > 1$."

Cipra, Barry, New proof makes light work of partial Latin squares, *Science* 265 (1 July 1994) 29.

In an $n \times n$ array, let each cell have an associated set of n different symbols available to choose from; different cells may offer different sets of symbols. In 1977, Jeff Dinitz, now at the University of Vermont, conjectured that in every such situation, it is always possible to choose a symbol for each cell so that no symbol appears twice in any row or column—thereby forming a *partial Latin square*. Fred Galvin (University of Kansas) has now given a *one-page* proof of the conjecture, by putting together an earlier result in the literature with 1992 work of Jeannette C.M. Janssen (Concordia University), who showed that $(n+1)$ choices at each cell suffice. Galvin's achievement shows the importance of researching the literature, since the *second* paper he looked up had the result that he needed.

Stewart, Ian, *Game, Set, and Math: Enigmas and Conundrums*, Basil Blackwell, 1989; viii + 191 pp, $19.95. ISBN 0-631-17114-2. *Another Fine Math You've Got Me Into ...*, W.H. Freeman, 1992; xi + 269 pp, $13.95 (P). ISBN 0-7167-2341-7.

Ian Stewart was inspired by Martin Gardner, who highly recommends him, and is now Gardner's successor as a columnist for *Scientific American*. Both of these books consist of columns from *Pour la Science*, the French national edition of *Scientific American*. Stewart writes a dozen columns a year, but the U.S. edition publishes only half (alternating them with "The Amateur Scientist"), while the French and Spanish editions publish all twelve. These books contain the extra columns, translated back into English, with the added French jokes deleted and the original English puns restored. So, unless you already subscribe to *Pour la Science*, in order to get your full helping of Stewart's marvelous mathematical dialogues, you must not only read *Scientific American* but also buy these volumes.

Peterson, I., Last word not yet in on Fermat's conjecture, *Science News* 145 (25 June 1994) 406-407.

As you read this, Andrew Wiles is continuing his lectures at Princeton on his approach to Fermat's Last Theorem and should finally be approaching the point in the proof where the gap lies. Will he be able to give the Last Word?

Cipra, Barry, Pinning down a treacherous border in logical statements, *Science* 264 (27 May 1994) 1249. Kirkpatrick, Scott, and Bart Selman, Critical behavior in the satisfiability of random Boolean expressions, *Science* 264 (27 May 1994) 1297–1301.

Any Boolean expression can be put into an equivalent *conjunctive normal form*, which has the form of a conjunction of clauses each of which is a disjunction of variables. Consider such an expression with N variables, M clauses, and k disjuncts in each clause. Is there some assignment of truth values to the variables that will make the expression true? The more variables, the harder it is to determine the satisfiability of the expression; but the ratio $\alpha = M/N$ of clauses to variables has a curious effect. For small α, most expressions are satisfiable; for large α, most are not. For both small and large α, the satisfiability algorithms run quickly, compared to intermediate values. In particular, there appears to be a threshold effect for the transition from satisfiability to unsatisfiability, which Kirkpatrick and Selman compare to the phase transitions of the spin glasses of theoretical physics.

Cipra, Barry A., "The magic words are squeamish ossifrage," *SIAM News* 27 (6) (July 1994) 1, 12–13. Kolata, Gina, 100 quadrillion calculations and, Eureka! problem solved, *New York Times* (National Edition) (27 April 1994) A11. Peterson, I., Team sieving cracks a huge number, *Science News* 145 (7 May 1994) 292–293. Taubes, Gary, Small army of code-breakers conquers a 129-digit giant, *Science* 264 (6 May 1994) 776–777.

The title of Cipra's article is the message that was encoded in 1977 using the RSA cryptosystem and a 129-digit integer called RSA-129. This past April, an international distributed computing effort factored the number and decoded the message. The main method, the quadratic sieve, led to a sparse matrix with half a million rows and columns, which in turn was converted into a matrix with only 186,000 rows and columns. The final step was to look for linear dependencies of the rows, some of which give rise to factors of the original number. A new tool, the number field sieve, is being geared up for trials on RSA-130, a 130-digit number in the same family. (An ossifrage, by the way, is a kind of hawk.)

Dewdney, A.K., *The (New) Turing Omnibus: 66 Excursions in Computer Science*, W.H. Freeman, 1993; xvi + 455 pp, \$24.95 (P). ISBN 0-7167-8271-5.

This book "is a vehicle of learning that visits the major landmarks of computer science. The tour stops at monuments of theory, visits major techniques, and travels the avenues of application." Each chapter treats a topic of computer science, and this new tourbus features six new chapters. Each chapter contains exercises, and a solutions manual is available. Every student of computer science needs a season pass on this bus!

Livingston, Charles, *Knot Theory*, MAA, 1993; xviii + 240 pp, \$31.50. ISBN 0-88385-027-3. Willerton, Simon, A topological tie-in, *Nature* (10 March 1994) 103–104.

Livingston's book and Willerton's too-brief survey sketch complement each other nicely. The book is an elegant exposition of knot theory as pure mathematics, detailing geometric, algebraic, and combinatorial techniques, with plenty of exercises. The article, on the other hand, considers knot theory as applied mathematics, from a topological point of view. Physicists are interested in knot theory because the probability of a particle traversing a knotted path in three-dimensional spacetime depends only on the type of knot. The article introduces a different perspective (knot space), a new development (Vassiliev invariants, from which the Jones polynomial discussed in the book is constructed), tools physicists use (Feynman path integral, Witten integral), and the expansion of a knot (as a sum of chord diagrams).

Windrip, Michael, S.A.T. increases the average score, by fiat, *New York Times* (11 June 1994) 1, 10.

Beginning April 1995, the Scholastic Aptitude Test will be "recalibrated" by adjusting the mean score to 500, from the present 424 verbal and 478 mathematics. Says a representative: "The question people will ask is, 'Aren't you just making kids feel better by giving them higher scores?' The answer is, absolutely, positively not. The performance that generates a 424 today will now generate a 500. The kid is no brighter, doesn't have any more bright answers. It's just the label is higher. Everyone will know."

Cipra, Barry, Mathematicians get an on-line fingerprint file, *Science* 265 (22 July 1994) 473.

No, we're not all getting fingerprinted. The title is journalese for N.J.A. Sloane's email service for identifying integer sequences. The first program (`sequences@research.att.com`) compares the sequence you supply (send a message like `lookup 1 2 3 4 5`) with a table of 5,000 entries; the second (`superseeker@research.att.com`) goes beyond the table to try to transform your sequence into one on the list or to find a formula that fits.

Toom, Andrei, A Russian teacher in America, *Journal of Mathematical Behavior* 12 (1993) 117–139.

An experienced instructor from abroad described to me last year how an associate dean where she was visiting intimidated her into raising grades, to a full letter grade above what she felt the students deserved (she taught in the associate dean's department, and some of his advisees complained to him about their grades; curiously, the head dean had spent the preceding year admonishing the faculty about grade inflation). Hence the impassioned remarks of Andrei Toom, a Russian immigrant, strike a resonant chord. From him, we learn how some foreign professors truly feel about U.S. students and our educational system, as distinct from what they do to get along or stay here (my friend needed a favorable recommendation from the associate dean). Toom speaks well of teaching graduate computer science at Boston University, but business calculus at a "huge state university" was discouraging. He found many students credential- and grade-directed, not learning-oriented, and U.S. education "molded by the pressures of those students who did not hide their contempt of the very idea of learning ... [Some students] seem to think that they BUY grades and PAY for them by learning. And they try to PAY as little as possible. ... In another state students complained about their mathematics teacher, another newcomer from Russia: 'We pay as much as others, but have to know more than they for the same grade.'" Most students could not solve routine algebra word problems and had no interest in learning to do so, much less in tackling nonstandard problems—just following recipes made them content. "I had to learn by trial and error, how much of elementary mathematics was taboo ... I could not imagine that students who take 'calculus' were not supposed to know trigonometry ... Thus, I could discuss the equation $y'' - y = 0$, but not $y'' + y = 0$." Toom emphasizes the importance to education of cultural and parental attitudes and priorities. "One foreigner, experienced in teaching Americans, advised me in a friendly manner: 'Listen, don't ask for trouble. Education in this country is not our concern.' Of course, he teaches in a much more productive way in his own country. ..." Toom's disappointing experience was with a subpopulation (the 25% of American students who major in business); but the attitudes that he points out are widespread; they arise in and are supported by pragmatism in U.S. culture and its "sibling society" (the term is Robert Bly's) subculture. (Thanks to Domenico Rosa of Teikyo Post University.)

Angier, Natalie, Why birds and bees, too, like good looks, *New York Times* (National Edition) (8 February 1994) B5, B8. Not just another beauty contest, B8.

Recent studies indicate that mirror-image symmetry is an element of sexual attraction among some insects and animals, perhaps because it is an indicator of good health or good genes. Among humans, a preliminary study of U.S. college students indicates both that students find symmetrical faces more attractive and that those with such faces lose their virginity earlier and have more sexual partners.

White, Alvin M. (ed.), *Essays in Humanistic Mathematics*, MAA, 1993; xii + 212 pp, $24 (P). ISBN 88385-089-3

"Humanistic mathematics is not just 'friendly math' or 'touchy-feely' math. It is mathematics with a human face." This volume is the culmination of seven years of informal meetings and an informal newsletter among mathematicians interested in making the student more of an inquirer and in "acknowledging the emotional climate of the activity of learning." The 22 essays here discuss the humanistic aspects of mathematics, the relations of mathematics to other elements of culture, the inner life of mathematics, teaching and learning experiences, and how to vivify the mathematics of the past.

Fauvel, John, Raymond Flood, and Robin Wilson (eds.), *Möbius and His Band: Mathematics and Astronomy in Nineteenth-Century Germany*, Oxford University Press, 1993; 172 pp, $29.95. ISBN 0-19-853969-X

August Möbius (1790–1868) took part in the rise of German mathematics. "It is precisely because he was not outstanding in any particular way, but was a serious, competent, professional scholar, that Möbius is such a good mirror of his time." This collection of six essays attempts to illuminate his times by focusing on him. One essay centers on details of his life, another on the German mathematical community, and a third on the revolution in astronomy during that time (Möbius was professor of astronomy at Leipzig). The other three treat Möbius's geometrical mechanics, the development of topology, and Möbius's modern legacy (M function, M transformations, M strip, M nets, M duality, M inversion formula, and barycentric coordinates; but no M theorem).

Lorimer, Peter, *Making Mathematical Models*, Dunmore Press, 1990; 32 pp, U.S. $4 (or equivalent), including postage (P) (from Peter Lorimer, Dept. of Mathematics, University of Auckland, P.O. Box 92019, Auckland, New Zealand).

This booklet contains instructions for using hat elastic and straws to construct the regular polyhedra, hypercube, snub cube, cuboid, various graphs (Petersen, Fano, Pappus, Desargues), and more. Most sections conclude with a question or exploration.

Riordan, Teresa, Patents: Method to speed computer processes, *New York Times* (National Edition) (10 January 1994) C2.

Dao-Long Chen and Robert D. Waldron of the NCR Corporation have received a patent (5,268,857) for a simplified version of the Newton-Raphson method that cuts by one-half the time to calculate a square root. (Recall that computer graphics makes much use of square roots in manipulating and presenting objects.) Now, for 17 years, you won't be able to use their method in a product—even if you discover it independently yourself—without paying to license the rights from NCR. Perhaps Chen and Waldron had similar motivation; this patent is apparently the third for a square-root algorithm.

NEWS AND LETTERS

LETTERS TO THE EDITOR

Dear Editor:

In its February 1994 issue, this MAGAZINE published an article by Underwood Dudley, entitled "Smith Numbers" [4], in which the author offered his explanation as to why the subject has generated the interest it has received. His conclusion, however, that professional mathematicians are drawn to the problem, or, by extension, any problem that could in some sense be classified as recreational, do so because they are *not able to do something of significance* widely misses the mark.

Interest in the Smith number problem has been basically of two kinds. First, there has been an interest in discovering algorithms for the construction of large Smith numbers and in discovering special classes of Smith numbers--results of this kind belong, essentially, to the category of mathematics known as "recreational" (although the dividing line between "recreational" and "serious" mathematics is not absolutely clear). Second, I, and, to some extent, the late Samuel Yates, have had an interest in certain more general problems of which the Smith number problem is a special case.

The question considered, in its full generality, has to do with the relationship of the digits of an integer in an arbitrary base to the digits of its prime factors. Because the problem involves a consideration of *digits*, the problem may appear to belong to recreational mathematics and be unworthy of the attention of professional mathematicians. Dickson's [3] chapter on the properties of the digits of numbers, the *SIAM J. of Appl. Math.* paper by Stolarsky [9], and some of its more than 60 references should convince the reader otherwise.

Professor Dudley's first objection is to any generalization of Smith numbers. He is critical of the fact that in [6], where I proved the existence of infinitely many Smith numbers, I defined the slightly more general "k-Smith number." The concept wasn't generalized for the sake of generalization, but rather because it was clear that my proof actually showed a more general relationship; the relationship is not devoid of interest so it was reasonable to establish the more general result.

With certain qualifications, I would agree that Smith numbers are not serious mathematics. However, I can't agree that "...those who investigate Smith numbers are not trying to penetrate deep into the secrets of integers; they are instead observing mere accidents of their representation in an arbitrary system." I mention the fact that I published an article in [7], establishing several results, of which I will state just one:

Let $b \geq 8$. If $S(b,m)$ denotes the sum of digits function base b, $m = p_1 \cdots p_k$ (p_i prime, not necessarily distinct), and $S_p(b, m) = \Sigma_{i=1}^{k} S(b, p_i)$, then the set

$$\{S(b, m)/S_p(b, m) \mid m \text{ an integer base } b\} \cap [0, 1]$$

is dense in the interval $[0, 1]$.

My understanding of the relationships necessary to prove the theorems in this article was, in large part, obtained through my investigation of Smith numbers, and I submit that some insight into the "secrets of the integers" was indeed required in order to prove the theorems.

It is important, because of the erroneous impression this article imparts, that I comment on the conclusion that the individuals who have been interested in Smith numbers are unable to do something of significance and therefore latch on to Smith numbers. Dudley's references list only Samuel Yates and me as authors (since 1983) of an article in which Smith numbers are considered and which is not clearly "recreational mathematics." Sam Yates, the author of four articles on Smith numbers, is no longer here to speak for himself. He was a man who, like most mathematicians (and many nonmathematicians), loved a good challenge, and, although he was not trained as a mathematician, he had a very thorough mastery of the mathematics in which he was interested. Anyone who doubts that should read his little book *Repunits and Repetends*; he was, in addition, the author of over 40 journal articles.

I would now like to make a few remarks concerning how it is possible to have the views expressed by Dudley in his concluding statements. My remarks will still hold if the reader replaces "Smith numbers" in Dudley's statements with "x" and lets the variable assume as a value any topic in mathematics in which he/she has had an interest.

Work in mathematics rarely can be judged by looking at the *problem*. All research mathematicians are aware that the result is often only as interesting as the method whereby the result was obtained. If a proof is "new" and "clever", or yields some unexpected insight, or introduces a method or approach that has applicability to other problems, then the cause of mathematics is advanced. We are all keenly aware that much good mathematics has come out of attempts to solve relatively "unimportant" problems. It is almost trite to mention how interest in Fermat's Last Theorem led to the development of the theory of algebraic numbers. On a smaller scale, consider how interest in the Farey sequence led to advances in approximation of irrationals. Or consider, from the more recent past, J.H.E. Cohn's interest in perfect squares in the Fibonacci sequence [1], which led to what has become known as Cohn's (simpler) approach to solving certain Diophantine equations (see, e.g. [2]). There are literally hundreds of examples of this sort, and most readers can, in fact, supply examples of their own illustrating my point. Mathematics is richer for those efforts that are motivated by curiosity and the fun and excitement of trying to understand an observed phenomenon; that activity should be encouraged.

> *It is ... a melancholy experience for a professional mathematician not to be able to do something of significance. The answer, then, to 'Why Smith numbers?' is 'We have to do* something.*'*

Finally, Dudley's statement suggests a profound lack of understanding of what motivates the research mathematician. An interest in problems that are not "significant" problems does not correlate with a lack of competency! Rather, it is associated with *curiosity* and the desire to unlock the hidden mysteries therein. It is simply not true that mathematicians who are able to "do something of significance" will show no interest in any of the many problems that are not themselves significant; quite the contrary. (Does anyone really believe that all of the contributions of Euler, Fermat, Eisenstein, and Gauss to number theory were the result of their interest in "significant" problems?)

The question "why Smith numbers?" has not been answered by Dudley. The answer has to do with the mathematician's innate curiosity, and the fact that there is enough substance there that some progress concerning the character or the distribution of these, or related concepts, is possible. Those who get caught up in the excitement of discovering previously unknown relationships in this problem, or in any of hundreds of other problems in mathematics, should not be dissuaded by the foolish suggestion that they are not able to do something more "significant." You are in good company! Aside from the fact that it is often impossible to tell whether interest in a minor problem will lead to something of considerable interest, one should keep in mind that the edifice called "mathematics" is being built a piece at a time; each builder contributes to its construction. And the block that each of us adds may have value in and of itself, or because of the tool we have used to shape it, or perhaps because of the idea/inspiration it gives to someone else who sees it.

Although I have not had an active interest in digital problems for several years, I do want to say one final word about problems relating to the Smith number problem. In 1989, Yates and I published in *Nieuw Archief voor Wiskunde* [5] an article in which we gave a history of the sum of digits function, with an emphasis on those relationships that could conceivably be of value in investigating a certain rather general set of integers (which includes Smith numbers, Mersenne primes, and the squares of Fermat primes, base 2). We obtained explicit formulas for S(b,m) for two classes of integers whose digits may not be known, and then proceeded to apply the theorems to one subset of our set. Several open questions are mentioned in the paper, and a table of values of $S_p(10,10^n\text{-}1)$ (notation defined above) for all n for which the repunit R_n has been completely factored is included. Since the paper is partially an expository paper, it includes an extensive bibliography containing not only references of direct relevance, but also those of peripheral interest. The paper is quite accessible to the nonspecialist (as is required by *NAvW* of their expository papers). If there are readers who are interested in digital problems, they may find this paper of interest.

Wayne L. McDaniel
University of Missouri-St. Louis
St. Louis, MO

REFERENCES

1. Cohn, J.H.E., On square Fibonacci numbers, etc., *Fibonacci Quarterly* 2 (1964), 109-113.
2. Cohn, J.H.E., Eight Diophantine equations, *Proc. London Math. Soc.* (3)16(1966), 153-166.
3. Dickson, L.E. *History of the Theory of Numbers*, Vol. 1, Chelsea (reprint), New York, 1952.
4. Dudley, U., Smith numbers, this MAGAZINE 67(1994), 62-65.
5. Hardy, G.H., *A Mathematician's Apology*, Cambridge University Press, London, 1967.
6. McDaniel, W.L., The existence of infinitely many k-Smith numbers, *Fibonacci Quarterly* 25(1987), 76-80.
7. McDaniel, W.L., Difference of the digital sums of an integer base b and its prime factors, *J. of Number Thy.* 31(1989), 91-98.
8. McDaniel, W.L. and S. Yates, The sum of digits function and its application to a generalization of the Smith number problem, *Nieuw Archief voor Wiskunde* 7(1989), 39-51.
9. Stolarsky, K.B., Power and exponential sums of digital sums related to binomial coefficient parity, *SIAM J. Appl. Math.* (4)32(1977), 717-730.

Dear Editor:

I am writing this letter with reference to the article "Smith Numbers," by Underwood Dudley (February 1994).

We should be glad that in the past no one had discouraged mathematicians by saying: "Those who are trying to determine solutions in nonzero integers of the equation $x^2 + y^2 = z^2, x^3 + y^3 = z^3$, etc., are not penetrating deep into the theory of the solution of polynomial diophantine equations in three variables; they are instead trying to observe the result of mere accidents of assigning the value 1 and -1 to some coefficients and 0 to others. This is nothing but recreational mathematics. However, it is ominous that some have now ventured to consider the generalization $x^n + y^n = z^n$. They are, in fact, people with limited talent." (On second thought, there must have been some detractors in every epoch of mathematics but most investigators must have advanced mathematics by ignoring them.)

The theory of matrices is about the observations of a representation that Arthur Cayley used "accidentally" for some equations of transformation and, if the above argument is used, it should be relegated to the realm of recreational mathematics. The same can be said about the theory of Fourier series, where the choice of (a basis of) sines and cosines was arbitrary. Many more examples can be given. In fact, all rigorous mathematics is about observations that can be made by "accidentally" assuming certain axioms from an unlimited pool of axioms. No single choice of (consistent and independent) axioms is "better" than another except for its historical relevance to the need at that period of time (which is, in a mathematical context, accidental).

When Michael Faraday, the inventor of the electric motor, exhibited his invention, which appeared to be a toy at that time, a woman (who was not well-read in science) asked, "What is the earthly purpose of this?" The inventor replied, "Madam, what is the earthly use of a newborn baby?"

R. Kit Kittappa
Millersville University
Millersville, PA

Dear Editor:

This is a postscript to Richard Guy's interesting note, "There Are Three Times as Many Obtuse-angled Triangles as There Are Acute-angled Triangles"[June 1993, pp.175-179]. This is apparently true in the plane but not in higher dimensional spaces. In *SIAM Reviews* problem 92-3* [March 1992, March, 1993], D. J. Newman asked to show that for "most continuous distributions", a random triangle is more probable to be obtuse. The following are some of the editorial comments on the problem.

Related to this problem is Problem 85-4* by Bates and Forest, *Math. Intelligencer* [8(1986)49; 9(1987)63], "Three points are chosen at random, independently, and uniformly with respect to surface content, on an n-sphere in E^{n+1}. What is the probability P_n that the plane triangle determined by the points is acute?" Michael P. Lamoureux gave an infinite series representation for P_n and the numerical results

$$P_1 = 0.25,\ P_2 = 0.5,\ P_3 \approx 0.6553,\ P_4 \approx 0.7571,\ P_5 \approx 0.8264,\ P_6 \approx 0.8746.$$

That $P_n \to 1$ as $n \to \infty$ was indicated by Bert Fristedt who noted that the three vectors from the center of the sphere to the three random points for large n will be close to mutually orthogonal vectors with probability close to 1. Hence the triangle they form will be close to an equilateral triangle with probability close to 1.

We note that if one is not restricted to continuous distributions, then by choosing three points at random, equally likely, from the vertices of a regular n-dimensional simplex, the probability the three points are vertices of an equilateral triangle is

$$(n + 1)(n)(n - 1)/(n + 1)^3 = (1 - 1/n)/(1 + 1/n)^2,$$

which approaches 1 as $n \to \infty$.

Murray S. Klamkin
University of Alberta
Edmonton, Canada

Carl B. Allendoerfer Award

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in MATHEMATICS MAGAZINE. The Awards are named for Carl B. Allendoerfer, a distinguished mathematician at the University of Washington and President of the Mathematical Association of America, 1959-60.

Joan P. Hutchinson
"Coloring Ordinary Maps, Maps of Empires, and Maps of the Moon"
Mathematics Magazine 66 (1993), 211-226.

"How can one say something fresh and interesting about a well-known and over-reported topic like the four color problem? Joan Hutchinson gives us a splendid reply with her article. By looking at maps of empires, maps of moon colonies, and at circuit boards that have the same graph-theoretic structure as certain maps, she considers situations in which the celebrated four color theorem is false and provides new information and insights about coloring problems.

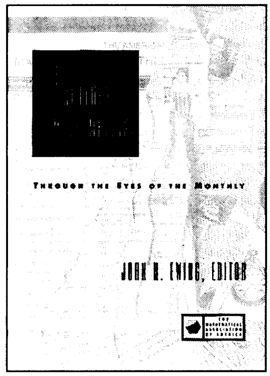
Professor Hutchinson's prose is crisp and direct, and the mathematics is essentially self-contained. All of the essential ideas are provided and explained in a clear and inviting fashion. The article offers both entertainment and enlightenment to students of mathematics and to seasoned faculty as well. It is an excellent model of expository writing."

Biographical Note

Professor Hutchinson writes: "I was first encouraged to color maps by my mathematician mother (who incidentally studied from a text by Allendoerfer and Oakley), but was introduced seriously to combinatorics, graph theory, and graph coloring by Professor Herbert Wilf, an expert guide and my thesis advisor at the University of Pennsylvania (1969-1973). While a graduate student and as a John Wesley Young Research Instructor at Dartmouth College (1973-75), I began research in topological and chromatic graph theory. Most of my learning and work in these areas took place in fruitful collaboration with Michael Albertson during my 15 years at Smith College. The actual writing of my *Mathematics Magazine* paper took place while I held one of the N.S.F. Visiting Professorships for Women at the University of Washington and as I began my current position in 1990 as Professor of Mathematics and Computer Science at Macalester College."

A Century of Mathematics
Through the Eyes of the Monthly

John Ewing, Editor



This is the story of American mathematics during the past century. It contains articles and excerpts from a century of the **Monthly**, giving the reader and opportunity to skim all one hundred volumes without actually opening them. It samples mathematics year by year and decade by decade. Along the way, readers can glimpse the mathematical community at the turn of the century, and the divisions between the mathematical communities of teachers and researchers. They read about the struggle to prevent colleges from eliminating mathematics requirements in the 1920's, the controversy about Einstein and relativity, the debates about formalism in logic, the immigration of mathematicians from Europe, and the frantic effort to organize as the war began. At the end of the war, they hear about new divisions between pure and applied mathematics, heroic efforts to deal with large numbers of new students in the universities, and the rise of federal funding for mathematicis. In more recent times, they see the advent of computers and computer science, the problems faced by women and minorities, and some of the triumphs of modern research.

This is a book about mathematics—about teaching and research, applied and pure, elite universities and community colleges. Browsing through its pages, readers see what has changed (the kinds of mathematics in fashion, for example) and what has stayed the same (our concern about teaching and our complaints about the deplorable state of our students).

This is a book about history—a sampling of history, meant to be savored rather than studied. For one hundred years, the Monthly has contained articles by some of the greatest mathematicians in the world, as well as articles by students and faculty from small midwestern colleges where those great names were barely known. This book gives a glimpse of both worlds. It tells a story rather than the details of history.

This is the story of a century of mathematics in America.

335 pp., Hardbound, 1994

ISBN 0-88385-457-0

List: $39.50 MAA Member: $32.00

Catalog Number CENTMA

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Qty.	Catalog Number	Price

Total $________

Name______________________________

Address ______________________________

City______________________________

State___ Zip Code________________

Payment ❑ Check ❑ VISA ❑ MASTERCARD

Credit Card No.______________________________

Signature______________________________

Exp. Date______________________________

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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