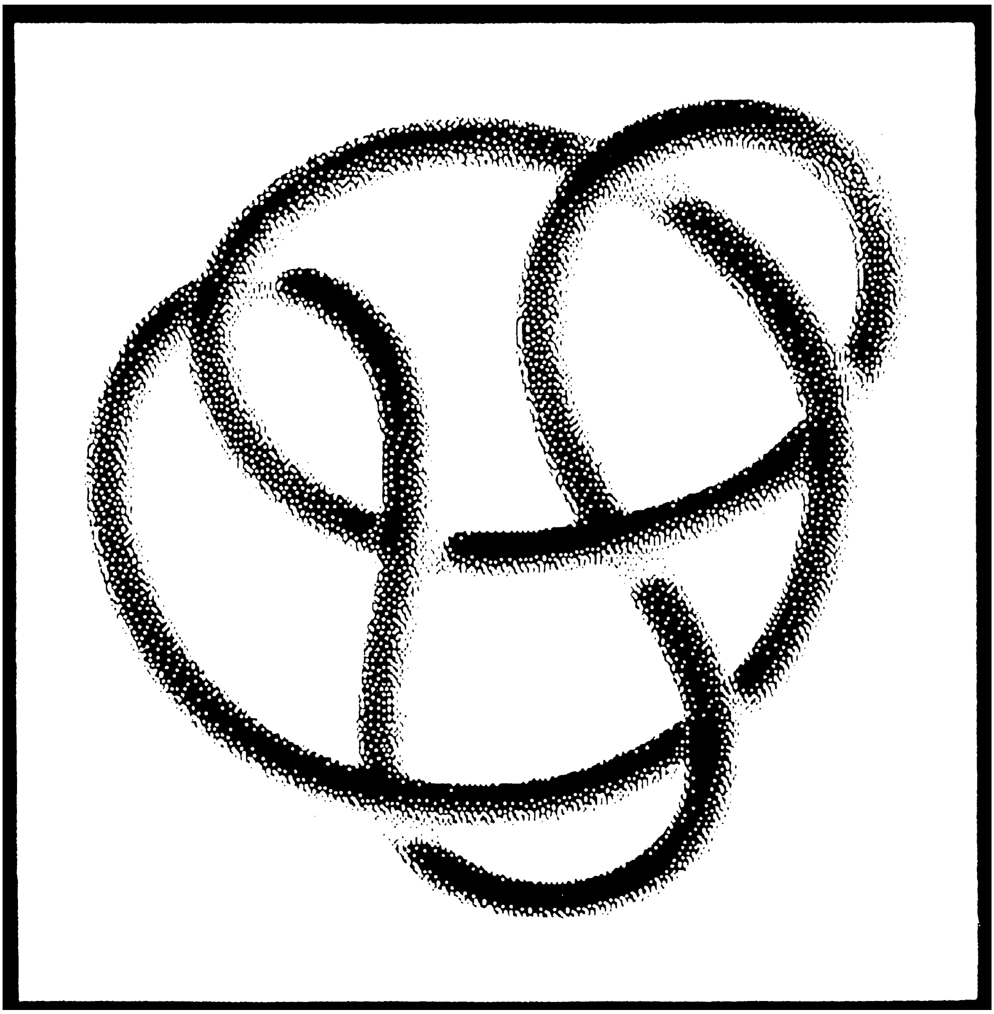
Vol. 61 No. 1 February 1988

MATHEMATICS MAGAZINE



Ishii

• Knots: Polynomial Invariants
• Love Affairs and Differential Equations
• Diagrams Venn and How

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

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

AUTHORS

Raymond Lickorish, an Englishman, is a product of the British educational system and, in particular, of the University of Cambridge. He is now the Cayley Lecturer in Pure Mathematics at that University. He is the Director of Studies at Pembroke College, where he is a Fellow, and he has recently survived being Chair of the Cambridge University Faculty Board of Mathematics. An enchantment with America, began with a year as a visiting Professor at the University of Wisconsin in Madison, in 1967, has continued happily with many short visits and with visiting posts at the University of California, both in Berkeley and in Santa Barbara. In research he specializes in geometric topology, the sort of topology where some intuitive visualization is at least possible; he has written research papers on 3-manifolds, on tangles, and on knots, and he supervises graduate work in that area. From time to time he has organized informal summer gatherings of topologists in Cambridge and one of them, in 1984, saw the genesis of some of the new results on link polynomials. During a discussion over tea one summer afternoon in the Mathematics Commons Room of the University of Cambridge with co-author **Kenneth C. Millett**, Professor of Mathematics at the University of California, Santa Barbara who was visiting Cambridge following a year-long sabbatical leave at the Institut des Hautes Etudes Scientifiques near Paris, a surprising similarity between the new Jones and much older Alexander polynomials was discovered. Thereupon began a research collaboration leading to this article describing their results and those of others. Professor Millett's research interests include the geometric topology of manifolds and the static and dynamic phenomena associated to them; parameterized families of geometric configurations, knots and knotting phenomena suggested by applications in the natural sciences, and analytic and topological properties of foliated structures in manifolds. He first became interested in topology while an undergraduate mathematics student at the Massachusetts Institute of Technology and wrote his Ph.D. thesis at the University of Wisconsin in Madison in topology. His research interests now include applications of knots to molecular biology and mathematical chemistry in addition to purely mathematical issues.

Vol. 61 No. 1 February 1988

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 $11 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The non-member/library subscription price is $28 per year. Bulk subscriptions (5 or more copies) are available to colleges and universities for classroom distribution to undergraduate students at a 41% discount ($6.50 per copy—minimum order $32.50).
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20077-9564. Back issues may be purchased, when in print, from P. and H. Bliss Company, Middletown, CT 06457. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.
Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, The Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

Second class postage paid at Washington, D.C. and additional mailing offices.
Postmaster: Send address changes to Mathematics Magazine Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20077-9564.
PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

The New Polynomial Invariants of Knots and Links

W. B. R. LICKORISH
University of Cambridge
Cambridge CB2 1SB
England

K. C. MILLETT*
University of California
Santa Barbara, CA 93106

The theory of knots and links is the analysis of disjoint simple closed curves in ordinary 3-dimensional space. It is the consideration of a collection of pieces of string in 3-space, the two ends of each string having vanished by being fastened together as in a necklace. Many examples can be seen in the diagrams that follow. If the strings can be moved around from one position to another those two positions are the same link or 'equivalent' links. Of course, during the movement no part of a string is permitted to pass right through another part in some supernatural fashion; the string is regarded as being extremely thin and pliable; it can stretch and there is no friction nor rigidity to be considered. As an example, FIGURE 1 shows two pictures of the same link, a famous link called the Whitehead link. Thus the problem of understanding knots and links is one of geometry and topology, and within those disciplines the subject has received considerable study during the last hundred or more years. Knot theory has been a real inspiration to both algebraic and geometric topology, and, conversely, the theoretical machinery of topology has been used to make vigorous attacks on knot theory. The principal problem has always been to find ways of deciding whether or not two links are equivalent. Confronted with two heaps of intertwined strings, how is one to know if one can move the first to the configuration of the second (without cheating and breaking the strings)? Algebraic topology provides some 'invariants,' but recently some entirely new methods have been discovered which are extraordinarily effective (though not infallible), and which, judged by the standards of most modern mathematics, are breathtakingly simple.

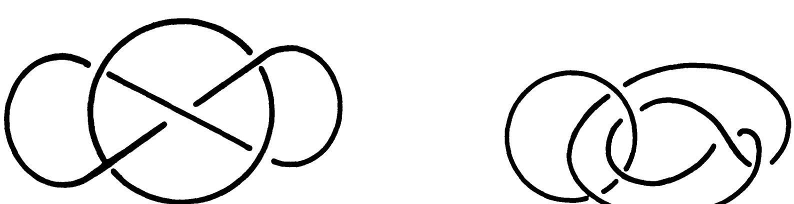


FIGURE 1

*Supported in part by National Science Grant DMS8503733.

The story begins in the spring of 1984. Professor V. F. R. Jones, now of the University of California at Berkeley, had for some years been studying operator algebras and trace functions on these algebras. It was pointed out to him that some of the formalism of his work closely resembled that of the well-known *braid group* of E. Artin [3]. This braid group can be used to study knots, and eventually Jones realized that, using his trace functions, he could define polynomials for knots and links which are *invariants* [9]. This means that to each configuration of pieces of string is associated a polynomial, and that if the string is moved (as described above) to a new position it still has the same polynomial. Thus, if calculation shows two heaps of string have two distinct polynomials, then it is not possible to move the strings from one position to the other. To get the idea of an invariant, consider what is probably the easiest of them all, namely the number of strings that make up a link; a link of two strings can never be deformed to one of three strings. Another polynomial invariant for links (discovered in about 1926 by J. Alexander [1]) was well known so, for a while, it was suspected that Jones' polynomial might be but some elementary manipulation of that polynomial. Soon however it was established that the Jones polynomial was entirely new, independent of all other known invariants. Strenuous efforts to understand the Jones polynomial have since been made by many mathematicians scattered around the world. The most amazing things about it are its simplicity and the fact that it exists at all. In retrospect it seems that several mathematicians during the last thirty years came exceedingly close to discovering Jones' polynomial and would surely have done so had they dreamt there was anything there to discover. A very simple complete proof of the existence of this polynomial appears in §3 below. By now, Jones' polynomial has been generalized two or three times, lengthy computer generated tabulations of examples have been produced, proofs have been explored and simplified, some correlations with algebraic topology have been found and a few geometric applications have been produced. Nevertheless, at the time of writing, there is still a feeling that these new ideas are not really understood, that they do not really fit in with more established theories, and that more generalizations and applications may be possible. Intense investigation continues.

§1. Basic Background

A little basic information about knots and links may allay some misunderstandings, but the confident will proceed to the next section. There are several excellent surveys of the subject prior to Jones' discovery; [2], [15], [17], [16] and [5] are accounts in (approximate) order of increasing mathematical sophistication. As already stated, a link is a finite collection of disjoint simple closed curves in 3-dimensional space $\mathbb{R}^3$, the individual simple closed curves being called the *components* of the link. A link of just one component is a *knot*. It is tacitly assumed that the closed curves are *piecewise linear*, that is that they consist of a finite number (probably very large) of straight line segments placed end to end. This is a technical restriction best ignored in practice; it does however ensure that an infinite number of kinks of any sort, possibly converging to zero size, never occurs. Restricting the components to being differentiable would do equally well. The orthogonal projection of $\mathbb{R}^3$ onto a plane $\mathbb{R}^2$ in $\mathbb{R}^3$ maps a link to a diagram of the type seen frequently in the pages that follow. The direction of that projection is always chosen so that, when in $\mathbb{R}^2$ projections of two distinct parts of the link meet, they do so transversally at a *crossing* as in FIGURE 2(i), never as in FIGURE 2(ii), (iii), or (iv). At a crossing it is indicated which of the two arcs corresponds to the

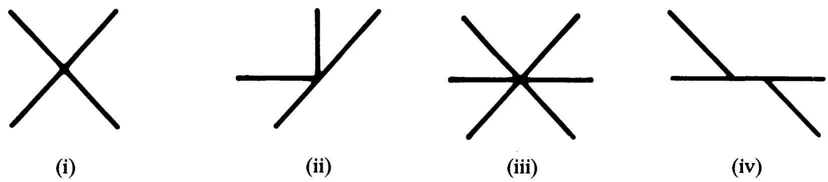


FIGURE 2

upper string, and which to the lower, by breaking the line of the lower one at the crossing. Such a planar diagram will be called a *projection* of the link.

A movement of a link from one position in $\mathbb{R}^3$ to another is called an *ambient isotopy*; that idea defines when two links are the same or '*equivalent*.' Such a movement changes the planar projections of a link. An established theorem states that two links are equivalent if and only if (any of) their projections differ by a sequence of the *Reidemeister moves* [16]. These moves, of types I, II and III, are those shown in FIGURE 3 (and their reflections), where, for each type, a small part of the projection is shown before and after the move; the remainder of the projection remains unchanged. It is clear that if two link projections so differ then the links are equivalent; the converse is established by a routine and inelegant proof. Thus to show that the two diagrams of FIGURE 1 represent the same link one can construct a sequence of these moves that allow the first diagram to evolve to the second. However, for two general link projections one has no idea whether many million moves may be required.

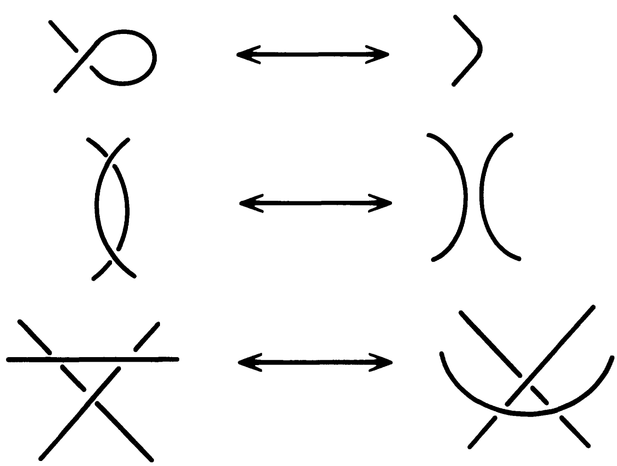


FIGURE 3

An *oriented* link is a link with a direction (usually indicated by an arrow) assigned to each component, so that each acquires a preferred way of travelling around it. Thus a link with n components has 2^n possible orientations. The two oriented links of FIGURE 4 are distinct, for one cannot be moved to the other sending the directions on the one link to those on the other (this is proved in what follows).

A knot is *unknotted* if it is equivalent to a knot that has a projection with zero crossings. Two oriented knots can be *summed* together as indicated in FIGURE 5; they are placed some way apart and a 'straight' band joins one to the other so that in the resultant sum the orientations match up. A knot, other than the unknot, is *prime* if it

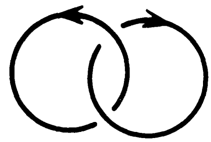


FIGURE 4(a)

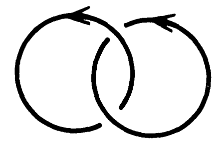


FIGURE 4(b)

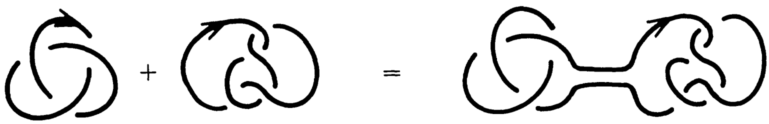


FIGURE 5

cannot be expressed as a sum of two knots neither of which is unknotted. Note that the sum of two oriented links of more than one component is not well defined unless it is specified which two components are to be banded together. As it is known that any knot is uniquely expressible as the sum of prime knots, listings of knots usually include only the prime knots.

If L is an oriented link, let ρL denote L with all its directions reversed, and let $\bar{L}$ be the reflection of L. When considering projections, this reflection is usually thought of as reflection in the plane of the paper, so that the projection of $\bar{L}$ comes from that of L by changing all underpasses to overpasses and *vice versa*. Thus from L can be created ρL, $\bar{L}$, and $\rho\bar{L} = \overline{\rho L}$, and these may be four distinct links, they may be the same in pairs, or all four may be the same. The trefoil knot 3_1 (see FIGURE 6) creates two pairs in this way, the figure of eight knot 4_1 has all four the same whilst 9_{32} has all four distinct.

Knot 3_1 Knot 9_{32}

(a) (b) (c)

FIGURE 6

Inherent in the idea that reflection can change a knot is the convention that the enveloping three-dimensional space $\mathbb{R}^3$ is oriented; it is equipped with a distinction between left-hand and right-hand screwing motions. Knot tables have traditionally listed prime knots according to the minimum number of crossings in a projection of the knot. Thus 7_4 denotes the fourth knot, in some traditional order, that needs seven and no more than seven crossings. The tables have deliberately ignored reflections and reversals, so that an entry may stand for as many as four knots if these orientations are taken into account. With these conventions knots have been classified up to thirteen crossings [19] with the help of computers and the following table has been produced.

#crossings	3	4	5	6	7	8	9	10	11	12	13
#knots	1	1	2	3	7	21	49	165	552	2176	9988

This totals 12965 knots.

There now follows a discussion of the new polynomial invariants of knots and links. It is not possible to restrict the discussion to knots alone, for many-component links are an integral part of the theory. The ideas will not here be developed in the order of their discovery but in an order that now seems simpler to understand.

§2. The Oriented Polynomial

The polynomials to be considered here are Laurent polynomials with *two* variables ℓ and m and with integer coefficients. A Laurent polynomial differs from the usual polynomials of high school only inasmuch as negative as well as positive powers of the variables may occur. One such polynomial $P(L)$ will be associated to each oriented link L. For example, to the link of FIGURE 1 will be assigned the polynomial

$$(-\ell^{-1}-\ell)m^{-1}+(\ell^{-1}+2\ell+\ell^{3})m-\ell m^{3}.$$

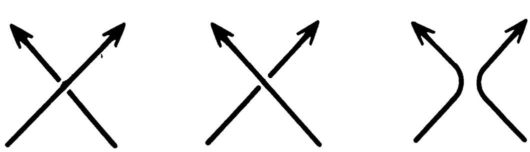
The result that encapsulates this was discovered almost simultaneously by four sets of authors [7] in the wake of Jones' first announcement ($P(L)$ *generalises* Jones' polynomial, see §3). It can be stated as follows:

THEOREM 1. *There is a unique way of associating to each oriented link L a Laurent polynomial $P(L)$, in the variables ℓ and m, such that equivalent oriented links have the same polynomial and*

(i) $P(unknot)=1$,

(ii) *if L_+, L_-, and L_0 are any three oriented links that are identical except near a point where they are as in FIGURE 7, then*

$$\ell P(L_+)+\ell^{-1}P(L_-)+mP(L_0)=0.$$



L_+ L_- L_0

FIGURE 7

In a projection of an *oriented* link the crossings are of two types; that of L_+ in FIGURE 7 is called *positive*, that of L_- is *negative*. This idea will be exceedingly important. In L_+, the direction of one segment can be thought of as pointing in the direction dictated by a right-hand screw motion along the direction of the other segment. All the orientations are needed to make this important distinction. Of course, the choice of which type of crossing is given which sign is but a convention.

The meaning of Theorem 1 becomes apparent as one uses it to make a few calculations. Shown in FIGURE 8 is a very elementary example of a triple of links L_+, L_-, and L_0.

FIGURE 8

The first two links are just pictures of the unknot, the first with a single positive crossing, the second with just a negative one. Formulae (i) and (ii) imply that $\ell 1 + \ell^{-1} 1 + mP(L_0) = 0$, from which one deduces that $P(L_0)$, the polynomial for the trivial link of two separated unknots, is $-(\ell^{-1} + \ell)m^{-1}$. Consider now the triple of links in FIGURE 9 (where it is the uppermost crossing that is to be considered).

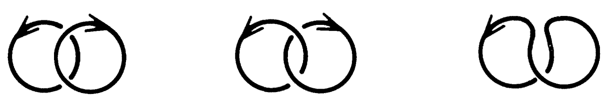


FIGURE 9

Here L_+ is the link whose polynomial has just been calculated, or at least it is equivalent to it and so has that same polynomial; L_0 is the unknot. Thus

$$-\ell(\ell^{-1} + \ell)m^{-1} + \ell^{-1}P(L_-) + m1 = 0,$$

so that the polynomial for the simple link L_-, the link of FIGURE 4(a), is $(\ell + \ell^3)m^{-1} - \ell m$. FIGURE 10 shows a third triple (look at the right-hand crossing). This yields the equation

$$\ell 1 + \ell^{-1}P(L_-) + m\{(\ell + \ell^3)m^{-1} - \ell m\} = 0.$$

Hence the polynomial of the left-hand version of the trefoil knot 3_1 (seen also in FIGURE 6(a)) is $-2\ell^2 - \ell^4 + \ell^2 m^2$.

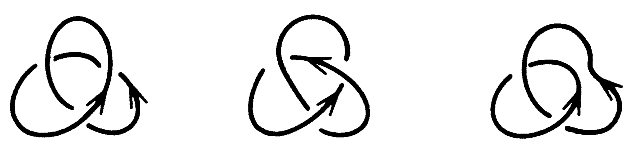


FIGURE 10

This proves that the trefoil is indeed knotted, for were it equivalent to the unknot it would, by Theorem 1, have 1 for its polynomial. Similarly the link of FIGURE 4(a) is not equivalent to the link consisting of two separated unknots.

Consider the method of the preceding calculation of the trefoil's polynomial. Attention was given to one crossing. Switching that crossing produced the unknot, nullifying it (to get 'L_0') produced a link with fewer crossings that had already been considered. This procedure works in general. Suppose that one is confronted with an oriented link L of n crossings and c components. Assume that one has already calculated the polynomials of all (relevant) oriented links of $n-1$ crossings; there are only finitely many of them. Then formula (ii) calculates $P(L)$ in terms of the polynomial of a modified L, namely L with some chosen crossing switched. However it is always possible to change L to U^c, the unlink of c unknots, by switching a subset of some s of the crossings. Thus, performing the switches one by one, using formula (ii) each time, $P(L)$ is calculated in terms of the polynomials of s links of fewer crossings (the 'L_0's') and of $P(U^c)$. However, it is an easy exercise to show, by induction on c, that $P(U^c) = (-(\ell^{-1} + \ell)m^{-1})^{c-1}$. This method of calculation will always work, and it is essentially the only known method of calculating these polynomials. It is nevertheless not a very welcome method for the length of calculation increases exponentially with the number of crossings of the link presentation.

Note that if *all* the diagrams in the above calculations were reflected in the plane of the paper this would change each positive crossing to a negative crossing and *vice versa*. Each L_+ would become an L_-. This would simply exchange the rôles played in the calculation by ℓ and ℓ^{-1}. Thus reflecting a link has the effect on its polynomial of interchanging ℓ and ℓ^{-1}. The left-hand trefoil of FIGURE 6(a) has polynomial $-2\ell^2 - \ell^4 + \ell^2 m^2$, so the right-hand trefoil's polynomial (FIGURE 6(b)) is $-2\ell^{-2} - \ell^{-4} + \ell^{-2}m^2$. These polynomials are obviously different so, by Theorem 1, the trefoils are inequivalent (this fact was tricky to prove until 1984 when Jones produced a version of this proof). Similarly the two oriented links of FIGURE 4 are distinct. For any polynomial P, let $\bar{P}$ denote P with ℓ and ℓ^{-1} interchanged (c.p. complex conjugation). The above discussion has demonstrated the following result.

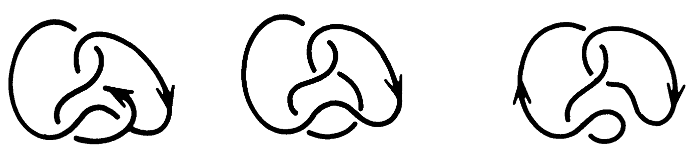
PROPOSITION 1. *For any oriented link* L, $\overline{P(L)} = P(\bar{L})$.

Consider now the triple of FIGURE 11. This yields

$$\ell P(4_1) + \ell^{-1}1 + m\left\{(\ell + \ell^3)m^{-1} - \ell m\right\} = 0$$

so that

$$P(4_1) = -\ell^{-2} - 1 - \ell^2 + m^2.$$



4_1 U

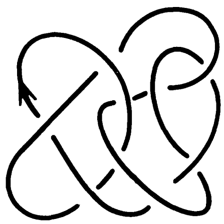
FIGURE 11

Here then the polynomial *is* symmetric with respect to ℓ and ℓ^{-1} and so it does not show 4_1 and $\bar{4}_1$ to be different, and, in fact, a little experimentation shows them to be the same. In practice the P-polynomial provides a very good test as to whether or not $L = \bar{L}$, but any hope that it might be an infallible test is dashed by the knot 9_{42} shown in FIGURE 12. It is known that $9_{42} \neq \bar{9}_{42}$ because a certain 'signature' invariant, from algebraic topology, is nonzero. However,

$$P(9_{42}) = (-2\ell^{-2} - 3 - 2\ell^2) + (\ell^{-2} + 4 + \ell^2)m^2 - m^4$$

and this is a self-conjugate polynomial.

It should be remarked that other *notations* can be used in the whole of this theory of polynomials for knots and links. For example, $P(L)$ can be taken to be a polynomial in three variables x, y, z with the vital defining formula being $xP(L_+) + yP(L_-) + zP(L_0) = 0$. However, the three variables are homogeneous variables as in projective planar geometry (there are still really only two variables), and the balance between ℓ and ℓ^{-1} is lost. Some authors also have a strong preference for some negative signs in the defining formula.



Knot 9_{42} FIGURE 12

Recall that ρL is obtained by L by reversing all its arrows. Unfortunately, $P(\rho L) = P(L)$, for changing L to ρL leaves the signs of all its crossings unchanged. Thus any calculation for $P(L)$ induces exactly the same calculation for $P(\rho L)$. (This means that for a knot, a link of one component, it is not really necessary to specify an orientation at all when thinking about the polynomial.) If, however, the directions of some, but not all, of the components of L are changed, then $P(L)$ can change in a rather drastic way that is not well understood. Examples occur in some of the polynomials of two-component links listed in the table at the end of this article.

A result concerning the behaviour of the polynomial under sums and 'distant' unions is as follows:

PROPOSITION 2. (i) $P(L_1 + L_2) = P(L_1)P(L_2)$;
(ii) $P(L_1 \cup L_2) = -(\ell + \ell^{-1})m^{-1}P(L_1)P(L_2)$.

In (i) $L_1 + L_2$ denotes the sum (see FIGURE 5) of oriented links using *any* component of L_1 to add to *any* component of L_2. As different choices may be made for these components, this leads easily to examples of distinct links having the same polynomial. For example, the two links in FIGURE 13 are distinct as their individual components are different knots. However, by Proposition 2(i) they both have polynomial

$$\{-2\ell^2 - \ell^4 + \ell^2 m^2\}\{-\ell^{-2} - 1 - \ell^2 + m^2\}\{(\ell^{-1} + \ell^{-3})m^{-1} - \ell^{-1}m\}.$$

In (ii) $L_1 \cup L_2$ denotes the union of L_1 and L_2 placed some distance apart from each other so that no part of L_1 crosses over or under part of L_2. Proposition 2 is significant because it relates simple geometry to the *product* of polynomials. This uses the multiplicative structure of polynomials; $P(L)$ is not just an array of coefficients but is a polynomial that may be used to multiply another polynomial! The proposition is easy to prove from Theorem 1.

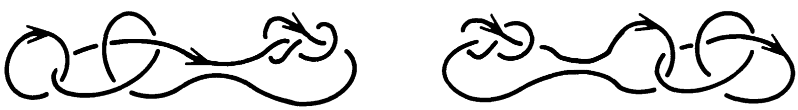


FIGURE 13

There is another way in which it is known that two oriented links will have the same polynomial. Deep in the geometric structure of link theory is the simple idea of decomposing a link using spheres that cut the link at four points. The dotted sphere of FIGURE 14 is an example. If the inside of that sphere is rotated through angle π (about the polar axis) the second diagram results. Such an operation is called *mutation*. Mutation never changes the P-polynomial of a link though it can well change the link, as indeed it does in FIGURE 14.

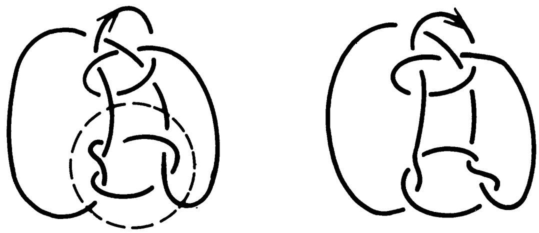


FIGURE 14

For a further example consider a *pretzel* knot as shown in FIGURE 15. The ith circle contains a twist of c_i crossings as indicated, each c_i being odd (to side-step any orientation difficulties). Mutation, with respect to the indicated ellipse, interchanges c_2 and c_3. As any permutation is the result of a sequence of such adjacent interchanges, the P-polynomial of the pretzel knot is independent of the ordering of the c_i's; in general the knot does change if that ordering changes.

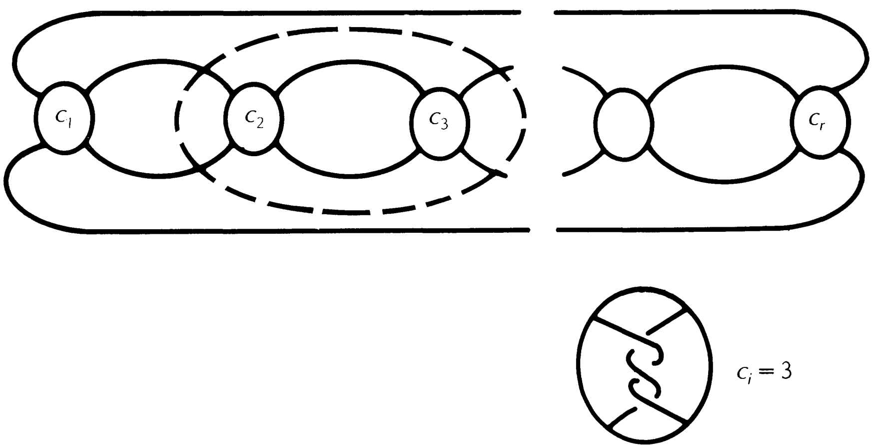


FIGURE 15

So far nothing has been said about a proof for Theorem 1. The proof in [**13**] consists of defining the polynomial with a lengthy argument of induction on the number of crossings of a presentation, showing that however a calculation (like those already discussed) is made the same polynomial results, and checking invariance under the Reidemeister moves. It is thus entirely combinatoric, but the induction argument needs delicate handling. Although other proofs differ in style and emphasis they all seem to use essentially the same combinatorics.

§3. The Jones Polynomial

Whereas the proof that $P(L)$ exists is a little arduous, an almost trivial proof of the existence of the polynomial of Jones has been found by L. H. Kauffman [**11**]. This proof, which must, in recent years, be one of the most remarkable discoveries of readily accessible mathematics, is outlined below.

The original polynomial of V. F. R. Jones associated with an oriented link L is denoted $V(L)$. It is a Laurent polynomial in the variable $t^{1/2}$, that being simply a symbol whose square is the symbol t. It satisfies

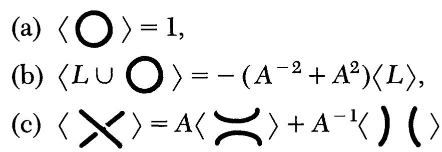
$$V(\text{unknot}) = 1$$
$$t^{-1}V(L_+) - tV(L_-) + (t^{-1/2} - t^{1/2})V(L_0) = 0,$$

where L_+, L_-, and L_0 are oriented links related as before. Thus $V(L)$ is obtained from $P(L)$ by the substitution

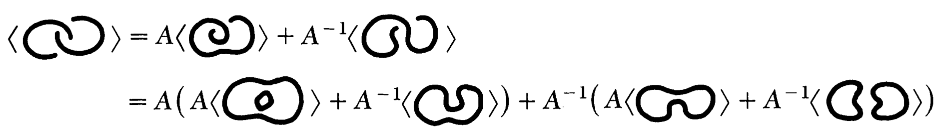
$$(\ell, m) = \left(it^{-1}, i(t^{-1/2} - t^{1/2})\right),$$

where $i^2 = -1$. As mentioned before, $P(L)$ was conceived as a generalisation of $V(L)$.

Begin all over again by considering *projections* (pictures) of *unoriented* links. For each such projection L define a Laurent polynomial $\langle L \rangle$ in one variable A by the following three rules that will shortly be explained:

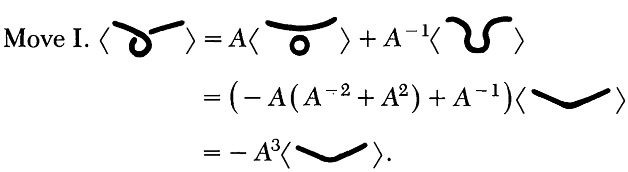


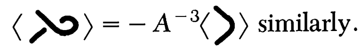
This $\langle L \rangle$ is called the *bracket polynomial* of L. Rule (a) states that 1 is the polynomial of the particular projection of the unknot that has no crossing at all. In rule (b) $L \cup \bigcirc$ denotes the projection that consists of L plus an extra component that contains no crossing. Rule (c) refers to three projections exactly the same, except near one point where they are as shown. The first projection of this triple shows a crossing, and in the other two that crossing has been destroyed. It should be noted that, given the picture of the crossing, one can distinguish between the two other pictures using the orientation of space: If, when moving along the underpass towards the crossing one swings to the right, up on to the overpass, one creates the picture of the link whose polynomial is multiplied by A in rule (c). No arrows are required for that. A simple example involving the use of all three rules is as follows:



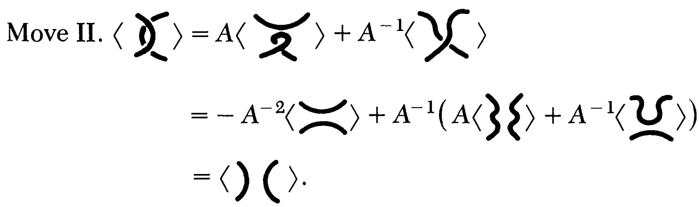
$$= -(A^2 + A^{-2})^2 + 2.$$

Here when calculating $\langle L \rangle$ there are no problems about making judicious choices of crossings to switch in order to maneuver towards an unknotted situation (as there were with the P-polynomial). Each use of rule (c) *reduces* the number of crossings in the projections until there are no crossings at all; then rules (b) and (a) finish the job of calculation. It is evident that the choice of the order in which the crossings are attacked is irrelevant, so that these rules do indeed define unambiguously a polynomial for each unoriented link projection. What remains to be done is to see if $\langle L \rangle$ is unchanged by the Reidemeister moves I, II and III of §1; if it is, then it is an invariant of real links in $\mathbb{R}^3$:

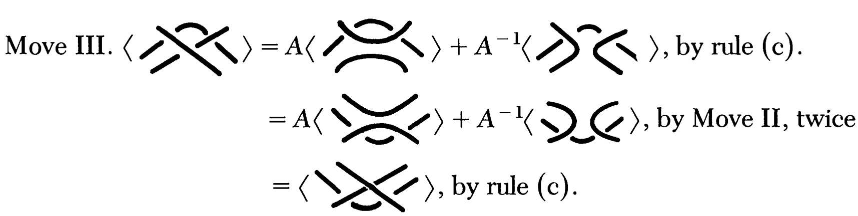




Thus the bracket polynomial *fails* to be invariant under Move I, and that is an exceedingly important observation.



Hence the bracket polynomial *is* invariant under Move II.



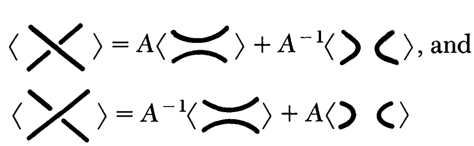
Hence there is also invariance under Move III.

Now give L an *orientation*. Let $w(L)$, the *writhe* of L, be the algebraic sum of the crossings of L, counting $+1$ for a positive crossing, and -1 for a negative crossing (for example, w(left-hand trefoil) $= -3$). Move I adds or subtracts one to $w(L)$, so $w(L)$ is certainly not invariant under that move, but it is (clearly) invariant under Moves II and III. Thus any combination of $w(L)$ and $\langle L\rangle$ will be invariant under Moves II and III, and their non-invariant behaviours under Move I *cancel* in the expression

$$X(L) = (-A)^{-3w(L)}\langle L\rangle.$$

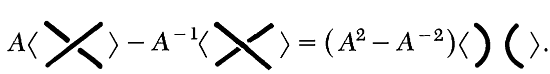
The above is a *complete proof* that $X(L)$ is a well defined invariant of oriented links.

For projections related in the usual way, rule (c) gives



(*).

Thus



Suppose that orientations can be chosen for these last three projections so that the arrows point approximately upwards (c.f. FIGURE 7); call them L_+, L_- and L_0. Then $w(L_{\pm}) = w(L_0) \pm 1$. Hence, substitution gives

$$A(-A)^3 X(L_+) - A^{-1}(-A)^{-3} X(L_-) = (A^2 - A^{-2}) X(L_0).$$

Writing $t^{-1/4} = A$ this becomes

$$t^{-1}X(L_+) - tX(L_-) + (t^{-1/2} - t^{1/2})X(L_0) = 0,$$

so that, under the substitution $A = t^{-1/4}$, $X(L)$ is the original Jones polynomial $V(L)$ for they satisfy the same defining formula.

No analogously simple proof is known for the existence of the P-polynomial; in a proof, the difficulty is to show that different chains of calculations *never* give different polynomials. In terms of distinguishing links the P-polynomial is more powerful than is the V-polynomial; two variables are better than one. For example, the knots in FIGURE 16 have the same V-polynomial but different P-polynomials.

FIGURE 16

There is a property of the Jones polynomial, a *reversing result*, that *seems* to have no analogue in terms of the P-polynomial. Suppose that k is one component of an oriented link L and that a new oriented link L^* is formed from L by reversing just the orientation of k. Of course, $\langle L \rangle = \langle L^* \rangle$, for the bracket polynomial disregards all orientations. Thus $V(L)$ and $V(L^*)$ are the same up to multiplication by some power of t (for each is $\langle L \rangle$ multiplied by a power of $A = t^{-1/4}$). The precise result is:

PROPOSITION 3. $V(L^*) = t^{-3\lambda}V(L)$, *where* 2λ *is the sum of the signs of the crossings of* k *with the other components of* $L - k$.

This λ is called the *linking number* of k and $L - k$. It is noteworthy that, though this result follows trivially from the approach of the bracket polynomial, it is by no means obvious when working from the (L_+, L_-, L_0)-definition of the V-polynomial.

The simplicity of Kauffman's approach to the V-polynomial has led to a much better understanding of that polynomial and to a most pleasing application ([**11**], [**14**], [**18**]) concerning alternating knots. A projection of a link is alternating if, when travelling along any part of the link, the crossings are encountered alternately over, under, over, under,.... In FIGURE 16 the four-crossing projection is alternating, the other is not alternating. The first thirty-one knots in the classical knot tables have alternating projections. A crossing in a link projection will be called *removable* if it is like the crossing in FIGURE 17; it could be removed by rotating the part of the link in the box labelled Y.

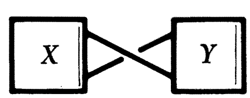


FIGURE 17

PROPOSITION 4. (See [**11**], [**14**] and [**18**].) *Let* L *be a connected oriented link projection of* n *crossings, then*

(i) $n \geqslant$ Spread $V(L)$, *where* spread $V(L)$ *is the difference between the maximum and the minimum degrees of* t *that appear in* $V(L)$;

(ii) $n =$ Spread $V(L)$ *if* L *is also alternating and has no removable crossing.*

It is easy to see, by inspection, if a knot projection is alternating and has no removable crossing. If it has these properties, and n crossings, the proposition implies that the knot can have no projection with fewer crossings. If there were a projection with $n-1$ crossings, then, by (i), $n-1 \geq$ Spread $V(L)$, which contradicts (ii). This solves a very old problem in knot theory; it has always been suspected that an alternating projection was the simplest available.

For the record, this section should include mention of the *Alexander polynomial* $\Delta(L)$ of an oriented link L. Like the Jones polynomial it is a Laurent polynomial in $t^{1/2}$, it can be defined in a similar way by $\Delta(\text{unknot}) = 1$, and

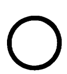
$$\Delta(L_+) - \Delta(L_-) + (t^{1/2} - t^{-1/2})\Delta(L_0) = 0.$$

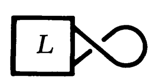
Thus $\Delta(L)$ is obtained from $P(L)$ by a substitution for the variables (though it was the similarity between this formula and that defining $V(L)$ that lead to the discovery of the P-polynomial). The Alexander polynomial has been known and developed for about sixty years [**1**]. It is discussed in the text books of knot theory, usually being defined in terms of the determinant of a certain matrix (though see [**6**]). The value of $\Delta(L)$ when $t = -1$ is called the *determinant* of the link, and this integer was one of the first link invariants to be studied. Although the Alexander polynomial is quite good at distinguishing knots, there do exist knots that it cannot distinguish from the unknot; an example is the pretzel knot (see Figure 15) for which $(c_1, c_2, c_3) = (-3, 5, 7)$; for this the Jones polynomial is certainly non-trivial. The Alexander polynomial is fairly well understood in terms of the machinery of algebraic topology (homology groups, fundamental groups, covering spaces, etc.). The same cannot be said for the Jones polynomial and its generalizations. Is there a knot K, other than the unknot, for which $V(K) = 1$? The answer to this is not known. If there is no such K then the Jones polynomial is the long sought elementary method of determining knottedness. There is no reason to suppose that it is so powerful an invariant, but computer searches have revealed no example of such a K, neither has understanding given any clue to finding a method by which such a K might be constructed.

§4. The Semioriented Polynomial

Although it may seem that the preceding sections contain many polynomials, only the P-polynomial and some specialisations of it occur. There has been discovered, however, another polynomial, the F-polynomial, that is similar in concept to the P-polynomial though the two are quite distinct. The way to define this F-polynomial is rather like the way in which $V(L)$ was derived from $\langle L \rangle$.

Firstly, for a *projection* L of an *unoriented* link, define a Laurent polynomial $\Lambda(L)$ in two variables a and x by the rules

(a) $\Lambda\big($  $\big) = 1$;

(b) $\Lambda\big($  $\big) = a\Lambda(L)$, $\Lambda\big($  $\big) = a^{-1}\Lambda(L)$, and $\Lambda(L)$ does not change when L is changed by a Reidemeister move of type II or type III;

(c) $\Lambda(L_+) + \Lambda(L_-) = x(\Lambda(L_0) + \Lambda(L_\infty))$

where L_+, L_-, L_0 and L_∞ are projections of unoriented links that are exactly the same except near a point where they are as shown in Figure 18.

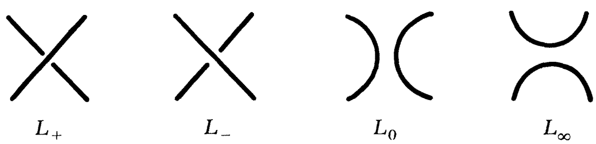


FIGURE 18

Notes. (a) This means that Λ is 1 for the projection of just one component which has no crossing.

(b) If a positive kink is removed, the Λ-polynomial is multiplied by a (or by a^{-1} for a negative kink). Thus the Λ-polynomial is not invariant under Reidemeister Move I.

(c) In the absence of orientations it is not clear which picture in FIGURE 18 should be called L_+ and which L_-, nor which is L_0 and which L_∞. Those ambiguities are irrelevant in the light of the symmetry of the formula of Rule (c).

It should be clear that the methods of calculation developed for the P-polynomial will work equally well in this new situation. The new situation is easier in that orientations do not (yet) appear, but more troublesome in that Rule (c) uses four pictures instead of three. As an excercise check that the Λ-polynomial of the two-component link projection with no crossing is $((a^{-1}+a)x^{-1}-1)$ while that of the usual projection of the left-hand trefoil knot is

$$-2a^{-1}-a+(1+a^2)x+(a^{-1}+a)x^2.$$

A proof that, for a given projection, different schemes of calculation always give the same polynomial requires a more complicated version of the inductive method mentioned at the end of §2 for the P-polynomial.

The Λ-polynomial is, as stated, invariant under the second and third of Reidemeister's moves. Its failure to be invariant under move I can easily be corrected (as was done for $\langle L\rangle$) if L is *now oriented*. As before let $w(L)$ be the sum of the signs of the crossings of the oriented link L.

THEOREM 2. *For any oriented link L, let*

$$F(L)=a^{-w(L)}\Lambda(L).$$

This F(L) is a well-defined invariant of oriented links in 3-space.

Tables of the P-polynomial and of this second two-variable polynomial have been produced by M. B. Thistlethwaite for the 12,965 knots in his tabulation of knot projections up to thirteen crossings. He works, of course, with a computer, that being all the more desirable for the F-polynomial; the occurrence of four rather than three diagrams in the defining formula does make calculations for F much more arduous than for P. The F-polynomials contain very many more terms than do the P-polynomials; for example for either knot in FIGURE 14 the P-polynomial has 14 terms, the F-polynomial has 45 terms. A few more examples appear in the tables at the end of this paper. The greater number of terms seems to mean, in practice, that two knots are more likely to be distinguished by F than by P.

The F-polynomial has a right to be called 'semioriented' because, although L must be oriented to define $F(L)$, changing the orientation only changes $F(L)$ by multiplication by a power of a. The relevant result is:

PROPOSITION 5. *Suppose L^* is obtained from L by reversing the orientation of a component k, then $F(L^*)=a^{4\lambda}F(L)$, where λ is the linking number of k with the other components of $L-k$.*

Compare this with Proposition 3; the proof here is much the same.

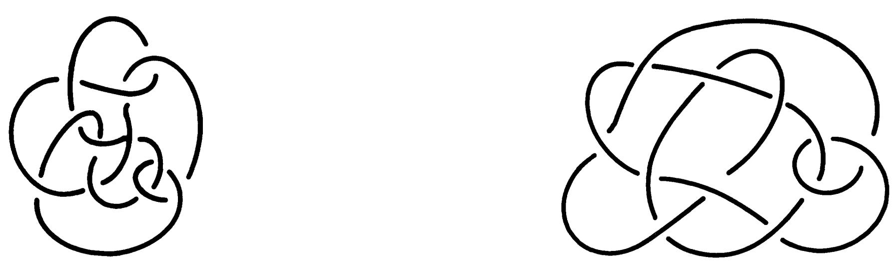
An interesting specialisation of $F(L)$ is obtained by the substitution $a=1$. The resultant Laurent polynomial $Q(L)$, in the one variable x, is called the *absolute* polynomial. This substitution makes all the subtleties of the above definition disappear, no orientations of any sort are required (hence the name 'absolute') and one can work entirely with links in $\mathbb{R}^3$ rather than projections in the plane. The Q-polynomial is simply defined by $Q(\text{unknot})=1$, and

$$Q(L_+)+Q(L_-)=x(Q(L_0)+Q(L_\infty)).$$

Chronologically this polynomial was discovered by Brandt, Lickorish, and Millett [4] and Ho [8] as an extension of the ideas of the P-polynomial, and Kauffman [**10**] explained how to insert the second variable 'a' to create the F-polynomial. The fact that the Q-polynomial uses no arrows, and $Q(L)=Q(\bar{L})$, makes this something of a recommended polynomial for beginners. Unfortunately the proofs that the Q and F polynomials are unambiguously defined are of almost the same form and complexity. In FIGURE 19 are examples that show that the P and F polynomials are independent in the sense that neither is hidden within the other, to be revealed by some subtle change of the variables.

8_8 10_{129}

These have the same P but different Q (and F) polynomials.



11_{255} 11_{257}

These have the same F but different Δ (and P) polynomials.

FIGURE 19

As might be expected, some basic properties of the F-polynomial are similar to those of the P-polynomial. This is summarized in the next result; it uses some of the notations from §2.

PROPOSITION 6. (i) $\overline{F(L)}=F(\bar{L})$, *where* $\bar{a}=a^{-1}$ *and* $\bar{x}=x$;
(ii) $F(L_1+L_2)=F(L_1)F(L_2)$;
(iii) $F(L_1\cup L_2)=((a^{-1}+a)x^{-1}-1)F(L_1)F(L_2)$;
(iv) *F is unchanged by mutation.*

A result that came as something of a surprise [12] was that *both* the P-polynomial *and* the F-polynomial contain the original polynomial of Jones, the V-polynomial. The result, which now has an easy proof, is:

Proposition 7. *For any oriented link L, the substitution*

$$(a, x) = \left(-t^{-3/4}, (t^{-1/4} + t^{1/4})\right)$$

reduces $F(L)$ to $V(L)$.

Proof. Adding together the two equations (*) from §3 gives

$$\langle \times \rangle + \langle \times \rangle = (A + A^{-1})(\langle)(\rangle + \langle \asymp \rangle).$$

Thus $\langle L \rangle$ satisfies exactly the same defining equations as $\Lambda(L)$ when $x = (A + A^{-1})$ and $a = -A^3$ (the latter arising from comparison of the effects of the first Reidemeister move on the two polynomials). Then just recall that it is the substitution $A = t^{-1/4}$ that produces the Jones polynomial.

§5. Calculations, Problems and Tables

Calculations of any of the polynomials mentioned in previous sections can be performed 'by hand' for links with few crossings or for those with some simple pattern. Sometimes the linear nature of the formulae that define these polynomials can be exploited in a most pleasing way. That idea can be illustrated using the Q-polynomial to avoid orientation complications. Suppose that a link T_n contains as part of it the n-crossing twist as shown in Figure 20 (where by convention $n = \infty$ is also permitted, as illustrated); if n is negative the twist goes the other way.

T_n T_∞

FIGURE 20

Focussing on one of these crossings produces a quadruple of links L_+, L_-, L_0 and L_∞ (as in the defining formula for the Q-polynomial), namely, T_n, T_{n-2}, T_{n-1} and T_∞. The defining formula gives

$$Q(T_n) + Q(T_{n-2}) = x(Q(T_{n-1}) + Q(T_\infty)).$$

Thus

$$\begin{bmatrix} Q(T_n) \\ Q(T_{n-1}) \\ Q(T_\infty) \end{bmatrix} = M \begin{bmatrix} Q(T_{n-1}) \\ Q(T_{n-2}) \\ Q(T_\infty) \end{bmatrix} = M^n \begin{bmatrix} Q(T_0) \\ Q(T_{-1}) \\ Q(T_\infty) \end{bmatrix}$$

where M is the matrix

$$\begin{bmatrix} x & -1 & x \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

As an exercise use this idea to calculate, in terms of the matrix M, the Q-polynomial of the link of FIGURE 21, where the twists have m and n crossings respectively.

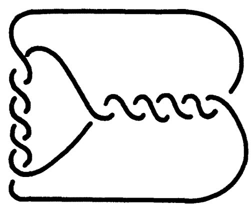


FIGURE 21

In the following exercises L is an oriented link with $c(L)$ components. The proofs are all performed using, in the relevant defining formulae, induction on the number of crossings in a projection.

Exercises

(i) $P(L)=1$ when $m=-(\ell+\ell^{-1})$.

(ii) In $P(L)$ the least power of m is $m^{1-c(L)}$.

(iii) $F(L)=(-2)^{c(L)-1}$ when $(a,x)=(1,-2)$.

(iv) $F(L)=(-1)^{c(L)-1}$ when $(a,x)=(i,x)$, $x\neq 0$.

(v) $V(L)=\Delta(L)$ when $t=-1$.

(vi) If $c(L)=1$, $F(L)=P(L)$ when $x=0=m$, and $a=\ell$.

Unanswered questions abound. Here are some of them.

Questions. (1) Can the new polynomials be defined without reference to diagrams in the plane?

(2) Can the P and F polynomials be defined in as simple a way as the bracket polynomial?

(3) Can any of the new polynomials for a link be calculated in one step (e.g., by means of a determinant) without working out many polynomials of simpler links?

(4) Is there some simple characterisation of what polynomials can arise as the P or F or V or Q polynomial of some link?

(5) Is there a nontrivial link of c components with the same polynomial (in the P, V, F, or Q sense) as the trivial unlink of c components?

(6) Does any of the new polynomials give information about the number of crossing switches needed to undo a knot (this is called its *unknotting number*)?

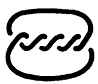
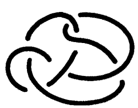
(7) Does there exist some grand master polynomial in which particular substitutions produce both P and F?

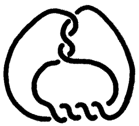
(8) Is there a "coloured" theory for P or F? This would be a theory that had more variables and which could distinguish, for example, a red trefoil linked with a blue unknot from a blue trefoil linked with a red unknot. There is such a variant of the Alexander polynomial.

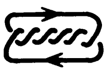
(9) Are there polynomials other than P and F that can be defined along the same lines as they are defined? Several attempts have been made but all have turned out to be subtle variants of the original two polynomials.

TABLES. Below are given tables of the P and F polynomials for a few knots and links of low numbers of crossings to give a feeling for what is involved.

F & P for knots $\leqslant 7$ crossings, for links $\leqslant 6$ crossings.

		P	**F**
3_1		$(-2\ell^2 - \ell^4) + \ell^2 m^2$	$(-2a^2 - a^4) + (a^3 + a^5)x + (a^2 + a^4)x^2$
4_1		$(-\ell^{-2} - 1 - \ell^2) + m^2$	$(-a^{-2} - 1 - a^2) + (-a^{-1} - a)x + (a^{-2} + 2 + a^2)x^2 + (a^{-1} + a)x^3$
5_1		$(3\ell^4 + 2\ell^6) + (-4\ell^4 - \ell^6)m^2 + \ell^4 m^4$	$(3a^4 + 2a^6) + (-2a^5 - a^7 + a^9)x + (-4a^4 - 3a^6 + a^8)x^2 + (a^5 + a^7)x^3 + (a^4 + a^6)x^4$
5_2		$(-\ell^2 + \ell^4 + \ell^6) + (\ell^2 - \ell^4)m^2$	$(-a^2 + a^4 + a^6) + (-2a^5 - 2a^7)x + (a^2 - a^4 - 2a^6)x^2 + (a^3 + 2a^5 + a^7)x^3 + (a^4 + a^6)x^4$
6_1		$(-\ell^{-2} + \ell^2 + \ell^4) + (1 - \ell^2)m^2$	$(-a^{-2} + a^2 + a^4) + (2a + 2a^3)x + (a^{-2} - 4a^2 - 3a^4)x^2 + (a^{-1} - 2a - 3a^3)x^3$ $+ (1 + 2a^2 + a^4)x^4 + (a + a^3)x^5$
6_2		$(2 + 2\ell^2 + \ell^4) + (-1 - 3\ell^2 - \ell^4)m^2 + \ell^2 m^4$	$(2 + 2a^2 + a^4) + (-a^3 - a^5)x + (-3 - 6a^2 - 2a^4 + a^6)x^2 + (-2a + 2a^3)x^3$ $+ (1 + 3a^2 + 2a^4)x^4 + (a + a^3)x^5$
6_3		$(\ell^{-2} + 3 + \ell^2) + (-\ell^{-2} - 3 - \ell^2)m^2 + m^4$	$(a^{-2} + 3 + a^2) + (-a^{-3} - 2a^{-1} - 2a - a^3)x + (-3a^{-2} - 6 - 3a^2)x^2 + (a^{-3} + a^{-1} + a + a^2)x^3$ $+ (2a^{-2} + 4 + 2a^2)x^4 + (a^{-1} + a)x^5$

7_1		$(-4\ell^6-3\ell^8)+(10\ell^6+4\ell^8)m^2$ $+(-6\ell^6-\ell^8)m^4+\ell^6m^6$	$(-4a^6-3a^8)+(3a^7+a^9-a^{11}+a^{13})x+(10a^6+7a^8-2a^{10}+a^{12})x^2+(-4a^7-3a^9+a^{11})x^3$ $+(-6a^6-5a^8+a^{10})x^4+(a^7+a^9)x^5+(a^6+a^8)x^6.$
7_2		$(-\ell^2-\ell^6-\ell^8)+(\ell^2-\ell^4+\ell^6)m^2$	$(-a^2-a^6-a^8)+(3a^7+3a^9)x+(a^2+3a^6+4a^8)x^2+(a^3-a^5-6a^7-4a^9)x^3+(a^4-3a^6-4a^8)x^4$ $+(a^5+2a^7+a^9)x^5+(a^6+a^8)x^6.$
7_3		$(-2\ell^{-8}-2\ell^{-6}+\ell^{-4})+(\ell^{-8}+3\ell^{-6}-3\ell^{-4})m^2$ $+(-\ell^{-6}+\ell^{-4})m^4$	$(-2a^{-8}-2a^{-6}+a^{-4})+(-2a^{-11}+a^{-9}+3a^{-7})x+(-a^{-10}+6a^{-8}+4a^{-6}-3a^{-4})x^2$ $+(a^{-11}-a^{-9}-4a^{-7}-2a^{-5})x^3+(a^{-10}-3a^{-8}-3a^{-6}+a^{-4})x^4$ $+(a^{-9}+2a^{-7}+a^{-5})x^5+(a^{-8}+a^{-6})x^6.$
7_4^{*}		$(-\ell^{-8}+2\ell^{-4})+(\ell^{-6}-2\ell^{-4}+\ell^{-2})m^2$	$(-a^{-8}+2a^{-4})+(4a^{-9}+4a^{-7})x+(2a^{-8}-3a^{-6}-4a^{-4}+a^{-2})x^2+(-4a^{-9}-8a^{-7}-2a^{-5}+2a^{-3})x^3$ $+(-3a^{-8}+3a^{-4})x^4+(a^{-9}+3a^{-7}+2a^{-5})x^5+(a^{-8}+a^{-6})x^6.$
7_5		$(2\ell^4-\ell^8)+(-3\ell^4+2\ell^6+\ell^8)m^2$ $+(\ell^4-\ell^6)m^4$	$(2a^4-a^8)+(-a^5+a^7+a^9-a^{11})x+(-3a^4+a^8-2a^{10})x^2+(-a^5-4a^7-2a^9+a^{11})x^3$ $+(a^4-a^6+2a^{10})x^4+(a^5+3a^7+2a^9)x^5+(a^6+a^8)x^6.$
7_6		$(1+\ell^2+2\ell^4+\ell^6)+(-1-2\ell^2-2\ell^4)m^2$ $+\ell^2m^4$	$(1+a^2+2a^4+a^6)+(a+2a^3-a^7)x+(-2-4a^2-4a^4-2a^6)x^2+(-4a-6a^3-a^5+a^7)x^3$ $+(1+a^2+2a^4+2a^6)x^4+(2a+4a^3+2a^5)x^5+(a^2+a^4)x^6.$
7_7		$(\ell^{-4}+2\ell^{-2}+2)+(-2\ell^{-2}-2-\ell^2)m^2+m^4$	$(a^{-4}+2a^{-2}+2)+(2a^{-3}+3a^{-1}+a)x+(-2a^{-4}-6a^{-2}-7-3a^2)x^2+(-4a^{-3}-8a^{-1}-3a+a^3)x^3$ $+(a^{-4}+2a^{-2}+4+3a^2)x^4+(2a^{-3}+5a^{-1}+3a)x^5+(a^{-2}+1)x^6.$

		P	F
0_1^2		$(-\ell^{-1}-\ell)m^{-1}$	$(a^{-1}+a)x^{-1}-1$
2_1^2		$(\ell+\ell^3)m^{-1}-\ell m$	$(-a-a^3)x^{-1}+a^2+(a+a^3)x$
4_1^2		$(-\ell^3-\ell^5)m^{-1}+(3\ell^3+\ell^5)m-\ell^3m^3$	$(a^3+a^5)x^{-1}-a^4+(-3a^3-2a^5+a^7)x+(a^4+a^6)x^2+(a^3+a^5)x^3$
4_1^2		$(-\ell^{-5}-\ell^{-3})m^{-1}+(\ell^{-3}-\ell^{-1})m$	$(a^{-5}+a^{-3})x^{-1}-a^{-4}+(-3a^{-5}-2a^{-3}+a^{-1})x+(a^{-4}+a^{-2})x^2+(a^{-5}+a^{-3})x^3$
5_1^2		$(-\ell^{-1}-\ell)m^{-1}+(\ell^{-1}+2\ell+\ell^3)m-\ell m^3$	$(a^{-1}+a)x^{-1}-1+(-2a^{-1}-4a-2a^3)x+(-1+a^4)x^2+(a^{-1}+3a+2a^3)x^3+(1+a^2)x^4$
6_1^2		$(\ell^5+\ell^7)m^{-1}+(-6\ell^5-3\ell^7)m+(5\ell^5+\ell^7)m^3$ $-\ell^5m^5$	$(-a^5-a^7)x^{-1}+a^6+(6a^5+4a^7-a^9+a^{11})x+(-3a^6-2a^8+a^{10})x^2+(-5a^5-4a^7+a^9)x^3$ $+(a^6+a^8)x^4+(a^5+a^7)x^5$
6_1^2		$(\ell^{-7}+\ell^{-5})m^{-1}+(-\ell^{-5}+\ell^{-3}-\ell^{-1})m$	$(-a^{-7}-a^{-5})x^{-1}+a^{-6}+(6a^{-7}+4a^{-5}-a^{-3}+a^{-1})x+(-3a^{-6}-2a^{-4}+a^{-2})x^2$ $+(-5a^{-7}-4a^{-5}+a^{-3})x^3+(a^{-6}+a^{-4})x^4+(a^{-7}+a^{-5})x^5$
6_2^2		$(\ell^{-7}+\ell^{-5})m^{-1}+(-\ell^{-7}-2\ell^{-5}+2\ell^{-3})m$ $+(\ell^{-5}-\ell^{-3})m^3$	$(-a^{-7}-a^{-5})x^{-1}+a^{-6}+(-2a^{-9}+3a^{-7}+3a^{-5}-2a^{-3})x+(-a^{-8}-2a^{-6}-a^{-4})x^2$ $+(a^{-9}-2a^{-7}-2a^{-5}+a^{-3})x^3+(a^{-8}+2a^{-6}+a^{-4})x^4+(a^{-7}+a^{-5})x^5.$
6_3^2		$(-\ell^3-\ell^5)m^{-1}+(2\ell^3-\ell^5-\ell^7)m$ $+(-\ell^3+\ell^5)m^3$	$(a^3+a^5)x^{-1}-a^4+(-2a^3-a^5-a^9)x+(-3a^6-3a^8)x^2+(a^3+a^9)x^3$ $+(a^4+3a^6+2a^8)x^4+(a^5+a^7)x^5$
6_3^2		$(-\ell^{-5}-\ell^{-3})m^{-1}+(2\ell^{-3}+\ell^{-1}+\ell)m-\ell^{-1}m^3$	$(a^{-5}+a^{-3})x^{-1}-a^{-4}+(-2a^{-5}-a^{-3}-a)x+(-3a^{-2}-3)x^2+(a^{-5}+a)x^3$ $+(a^{-4}+3a^{-2}+2)x^4+(a^{-3}+a^{-1})x^5$

REFERENCES

1. J. W. Alexander, Topological invariants of knots and links, *Trans. Amer. Math. Soc.* 30 (1928), 275–306.
2. C. W. Ashley, *The Ashley Book of Knots*, Faber and Faber, 1947.
3. J. S. Birman, Braids, links and mapping class groups, *Ann. Math. Studies* 82 (1974).
4. R. D. Brandt, W. B. R. Lickorish, and K. C. Millett, A polynomial invariant for unoriented knots and links, *Invent. Math.* 84 (1986), 563–573.
5. G. Burde and H. Zieschang, *Knots*, de Gruyter, 1985.
6. J. H. Conway, An enumeration of knots and links, *Computational Problems in Abstract Algebra* (ed. J. Leech), Pergamon Press, 1969, 329–358.
7. P. Freyd, D. Yetter, J. Hoste, W. B. R. Lickorish, K. Millett, and A. Ocneanu, A new polynomial invariant of knots and links, *Bull. Amer. Math. Soc.* 12 (1985), 239–246.
8. C. F. Ho, A new polynomial for knots and links—preliminary report, *Abstracts Amer. Math. Soc.* 6, 4 (1985), 300.
9. V. F. R. Jones, A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.* 12 (1985), 103–111.
10. L. H. Kauffman, An invariant of regular isotopy (to appear).
11. ______, State models and the Jones polynomial, *Topology* 26 (1987), 395–407.
12. W. B. R. Lickorish, A relationship between link polynomials, *Math. Proc. Cambridge Philos. Soc.*, (1986), 100, 109–112.
13. W. B. R. Lickorish and K. C. Millett, A polynomial invariant of oriented links, *Topology* 26 (1987), 107–141.
14. K. Murasugi, Jones polynomials and classical conjectures in knot theory, *Topology* 26 (1987), 187–194.
15. L. Neuwirth, The theory of knots, *Scientific American* (June 1979), 84–96.
16. K. Reidemeister, *Knotentheorie* (reprint), Chelsea, New York, 1948.
17. D. Rolfsen, *Knots and Links*, Publish or Perish, Wilmington, Delaware, 1976.
18. M. B. Thistlethwaite, A spanning tree expansion of the Jones polynomial, *Topology* 26 (1987), 296–309.
19. ______, Knot tabulations and related topics, In *Aspects of Topology* (Ed. I. M. James and E. H. Kronheimer) L.M.S. Lecture Notes 93 (1985), 1–76.

Four Mathematical Clerihews

STEVEN CUSHING
Stonehill College
North Easton, MA 02357

1. Pythagoras
   Did stagger us
   And our reason encumber
   With irrational number.

2. Kurt Gödel
   Created a hurdle
   For the truths of a system:
   You just can't list'em!

3. Wily Fermat propounded,
   "Many will be confounded
   At the thought that my theorem
   Is really quite near'em."

4. Said Alfred Tarski:
   "Talk of 'truth' is a farce. Key
   To getting it right
   Is to know 'Snow is white.'"

NOTES

A Coin-Tossing Problem and Some Related Combinatorics

ROGER H. MORITZ
ROBERT C. WILLIAMS
Alfred University
Alfred, NY 14802

In this note we analyze a simple coin-tossing game. The arguments use only the most elementary facts in probability; the reader will, however, have to invest a little patience—say, 10 or 15 minutes of undivided attention—to follow some of the combinatorial reasoning. The problem is the following:

> Players $A_1, A_2, \ldots, A_n$ each toss, in turn, a (possibly biased) coin. If a player tosses a head, he goes out of the game and doesn't toss again. The remaining players continue to toss until all go out. For any permutation σ of the n players, find the probability that the players will go out in order σ.

Let S_n denote the set of all permutations of n players. For small values of n and a specific $\sigma \in S_n$, the stated problem is not difficult. The general problem is a bit harder. The solution, however, is concise and appealing, and leads to an unusual way of partitioning S_n. It also leads to an interesting problem in combinatorics.

The Solution

We first establish our notation. We set $p = P(\text{head})$ and $q = 1 - p = P(\text{tail})$. When referring to a particular outcome of our game, we will express the result relative to the order in which the players toss their coins. For example, the permutation $\sigma = [BDCA]$ refers to a situation in which $n = 4$ players toss in the order A, B, C, D, with B being the first to achieve a head, D the second, and so on. Note that this representation of a permutation is not at all the same as cycle notation (in which the above outcome would be written (124)). Rather, our representation is just the bottom row in the standard symbol $\begin{pmatrix} ABCD \\ BDCA \end{pmatrix}$. In addition, when referring to the probability $P(\sigma)$ of a particular outcome

$$\sigma = \left[A_{i_1}A_{i_2} \cdots A_{i_n}\right],$$

we delete the parentheses and simply write

$$P\left[A_{i_1}A_{i_2} \cdots A_{i_n}\right].$$

Let's handle the simple case of $n = 2$, not so much for its ease but because the argument that is employed will generalize. There are, of course, only 2 possible outcomes, $[AB]$ and $[BA]$. To find the probability of $[AB]$, we observe that player A on his first attempt will toss either a head or a tail. In the first case outcome $[AB]$ is determined. In the second instance player B will also have to toss tails and, should this happen, we are right back where we started. Thus,

$$P[AB] = p + q^2 P[AB].$$

Solving for $P[AB]$, we get

$$P[AB] = p/(1-q^2) = (1-q)/(1-q^2) = 1/(1+q).$$

It then follows that

$$P[BA] = 1 - P[AB] = q/(1+q).$$

Assuming the general pattern isn't yet apparent, let's try $P[BAC]$. On the initial round of tosses, A must fail. If B succeeds, then A and C remain, with C tossing first, and we want them to go out in order AC. The probability of this is simply $P[BA]$, which we now know. If, on the other hand, B should toss tails then C will need to as well. Once this occurs we are just starting over again. Thus,

$$P[BAC] = qpP[BA] + q^3 P[BAC].$$

Consequently,

$$P[BAC] = qpP[BA]/(1-q^3) = qP[BA]/(1+q+q^2) = \frac{q^2}{(1+q)(1+q+q^2)}.$$

Applying similar reasoning to the other elements of S_3, we find

$$P[ABC] = \frac{1}{(1+q)(1+q+q^2)},$$

$$P[ACB] = P[BCA] = \frac{q}{(1+q)(1+q+q^2)},$$

$$P[BAC] = P[CAB] = \frac{q^2}{(1+q)(1+q+q^2)},$$

$$P[CBA] = \frac{q^3}{(1+q)(1+q+q^2)}.$$

Certainly these results seem reasonable in that $[ABC]$ is obviously the most likely outcome, $[CBA]$ the least likely.

In our analysis of $P[BAC]$, we needed the probability of the permutation $[BA]$, since that is essentially what must occur after player B leaves the game. We will be referring later to this derived permutation; let us denote it by σ'. In general, for $\sigma \in S_n$, σ' is the element of S_{n-1} that must occur after the first player demanded by σ tosses a head. Note that if player A_j is the first to go out according to σ, then the remaining $n-1$ players toss in the order $A_{j+1}, A_{j+2}, \ldots, A_n, A_1, \ldots, A_{j-1}$. Thus, for instance, if $\sigma = [CEADB] \in S_5$, then $\sigma' = [BCAD]$ because after C leaves the game, D will toss first (so we will now call him A), then E will toss (so we call him B; he is the next to toss heads according to σ and, therefore, the first player named in σ'), etc.

In order to state the solution of the problem concisely, we need a bit of notation. For any permutation $\sigma \in S_n$, we define the norm of σ, $|\sigma|$, to be the *minimum* number of players who must toss tails in order for σ to be the order in which the players go out. In general, for

$$\sigma = \left[A_{i_1}A_{i_2}\cdots A_{i_n}\right],$$

we denote $|\sigma|$ by $|A_{i_1}A_{i_2}\cdots A_{i_n}|$. For example, with $n=3$, $|BAC|=2$, because A must toss tails to start and then, after B tosses heads, C must toss tails. The reader may confirm that $|CEADB|=4$. Note the general formulas

$$|A_1A_2A_3\cdots A_n|=0 \quad \text{and} \quad |A_nA_{n-1}A_{n-2}\cdots A_1|=\frac{n(n-1)}{2}.$$

It is also clear that

$$0\leqslant|\sigma|\leqslant\sum_{i=1}^{n-1} i=\frac{n(n-1)}{2} \text{ for all } \sigma\in S_n.$$

Let us denote this maximum value of the norm, namely, $n(n-1)/2$, by $M(n)$. Also, for any permutation σ, there is a simple relation between its norm and that of its derived permutation σ'. If player A_j is the first to toss heads according to σ, then σ and σ' call for the same sequence except for the $j-1$ failures preceding A_j's success. Therefore,

$$|\sigma'|=|\sigma|-(j-1).$$

Finally, for $n=1,2,\ldots,$ define the polynomial

$$D_n(x)=(1+x)(1+x+x^2)\cdots(1+x+x^2+\cdots+x^{n-1}).$$

Note that the degree of $D_n(x)$ is $1+2+\cdots+(n-1)=n(n-1)/2=M(n)$.

We now state the solution to our coin-tossing problem.

THEOREM 1. *For* $\sigma\in S_n$, $P(\sigma)=q^{|\sigma|}/D_n(q)$.

Proof. (By induction on n). We have already directly verified the formula for $n=2$; we assume it true for all permutations of $n-1$ players. Let $\sigma\in S_n$ and player A_j be the first to toss heads according to σ; σ' is the derived permutation of σ, as previously defined. We now reason as we have before: in order for the players to go out in order σ, players 1, 2,..., $j-1$ must all fail. At this point player j will either succeed or fail. If he succeeds and leaves the game we are left with sequence σ' to be accomplished. If he fails, then all n players will need to fail, at which point we face the same problem we had at the start. That is,

$$P(\sigma)=q^{j-1}pP(\sigma')+q^nP(\sigma),$$

so that

$$P(\sigma)=\frac{q^{j-1}pP(\sigma')}{1-q^n}=\frac{q^{j-1}P(\sigma')}{1+q+q^2+\cdots+q^{n-1}}.$$

By our inductive assumption,

$$P(\sigma')=\frac{q^{|\sigma'|}}{D_{n-1}(q)}.$$

Since $|\sigma'|=|\sigma|-(j-1)$, we have that

$$P(\sigma)=\frac{q^{j-1}q^{(|\sigma|-(j-1))}}{[D_{n-1}(q)](1+q+q^2+\cdots+q^{n-1})}=\frac{q^{|\sigma|}}{D_n(q)}.$$

As a computational example the reader may check that, with $p = 1/2$, $P[CADBE] = (1/2)^6/D_5(1/2) \doteq 0.00164$. The alternate form

$$P(\sigma) = \frac{p^{n-1}q^{|\sigma|}}{(1-q^2)\cdots(1-q^n)},$$

may be more convenient for computation. In the present example, it yields

$$P[CADBE] = 2^{-10}\left(\frac{4}{3}\right)\left(\frac{8}{7}\right)\left(\frac{16}{15}\right)\left(\frac{32}{31}\right),$$

which, by the way, gives an immediate feel for the size of $P[CADBE]$.

The Numbers $\left\langle {n \atop k} \right\rangle$

Theorem 1 tells us that for $\sigma_1, \sigma_2 \in S_n$, $P(\sigma_1) > P(\sigma_2)$ if and only if $|\sigma_1| < |\sigma_2|$, and that $P(\sigma_1) = P(\sigma_2)$ if and only if $|\sigma_1| = |\sigma_2|$. Taking equal probabilities as an equivalence relation on the set S_n, we induce a partition of this set. Let $S_{n,k} = \{\sigma \in S_n\colon |\sigma| = k\}$. For example, in S_4, we have:

$$\begin{aligned}
S_{4,0} &= \{[ABCD]\},\\
S_{4,1} &= \{[BCDA],[ACDB],[ABDC]\},\\
S_{4,2} &= \{[CDAB],[BDAC],[BCAD],[ACBD],[ADBC]\},\\
S_{4,3} &= \{[DABC],[CABD],[CDBA],[BACD],[BDCA],[ADCB]\},\\
S_{4,4} &= \{[BADC],[CADB],[CBDA],[DBCA],[DACB]\},\\
S_{4,5} &= \{[CBAD],[DBAC],[DCAB]\},\\
S_{4,6} &= \{[DCBA]\}.
\end{aligned}$$

Several questions arise naturally at this point. Is there some other way to characterize this partition of S_n? Is there a more algebraic method for computing $|\sigma|$? Is there a simple formula for the number of permutations of n players having norm k? These are, to our knowledge, unsolved problems. However, with regard to the last question, there are some things we can say.

For $n = 1, 2, 3, \ldots$ and $k = 0, 1, \ldots, M(n) = n(n-1)/2$, let $\left\langle {n \atop k} \right\rangle$ denote the number of permutations in S_n with norm k; that is, the number of elements in set $S_{n,k}$. For later convenience we specify that for $k < 0$ or $k > M(n)$, $S_{n,k} = \phi$ and $\left\langle {n \atop k} \right\rangle = 0$. TABLE 1 contains some values of $\left\langle {n \atop k} \right\rangle$. Although we do not have an explicit formula for $\left\langle {n \atop k} \right\rangle$, we can, nonetheless, prove four elementary properties of these numbers. Strikingly, these properties parallel well-known properties of the binomial coefficients.

THEOREM 2 (Row Sum Property). $\displaystyle\sum_{k=0}^{M(n)} \left\langle {n \atop k} \right\rangle = n!.$

Proof. This is immediate from the definition of $\left\langle {n \atop k} \right\rangle$.

THEOREM 3 (Additive Recursion Property). $\displaystyle\left\langle {n \atop k} \right\rangle = \sum_{j=k-n+1}^{k} \left\langle {n-1 \atop j} \right\rangle.$

(This simply says that the entry in the kth column, nth row of TABLE 1 is the sum of the (at most) n entries of row $n-1$, starting in column k and moving to the left.)

TABLE 1

Values of $\left\langle {n \atop k} \right\rangle$

n \ k	0	1	2	3	4	5	6	7	8	9	10	11	12	...
1	1													
2	1	1												
3	1	2	2	1										
4	1	3	5	6	5	3	1							
5	1	4	9	15	20	22	20	15	9	4	1			
6	1	5	14	29	49	71	90	101	101	90	71	49	29	...
7	1	6	20	49	98	169	259	359	455	531	573	573	531	...
.														
.														
.														

Proof. Consider an arbitrary $\sigma \in S_{n,k}$ (i.e., $|\sigma| = k$) and let player A_j be the first to toss a head according to σ. Recall that $|\sigma'| = |\sigma| - (j-1) = k - j + 1$ where $\sigma' \in S_{n-1}$ is the derived permutation of σ. Since $1 \leqslant j \leqslant n$, we see that for any $\sigma \in S_{n,k}$ having A_j as initial player out, its derived permutation belongs to the set $S_{n-1,k-j+1}$. This establishes a one-to-one correspondence between elements of $S_{n,k}$ and the disjoint union

$$S_{n-1,k-n+1} \cup S_{n-1,k-n+2} \cup S_{n-1,k-n+3} \cup \cdots \cup S_{n-1,k}.$$

(The preceding argument was suggested by our colleague Roger Douglass.)

THEOREM 4 (Symmetry Property). $\left\langle {n \atop k} \right\rangle = \left\langle {n \atop M(n) - k} \right\rangle.$

Proof. Let us agree to call players A_i and A_j symmetric if $i + j = n + 1$. We will write $\hat{A}_i$ to denote the player symmetric with player A_i. Correspondingly, permutations $\sigma_1, \sigma_2 \in S_n$ will be called symmetric with one another if the first player named in σ_1 is symmetric with the first named in σ_2, and the second players named are symmetric, and so on. Denoting the permutation symmetric to σ by $\hat{\sigma}$, we have simply that if

$$\sigma = \left[A_{i_1} A_{i_2} A_{i_3} \cdots A_{i_n}\right],$$

then

$$\hat{\sigma} = \left[\hat{A}_{i_1} \hat{A}_{i_2} \hat{A}_{i_3} \cdots \hat{A}_{i_n}\right].$$

The norms of symmetric permutations obviously satisfy the relation

$$|\sigma| + |\hat{\sigma}| = M(n),$$

so that

$$|\hat{\sigma}| = M(n) - |\sigma|.$$

This establishes a canonical bijection between the sets $S_{n,k}$ and $S_{n,M(n)-k}$.

THEOREM 5 (Generating Function Property). *$D_n(x)$ is the generating function for the numbers $\left\langle {n \atop k} \right\rangle$ in the sense that $\left\langle {n \atop k} \right\rangle$ is the coefficient of x^k in the expansion of the polynomial $D_n(x)$.*

Proof. For $0 < q < 1$,

$$1 = \sum_{\sigma \in S_n} P(\sigma) = \sum_{\sigma \in S_n} \frac{q^{|\sigma|}}{D_n(q)} = \frac{1}{D_n(q)} \sum_{k=0}^{M(n)} \left\langle {n \atop k} \right\rangle q^k.$$

Thus, $D_n(q) = \sum_{k=0}^{M(n)} \left\langle {n \atop k} \right\rangle q^k$, for $0 < q < 1$. For polynomial functions, this can only happen if they are identical.

As mentioned earlier, we do not know of a simple formula for the numbers $\left\langle {n \atop k} \right\rangle$. It would also be interesting to know if these numbers arise in other contexts.

REFERENCE

1. Curtis Cooper, Geometric series and a probability problem, *Amer. Math. Monthly* 93 (1986), 126–127.

The Game of Quatrainment

T. E. GANTNER
The University of Dayton
Dayton, OH 45469

The concept of a finite-state machine is important in a discrete mathematics course because of its role in the study of formal languages. Typical textbook examples often have just a few states and inputs: for example, parity check machines and binary adders. It is desirable to have a few more-complex examples to present to the student as supplementary material; at the same time, such examples should be on a fairly elementary level. One such example, Think-A-Dot, was discussed in detail in [**1**]. The example that we discuss here is the game of Quatrainment [**2**], which may be implemented on a variety of home computers, but which may equally well be played on ordinary graph paper. It may be completely analyzed using such elementary tools as: mathematical modeling, Boolean matrices, modulo 2 arithmetic, and algorithms. It is also interesting as an example of matrix theory over a finite field.

One begins playing Quatrainment by placing a random pattern of X's in each of two 4-by-4 grids called A and B. For example:

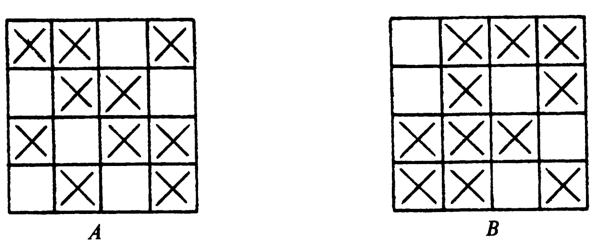


FIGURE 1

The object of the game is to modify the pattern of A, using a finite sequence of moves, so that it matches the pattern of B. Each move is made by selecting one of the 16 cells of A, which we label from 0 to 15, starting with the upper left corner, as indicated in FIGURE 2.

0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

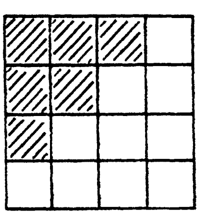
FIGURE 2

Each cell of A has one of two states: it is marked by an X, or else it is blank. Since there are 4 corner cells, 4 center cells, and 8 edge cells, there are three types of moves, which are defined by the following rules:

If a corner cell is selected, then reverse the states of the six cells in the triangle at that cell (FIGURE 3a).

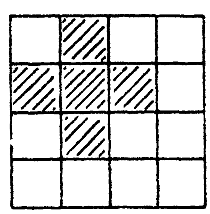
If a center cell is selected, then reverse its state as well as the states of its four neighbors (FIGURE 3b).

If an edge cell is selected, then reverse the states of its three neighbors (FIGURE 3c).



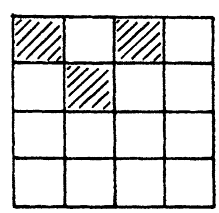
Choosing Cell 0

FIGURE 3a



Choosing Cell 5

FIGURE 3b



Choosing Cell 1

FIGURE 3c

In addition, we could suppose that the game is played on a computer and that the computer keeps track of the elapsed time as well as the number of moves. The task should be accomplished in minimum time using the least number of moves. We show below that there is always a unique solution using a minimum number of moves.

The Finite State Machine and Its Matrix Representation

As a finite state machine, each cell of A is either blank or is marked with an X, so A is capable of having 2^{16} distinct states. Each of these states may be represented as a 4-by-4 matrix of 0's and 1's; an entry 1 occupies the position of each cell that is marked with an X, and an entry 0 occupies the position of a blank cell. Thus the set of states is just the set $\mathscr{S}$ of 4-by-4 matrices with entries 0 or 1. The inputs of the finite state machine are the cell numbers $I = \{0, 1, 2, \ldots, 15\}$, and for each state S and input $k \in I$, the next state $f_k(S)$ is the state derived from S by selecting the input cell $k \in I$. For example, if $k=0$ and if S is the state A of FIGURE 1, then

$$S=\begin{pmatrix}1 & 1 & 0 & 1\\ 0 & 1 & 1 & 0\\ 1 & 0 & 1 & 1\\ 0 & 1 & 0 & 1\end{pmatrix}, \quad \text{and} \quad f_0(S)=\begin{pmatrix}0 & 0 & 1 & 1\\ 1 & 0 & 1 & 0\\ 0 & 0 & 1 & 1\\ 0 & 1 & 0 & 1\end{pmatrix}.$$

We may define output functions, if desired, to be identical with the next-state functions.

We further note that the next-state functions can be described in terms of matrix addition. For example, associated with the moves that correspond to cell choices 0, 5, 1 of FIGURE 3, we have matrices

$$M_0=\begin{pmatrix}1 & 1 & 1 & 0\\ 1 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \qquad M_5=\begin{pmatrix}0 & 1 & 0 & 0\\ 1 & 1 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \qquad M_1=\begin{pmatrix}1 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}.$$

Similarly we have a matrix M_k for each $k \in I$. If we add the entries of matrices using modulo 2 arithmetic, then we have

$$f_k(S)=S+M_k, \quad \text{for all } k \in I \quad \text{and} \quad S \in \mathscr{S}.$$

Using the fact that matrix addition is both associative and commutative, two facts are now evident:

$$(f_k \circ f_k)(S)=S+M_k+M_k=S, \quad \text{so } f_k \circ f_k = id, \tag{1}$$

where id is the identity function on $\mathscr{S}$, and

$$(f_k \circ f_j)(S)=S+M_j+M_k=S+M_k+M_j=(f_j \circ f_k)(S), \quad \text{so } f_k \circ f_j = f_j \circ f_k. \tag{2}$$

Thus the semigroup I^* of the machine consisting of all finite input sequences is actually a commutative group in which each input element is its own inverse. Also, associated with each word $k_1 k_2 \cdots k_n \in I^*$ is the matrix

$$h(k_1 k_2 \cdots k_n)=M_{k_1}+M_{k_2}+\cdots+M_{k_n},$$

and this formula defines an isomorphism of I^* with the additive subgroup (or linear subspace) $\mathscr{M}$ of $\mathscr{S}$ that is generated by the selection matrices $M_0, M_1, \ldots, M_{15}$.

A Solution Algorithm

In terms of finite state machines, the object of Quatrainment may be rephrased as follows. Given two states A and B, determine a sequence $k_1 k_2 \ldots k_n$ of inputs such that $(f_{k_n} \circ \cdots \circ f_{k_2} \circ f_{k_1})(A)=B$. A solution algorithm exists if we can find a set of input sequences, each of which will change the state of just one chosen cell. By rotation and reflection, it suffices to obtain a sequence that will change the state of just one corner cell, center cell, or edge cell. For example, starting with the zero-matrix state, we can change the state of cell 0 by the sequence of moves indicated below in (3). Similarly, the sequence (4) will change the state of cell 5, and the sequence (5) will change the state of cell 1.

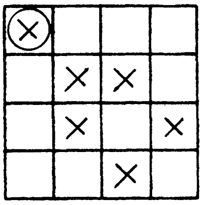
$$
\begin{gathered}
\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_0}\begin{pmatrix}1&1&1&0\\1&1&0&0\\1&0&0&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_5}\begin{pmatrix}1&0&1&0\\0&0&1&0\\1&1&0&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_6}\begin{pmatrix}1&0&0&0\\0&1&0&1\\1&1&1&0\\0&0&0&0\end{pmatrix}\\
\xrightarrow{f_9}\begin{pmatrix}1&0&0&0\\0&0&0&1\\0&0&0&0\\0&1&0&0\end{pmatrix}\xrightarrow{f_{11}}\begin{pmatrix}1&0&0&0\\0&0&0&0\\0&0&1&0\\0&1&0&1\end{pmatrix}\xrightarrow{f_{14}}\begin{pmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}
\end{gathered}
\tag{3}
$$

$$
\begin{gathered}
\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_3}\begin{pmatrix}0&1&1&1\\0&0&1&1\\0&0&0&1\\0&0&0&0\end{pmatrix}\xrightarrow{f_5}\begin{pmatrix}0&0&1&1\\1&1&0&1\\0&1&0&1\\0&0&0&0\end{pmatrix}\xrightarrow{f_6}\begin{pmatrix}0&0&0&1\\1&0&1&0\\0&1&1&1\\0&0&0&0\end{pmatrix}\\
\xrightarrow{f_7}\begin{pmatrix}0&0&0&0\\1&0&0&0\\0&1&1&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_9}\begin{pmatrix}0&0&0&0\\1&1&0&0\\1&0&0&0\\0&1&0&0\end{pmatrix}\xrightarrow{f_{12}}\begin{pmatrix}0&0&0&0\\0&1&0&0\\0&1&0&0\\1&0&1&0\end{pmatrix}\\
\xrightarrow{f_{13}}\begin{pmatrix}0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}
\end{gathered}
\tag{4}
$$

$$
\begin{gathered}
\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_0}\begin{pmatrix}1&1&1&0\\1&1&0&0\\1&0&0&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_2}\begin{pmatrix}1&0&1&1\\1&1&1&0\\1&0&0&0\\0&0&0&0\end{pmatrix}\xrightarrow{f_3}\begin{pmatrix}1&1&0&0\\1&1&0&1\\1&0&0&1\\0&0&0&0\end{pmatrix}\\
\xrightarrow{f_4}\begin{pmatrix}0&1&0&0\\1&0&0&1\\0&0&0&1\\0&0&0&0\end{pmatrix}\xrightarrow{f_8}\begin{pmatrix}0&1&0&0\\0&0&0&1\\0&1&0&1\\1&0&0&0\end{pmatrix}\xrightarrow{f_{13}}\begin{pmatrix}0&1&0&0\\0&0&0&1\\0&0&0&1\\0&0&1&0\end{pmatrix}\\
\xrightarrow{f_{14}}\begin{pmatrix}0&1&0&0\\0&0&0&1\\0&0&1&1\\0&1&1&1\end{pmatrix}\xrightarrow{f_{15}}\begin{pmatrix}0&1&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}
\end{gathered}
\tag{5}
$$

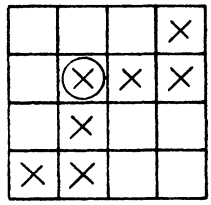
The above input sequences are most easily remembered in terms of the patterns indicated in FIGURE 4. If we select the six cells checked in FIGURE 4a, we change the state of only cell 0. Similarly, FIGURE 4b indicates the cell choices needed to change the state of only cell 5, and Figure 4c gives the choices needed to change only cell 1.

Thus a solution always exists. Simply change the state of each cell of A that does not match the corresponding cell of B by using a finite sequence of inputs as indicated by the above patterns, and the task is done. However, if we are required to change several cells of A, then in all likelihood some of the input choices will be made several times; but this is not necessary, because each input choice is its own group inverse.



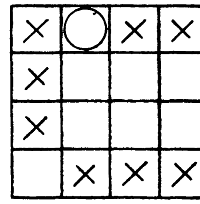
Change a Corner

FIGURE 4a



Change a Center

FIGURE 4b



Change an Edge

FIGURE 4c

Thus, each input need be used just once or not at all, and so there should be a solution which uses at most 16 moves.

A Minimal Algorithm

Let us look again at our initial example, where

$$A=\begin{pmatrix}1&1&0&1\\0&1&1&0\\1&0&1&1\\0&1&0&1\end{pmatrix}, \quad \text{and} \quad B=\begin{pmatrix}0&1&1&1\\0&1&0&1\\1&1&1&0\\1&1&0&1\end{pmatrix}.$$

Given two such states, we construct the matrix M according to the following rule: an entry of M is 0 if the corresponding entries of A and B are the same, otherwise it is 1. In other words, $M=A+B$. Since entry addition is modulo 2, this means that $A+M=B$. In our example:

$$M=\begin{pmatrix}1&0&1&0\\0&0&1&1\\0&1&0&1\\1&0&0&0\end{pmatrix}.$$

The question now is this: Do there exist distinct inputs $0 \leqslant k_1 < k_2 < \cdots < k_n \leqslant 15$ such that $M = M_{k_1} + M_{k_2} + \cdots + M_{k_n}$? In the terminology of vector spaces, do $M_0, M_1, \ldots, M_{15}$ form a basis for $\mathscr{S}$? The answer is yes. Above we saw how to express each of the standard basis matrices $E_0, E_1, \ldots, E_{15}$ in terms of $M_0, M_1, \ldots, M_{15}$, where E_k has a 1 in cell k and all of its other entries are zero. Thus $M_0, M_1, \ldots, M_{15}$ generate all of $\mathscr{S}$; and the linear dimension of $\mathscr{S}$ over the field of integers modulo 2 is 16, so they form a basis for $\mathscr{S}$. Since they do form a basis, each such $M \in \mathscr{S}$ is expressible *uniquely* as a sum of at most 16 of the selection matrices $M_0, M_1, \ldots, M_{15}$. Thus there is a unique solution having a minimum number of moves. In the above example, it is not difficult to verify that

$$M = M_0 + M_1 + M_3 + M_4 + M_7 + M_9 + M_{10} + M_{12}.$$

Unfortunately, from a practical point of view, in order to express an M as a unique sum of the matrices $M_0, M_1, \ldots, M_{15}$, one must either solve a 16-by-16 linear system, or one must make a chart like those of FIGURE 4 for each entry of A that must be changed and then determine from these charts which cell choices must be made an odd number of times. Both of these procedures are time consuming; keep in mind that the object of the game is to do the task in as few moves as possible *and* in minimal time. Therefore, it is still a challenging game for humans.

Some Modifications

We may try to play games with the game itself either by generalizing it to a larger (even higher-dimensional) grid, or by modifying the effect of the various cell choices. If we keep the same grid, there are two obvious cell choice modifications that may be made:

(a) Make an edge cell selection reverse the states of the selected cell as well as its three neighbors; for example, for cell 1 we would have

$$M_1 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{instead of } M_1 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

(b) Make the corner and center cell selections behave in the chess-board way that the edge cell selections behave, that is, do not change the states of two neighboring cells.

For example, for cells 0 and 5, we would have

$$M_0 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{instead of } M_0 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$M_5 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{instead of } M_5 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Unfortunately neither of these modified games is solvable in general because in neither case is the subspace $\mathscr{M}$ of $\mathscr{S}$ equal to all of $\mathscr{S}$. Thus there will exist states A and B so that $A + B = M \notin \mathscr{M}$, hence M will not be obtainable by any finite sequence of cell choices; it is a nice exercise for the reader to find such states. Here are the reasons for the claim.

In modification (a), it is easily verified that

$$M_0 + M_3 = M_4 + M_7, \qquad M_0 + M_{12} = M_1 + M_{13},$$
$$M_3 + M_{15} = M_2 + M_{14}, \qquad M_{12} + M_{15} = M_8 + M_{11},$$

so that $\mathscr{M}$ is generated by the 12 matrices $M_2, M_3, \dots, M_{12}, M_{13}$; and these are linearly independent, so the dimension of $\mathscr{M}$ is 12.

In modification (b), it is easily seen that

$$M_5 + M_{10} = M_2 + M_7 + M_8 + M_{13},$$
$$M_6 + M_9 = M_1 + M_4 + M_{11} + M_{14},$$

so, for example, M_5 and M_6 can be eliminated to obtain a basis for $\mathscr{M}$; in this case the dimension of $\mathscr{M}$ is 14.

REFERENCES

1. John A. Beidler, Think-a-dot revisited, this MAGAZINE 46 (May-June, 1973), 128–136.
2. Sean Puckett, Quatrainment, *COMPUTE! Magazine* 6 (February, 1984), 76–91.

Love Affairs and Differential Equations

STEVEN H. STROGATZ
Harvard University
Cambridge, MA 02138

The purpose of this note is to suggest an unusual approach to the teaching of some standard material about systems of coupled ordinary differential equations. The approach relates the mathematics to a topic that is already on the minds of many college students: the time-evolution of a love affair between two people. Students seem to enjoy the material, taking an active role in the construction, solution, and interpretation of the equations.

The essence of the idea is contained in the following example.

Juliet is in love with Romeo, but in our version of this story, Romeo is a fickle lover. The more Juliet loves him, the more he begins to dislike her. But when she loses interest, his feelings for her warm up. She, on the other hand, tends to echo him: her love grows when he loves her, and turns to hate when he hates her.

A simple model for their ill-fated romance is

$$dr/dt = -aj, \qquad dj/dt = br,$$

where

$$r(t) = \text{Romeo's love/hate for Juliet at time } t$$
$$j(t) = \text{Juliet's love/hate for Romeo at time } t.$$

Positive values of r, j signify love, negative values signify hate. The parameters a, b are positive, to be consistent with the story.

The sad outcome of their affair is, of course, a neverending cycle of love and hate; their governing equations are those of a simple harmonic oscillator. At least they manage to achieve simultaneous love one-quarter of the time.

As one possible variation, the instructor may wish to discuss the more general second-order linear system

$$dr/dt = a_{11}r + a_{12}j$$
$$dj/dt = a_{21}r + a_{22}j,$$

where the parameters a_{ik} $(i, k = 1,2)$ may be either positive or negative. A choice of sign specifies the romantic style. As named by one of my students, the choice a_{11}, $a_{12} > 0$ characterizes an "eager beaver"—someone both excited by his partner's love for him and further spurred on by his own affectionate feelings for her. It is entertaining to name the other three possible styles, and also to contemplate the romantic forecast for the various pairings. For instance, can a cautious lover ($a_{11} < 0$, $a_{12} > 0$) find true love with an eager-beaver?

Additional complications may be introduced in the name of realism or mathematical interest. Nonlinear terms could be included to prevent the possibilities of unbounded passion or disdain. Poets have long suggested that the equations should be nonautonomous ("In the spring, a young man's fancy lightly turns to thoughts of love"—Tennyson). Finally, the term "many-body problem" takes on new meaning in this context.

Diagrams Venn and How

J. CHRIS FISHER
E. L. KOH
University of Regina
Regina, Canada S4S 0A2

BRANKO GRÜNBAUM
University of Washington
Seattle, WA 98195

The fascinating drawing by Boyd [1] shows a Venn diagram consisting of five rectangles. It suggested to us the desirability of finding *aesthetically pleasing* Venn diagrams with arbitrarily many sets, possessing moreover the property that the appropriate distribution of regions is *obvious on sight*. It is easy to see that Boyd's method cannot work for more than five sets. Instead, we propose here another simple construction of Venn diagrams—one that may be of particular interest to students taking a course in finite mathematics. Our presentation makes use of base 2 arithmetic, the notion of absolute value, and a simple mathematical induction; on the other hand, each of these devices can be eliminated either by appropriate explanations, or by simple handwaving. A general discussion of Venn diagrams, including historic and bibliographic references, can be found in [2] and [3].

It is generally (though not universally) agreed that in a Venn diagram for n sets

a) *the boundary of each set should be a simple closed curve in the plane*, and

b) *these curves should determine precisely* 2^n *regions, each one a connected set of the form* $X_1 \cap X_2 \cap \cdots \cap X_n$ *where* X_j *is chosen to be either the interior or the exterior of the jth set.*

We shall understand Venn diagrams in this sense, adding to these conditions the requirement that the boundary of each set should be easily distinguishable by sight from the boundaries of the $n-1$ other sets. We require further that the boundaries be easily constructed polygons.

We begin with a rule for labelling the regions in Venn diagrams. First the sets of a given Venn diagram are labelled by integers from 1 to n. (In the illustrative example shown in FIGURE 1 we have $n=6$ and have used Roman numerals for the set labels.) Then each of the 2^n regions is labelled by the binary number $a_n \ldots a_3 a_2 a_1$, where $a_j = 1$ if the region is in the set labelled j and $a_j = 0$ if the region is in the complement of the set j. For brevity, we write this binary label in the decimal notation; thus 9 stands for 001001, and is the label of the region contained in the sets 1 and 4 and outside the sets 2, 3, 5, and 6. Note that with this labelling the difference between labels of two regions adjacent along the boundary of the jth set is 2^{j-1}. Such a labelling is useful in various situations (cf. [4]).

Now for the promised construction of a Venn diagram for n sets. First let's define the *zigzag function* $y = C(x)$ as a periodic function with period 2 by setting $C(x) = |x-1| - 1/2$ for $0 \leqslant x \leqslant 2$. For $j = 1,2,3\ldots$ let the set V_j be bounded by $x=0$, $x=1$, $y=-1/2$ and the zigzag

$$y_j = C(2^{j-1}x)/2^{j-1}.$$

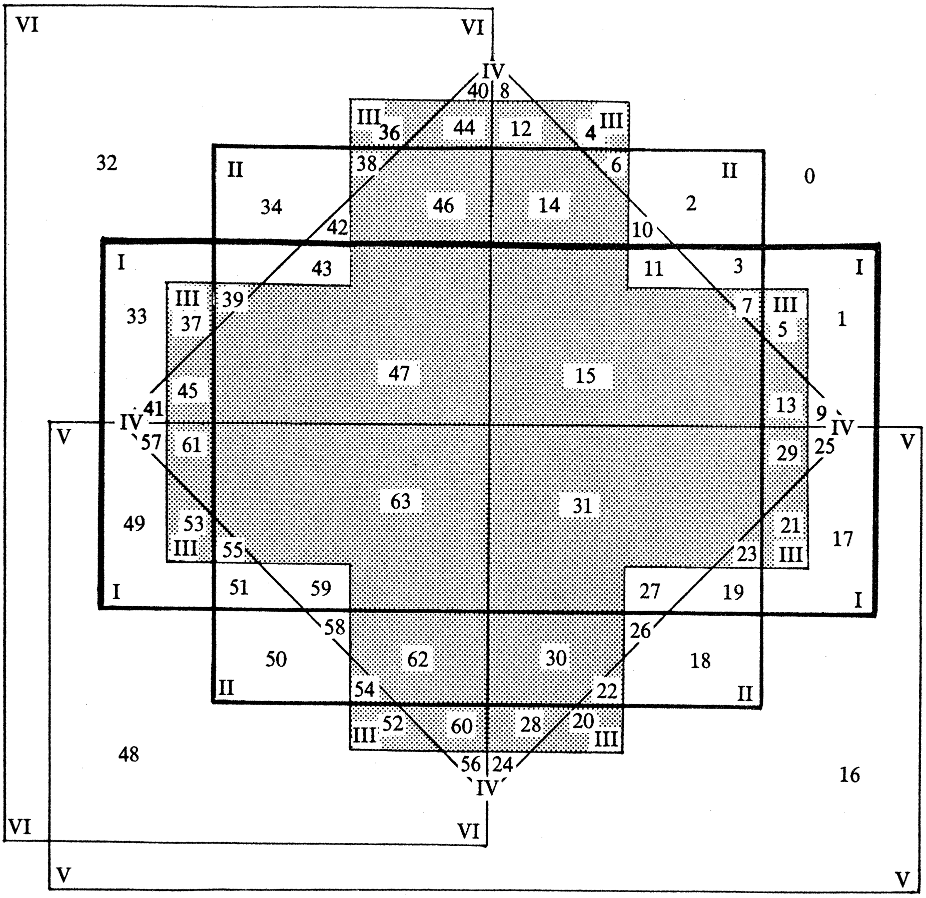
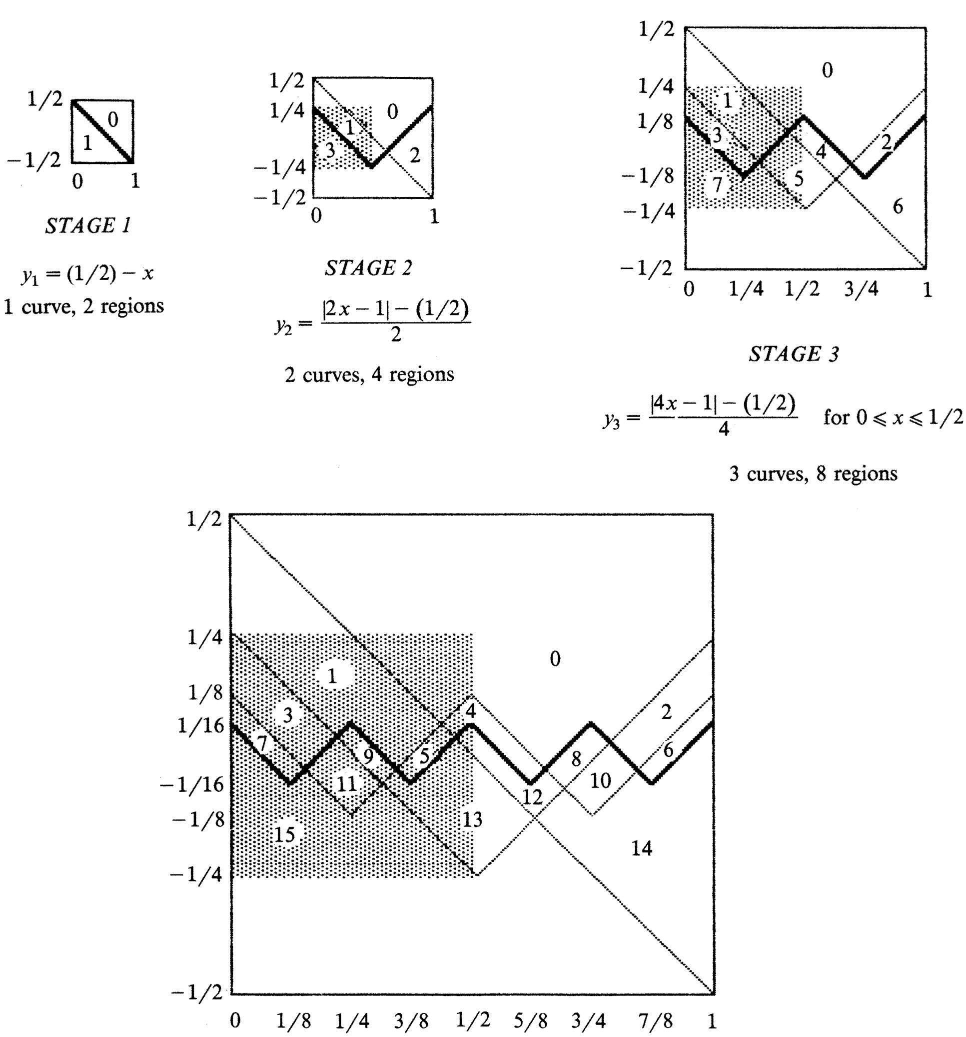


FIGURE 1.
A Venn diagram for 6 sets. Each region is given the decimal number that equals $a_6a_5a_4a_3a_2a_1$ in base 2, where $a_j = 1$ when the region is in set j and 0 otherwise.

That is, V_j is the portion of a unit square below the zigzag that starts at the point $(0, 2^{-j})$ and alternates going down then up between the horizontal lines $y = \pm(1/2^j)$ until it reaches the vertical line $x = 1$ (as in FIGURE 2). Except for y_1, the zigzags end at the point $(1, 2^{-j})$. We claim that *for each $n \geqslant 1$, the first n sets form a Venn diagram.*

Our proof of the claim uses induction to show that y_N cuts in two each of the 2^{N-1} regions formed by the previous zigzags. Specifically, we assume that the zigzags y_j, $j = 1, 2, \ldots, N-1$, determine a Venn diagram of 2^{N-1} regions. We label these regions from 0 to 2^{N-1} as in our opening exercise. Scale that diagram to half its size and place it in the middle of the left-hand side of the unit square (like the shaded regions in each stage of FIGURE 2). Flip the unit square horizontally (over the vertical line through its center) to obtain the right half of the new Venn diagram. (Note that for each j, y_j on the left is joined to the copy of itself on the right to become y_{j+1}.) Label each region on the right by multiplying the previous label by 2, and relabel the regions on the left by adding 1 to the corresponding region on the right. Had we used the vertical line $x = 1/2$ to define the first set, then the induction would be complete since that line



STAGE 1

$y_1 = (1/2) - x$

1 curve, 2 regions

STAGE 2

$$y_2 = \frac{|2x-1| - (1/2)}{2}$$

2 curves, 4 regions

STAGE 3

$$y_3 = \frac{|4x-1| - (1/2)}{4} \quad \text{for } 0 \leqslant x \leqslant 1/2$$

3 curves, 8 regions

STAGE 4

$$y_4 = \frac{|8x-1| - (1/2)}{8} \quad \text{for } 0 \leqslant x \leqslant 1/4$$

4 curves, 16 regions

FIGURE 2.
The construction of a Venn diagram for n sets, $n = 1, 2, 3, 4$. Each set is defined to be the portion of the unit square below its zigzag boundary. The functions y_j are periodic with period 2^{1-j}.

separates the even regions on the right (the complement of V_1) from the odd regions (contained in V_1) on the left. Instead we used the diagonal of the unit square, namely $y_1 = (1/2) - x$, to simplify our construction. One must therefore check that no zigzags meet in the portion of the square that lies between y_1 and $x = 1/2$. Thus y_N indeed passes through all 2^{N-1} regions of the previous stage, so that the Nth stage has 2^N regions, and the induction is complete.

We close with some observations.

1. Note how FIGURE 1 has been derived from the general construction with $n = 3$ —it was achieved using a rotation and a couple of reflections combined with a touch of artistic license.

2. As an alternative to our zigzags, $\cos \pi x$ could be used instead of $C(x)$. The upper boundary of V_j would then be $y_j = 2^{-j}(\cos 2^{j-1}\pi x)$, $j = 1, 2, \dots$.

3. Using analogous methods, one could construct Venn diagrams on surfaces other than the plane. For example, to interpret the squares of FIGURE 2 as diagrams on a sphere simply identify the top edge ($y = 1/2$) with the North Pole, the bottom edge ($y = -1/2$) with the South Pole, the right and left edges ($x = 0$ and $x = 1$) with the 180th meridian, and the diagonal y_1 with the 0th meridian. The other zigzags then undulate back and forth about the equator, producing the correct number and placement of regions. (See FIGURE 3.)

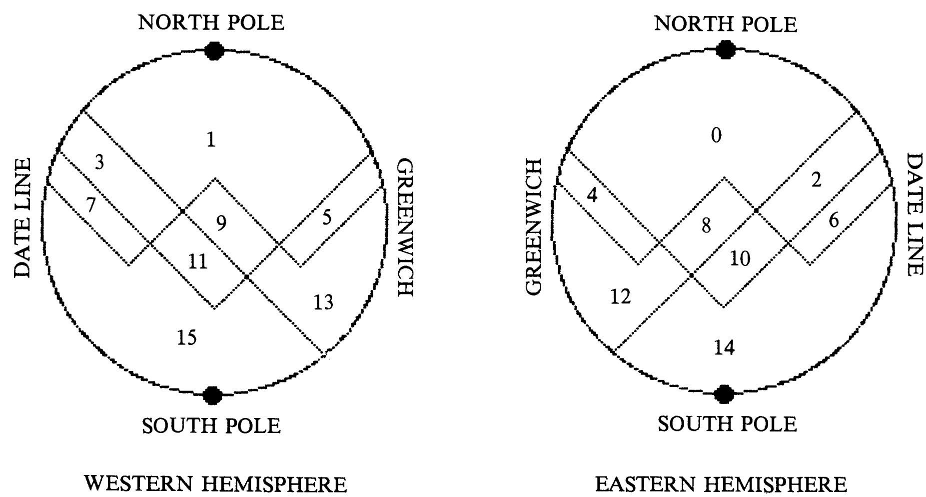


FIGURE 3.
A schematic representation of a Venn diagram on a sphere.

4. Our zigzags suggest a "k-fold Venn problem," which apparently has not been considered in the literature. A *k-fold Venn diagram* is defined exactly as an ordinary Venn diagram, except that each of the intersections $X_1 \cap X_2 \cap \cdots \cap X_n$ is required to consist of precisely k connected components. Thus the usual Venn diagrams are 1-fold. For $n, k \geqslant 2$ start with a convex $(2k)$-gon and replace every other side by a family of zigzags as in FIGURE 4. The zigzags fail to be a k-fold Venn diagram only because the region exterior to all sets, namely 0, consists of just one component, as does the region $2^n - 1$ common to all sets. But this shortcoming can be remedied as in FIGURE 4: on $k - 1$ of the zigzag bunches stretch the 1 region through the common exterior 0 to overlap the $2^n - 2$ region in a new $2^n - 1$, thus creating just the right number of regions of all kinds. This unfortunately destroys the symmetry—but the existence of the unbounded region shows that one cannot expect all k components of each region to be all alike. However, it is an open question whether symmetric k-fold diagrams can be constructed on the sphere or on other surfaces.

5. In our construction some points (and segments) belong to the boundary of several of the sets. However, it is easy to modify the construction so that *simple* Venn

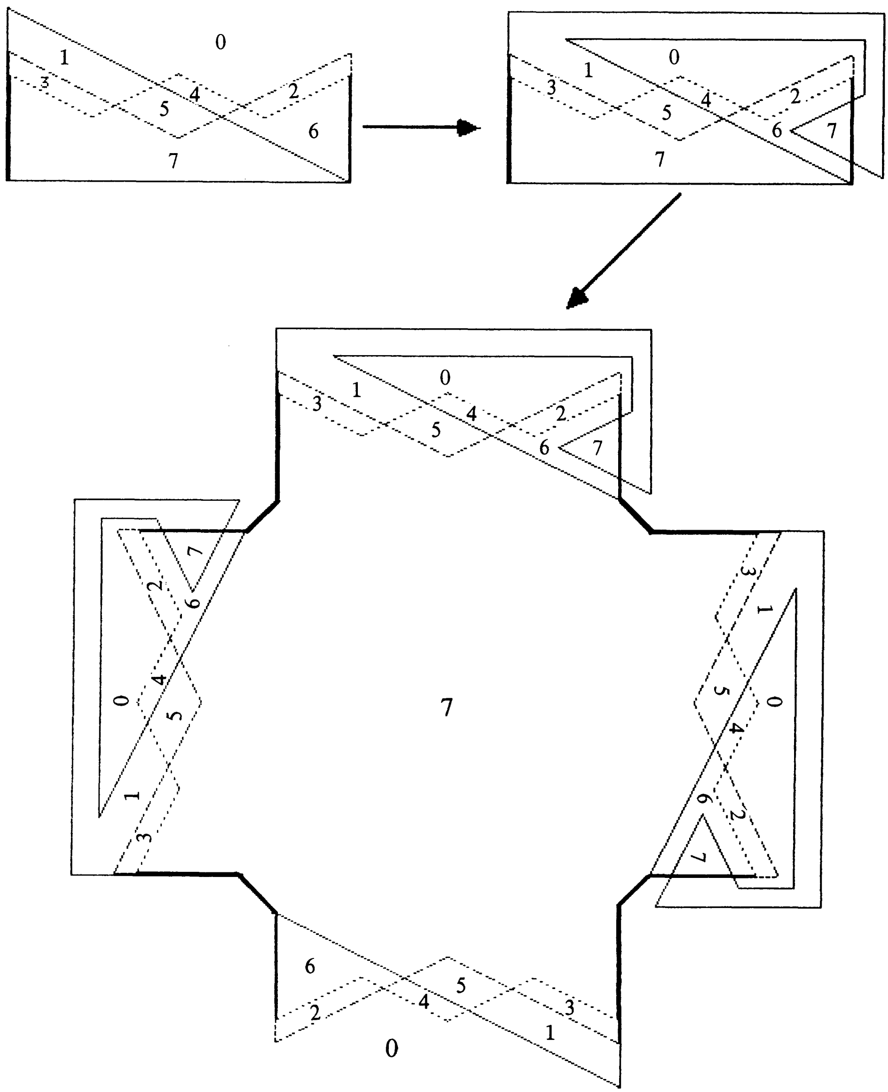


FIGURE 4.
The construction of a 4-fold Venn diagram for 3 sets.

diagrams are obtained (that is, diagrams in which no point belongs to the boundary of three sets, and no segment belongs to two boundaries). One way of achieving this is by expanding the sets V_j in different ratios, all sufficiently near 1, from the center of the square.

REFERENCES

1. A. V. Boyd, Venn diagram of rectangles, this MAGAZINE 58 (1985), 251.
2. B. Grünbaum, Venn diagrams and independent families of sets, this MAGAZINE 48 (1975), 12–22.
3. ______, The construction of Venn diagrams, *College Math. J.* 15 (1984), 238–247.
4. Peter Winkler, Venn diagrams: some observations and an open problem, *Congr. Numer.* 45 (1984), 267–274.

A Formula that Produces All, and Nothing But, Irreducible Polynomials in $\mathbb{Z}_p[x]$

MARKUS NIJMEIJER
KLS College of Education
Sittard, The Netherlands

MIKE STARING
Colegio Arubano
Oranjestad, Aruba

In a stimulating short note [3], Keith Devlin stated: "Given positive whole numbers M and N, calculate

$$K = M(N+1) - (N!+1)$$

and then calculate

$$P = \frac{1}{2}(N-1)\big(|K^2-1| - (K^2-1)\big) + 2.$$

For any values of M and N, the value of P you obtain from the above formula will be a prime number. Moreover, every prime number will be a value of P for some values of M and N."

Some analysis would show that the formula is a consequence of Wilson's theorem: $(p-1)! = p-1 (\bmod p)$ iff p is a prime. (For three different proofs of this last theorem, see [1], [4], and [5].) In this note we will provide a generalization of Wilson's theorem to the domain of irreducible polynomials. From this generalized theorem, a formula generating all irreducible polynomials in $\mathbb{Z}_p[x]$ will be derived.

The Formula

For the purpose of this note it will be useful to reformulate the prime number generating function to a formula containing only one variable. This will prove to be not very hard. Let $K(\bmod N)$ denote the nonnegative remainder of the division of K by N. Suppose n is a composite number, and $(n-1)! \neq 0 \ (\bmod n)$. Then there are numbers a and b with $n = ab$, whereas the product $(n-1)! = (n-1)(n-2)\cdots 2\cdot 1$ cannot contain both a and b as separate factors. Hence, $a = b$ and a is a prime. So, if n is composite and $(n-1)! \neq 0 \ (\bmod n)$, then there is a prime a with $n = a^2$. In case $a = 2$, we find $(n-1)! = (2^2-1)! = 3! = 6 = 2 \ (\bmod n)$; the case that a is an odd prime leads to a contradiction because then

$$(n-1)! = 1\cdot 2\cdots \underline{\underline{a}}\cdot(a+1)\cdots \underline{\underline{2a}}\cdots(a^2-1) = 0 \ (\bmod a^2) = 0 \ (\bmod n).$$

From this, and from Wilson's theorem, it follows

$$(p-1)! = \begin{cases} p-1 \ (\bmod p) \text{ iff } p \text{ is a prime} \\ 2 \ (\bmod p) \text{ iff } p = 4 \\ 0 \ (\bmod p) \text{ iff } p \text{ is composite and } p > 4. \end{cases}$$

Hence,

$$\{(p-1)!\}^2 \pmod p = \begin{cases} 1 \text{ iff } p \text{ is a prime} \\ 0 \text{ iff } p \text{ is composite.} \end{cases}$$

Now define for each $N > 2$

$$f(N) = \{(N-1)!\}^2 (\operatorname{mod} N) * N + \left(1 - \{(N-1)!\}^2 (\operatorname{mod} N)\right) * 2,$$

then $f(N) = 2$ if N is composite, and $f(N) = N$ if N is a prime, so the function f generates the primes and nothing but the primes.

In order to derive a similar formula that generates the irreducible polynomials in $\mathbb{Z}_p[x]$, we will need the following:

LEMMA 1. *Suppose p is a prime and $P(x)$ is a polynomial of degree n in $\mathbb{Z}_p[x]$. Let $Q(x)$ denote the product of all nonzero elements of the ring $\mathbb{Z}_p[x]/(P(x))$. Then $Q(x) = -1 \pmod{P(x)}$ iff $P(x)$ is irreducible.*

The following proof of this lemma is the "simplest," although not the shortest, proof we know. Shortcuts, involving some algebraic theorems, are not difficult to find.

Proof. Case A: Suppose $p = 2$.

(i) In a group $\ulcorner\{e, a_1, a_2, \ldots, a_{2m}\}, \cdot\urcorner$ with an odd number of elements, the congruence $u \cdot u = e$ has exactly one solution. For, suppose besides e there is an element a_i with $a_i \cdot a_i = e$, then $\ulcorner\{e, a_i\}, \cdot\urcorner$ would be a subgroup of the original group. In that case—according to Lagrange—the number of elements of the original group would have to be even instead of odd: contradiction! As a consequence the product of all elements of a commutative group $\ulcorner\{e, a_1, \ldots, a_{2m}\}, \cdot\urcorner$ equals e, because the factors in $e \cdot a_1 \cdot a_2 \cdots a_{2m}$ can be rearranged in such a way that each a_i is situated next to its inverse element, which—according to the foregoing—cannot be e, or a_i itself. Now, suppose $P(x)$ is an irreducible polynomial of degree n in $\mathbb{Z}_2[x]$, then $\mathbb{Z}_2[x]/(P(x))$ is a field with $(2^n - 1)$ nonzero elements. Following from what is stated above, $Q(x) = 1$. As $1 = -1$ holds in $\mathbb{Z}_2$, we may conclude: If $P(x)$ is irreducible in $\mathbb{Z}_2[x]$, then $Q(x) = -1 \pmod{P(x)}$.

(ii) On the other hand, suppose the identity $Q(x) = -1$ holds in the ring $\mathbb{Z}_2[x]/(P(x))$. In that case there is a polynomial $K(x)$ with $K(x) \cdot P(x) = Q(x) + 1$. The right-hand side of this identity cannot be divided by any polynomial of degree k with $1 \leqslant k < n$, because of the definition of Q. Hence, the left-hand side cannot either. Therefore, $P(x)$ is irreducible. So: If $Q(x) = -1 \pmod{P(x)}$ in $\mathbb{Z}_2[x]$, then $P(x)$ is irreducible in $\mathbb{Z}_2[x]$.

Case B: Suppose p is an odd prime.

(i) If $P(x)$ is an irreducible polynomial of degree n in $\mathbb{Z}_p[x]$, then $\mathbb{Z}_p[x]/(P(x))$ is a finite field. In this finite field the quadratic congruence $u^2 = -1$ has exactly two solutions: $u = 1$ and $u = p - 1$. Therefore, the factors of the product $Q(x)$ can be rearranged, putting each factor that is unequal to 1 or $p - 1$ next to its inverse element in $\mathbb{Z}_p[x]/(P(x))$. It follows that $Q(x) = 1 \cdot (p-1) = -1 \pmod{P(x)}$.

(ii) The proof that if $Q(x) = -1 \pmod{P(x)}$ in $\mathbb{Z}_p[x]$, then $P(x)$ is irreducible, is quite similar to case A(ii).

An analysis of the reducible polynomials in $\mathbb{Z}_p[x]$ in the same way as for the composite numbers yields: $Q(x) \neq 0 \pmod{P(x)}$ only if $p = 2$ and the degree of $P(x)$

equals 2. Hence,

$$Q(x)=\begin{cases}-1 \pmod{P(x)} \text{ iff } P(x) \text{ is irreducible}\\ x \pmod{x^2} \text{ in } \mathbb{Z}_2[x]\\ x+1 \pmod{x^2+1} \text{ in } \mathbb{Z}_2[x]\\ 0 \pmod{P(x)} \text{ for all other reducible polynomials in}\\ \qquad \mathbb{Z}_2[x], \text{ and for all reducible polynomials in}\\ \qquad \mathbb{Z}_p[x] \text{ where } p \text{ is an odd prime.}\end{cases}$$

From this it can simply be derived that for all primes p

$$Q^2(x)=\begin{cases}1 \pmod{P(x)} \text{ if } P(x) \text{ is irreducible in } \mathbb{Z}_p[x]\\ 0 \pmod{P(x)} \text{ if } P(x) \text{ is reducible in } \mathbb{Z}_p[x].\end{cases}$$

The final ingredient for our formula is taken from the solution by A. Blokhuis of an old problem already mentioned and solved by J. A. Serret in 1879 [2].

LEMMA 2. *For all primes* p, $P(x)=x^p-x-1$ *is an irreducible polynomial in* $\mathbb{Z}_p[x]$.

Proof. In case $p=2$, the statement is obvious, as the reducible polynomials of degree 2 are x^2, x^2+x, and x^2+1. So, let p be an odd prime, and suppose $P(x)$ is reducible. Then there are nonconstant polynomials $G(x)$ and $H(x)$ in $\mathbb{Z}_p[x]$ with $P(x)=x^p-x-1=G(x)H(x)$. Without loss of generality, we may assume that $G(x)$ has the lowest degree that is possible. Because of the fact that

$$x^p-x=x(x-1)(x-2)\cdots(x-(p-1)),$$

the polynomial x^p-x-1 cannot contain any linear factors; hence the degree of G is 2 or more. As for all $i\in\mathbb{Z}_p$ the identity $P(x+i)=P(x)$ holds, we see that

$$G(x+i)H(x+i)=G(x)H(x)=P(x)$$

for all i. Define $G_i(x)=G(x+i)$, for all $i\in\mathbb{Z}_p$, then each $G_i(x)$ is an irreducible polynomial (with the same degree as that of $G(x)$) that divides $P(x)$. Therefore

$$\prod_{i=0}^{p-1}G_i(x)$$

divides $P(x)$. Because the degree of the polynomial

$$\prod_{i=0}^{p-1}G_i(x)$$

equals p times the degree of $G(x)$, and—on the other hand—cannot exceed the degree of $P(x)$ (because it divides $P(x)$), the degree of $G(x)$ must be 1. This contradicts our earlier conclusion that the degree of $G(x)$ must be 2 or more. Hence $P(x)$ is irreducible.

The final step is to combine the consequences of the foregoing into a single formula: Let p be a prime and let A_p denote the collection of all polynomials of degree 2 or more in $\mathbb{Z}_p[x]$, then the formula:

$$F(P(x)) = Q^2(x) \;(\mathrm{mod}\, P(x)) * P(x) + (1 - Q^2(x)) \;(\mathrm{mod}\, P(x)) * (x^p - x - 1), \qquad (P(x) \in A_p)$$

generates all irreducible polynomials in $\mathbb{Z}_p[x]$.]

REFERENCES

1. A. J. Beiler, *Recreations in the Theory of Numbers*, 2nd ed., Dover, New York, 1966, pp. 49–53.
2. A. Blokhuis, Solution of Problem 699, *Nieuw Arch. Wisk.*, 4e serie, deel 3 (maart 1985), 139.
3. K. Devlin, Prime Beef, *Mathematical Spectrum* 18 (1985), 33–34.
4. W. J. Gilbert, *Modern Algebra with Applications*, Wiley and Sons, 1976, p. 218.
5. E. Staring and M. Staring, Further support for a "physical" interpretation of the primes, *Amer. J. Physics* 51 (May 1983), 462–463.

Sums of Irrational Square Roots Are Irrational

GREGG N. PATRUNO
353 Giffords Lane
Staten Island, NY 10308

Square roots of integers were the first irrational numbers to be discovered, and the object of much attention among the ancient Greeks. Some of the most tantalizing irrational numbers found since then are the sums of irrational square roots: it is easy to believe that a number like $\sqrt{2} + \sqrt{5} + \sqrt{17} + \sqrt{23}$ is irrational, but not nearly as easy to prove it. Indeed, it was not until 2000 years later that modern algebra succeeded.

The natural generalization is usually presented as an exercise in the theory of extension fields:

If $p_1, p_2, \dots, p_n$ and $\sqrt{p_1} + \sqrt{p_2} + \cdots + \sqrt{p_n}$ are all rational, then so is each $\sqrt{p_i}$.

(Along these lines, see Theorem 4.7 in [2] or Theorem 5.5.1 in [1].)

In contrast, this note establishes by entirely elementary means a slightly stronger result, characterizing all sums and differences of square roots whose combination is rational:

If $a_1^2, a_2^2, \dots, a_n^2$ and $s = a_1 + a_2 + \cdots + a_n$ are rational, then either a_1 is rational or one of the subseries $a_1 + \Sigma a_i$ vanishes.

(The final provision accounts for exceptional cases like $s = \sqrt{2} + 1 - \sqrt{2}$, where the irrationality of a_1 is "neutralized" by the addition of a_3.)

The goal of our proof is to express a_1 as a combination of sums, products, and quotients of known rational numbers. In particular, a_1 will be written as a rational function of $s, a_1^2, a_2^2, \dots, a_n^2$ having integer coefficients:

$$a_1 = \frac{G(s, a_1^2, a_2^2, \dots, a_n^2)}{H(s, a_1^2, a_2^2, \dots, a_n^2)}.$$

When $s = \sqrt{5} + \sqrt{3} + \sqrt{2}$, for example, the numerical identity constructed by the method below gives an immediate, compelling proof that s is irrational:

$$\sqrt{5} = \frac{s}{4} + \frac{5}{s} - \frac{6}{s^3}.$$

(The author gives G/H explicitly for $n = 2$ and $n = 3$ in [3].)

An equivalent formulation of the problem is to construct a polynomial

$$F(x, a_1, a_2, \dots, a_n) = G(x, a_1^2, a_2^2, \dots, a_n^2) - a_1 \cdot H(x, a_1^2, a_2^2, \dots, a_n^2) \tag{1}$$

which vanishes when $x = a_1 + a_2 + \cdots + a_n$.

First, we know F must have a factor of $(x - a_1 - a_2 - \cdots - a_n)$. Next, we make sure that F includes $a_2, \dots, a_n$ only as even powers by providing every factor of the form $(\cdots - a_k \cdots)$ with a matching factor of the form $(\cdots + a_k \cdots)$. Thus we take

$$F(x, a_1, a_2, \dots, a_n) = \prod (x - a_1 \pm a_2 \pm \cdots \pm a_n), \tag{2}$$

where the product ranges over all combinations of plus and minus signs.

The even-powered terms in a_1 can be gathered into a polynomial $G(x, a_1^2, a_2^2, \dots, a_n^2)$, the odd terms can be grouped in the form $-a_1 \cdot H(x, a_1^2, a_2^2, \dots, a_n^2)$, and clearly all coefficients are integers. Therefore, F has the form (1) we are seeking, and

$$a_1 = \frac{G(s, a_1^2, a_2^2, \dots, a_n^2)}{H(s, a_1^2, a_2^2, \dots, a_n^2)},$$

a quotient of rational numbers.

Finally, we determine the exceptional cases for which the denominator H vanishes. Equation (1) implies

$$F(s, a_1, a_2, \dots, a_n) - F(s, -a_1, a_2, \dots, a_n) = -2a_1 \cdot H(s, a_1^2, a_2^2, \dots, a_n^2)$$

and (2) gives $F(s, a_1, a_2, \dots, a_n) = 0$, consequently

$$\begin{aligned} H(s, a_1^2, a_2^2, \dots, a_n^2) &= \frac{1}{2a_1} F(s, -a_1, a_2, \dots, a_n) \\ &= \frac{1}{2a_1} \prod (s + a_1 \pm a_2 \pm \cdots \pm a_n) \\ &= \frac{1}{2a_1} \prod [2a_1 + (a_2 \pm a_2) + \cdots + (a_n \pm a_n)] \\ &= \prod_{\substack{T \subset \{a_2, \dots, a_n\} \\ T \neq \varnothing}} \left[2\left(a_1 + \sum_{a_i \in T} a_i\right)\right]. \end{aligned}$$

We see that H only equals zero when one of the subseries $a_1 + \Sigma a_i$ vanishes; therefore $a_1 = G/H$ is rational in all other cases.

The author thanks Steven R. Bell of Purdue University for valuable conversations during the preparation of the manuscript, and Murray S. Klamkin of the University of Alberta for raising the topic as a challenging problem.

REFERENCES

1. I. N. Herstein, *Topics in Algebra*, 2nd ed., John Wiley and Sons, New York, 1975.
2. I. Niven, *Irrational Numbers*, Carus Mathematical Monographs of the MAA, no. 11, 1956.
3. G. Patruno, Student Proposal 3 and Solution, The Olympiad Corner, ed. M. S. Klamkin, *Crux Mathematicorum*, 89 (Nov. 1982).

PROBLEMS

Proposals

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

BRUCE HANSON, *associate editor*
St. Olaf College

To be considered for publication, solutions should be received by July 1, 1988.

1287. *Proposed by C. S. Gardner, Austin, Texas.*

Let P be an interior point of the rectangle $ABCD$. Draw lines through A, B, C, D perpendicular to PA, PB, PC, PD respectively. Show that the area of the convex quadrilateral enclosed by these four lines is equal to or greater than twice the area of the rectangle. When do we have equality?

1288. *Proposed by Leroy F. Meyers, The Ohio State University, Columbus.*

Find all differentiable real-valued functions f defined on the entire real line such that $f(x)f'(x) = 0$ for all real x.

1289. *Proposed by M. S. Klamkin, University of Alberta, Canada.*

Two identical beads slide on two straight wires intersecting at right angles. If the beads start from rest in any position other than the intersection point of the wires and attract each other in an arbitrary mutual fashion but also subject to a drag proportional to the speed, show that the beads will arrive at the intersection simultaneously.

1290. *Proposed by Edward T. H. Wang, Wilfrid Laurier University, Canada.*

For positive integers n and r, let $\left\langle {n \atop r} \right\rangle = \binom{n+r-1}{r}$. Find a closed form expression for

$$\sum_{r=1}^{k} r\left\langle {n \atop r} \right\rangle$$

where k denotes a positive integer.

1291. *Proposed by Mihály Bencze, Braşov, Romania.*

Let n be a positive integer, and let $\mathbf{A}$ be an $m \times m$ matrix of real numbers such that $\mathbf{A}^{2n+1} = \mathbf{A} - \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Prove that $(-1)^m(\det \mathbf{A}) > 0$.

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

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

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

Quickies

Answers to the Quickies are on page 58.

Q728. *Proposed by Martin C. Tangora, University of Illinois, Chicago (part a), and Bill Olk (student), University of Wisconsin, Madison (part b).*

a. Find the points where the graphs of $y = e^x$ and $y = \ln x$ are closest together.

b. Find that logarithmic function $y = \log_a x$, $a > 1$, which intersects its inverse at only one point.

Q729. *Proposed by M. S. Klamkin, University of Alberta.*

Determine the extreme values of

$$S = \frac{x+1}{xy+x+1} + \frac{y+1}{yz+y+1} + \frac{z+1}{zx+z+1}$$

where $xyz = 1$ and $x, y, z \geq 0$.

Q730. *Proposed by R. S. Luthar, University of Wisconsin Center, Janesville.*

Prove that an integer is expressible as a sum of two squares if and only if its cube is expressible as a sum of two squares.

Solutions

Least Common Multiple of $\{1, 2, \ldots, n\}$ **December 1986**

1252. *Proposed by Sydney Bulman-Fleming and Edward T. H. Wang, Wilfrid Laurier University, Canada.*

Let $L(n)$ denote the least common multiple of $\{1, 2, \ldots, n\}$. Color each positive integer red or blue according to the following rule: number 1 is red, and for $n > 1$, n has the same color as $n - 1$ if and only if $L(n) = L(n-1)$.

a. Find all instances of four consecutive integers with alternating colors.

b. Show that there are arbitrarily long sequences of consecutive integers with the same color.

c. Show that there are arbitrarily long sequences of consecutive red integers.

Solution by Roger B. Eggleton, University of Newcastle, Australia.

If n is a prime-power p^α, where p is prime and α is a positive integer, then $L(n) = pL(n-1)$ since $p^{\alpha-1} < n$ ensures that $p^{\alpha-1}$ divides $L(n-1)$. On the other hand, if n is not a prime-power, it is greater than every prime-power which divides it, so $L(n) = L(n-1)$. Thus color changes coincide with prime-powers.

a. If $a, a+1, \ldots, a+r$ is a run of consecutive positive integers with alternating colors, then $a+1, \ldots, a+r$ must be prime-powers. If $r \geqslant 2$, a multiple of 2 is present; if $r \geqslant 3$, a multiple of 3 is present. Since they are prime-powers, $a+1, \ldots, a+r$ must contain a power of 2 adjacent to a power of 3 if $r \geqslant 3$. It is well known that $8, 9$ are the largest integers of this form (e.g., see Ecklund and Eggleton, this MAGAZINE, 48 (1975) 277–281). Thus, $1, 2, 3, 4, 5$ and $6, 7, 8, 9$ are the only maximal runs of at least 4 consecutive integers with alternating colors.

b. For any n, let $P(n)$ be the product of all distinct primes $p \leqslant n$. If $2 \leqslant r \leqslant n$, then r divides $L(n)$, so $L(n)P(n)+r$ is a multiple of r. However, it is not a prime-power, for if p^{α} is a maximal prime-power factor of r then $p^{\alpha+1}$ is a factor of $L(n)P(n)$, so that $L(n)P(n)+r$ is greater than r but has p^{α} as a maximal prime-power factor. Therefore, $\{L(n)P(n)+r\colon 1 \leqslant r \leqslant n\}$ is a monochromatic run of n positive integers.

c. For any n, fix two distinct primes $p, q > n$. Since $L(n)P(n)$ and pq are coprime, there is a positive integer k for which $kL(n)P(n) \equiv 1 \pmod{pq}$. If k_0 is the smallest such k, then $kL(n)P(n)-1$ is a positive multiple of pq whenever $k \equiv k_0 \pmod{pq}$ and $k \geqslant k_0$. Then $kL(n)P(n)+1$ is the only possible prime-power among the $2n+1$ consecutive integers $\{kL(n)P(n)+r\colon\ -n \leqslant r \leqslant n\}$. By Dirichlet's Theorem, there are infinitely many primes in the arithmetic sequence $\{kL(n)P(n)+1\colon\ k>0,\ k \equiv k_0 \pmod{pq}\}$. Let $k_1L(n)P(n)+1$ be any such prime. The $\{k_1L(n)P(n)+r\colon -n \leqslant r \leqslant n\}$ comprises a monochromatic run of $n+1$ integers, followed by an opposite colored monochromatic run of n integers.

Also solved by Robert L. Doucette (student), Jerrold W. Grossman, David Morin (student) and Fritz Grobe (student), Bjorn Poonen (student), Kiran Lall Shrestha (student, Nepal), and the proposers.

Decreasing Function **December 1986**

1253. *Proposed by Gerald A. Heuer, Concordia College, Minnesota (as corrected).*

Suppose that x and y are related by $e^{-x}-e^{-y}=e^{-2}$. Prove that

$$\frac{1}{x}-\frac{1}{y}$$

is decreasing for x in the open interval $(0, 2-\ln 2)$.

Solution by Barry Brunson, Western Kentucky University.

Let $f(x)=1/x-1/y$, where $y=y(x)$ is defined implicitly as a function of x by $e^{-x}-e^{-y}=e^{-2}$. Implicit differentiation yields

$$\frac{df}{dx}=\frac{x^2y'-y^2}{x^2y^2}.$$

From the relation between x and y we obtain $y'=e^{y-x}$, and thus we must prove that $x^2e^y<y^2e^x$ for $0<x<2-\ln 2$, or equivalently, that $y-2\ln y<x-2\ln x$ for $0<x<2-\ln 2$.

The relation between x and y makes it clear that $x<y$, and it is easy to check that $x<2-\ln 2$ implies that $y<2$. But now the result follows since $t-2\ln t$ is decreasing for $0<t<2$.

Also solved by Seung Jin Bang (Korea), Chico Problem Group, Thomas P. Dence, Milton P. Eisner, Thomas E. Elsner, P. L. Hon (Hong Kong), Lars Höglund (Sweden), Mary S. Krimmel, L. Kuipers (Switzerland), Beatriz Margolis (France), Walter I. Nissen, Jr., P. Ramankutty (New Zealand), Kennard Reed, Jr., Volkhard Schindler (East Germany), Harvey Schmidt, Jr., J. M. Stark, David R. Stone, Ragnar Dybvik (Norway), W. R. Utz, and Harry Weingarten.

Continued Fraction **December 1986**

1254. *Proposed by Ambati Jaya Krishna (student), Johns Hopkins University, and Mrs. Gomathi S. Rao, Orangeburg, South Carolina.*

Find the value of the continued fraction

$$1+\frac{2}{3+}\,\frac{4}{5+}\,\frac{6}{7+}\cdots.$$

I. *Solution by Robert Heller, Mississippi State University.*

If $A_0=0$, $A_1=B_0=B_1=1$, and for $p\geqslant 1$,

$$A_{p+1}=(2p+1)A_p+2pA_{p-1},$$
$$B_{p+1}=(2p+1)B_p+2pB_{p-1},$$

then, $(A_n/B_n)_{n=1}^{\infty}$ is the sequence of approximants of the continued fraction

$$\frac{1}{1+}\,\frac{2}{3+}\,\frac{4}{5+}\,\frac{6}{7+}\cdots.$$

(Note that this continued fraction is the reciprocal of the continued fraction of the problem.)

It is straightforward to verify by mathematical induction that for each positive integer n, $A_n+B_n=2^n n!$ and $B_n=2^n n! S_n$, where

$$S_n=\sum_{k=0}^{n}\frac{(-1/2)^k}{k!}.$$

Now since

$$\frac{A_n+B_n}{B_n}=\frac{1}{S_n}$$

and $\lim_{n\to\infty}S_n=e^{-1/2}$, we have

$$\lim_{n\to\infty}\frac{A_n}{B_n}=\sqrt{e}-1.$$

Consequently, the value of the continued fraction as in the problem is $\dfrac{1}{\sqrt{e}-1}$.

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

We shall show the more general result that if each $x_i\geqslant k>1$, then the continued fraction

$$(x_1-1)+\frac{x_1}{x_2-1+}\,\frac{x_2}{x_3-1+}\,\frac{x_3}{x_4-1+}\cdots$$

converges to the value

$$\frac{1-\dfrac{1}{x_1}+\dfrac{1}{x_1x_2}-\dfrac{1}{x_1x_2x_3}+\cdots}{\dfrac{1}{x_1}-\dfrac{1}{x_1x_2}+\dfrac{1}{x_1x_2x_3}-\cdots}.$$

The given instance has $x_i=2i$ which leads to the value $1/(\sqrt{e}-1)$. More generally, we note that $x_i = ci$ leads to the value $(e^{1/c} - 1)^{-1}$. The case $c = 1$ is especially appealing:

$$\frac{1}{1+}\frac{2}{2+}\frac{3}{3+}\cdots = (e-1)^{-1}.$$

Any finitely truncated continued fraction of the form

$$y_0 + \frac{x_1}{y_1+}\frac{x_2}{y_2+}\cdots+\frac{x_n}{y_n}$$

has a rational value a_n/b_n, and we may use matrix multiplication to determine that

$$\begin{pmatrix} a_{n-1} & a_n \\ b_{n-1} & b_n \end{pmatrix} = \begin{pmatrix} 1 & y_0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & x_1 \\ 1 & y_1 \end{pmatrix}\begin{pmatrix} 0 & x_2 \\ 1 & y_2 \end{pmatrix}\cdots\begin{pmatrix} 0 & x_{n-1} \\ 1 & y_{n-1} \end{pmatrix}\begin{pmatrix} 0 & x_n \\ 1 & y_n \end{pmatrix}.$$

For each n, we may verify this identity by evaluating the products from right to left just as the continued fraction is evaluated from right to left. Thus this identity implies that

$$\begin{pmatrix} a_{n-1} & a_n \\ b_{n-1} & b_n \end{pmatrix} = \begin{pmatrix} a_{n-2} & a_{n-1} \\ b_{n-2} & b_{n-1} \end{pmatrix}\begin{pmatrix} 0 & x_n \\ 1 & y_n \end{pmatrix}.$$

Focusing on the right-hand column gives recursions

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = \begin{pmatrix} a_{n-2} & a_{n-1} \\ b_{n-2} & b_{n-1} \end{pmatrix}\begin{pmatrix} x_n \\ y_n \end{pmatrix}.$$

Now the hypothesis $y_{i-1} = x_i - 1$ allows us to show by induction that

$$a_n = \left(\prod_{i=1}^{n+1} x_i\right)\left(1 - \frac{1}{x_1} + \frac{1}{x_1x_2} - \cdots + (-1)^{n+1}\frac{1}{x_1x_2\ldots x_{n+1}}\right)$$

and

$$b_n = \left(\prod_{i=1}^{n+1} x_i\right)\left(\frac{1}{x_1} - \frac{1}{x_1x_2} + \cdots + (-1)^{n}\frac{1}{x_1x_2\ldots x_{n+1}}\right).$$

Finally, dividing a_n by b_n and letting n go to infinity gives the formula claimed above. Since each $x_i \geqslant k > 1$, both the numerator and denominator converge by the ratio test.

III. *Solution by Paul Bracken, Toronto.*

We calculate

$$\sum = \frac{2}{3+}\frac{4}{5+}\frac{6}{7+}\cdots.$$

It isn't difficult to show that this converges. The nth order convergent for Σ can be expressed as a ratio of the quantities A_n and B_n where these will satisfy the following recursion formulas:

$$A_{n+2} = (2n+3)A_{n+1} + (2n+2)A_n, \tag{1a}$$
$$B_{n+2} = (2n+3)B_{n+1} + (2n+2)B_n, \tag{1b}$$

for $n = 0, 1, 2, \ldots,$ with

$$A_0 = 1, \qquad A_1 = 0, \tag{2a}$$

$$B_0 = 0, \qquad B_1 = 1. \tag{2b}$$

These recursion formulas suggest a second-order differential equation, and would provide a method of calculating the continued fraction if the differential equation can also be solved in closed form. Let us define a function

$$y(x) = \sum_{n=0}^{\infty} \frac{c_n}{n!} x^n,$$

where c_n satisfies a relation such as (1). Then

$$y'' = \sum_{n=0}^{\infty} \frac{c_{n+2}}{n!} x^n = \sum_{n=0}^{\infty} \left((2n+3)c_{n+1} + (2n+2)c_n\right) \frac{x^n}{n!}.$$

Since

$$\sum_{n=0}^{\infty} nc_n \frac{x^n}{n!} = \sum_{n=0}^{\infty} c_{n+1} \frac{x^{n+1}}{n!} = xy'; \qquad \sum_{n=0}^{\infty} nc_{n+1} \frac{x^n}{n!} = xy''$$

we obtain

$$y'' = 2xy'' + (3+2x)y' + 2y.$$

Now, introduce a new variable $t = \frac{1}{2} - x$, we have $y(\frac{1}{2} - t) \equiv g(t)$ and

$$tg''(t) + (2-t)g'(t) - g(t) = 0.$$

The general solution to this equation is $g(t) = a(e^t/t) + b(1/t)$, or

$$f(x) = c\left(\frac{e^{-x}}{1-2x}\right) + d\left(\frac{1}{1-2x}\right).$$

The constants c and d are calculated for the cases (2a) and (2b). First, put $A(x) = y(x)$ when $c_n \equiv A_n$; then

$$A_0 = c + d = 1, \qquad A_1 = -c + 2c + 2d = 0,$$

so

$$A(x) = \frac{2e^{-x}}{1-2x} - \frac{1}{1-2x},$$

and for $B(x) = y(x)$ when $c_n \equiv B_n$, we have

$$B_0 = c + d = 0, \qquad B_1 = c + 2d = 1,$$

so

$$B(x) = \frac{-e^{-x}}{1-2x} + \frac{1}{1-2x}.$$

Now

$$B'(x) = B_1 + B_2 x + \cdots, \qquad B_m > 0 \ (m \geqslant 1),$$

and $\lim_{n\to\infty} A_n/B_n = \Sigma$. Moreover, replacing x by $s/2$, then from the form of $B(x)$ we must have $\lim_{s\to 1} B'(s/2) = \infty$, and so a theorem of E. Cesaro (see, for example,

T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series*, Macmillan, 1965, 149–150) can be used to calculate a value for Σ. Writing

$$\frac{A_n}{B_n} = \frac{n\dfrac{A_n}{n!}\dfrac{1}{2^{n-1}}}{n\dfrac{B_n}{n!}\dfrac{1}{2^{n-1}}}$$

we have

$$\lim_{n\to\infty}\frac{A_n}{B_n} = \lim_{n\to\infty}\left(\frac{n\dfrac{A_n}{n!}\dfrac{1}{2^{n-1}}}{n\dfrac{B_n}{n!}\dfrac{1}{2^{n-1}}}\right) = \lim_{s\to 1}\frac{A'(s/2)}{B'(s/2)}.$$

Consequently,

$$\sum = \lim_{n\to\infty}\frac{A_n}{B_n} = \frac{2-\sqrt{e}}{\sqrt{e}-1},$$

where the closed forms for $A(x)$ and $B(x)$ are used to calculate the derivatives. The required continued fraction is given by

$$1+\frac{2}{3+}\frac{4}{5+}\cdots = \frac{1}{\sqrt{e}-1}.$$

Also solved by Rich Bauer, J. C. Binz (Switzerland), Nirdosh Bhatnagar, J. Heuver (Canada), Dixon Jones, L. Kuipers (Switzerland), Kee-wai Lau (Hong Kong), Leo Michelotti, Michael J. Poris, Ira Rosenholtz, Volkhard Schindler (East Germany), Shailesh A. Shirali (India), M. Vowe (Switzerland), Charles H. Webster, John W. Wrench, Jr., and the proposers.

A direct and concise solution by G. A. Edgar, The Ohio State University, appears in The Journal of Recreational Mathematics 7 (1974), 152–153.

Odd Coefficients in Multinomial Expansion **December 1986**

1255. *Proposed by Harry D. Ruderman, Lehman College, The Bronx, New York.*

Prove that the number of odd coefficients in the expansion of

$$(x_1 + x_2 + \cdots + x_t)^n$$

is t^d, where d is the sum of the digits in the binary representation of n.

I. *Solution by Leroy F. Meyers, The Ohio State University.*

An easy induction shows that

$$(x_1 + \cdots + x_t)^{2^s} \equiv x_1^{2^s} + \cdots + x_t^{2^s} \pmod 2 \tag{1}$$

for $s = 0, 1, 2, \ldots$. Now let n be a nonnegative integer whose binary expansion, with only the nonzero terms listed, is

$$n = 2^{s_1} + 2^{s_2} + \cdots + 2^{s_d},$$

where $0 \leqslant s_1 < s_2 < \cdots < s_d$. Then by (1) we have

$$(x_1+\cdots+x_t)^n=(x_1+\cdots+x_t)^{2^{s_1}}\cdots(x_1+\cdots+x_t)^{2^{s_d}}$$
$$\equiv\left(x_1^{2^{s_1}}+\cdots+x_t^{2^{s_1}}\right)\cdots\left(x_1^{2^{s_d}}+\cdots+x_t^{2^{s_d}}\right)\pmod 2.$$

Now all the terms obtained by multiplying out the last line are distinct, and there are exactly t^d such terms, each with coefficient 1. This completes the proof.

II. *Solution by Yuval Peres, Ramat Hasharon, Israel.*

For a positive integer m, let $d(m)$ denote the sum of the digits in the binary expansion of m, and $b_i(m)$ the ith digit in the binary expansion of m. We use induction on t.

For $t=2$, we know by Lucas Theorem (see Don Knuth, *The Art of Computer Programming*, (1) *Fundamental Algorithms*, 1973, p. 68) that

$$\binom{n}{k}\equiv\prod_{i\geqslant 0}\binom{b_i(n)}{b_i(k)}\pmod 2.$$

Thus $\binom{n}{k}$ is odd if and only if $b_i(k)\leqslant b_i(n)$ for all i. The number of such k's is the number of subsets of a set of $\Sigma_i b_i(n)=d$ elements, and this is 2^d.

For the inductive step, we have

$$(x_1+\cdots+x_{t+1})^n=\sum_{k=0}^{n}\binom{n}{k}(x_1+\cdots+x_t)^k x_{t+1}^{n-k}.$$

An integer k contributes odd coefficients to the expansion if and only if $\binom{n}{k}$ is odd, in which case (by induction) it contributes $t^{d(k)}$ odd coefficients. Since

$$|\{k\colon b_i(k)\leqslant b_i(n)\text{ for all }i, d(k)=c\}|=\binom{d}{c},$$

there are altogether

$$\sum_{c=0}^{d}\binom{d}{c}t^c=(t+1)^d$$

odd coefficients.

III. *Solution by David Singmaster, South Bank Polytechnic, London, England.*

Let p be a prime, and suppose that p^e divides $\binom{n}{k_1,\ldots,k_t}$ but p^{e+1} does not. Then e is equal to the total amount carried in the addition of $k_1,k_2,\ldots,k_t$ when each of these are represented in the p-ary number system. (See David Singmaster, *A Collection of Manuscripts Related to the Fibonacci Sequence*, 18th anniversary volume of the Fibonacci Association, 1980, 98–113.)

It follows that if $n=\Sigma_i a_i p^i$, $k_j=\Sigma_i b_{ij}p^i$ are p-ary representations, then p does not divide $\binom{n}{k_1,\ldots,k_t}$ if and only if $a_i=\Sigma_j b_{ij}$ for each i. In this case, the $\{b_{ij}\}_j$ form an ordered partition of a_i into t nonnegative integers for each i. This can be done in $\binom{a_i+t-1}{t-1}$ ways. Thus, the number of coefficients in the multinomial expression $(x_1+\cdots+x_t)^n$ which are not divisible by p is

$$\prod_i\binom{a_i+t-1}{t-1}.$$

(For another proof of this result, see F. T. Howard, "The Number of Multinomial Coefficients Divisible by a Fixed Power of a Prime," *Pacific Journal of Mathematics*

80 (1974), 99–108.)

In the case of $p=2$, the a_i's are either 0 or 1, and the product is $\prod_{i=1}^{d} t\ (=t^d)$.

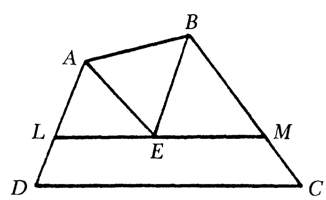
Also solved by Seung Jin Bang (Korea), J. C. Binz (Switzerland), Amir Daneshgar (student, Iran), Alberto Facchini (Italy), Edward H. Grossman, F. T. Howard, Ambati Jaya Krishna (student) and Mrs. Gomathi S. Rao and A. Murali Mohan Rao, L. Kuipers (Switzerland), E. Lee, Eugene Levine, Rhodes Peele, William P. Wardlaw, Gordon Williams, and the proposer.

Crux Mathematicorum *(then called* Eureka) 2 (1976), 34–36, *gives references and a chronology for this problem. Related material is found in a number of sources; one of the most recent and accessible is a paper by P. Goetgheluck: "Computing Binomial Coefficients,"* The American Mathematical Monthly 94 (1987), 360–365.

Cyclic Quadrilateral **December 1986**

1256. *Proposed by R. S. Luthar, University of Wisconsin Center, Janesville.*

Let $ABCD$ be a cyclic quadrilateral, let the angle bisectors at A and B meet at E, and let the line through E parallel to side CD intersect AD at L and BC at M. Prove that $LA+MB=LM$.



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

For arbitrary angles α, β,

$$\sin(2\alpha)+\sin(2\beta)=2\sin(2\beta-\alpha)\cos\alpha+2\sin(2\alpha-\beta)\cos\beta. \tag{1}$$

In the quadrilateral $ABML$, let $2\alpha, 2\beta$ be the angles at A, B, and therefore $180^\circ-2\beta$, $180^\circ-2\alpha$ are the angles at L and M respectively. We see that $\angle LEA=2\beta-\alpha$, $\angle BEM=2\alpha-\beta$. If d denotes the common distance of the point E from the sides LE, AB, BM, we obtain

$$\begin{aligned}\frac{LE+EM-LA-MB}{d}&=\frac{1}{\sin(2\beta)}+\frac{1}{\sin(2\alpha)}-\frac{\sin(2\beta-\alpha)}{\sin\alpha\sin(2\beta)}-\frac{\sin(2\alpha-\beta)}{\sin\beta\sin(2\alpha)}\\&=\frac{\sin(2\alpha)+\sin(2\beta)}{\sin(2\alpha)\sin(2\beta)}\\&\qquad-2\frac{\sin(2\beta-\alpha)\cos\alpha+\sin(2\alpha-\beta)\cos\beta}{\sin(2\alpha)\sin(2\beta)}\\&=0\qquad(\text{using}(1)),\end{aligned}$$

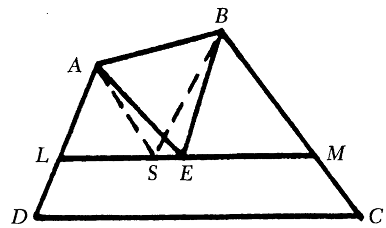
and, therefore, $LM=LA+MB$.

II. *Solution by H. Demir and C. Tezer, Middle East Technical University, Ankara, Turkey.*

Let $\angle DAB=2\alpha$, $\angle ABC=2\beta$, $\angle BCD=2\gamma$, $\angle CDA=2\delta$. Clearly, $\angle ELA=2\delta$, $\angle BME=2\gamma$, and $\alpha=\frac{\pi}{2}-\gamma$, $\beta=\frac{\pi}{2}-\delta$. We'll assume that $ABCD$ is convex and

$\alpha > \beta$.

Choose a point S on LM on the same side of AD as M such that $|LS| = |LA|$ (see figure).

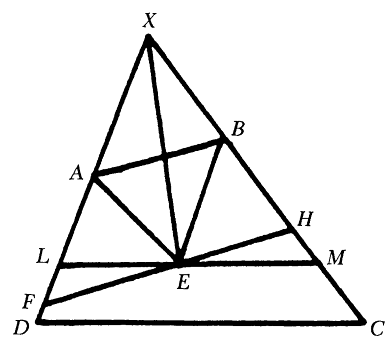


Obviously, $\angle ASL = \angle LAS = \beta$. Therefore, $ASEB$ is a cyclic quadrilateral. As $\angle LAS = \beta < \alpha = \angle LAE$, it follows that S is between L and E.

On the other hand, $\angle SBM = \angle SBE + \angle EBM = \angle SAE + \angle EBM = \angle LAE - \angle LAS + \beta = \alpha - \beta + \beta = \alpha = \angle BSM$. Consequently, MBS is an isosceles triangle and $|MS| = |MB|$. Therefore, $|LM| = |LS| + |SM| = |LA| + |MB|$.

III. *Solution by John P. Hoyt, Lancaster, Pennsylvania.*

Produce DA and CB to meet at X. Draw FH parallel to AB. Draw XE (see figure).



Since E is the intersection of two exterior angles of triangle XAB, XE is the bisector of $\angle AXB$. Triangles MLX and HFX are congruent because they have equal angles and a common angle bisector. The equal angles follow from the fact that the opposite angles of a cyclic quadrilateral are supplementary. Hence $ME = FE$, $HE = LE$, and $HM = LF$. Since FH is parallel to AB, and AE bisects $\angle DAB$, $\angle FAE = \angle AEF$. Thus, triangle FAE is isosceles, and $AF = FE$. Similarly, $BH = EH$.

The rest follows easily: $LA + MB = (AF - LF) + (BH + HM) = (AF + BH) + (HM - LF) = AF + BH = FE + EH = LM$.

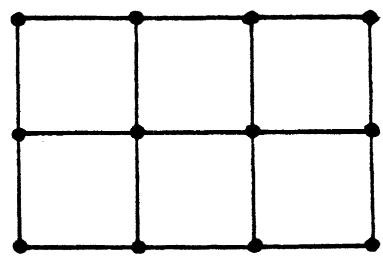
Also solved by Frank Allen, Farid G. Bassiri (student), Andreas Bender (student, Switzerland), Nirdosh Bhatnagar, David Earnshaw (Canada), Howard Eves, Herta T. Freitag, Richard A. Gibbs, J. T. Groenman (Netherlands), Michael B. Handelsman, P. L. Hon (Hong Kong), King Jamison, Geoffrey A. Kandall, Tsz-Mie Ko (student), Mary S. Krimmel, L. Kuipers (Switzerland), Kee-wai Lau (Hong Kong), J. C. Linders (The Netherlands), David Morin (student, four solutions), Anna Michaelides Penk, Farhood Pouryoussefi (student, Iran), Harry D. Ruderman, Kiran Lall Shrestha (Nepal), J. M. Stark, M. Vowe (Switzerland), Harry Weingarten, and Brent Young (student).

Most of the solutions were based on trigonometric arguments (an impressive variety of trigonometric identities). A brillant, purely geometric, solution, due to Gregg Patruno (U.S.A.), appears in Murray Klamkin's International Mathematical Olympiads, 1978–1985, *New Mathematical Library, No.* 31, *MAA (solution to Problem* 1, 1985).

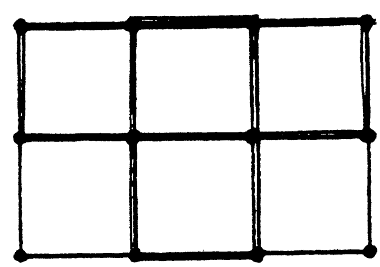
Constructing a Grid with Rectangles **February 1987**

1257. *Proposed by Bruce Reznick, University of Illinois, Urbana-Champaign.*

Let $G(r, s)$ denote the grid of rs points arranged in r rows and s columns with adjacent points connected, $r, s \geqslant 2$. (See illustration of $G(3,4)$.)



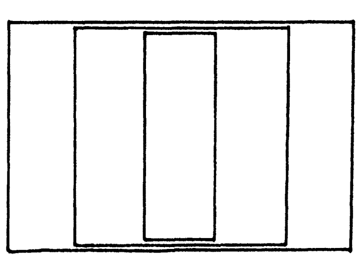
Let $R(r, s)$ denote the minimum number of rectangles needed to cover the edges and vertices of $G(r, s)$. The accompanying illustration shows that $R(3,4) \leqslant 3$.



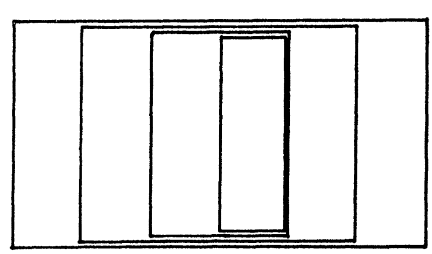
It is easy to check that $G(3,4)$ cannot be covered by two rectangles, so $R(3,4)=3$. Find a simple expression for $R(r, s)$.

Solution by the proposer.

We shall show that $R(r, s)=\lceil r/2\rceil+\lceil s/2\rceil-1$, where $\lceil t\rceil$ is the least integer $\geqslant t$. First, consider the following construction. Take $\lceil s/2\rceil$ nested rectangles of full height and with left and right sides coming in. (The last rectangle would then be a vertical line if s is odd; in this case, take any rectangle of full height with this line as a side; see illustration.)



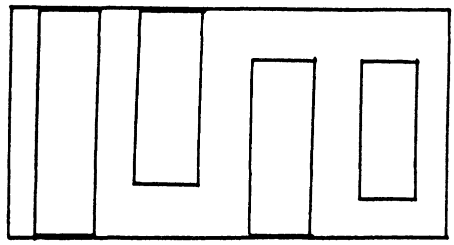
$s=6$



$s=7$

In this way, we have accounted for all vertical edges in $G(r, s)$ as well as the full top and bottom horizontal edges. This leaves $r-2$ full horizontal edges uncovered, and the same construction, rotated by $\pi/2$, covers these with $\lceil (r-2)/2\rceil$ nested rectangles. Thus, $R(r, s) \leqslant \lceil r/2\rceil+\lceil s/2\rceil-1$.

On the other hand, suppose $G \equiv G(r, s)$ is covered by $R \equiv R(r, s)$ rectangles, of which a have their top and bottom on the top and bottom of G, b have their top on the top of G only, c have their bottom on the bottom of G, and d have neither top nor bottom on the top or bottom of G (see illustration).



First,

$$a + b + c + d = R. \tag{1}$$

Since there are $r - 2$ different internal horizontal edges, there must be at least $r - 2$ internal horizontal sides from the rectangles; that is,

$$b + c + 2d \geqslant r - 2. \tag{2}$$

There are also s vertical edges running from the top row to the next row down, so at least s edges must hit the top row. Thus, counting edges which run down from the top,

$$2a + 2b \geqslant s, \quad \text{hence } 2a + 2b \geqslant 2\lceil s/2 \rceil. \tag{3}$$

A similar argument applied to the vertical edges running from the bottom row to the next row up shows that

$$2a + 2c \geqslant s, \quad \text{hence } 2a + 2c \geqslant 2\lceil s/2 \rceil. \tag{4}$$

Averaging (3) and (4), $2a + b + c \geqslant 2\lceil s/2 \rceil$, and so, combining with (2), $2a + 2b + 2c + 2d \geqslant 2\lceil s/2 \rceil + r - 2$, or equivalently, $R(r, s) = a + b + c + d \geqslant \lceil s/2 \rceil + r/2 - 1$. Since $R(r, s)$ is an integer,

$$R(r, s) \geqslant \lceil \lceil s/2 \rceil + r/2 - 1 \rceil = \lceil s/2 \rceil + \lceil r/2 \rceil - 1,$$

completing the proof.

Also solved by Robert Bernstein, Roger B. Eggleton, Charles H. Frick, Richard A. Gibbs, Mark D. Meyerson, Clarence R. Perisho, Kiran Lall Shrestha (Nepal), Alvin Tirman, and William P. Wardlaw.

Integers with Nonconsecutive One in Binary Notation **February 1987**

1258. *Proposed by Jerrold W. Grossman, Oakland University, Michigan.*

Let $s_1, s_2, s_3, \ldots$ be the sequence of positive integers (in increasing order) whose binary representation contains no two consecutive ones. The sequence begins (in binary notation) $1, 10, 100, 101, \ldots$. How can one efficiently determine s_n from n and vice versa?

Solution by David Callan, University of Bridgeport, Connecticut.

The following facts concerning the Fibonacci sequence ($F_1 = 1$, $F_2 = 2$, and for $n \geqslant 0$, $F_{n+2} = F_{n+1} + F_n$) are well known and easily proved (e.g., see Don Knuth, *The Art of Computer Programming*, (1) *Fundamental Algorithms* (1973), p. 85).

1. $F_{2n+1} = 1 + \sum_{k=1}^{n} F_{2k}$; $F_{2n} = 1 + \sum_{k=1}^{n} F_{2k-1}$.
2. F_n exceeds any sum of distinct "nonconsecutive" (i.e., nonadjacent terms of the sequence) lower indexed Fibonacci numbers.
3. Every positive integer n can be expressed uniquely as a sum of distinct, "nonconsecutive" Fibonacci numbers.

Let S denote the given sequence. Define Φ: $S \to N^+$ by $\Phi(s_n) = \sum_i F_i$ where i runs over the positions (right to left) occupied by ones in the binary representation of n; for example, $\Phi(1010) = F_2 + F_4$. By the results quoted in the preceding paragraph, Φ is an order preserving bijection, and it follows that $\Phi(s_n) = n$. This mapping gives an efficient algorithm for passing from s_n to n and vice versa.

Also solved by Seung Jin Bang (Korea), Robert E. Bernstein, George Crofts, M. G. Deshbande, Richard A. Gibbs, William Gratzer, L. Kuipers (Switzerland), C. Peter Lawes, William Mixon, H. C. Morris, Clarence R. Perisho, Herman Roelants (Belgium), Mike Pinter, Shailesh Shirali (India), J. M. Stark, John T. Ward, William P. Wardlaw, and the proposer.

Answers

Solutions to the Quickies on p. 47

A728. a. By symmetry, we need the point on $y = e^x$ that is closest to the line $y = x$. This point has its tangent parallel to the line $y = x$. Hence the point must satisfy $(d/dx)e^x = e^x = 1$, and thus $x = 0$. It follows that the closest points are $(0,1)$ on the one curve and $(1,0)$ on the other.

b. Let p be the x-coordinate of the intersection. The curve $y = \log_a x$ must intersect $y = a^x$ on $y = x$, and the two must be tangent at p with slope 1. Thus, $p = \log_a p$, $p = a^p$, and $(1/p)(1/\ln a) = 1$. Then $1/\ln a = p = \log_a p = (\ln p)/(\ln a)$, so $\ln p = 1$, and $p = e$. It follows that $\ln a = 1/e$ and $a = e^{1/e}$.

A729. Let $A = x/(xy + x + 1)$, $B = y/(yz + y + 1)$, $C = z/(zx + z + 1)$. Then,

$$\begin{aligned}
A &= x/(1/z + x + 1) = zx/(zx + z + 1) = 1/(yz + y + 1),\\
B &= y/(1/x + y + 1) = xy/(xy + x + 1) = 1/(zx + z + 1),\\
C &= z/(1/y + z + 1) = yz/(yz + y + 1) = 1/(xy + x + 1)
\end{aligned}$$

Therefore, $3(A + B + C) = 3$ and S has the constant value of 2.

An alternate and quicker solution is

$$S = \frac{x+1}{xy+x+1} + \frac{y+1}{yz+y+1} + \frac{z+1}{zx+z+1}$$

$$= \frac{x+1}{xy+x+1} + \frac{y+1}{\frac{1}{x}+y+1} + \frac{\frac{1}{xy}+1}{\frac{1}{y}+\frac{1}{xy}+1} = 2.$$

A related open problem is to find the extremes of S if the condition $xyz = 1$ is replaced by $xyz = a$.

A730. This is an immediate consequence of the fact that an integer is expressible as a sum of two squares if and only if each of its prime factors that are congruent to 3 (mod 4) occur with an even exponent.

Comments

1235. (Proposed February 1986).

No solution was received for part b, which was to find, if possible, a *polynomial* in two variables with "two mountains without a valley in between." *Roy O. Davies* (The University of Leicester, England), gives the following example: Take

$$f(x,y) = -\left(x^2y - x - 1\right)^2 - \left(x^2 - 1\right)^2.$$

We have

$$\begin{aligned} f_x &= -2\left(x^2y - x - 1\right)(2xy - 1) - 4x\left(x^2 - 1\right), \\ f_y &= -2\left(x^2y - x - 1\right)x^2. \end{aligned}$$

If (x, y) is a critical point, then $x \neq 0$, because $f_x(0, y) = -2$. Therefore, from $f_y = 0$ we have $x^2y - x - 1 = 0$. Then, from $f_x = 0$ we conclude that $-4x(x^2 - 1) = 0$. It follows that either $x = 1$ and $y = 2$, or, $x = -1$ and $y = 0$. Also, $f_{xx}f_{yy} - (f_{xy})^2 = 16$ and $f_{yy} = 2$ at both $(1,2)$ and $(-1,0)$. Thus, the only critical points are two strict local maxima.

Q720. (April 1987).

The problem established that if A is a *nonsingular* $n \times n$ matrix then $\operatorname{adj}(\operatorname{adj}(A)) = (\det A)^{n-2}A$. *Edward T. H. Wang* (Wilfred Laurier University) and *Seung Jin Bang* (Seoul, Korea) independently showed that this identity holds even when A is *singular*. (Here, when $n = 2$ and $\det A = 0$, we define $0^0 = 1$.) The following argument is due to Seung Jin Bang.

If $0 < |\varepsilon| \ll 1$ then $A + \varepsilon I$ is invertible, so by **Q720**, we have

$$\operatorname{adj}(\operatorname{adj}(A + \varepsilon I)) = (\det(A + \varepsilon I))^{n-2}(A + \varepsilon I).$$

Note that the entries of the matrices on both sides of the equation are real *polynomials* in ε and hence continuous. Equating the entries in the corresponding positions and taking the limit, letting $\varepsilon \to 0^+$, yields the desired result.

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.

Steen, Lynn Arthur, Who still does math with paper and pencil?, *The Chronicle of Higher Education* (14 October 1987) A48.

"Each term, an army of three million students labors with primitive tools to master the art of digging and filling intellectual ditches: Instead of using shovels and pick axes, these students use paper and pencil to perform millions of repetitive calculations in algebra, calculus, and statistics. Mathematics, the queen of the sciences, has become the serf of the curriculum. ... All that is going to change in the next two or three years. ... Mathematics-speaking machines are about to sweep the campuses. embodied both as computer disks and as pocket calculators. ... What can be said with certainty ... is that the era of paper-and-pencil mathematics is over."

Walker, Jearl, The amateur scientist: Now there is Rubik's Magic, a new puzzle that provides a study in permutation operators, *Scientific American* 257:4 (October 1987) 170-173, 184.

Analyzes Rubik's Magic puzzle in terms of states and transition operators between them.

Dewdney, A. K., Computer recreations: "After MAD": A computer game of nuclear strategy that ends in a Prisoner's Dilemma, *Scientific American* 257:4 (October 1987) 174-177, 184.

Describes a computer game of nuclear strategy designed as a "research tool to probe the effects of technological change on the strategic relation between the U.S. and the Soviet Union." The game incorporates a tree of 2×2 matrices that span the period from "Early MAD [Mutually Assured Destruction]" to 45 years after, during which weapons become increasingly accurate and the temptation to launch a first strike grows. The latter feature is reflected in the increasing instability of the corresponding game matrices, all of the Prisoner's Dilemma type. Experimental results are not encouraging: "Players generally did not wait for game matrices to become unstable before going to war."

Stewart, Ian, The symplectic camel, *Nature* 329 (3 September 1987) 17-18.

V. I. Arnold (Moscow State) has laid down a "manifesto" of symplectic mathematics: "As each skylark must display its comb, so every branch of mathematics must finally display symplectization"—that is, its concepts have analogies in symplectic geometry, a type of geometry important in classical mechanics. Symplectic geometry corresponds to the bilinear form $x_1y_2 - x_2y_1$, which is the area of the parallelogram formed by the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Thus, in the plane a symplectic mapping is one that preserves area. In higher dimensions, symplectic mappings preserve volume; but being symplectic imposes more restrictions than just that. Says Stewart: "We are witnessing just the tip of the symplectic iceberg."

Ralston, Anthony, Let them use calculators, *Technology Review* (August-September 1987) 30-31.

Past justifications for teaching proficiency in paper-and-pencil arithmetic to young children (useful in everyday life, necessary for higher math, it can be done) no longer apply; the level of arithmetic skill we attempt to impose is far greater than necessary. What children should learn instead: simple mental arithmetic, estimation, and how to decide which arithmetic operation applies to a particular problem. The curricular implications of such change are staggering, and the vested interests enormous.

Gillman, Leonard, *Writing Mathematics Well: A Manual for Authors*, MAA, 1987; ix + 49 pp (P).

This manual contains much useful advice, not only for authors intending to submit manuscripts to the MAA, but for authors of any mathematical manuscript. (Unfortunately, the mathematical community still seems stuck with the inconvenient system of citing references by using numbers, instead of, for example, using the author-date system; anyone who has had to add or delete a numbered reference knows the pain involved.)

Steen, Lynn Arthur (ed.), *For All Practical Purposes: Introduction to Contemporary Mathematics*, Freeman, 1988; xii + 450 pp, $28.95.

"This book is designed for use in a one-term course in liberal arts mathematics or in courses that survey mathematical ideas. ... [W]e aim to develop in the reader strong conceptual understanding and appreciation, not computational expertise." The book represents an answer to the question. "What would you teach students if they took only one semester of math during their entire college career?" The 21 chapters cover management science, statistics, social choice, size and shape, and computers. Exercises (some with answers) are included, and an instructor's guide is available. There is an accompanying series of 26 half-hour television programs, currently being broadcast on PBS and also available for separate purchase. (Truth-in-reviewing revelation: I wrote part of this beautiful book.)

Barnett, Arnold, Misapplications reviews: High road to glory, *Interfaces* 17:5 (September-October 1987) 51-54.

How can we impress the public that mathematics is relevant? One marvelous way, of course, is to demonstrate its usefulness in matters of national urgency—such as the obsession with maximizing frequent-flyer air-travel miles. The author formulates a problem he personally encountered and solves it using integer programming, thus illustrating his objective of "devising a way to inform the world how central we are to its functioning."

Stewart, Ian, The arithmetic of chaos, *Nature* 329 (22 October 1987) 670-671.

"For at least a decade, evidence of the applicability of number theory has been accumulating." Stewart cites as the most recent evidence work by I. Percival and F. Vivaldi (Queen Mary College) applying methods from algebraic number theory to "one of the most flourishing, fascinating and fashionable research fields of mathematical physics—chaotic dynamics."

Bowen, Bruce, Calculating apes: When the chips are down, do chimpanzees sum quantities in a simple way linked to the human capacity for counting? *Science News* 131 (23 May 1987) 334-335.

Can chimps add? Well: maybe, maybe not. Research shows that they may be able to subitize, that is, perceive at a glance the number of items present, and even combine and compare results from subitizings. Stay tuned.

Pólya, George, *The Pólya Picture Album: Encounters of a Mathematician*, edited by G. L. Alexanderson, Birkhäuser, 1987; 160 pp, $35.

Annotated photographs of mathematicians from Pólya's photo album, with a biographical sketch of Pólya by Alexanderson.

Lay-Yong, Lam, The conceptual origins of our numeral system and the symbolic form of algebra, *Archive for History of Exact Sciences* 36:3 (1986) 183-195.

The ancient Chinese developed a decimal place-value numeral system and the notion of symbolic concepts in algebra. The ingenuity of their bamboo-rod notation "cannot be overrated," says Lay-Yong, who also asks the difficult question: Was the Chinese concept of decimal place-value numeration transmitted to India rather than originating there independently? "[T]he possibility ... cannot be ruled out."

Peterson, Ivars, Twists of space, *Science News* 132 (24 October 1987) 264-266.

Account of a collaboration of an artist, a mathematician, and a computer scientist to visualize "exotic" topological forms and homotopies between them. The surfaces involved are Steiner's Roman surface (which features a Moebius strip sewn to a disk) and Boy's surface.

Hoffman, Paul, The man who loves only numbers, *Atlantic Monthly* (November 1987) 60-74.

You guessed it—another freelance writer has discovered Paul Erdős; eccentricity sells. Mathematicians will find this article more informative than most on Erdős; the general public will reconfirm its suspicions of mathematicians.

Cole, K. C., A theory of everything, *New York Times Magazine* (18 October 1987) 20-28.

Feature article on (super)string theory, with little exposition of the theory and a full-page color photo of its leading proponent. What we have here, in other words, is science popularized in cult-of-personality style. The point is made, however, that string theory has forged a "new liaison" between physicists and mathematicians.

Gardner, Martin, *Time Travel and Other Mathematical Bewilderments*, Freeman, 1987; ix + 295 pp, $17.95, $12.95 (P).

Here is the twelfth collection of Martin Gardner's "Mathematical Games" columns from Scientific American, with 22 columns updated on the basis of readers' letters. Included is the famous April Fools' column of 1975, which indeed fooled many into believing that the four color problem had been resolved in the negative.

Traub, Joseph F., et al. (eds.), *Annual Review of Computer Science*, Volume 2, 1987, Annual Reviews Inc., 1987; vii + 565 pp, $39.

Nine of the 17 essays in this annual treat aspects of artificial intelligence; others address such topics as computer algebra algorithms, the state of linear programming, algorithmic geometry of numbers (including integer programming), and computer applications in education.

Phillips, Esther R. (ed.), *Studies in the History of Mathematics*, MAA Studies in Mathematics vol. 26, MAA, 1987; 308 pp, $32.

Ten unrelated essays on topics in the history of modern and contemporary mathematics, by leading historians of mathematics. Topics include the discovery of non-euclidean geometry, Dedekind's invention of ideals, the emergence of first-order logic, the separation of variables technique in differential equations, and the beginnings of Italian algebraic geometry.

NEWS & LETTERS

LETTERS TO THE EDITOR

Dear Editor:

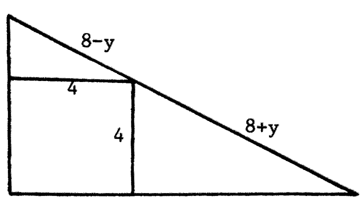
I would like to thank Ruma Falk of The Hebrew University Department of Psychology for pointing out two mistakes in my June 1987 note on Murphy's Law. On page 161, line 26, a reference is made to the "third" column of the table on page 160. This should have been "fourth" column. Also, on page 161, line 29 starts with "chances of the first mishap." This should have been "chances of only a single mishap." I regret these errors and again would like to thank Dr. Falk.

Gene G. Garza
Department of Mathematics
University of Montevallo
Montevallo, AL 35115

Dear Editor:

May I suggest the following neater solution to Monte Zerger's "Ladder Problem" (vol. 60, no. 4, pp. 239-242).

The figure can be alternatively parameterised as:



whence

$$\left(\frac{4}{8-y}\right)^2 + \left(\frac{4}{8+y}\right)^2 = 1 ,$$

which then gives the result.

Oliver D. Anderson
Department of Statistics
Temple University
338 Speakman Hall (006-00)
Philadelphia, PA 19122

ANNOUNCEMENTS

BOSTON WORKSHOP FOR MATHEMATICS FACULTY

The Second Boston Workshop will be held at Wellesley College on August 5-8, 1988, prior to the Centennial Celebration in Providence. The goal is to strengthen the teaching of calculus, applied linear algebra, differential equations, and numerical methods. For information about presentations, discussions, and housing contact Professor Gilbert Strang, Room 2-240 M.I.T., Cambridge, MA 02139.

INTERNATIONAL CONGRESS ON MATHEMATICAL EDUCATION - TRAVEL GRANTS

Mathematics educators, including precollege classroom teachers, may qualify for travel grants for the Sixth International Congress on Mathematical Education. The Congress takes place from July 27 - August 3, 1988 in Budapest, Hungary. Information and applications for travel grants can be obtained by writing:

Dept. E
National Council of Teachers of Mathematics
1906 Association Drive
Reston, VA 22091

Applications must be received by March 1, 1988.

ERRATA

In the Letter to the Editor by Branko Grünbaum (June 1987), the second *pgg* should have been *pg*.

The following name was inadvertently omitted from the Referee Acknowledgement list (60 (1987) xxx):

Katherine Heinrich, *Simon Fraser University*.

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

COVER

Cover art by Robert Ishi

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