MATHEMATICS MAGAZINE

CONTENTS

CODEN: MAMGA8

VOLUME 48 • MAY 1975 • NUMBER 3

MATHEMATICS MAGAZINE

GERHARD N. WOLLAN, *EDITOR*

EDITORIAL CORRESPONDENCE should be sent to the EDITORS-ELECT, J. ARTHUR SEEBACH, JR., and LYNN A. STEEN, Department of Mathematics, St. Olaf College, Northfield, Minnesota 55057. Articles should be typewritten and triple-spaced on $8\frac{1}{2}$ by 11 paper. The greatest possible care should be taken in preparing the manuscript. Authors should submit two copies and keep one copy for protection against possible loss of the manuscript. Figures should be drawn on separate sheets in India ink. They should be appropriately lettered and of a suitable size for photographing.

REPRINT PERMISSION should be requested from LEONARD GILLMAN, Mathematical Association of America, University of Texas, Austin, Texas 78712. (See also the copyright notice below.)

NOTICE OF CHANGE OF ADDRESS and other subscription correspondence should be sent to the Executive Director, A. B. WILLCOX, Mathematical Association of America, Suite 310, 1225 Connecticut Avenue, N. W., Washington, D. C. 20036.

ADVERTISING CORRESPONDENCE should be addressed to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo (University of Buffalo), Buffalo, New York 14214.

THE MATHEMATICS MAGAZINE is published by the Mathematical Association of America at Washington, D.C., bi-monthly except July-August. Ordinary subscriptions are $10 per year. Members of the Mathematical Association of America and of Mu Alpha Theta may subscribe at the special rate of $7.00. Back issues may be purchased, when in print, for $2.00.

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

FIBONACCI ANALOGS OF THE CLASSICAL POLYNOMIALS

A. G. SHANNON, The New South Wales Institute of Technology

Introduction. In this note we define a generalized Fibonacci polynomial analogous to the classical polynomials of Bernoulli, Euler and Hermite. We then use it to obtain a combinatorial result for the associated generalized Fibonacci numbers. It is an advantage, though not an essential prerequisite, for the reader to have had some prior acquaintance with Fibonacci numbers and classical polynomials, a number of papers on which have appeared in this journal.

We define generalized Fibonacci number sequences, $\{u_n\}$ and $\{v_n\}$:

$$\begin{aligned} u_n &= \sum_{j=1}^{r} (-1)^{j+1} P_j u_{n-j} & n > 0 \\ u_n &= 1 & n = 0 \\ u_n &= 0 & n < 0 \end{aligned}$$

and

$$\begin{aligned} v_n &= \sum_{j=1}^{r} (-1)^{j+1} P_j v_{n-j} & n \geqq r \\ v_n &= \sum_{j=1}^{r} \alpha_j^n & 0 \leqq n < r \\ v_n &= 0 & n < 0 \end{aligned}$$

where the P_j are arbitrary integers and the α_j are the roots, assumed distinct, of the auxiliary equation

$$x^r - \sum_{j=1}^{r} (-1)^{j+1} P_j x^{r-j} = 0,$$

which is associated with the homogeneous, linear recurrence relations of arbitrary order which the generalized Fibonacci numbers satisfy.

For example, when $r = 2$ we have

$$u_n = P_1 u_{n-1} - P_2 u_{n-2}$$

with $u_0 = 1$, $u_1 = P_1$, $u_2 = P_1^2 - P_2$ and so on. These are often referred to as the Lucas fundamental numbers. When $P_1 = -P_2 = 1$ we get the Fibonacci numbers, f_n. The v_n correspond to the Lucas primordial numbers since when $r = 2$, we have $v_0 = 2$, $v_1 = \alpha_1 + \alpha_2 = P_1$, and so on. When $P_1 = -P_2 = 1$ these become the ordinary Lucas numbers, L_n. (The interested reader can find more of their properties in Hoggatt [**1**].)

We need the result that

$$v_n = \sum_{j=1}^{r} \alpha_j^n \quad \text{for all } n \geqq 0.$$

Proof.

$$\begin{aligned} v_n &= \sum_{j=1}^{r} (-1)^{j+1} P_j v_{n-j} \\ &= \sum_{j=1}^{r} (-1)^{j+1} P_j \sum_{i=1}^{r} \alpha_i^{n-j} \\ &= \sum_{i=1}^{r} \alpha_i^{n-r} \sum_{j=1}^{r} (-1)^{j+1} P_j \alpha_i^{r-j} \\ &= \sum_{i=1}^{r} \alpha_i^{n-r} \alpha_i^{r} \\ &= \sum_{i=1}^{r} \alpha_i^{n}, \text{ as required.} \end{aligned}$$

The ordinary generating function for $\{u_n\}$ is given (formally) by

$$\sum_{n=0}^{\infty} u_n x^n = \prod_{j=1}^{r} (1 - \alpha_j x)^{-1}.$$

Proof. If $u(x) = \sum_{n=0}^{\infty} u_n x^n$, then $(1 - P_1 x + P_2 x^2 - \cdots + (-1)^r P_r x^r) u(x)$

$$\begin{aligned} &= u_0 + (u_1 - P_1 u_0)x + (u_2 - P_1 u_1 + P_2 u_0)x^2 + \cdots \\ &= 1, \text{ and the result follows.} \end{aligned}$$

We now use these two results to establish

$$\sum_{n=0}^{\infty} u_n x^n = \exp\left(\sum_{m=1}^{\infty} v_m x^m / m \right).$$

Proof. $u(x) = \prod_{j=1}^{r} (1 - \alpha_j x)^{-1}$, $\log(u(x)) = \log \prod_{j=1}^{r} (1 - \alpha_j x)^{-1}$

$$\begin{aligned} &= -\log \prod_j (1 - \alpha_j x) \qquad \text{if } |\alpha_j x| < 1, \\ &\qquad\qquad\qquad\qquad\qquad j = 1, 2, \cdots, r, \\ &= -\sum_j \log(1 - \alpha_j x) \\ &= \sum_j \sum_{m=1}^{\infty} \frac{1}{m} \alpha_j^m x^m \\ &= \sum_{m=1}^{\infty} \frac{1}{m} \left(\sum_j \alpha_j^m \right) x^m \\ &= \sum_{m=1}^{\infty} \frac{1}{m} v_m x^m . \end{aligned}$$

Thus

$$u(x) = \exp\left(\sum_{m=1}^{\infty} v_m x^m / m \right).$$

Generalized Fibonacci polynomials. The last result suggests a definition for generalized Fibonacci polynomials $u_n(x)$:

$$\sum_{n=0}^{\infty} u_n(x)t^n/n! = \exp\left(xt + \sum_{m=1}^{\infty} v_m t^m/m\right)$$

so that

$$u_n(0) = u_n \cdot n!$$

and

$$\begin{aligned}\sum_{n=0}^{\infty} u_n(x)t^n/n! &= e^{xt}\exp\left(\sum_{m=1}^{\infty} v_m t^m/m\right)\\ &= e^{xt}\sum_{n=0}^{\infty} u_n t^n\\ &= e^{xt}\sum_{n=0}^{\infty} u_n(0)t^n/n!\end{aligned}$$

which are like the conditions

$$\begin{aligned}\sum_{n=0}^{\infty} H_n(x)t^n/n! &= e^{2xt}\sum_{n=0}^{\infty} H_n(0)t^n/n!,\\ \sum_{n=0}^{\infty} B_n(x)t^n/n! &= e^{xt}\sum_{n=0}^{\infty} B_n(0)t^n/n!,\\ \sum_{n=0}^{\infty} E_n(x)t^n/n! &= e^{xt}\sum_{n=0}^{\infty} E_n(0)t^n/n!,\end{aligned}$$

for the polynomials of Hermite, Bernoulli and Euler respectively.

The generalized Fibonacci polynomials are Appell.

Proof. In differentiation of series we assume that conditions of continuity and uniform convergence are satisfied in the appropriate closed intervals.

$$\begin{aligned}\frac{\partial}{\partial x}\sum_{n=0}^{\infty} u_n(x)t^n/n! &= t\exp\left(xt + \sum_{m=1}^{\infty} v_m t^m/m\right)\\ &= t\sum_{n=0}^{\infty} u_n(x)t^n/n!\\ &= \sum_{n=1}^{\infty} n\, u_{n-1}(x)t^n/n!\end{aligned}$$

which gives

$$u_n'(x) = n\, u_{n-1}(x), \quad n = 1,2,3,\cdots$$

which is the Appell set criterion.

These generalized Fibonacci polynomials are not orthogonal since Shohat [**2**] has proved that the only system of orthogonal polynomials which is an Appell polynomial sequence is that which is reducible to Hermite polynomials by a linear transformation. That the generalized Fibonacci numbers cannot be so reduced can be deduced from the following result which can also be used to prove some congruences mentioned below:

$$u_{n+1}(x) = x\, u_n(x) + \sum_{j=0}^{n} v_{j+1} n_j u_{n-j}(x)$$

where n_j is the falling factorial coefficient

$$n_j = n(n-1)\cdots(n-j+1).$$

Proof.

$$\frac{\partial}{\partial t} \sum_{n=0}^{\infty} u_n(x) \frac{t^n}{n!} = \sum_{n=0}^{\infty} u_{n+1}(x) \frac{t^n}{n!},$$

and

$$\begin{aligned}
\frac{\partial}{\partial t} \sum_{n=0}^{\infty} u_n(x) \frac{t^n}{n!} &= \frac{\partial}{\partial t} \exp\left(xt + \sum_{m=1}^{\infty} v_m \frac{t^m}{m}\right) \\
&= \exp\left(xt + \sum_{m=1}^{\infty} v_m \frac{t^m}{m}\right) \frac{\partial}{\partial t}\left(xt + \sum_{m=1}^{\infty} v_m \frac{t^m}{m}\right) \\
&= \left(x + \sum_{m=1}^{\infty} v_{m+1} t^m\right) \exp\left(xt + \sum_{m=1}^{\infty} v_m \frac{t^m}{m}\right) \\
&= \left(x + \sum_{m=0}^{\infty} v_{m+1} t\right) \sum_{n=0}^{\infty} u_n(x) t^n / n! \\
&= \sum_{n=0}^{\infty} x u_n(x) \frac{t}{n!} + \sum_{m=0}^{\infty} v_{m+1} t^m \sum_{n=0}^{\infty} u_n(x) \frac{t^n}{n!} \\
&= \sum_{n=0}^{\infty} x u_n(x) \frac{t^n}{n!} + \\
&\quad + \sum_{n=0}^{\infty} \sum_{j=0}^{n} v_{j+1} n^j u_{n-j}(x) \frac{t^n}{n!}
\end{aligned}$$

which gives the required result.

From this we get when $x = 0$ that $(n+1)!u_{n+1} = n!v_1 u_n + \sum_{j=1}^{n} v_{j+1} n! u_{n-j}$ since $n! = n_j(n-j)!$

Thus

$$(n+1)u_{n+1} = \sum_{j=0}^{n} v_{j+1} u_{n-j}$$

which is a relationship between the primordial and fundamental sequences. For example, when $r = 2$, we get for the Fibonacci and Lucas numbers that

$$(n+1)f_{n+1} = \sum_{j=0}^{n} L_{j+1}f_{n-j};$$

$$n = 3,\ (n+1)f_{n+1} = 4f_4 = 20,$$

and

$$\begin{aligned}\sum_{j=0}^{3} L_{j+1}f_{3-j} &= L_1f_3 + L_2f_2 + L_3f_1 + L_4f_0 \\ &= 1\times 3 + 3\times 2 + 4\times 1 + 7\times 1 \\ &= 20.\end{aligned}$$

The Bernoulli polynomials can be written in the suggestive form

$$B_n(x) = (x + B(0))^n$$

wherein it is understood that after expansion of the right member, a^k is replaced by a_k. So too

$$u_n(x) = (x + u(0))^n.$$

Proof.

$$\begin{aligned}\sum_{n=0}^{\infty} u_n(x)t^n/n! &= \exp(xt)\exp\left(\sum_{m=1}^{\infty} v_m t^m/m\right) \\ &= \sum_{k=0}^{\infty} x^k t^k/k! \sum_{j=0}^{\infty} u_j t^j \\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{n!}{k!}u_{n-k}x^k t^n/n!\end{aligned}$$

so that on equating coefficients of t^n we find

$$\begin{aligned}u_n(x) &= \sum_{k=0}^{n} \frac{n!}{k!}u_{n-k}x^k \\ &= \sum_{k=0}^{n} \binom{n}{k} u_{n-k}(0)x^k\end{aligned}$$

which is what we seek. We shall use this result later.

Any polynomial can in fact be expressed in a series of these generalized Fibonacci polynomials.

Proof.

$$\begin{aligned}\exp xt &= \exp\left(-\sum_{m=1}^{\infty} v_m t^m/m\right) \sum_{n=0}^{\infty} u_n(x)\,t^n/n! \\ &= \left(1 + \sum_{j=1}^{r} (-1)^j P_j t^j\right) \sum_{n=0}^{\infty} u_n(x)\,t^n/n!\end{aligned}$$

If $P_0 \equiv 1$, then

$$\sum_{n=0}^{\infty} x^n t^n/n! = \sum_{n=0}^{\infty} \sum_{j=0}^{r} u_{n-j}(x) P_j (-1)^j n_j \frac{t^n}{n!}.$$

Equate coefficients of t^n and

$$x^n = \sum_{j=0}^{r} (-1)^j P_j n_j u_{n-j}(x).$$

The first few generalized Fibonacci polynomials are then

$$\begin{aligned} u_0(x) &= & &= u_0, \\ u_1(x) &= x + P_1 & &= u_0 x + u_1, \\ u_2(x) &= x^2 + 2P_1 x + 2(P_1^2 - P_2) & &= u_0 x^2 + 2u_1 x + 2u_2, \\ u_3(x) &= x^3 + 3P_1 x^2 + 6(P_1^2 - P_2)x + 6(P_1^3 - 2P_1P_2 + P_3) \\ &= u_0 x^3 + 2u_1 x^2 + 6u_2 x + 6u_3. \end{aligned}$$

Since $n_{n-k} = n(n-1)_{n-1-k}$, the coefficient array follows the pattern:

1						
1	1					
1	2	2				
1	3	6	6			
1	4	12	24	24		
1	5	20	60	120	120	
1	6	30	120	360	720	720.

A combinatorial result. A composition of the positive integer n is a vector $(a_1, a_2, \cdots, a_k)$, the components of which are the positive integers such that $a_1 + a_2 + \cdots + a_k = n$. If the vector has order k, then the composition is a k-part composition. In what follows $\gamma(n)$ will indicate summation over all the compositions $(a_1, a_2, \cdots, a_k)$ of n, the number of components being variable.

Let

$$w_n = \sum_{\gamma(n)} \frac{(-1)^{k-1}}{k} u_{a_1} \cdots u_{a_k}.$$

Then

$$\begin{aligned} \sum_{n=1}^{\infty} w_n x^n &= \sum_{n=1}^{\infty} \sum_{\gamma(n)} \frac{(-1)^{k-1}}{k} u_{a_1} \cdots u_{a_k} x^n \\ &= \sum_{k=1}^{\infty} - \left(- \sum_{n=1}^{\infty} u_n x^n \right)^k / k \end{aligned}$$

$$= \ln\left(1 + \sum_{n=1}^{\infty} u_n x^n\right),$$

$$= \ln\left(\sum_{n=0}^{\infty} u_n x^n\right)$$

or

$$\sum_{n=0}^{\infty} u_n x^n = \exp\left(\sum_{n=1}^{\infty} w_n x^n\right),$$

which is satisfied by $w_n = v_n/n$ from above. Thus

$$v_n = \sum_{\gamma(n)} (-1)^{k-1} \frac{n}{k} f_{a_1} \cdots f_{a_k}.$$

When $r = 2$, $P_{21} = -P_{22} = 1$, we obtain

$$L_n = \sum_{\gamma(n)} (-1)^{k-1} \frac{n}{k} u_{a_1} \cdots u_{a_2}.$$

For example,

$$L_1 = 1 = f_1;$$

$$L_2 = 3$$

$$= -\frac{2}{2} f_1 f_1 + \frac{2}{1} f_2$$

$$= -1 + 4;$$

$$L_3 = 4$$

$$= -\frac{3}{2} f_1 f_2 - \frac{3}{2} f_2 f_1 + \frac{3}{1} f_3 + \frac{3}{3} f_1 f_1 f_1$$

$$= -3 - 3 + 9 + 1.$$

Conclusion. These generalized Fibonacci polynomials can be used to develop other properties: for instance, it can be established that $u_{n+tm}(x) \equiv u_n(x)(u_m(x))^t \pmod{m}$ from which it follows that

$$u_n u_m \equiv \binom{n+m}{m} u_{n+m} \pmod{m}.$$

For example, when $r = 2$, $P_{21} = -P_{22} = 1$, we have

$$f_3 f_4 = 3 \times 5 \equiv 3 \pmod{4}$$

and

$$\binom{7}{4} f_7 = 35 \times 21 \equiv 3 \pmod{4}.$$

The proof of the congruence for the polynomials is left to the interested reader.

More easily, other analogs of properties of the classical polynomials can be developed with a view toward obtaining results for generalized Fibonacci numbers. The generalized Fibonacci polynomials can also be related to the classical polynomials more directly; for instance, in the case of the Hermite polynomials we have

$$\sum_{m=0}^{\infty} H_m(x)u_n(y)t^m/m! = \exp(2xyt - y^2t^2) \sum_{n=0}^{\infty} H_n(x - yt)u_n(0)t^n/n!$$

Proof. We use the well-known result that

$$\sum_{n=0}^{\infty} H_{m+n}(x)y^n/n! = \exp(2xy - y^2)H_m(x-y).$$

$$\exp(2xyt - y^2t^2) \sum_{n=0}^{\infty} H_n(x - yt)u_n(0)t^n/n!$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} H_{m+n}(x)u_n(0)y^m t^{m+n}/m!n!$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} H_{m+n}(x)u_n(0)\binom{m+n}{n} y^m t^{m+n}/(m+n)!$$

$$= \sum_{m=0}^{\infty} H_m(x)\left(\sum_{n=0}^{m}\binom{m}{n} u_n(0)y^{m-n}\right)t^m/m!$$

$$= \sum_{m=0}^{\infty} H_m(x)u_m(y)t^m/m! \text{ as required.}$$

This is a bilateral generating function for the Hermite polynomials and is an illustration of Corollary 5 of Singhal and Srivastava [3].

As an exercise the reader might like to prove a relation between these generalized Fibonacci numbers and the Bernoulli numbers:

$$u_{n-1} = \sum_{k=0}^{n} \frac{B_{n-k}(0)}{(n-k)!} \frac{\Delta u_k(0)}{k!}$$

where $\Delta u_n(x) = u_n(x+1) - u_n(x)$.

These polynomials then can be used to establish new results for generalized Fibonacci-type numbers or new insights into properties of the classical polynomials. It does not require much imagination to extend these results to obtain some reasonably elegant formulas.

References

1. V. E. Hoggatt, Jr., Fibonacci and Lucas Numbers, Houghton Mifflin, Boston, 1969.

→ J. Shohat, The relation of the classical orthogonal polynomials to the polynomials of Appell, Amer. J. Math., 58 (1936) 453–464.

3. J. P. Singhal and H. M. Srivastava, A class of bilateral generating functions for certain classical polynomials, Pacific J. Math., 42 (1972) 755–762.

GENERAL SOLUTION TO THE OCCUPANCY PROBLEM WITH VARIABLY SIZED RUNS OF ADJACENT CELLS OCCUPIED BY SINGLE BALLS

ROGER W. PEASE, JR., G. & C. Merriam Company, Springfield, Massachusetts

Consider the occupancy problem in which (1) there is a row of n cells of which m cells are occupied by one ball each, (2) there are runs of $r_1, r_2, \cdots, r_k$ adjacent occupied cells with $\sum_{i=1}^{k} r_i = m$ that occur in order from one end of the row to the other, and (3) it is desired to find how many different arrangements of occupied and unoccupied cells there are for the two cases in which there is no restriction on the number of empty cells between runs and in which there is at least one empty cell between each pair of runs.

A particular application of this occupancy problem that is easy to grasp involves the seating of theater parties in a row of seats. If parties of size $p_1, p_2, \cdots, p_k$ with $\sum_{i=1}^{k} p_i = m$ are seated in order in a row of n seats, how many arrangements of occupied and unoccupied seats are there if (1) there need not be a vacant seat between parties, and (2) there is at least one vacant seat between parties?

The case in which the parties are all of the same size was solved by an instructive use of recursion by Wiggins [1] including the cases in which each party was composed of two individuals and of m individuals. His model involves the counting of dyadic numbers consisting of 1's representing cell partitions and intercalated 0's representing balls occupying cells. The method is related to the one used in Feller [2] to derive the formula for cases in which multiple occupancy of cells is permitted and which considers the number of ways in which the $n-1$ partitions between cells or alternatively the balls can be arranged, one to a place, in a total number of places equal to the sum of the $n-1$ partitions between cells and the balls. However, since at most one ball is permitted to reside in each cell in the problems considered here, the number of cells n figures prominently in the discussion rather than the number of partitions between cells $n-1$.

The intuitive insight that leads to solution of the general case is that the runs of adjacent occupied cells can be considered as occupying single cells if the number of balls is reduced by the difference between the number of balls and the number of runs. The intuitive answer is immediate. In the theater party application let the number of seats be n, the total number of people be m, and the number of parties be k. The number of arrangements of occupied and unoccupied seats is the number of ways in which the $n-m$ empty seats can be chosen from the adjusted number of seats $n-m+k$. The answer is $\binom{n-m+k}{n-m}$. Note that the number in each party is not required for solution. Thus, five parties of sizes 100, 5, 5, 5, 5 will generate the same number of arrangements of occupied and unoccupied seats as five parties of sizes 16, 20, 24, 28, and 32 provided the parties are seated in a particular order.

This result will be made rigorous first by a proof that is based on the intuition itself and sets up a one-to-one mapping onto a set of arrangements that we know how to count and second by a proof that involves recursion. The one-to-one onto mapping required in the first proof is obtained in the following way. Since there are no vacant seats within parties, each arrangement can be identified and is completely determined by the ordered set of seat numbers of the leftmost party members $\{t_1,\cdots,t_k\}$ where the t_i's are counted from the left end of the row. If $\{p_i: i = 1,\cdots,k\}$ is the set specifying the number of individuals in each party, the following statements hold for any particular arrangement:

$$t_1 \geqq 1,$$

$$t_{i+1} \geqq t_i + p_i,$$

$$t_k \leqq n-(p_k-1).$$

The intuitive process apparently involved forming the conviction that there was a one-to-one correspondemce between these arrangements and the arrangements of k ordered objects in a row of cells of length $n-m+k$. Each such arrangement is specified by a set $\{s_1,\cdots,s_k\}$ where s_i is the number of the ith occupied cell counting from the left. The one-to-one correspondence is specified by the transformation

$$s_1 = t_1,$$

$$s_{i+1} = t_{i+1} - \sum_{j=1}^{i}(p_j-1),$$

$$s_k = t_k - m + p_k + k - 1.$$

Each t_i uniquely determines an s_i and vice versa. Each set of t_i's uniquely determines a set of s_i's and vice versa. Therefore, the mapping is one-to-one and onto, and the two sets of arrangements have the same cardinality, namely $\binom{n-m+k}{k}$.

The artificial proof is modeled after that of Wiggins [1] for the cases in which the parties are all of the same size. Let $N_{n,m}^{(k)}$ be the number of arrangements of occupied and unoccupied seats for k parties arranged in a particular order with varying numbers of unoccupied seats between them, n the total number of seats, and m the number of people.

The case in which there is one party is trivial. The leftmost member of the party can occupy any of $n-m+1$ seats so that

$$N_{n,m}^{(1)} = n-m+1 = \binom{n-m+1}{n-m}. \tag{1}$$

Now consider the case with two parties. If a second party with m_2 members is seated to the left of the first party with m_1 members, the total number of arrangements will be the sum of the ways of seating the first party as the leftmost seat occupied by the second party moves from left to right.

$$(2) \quad N^{(2)}_{n,m_1+m_2} = N^{(1)}_{n-m_2\ m_1} + N^{(1)}_{n-m_2-1,m_1} + \cdots + N^{(1)}_{m_1,m_1}$$

$$= \sum_{i=0}^{n-m_1-m_2} N^{(1)}_{n-m_2-i,m_1} = \sum_{i=0}^{n-m_1-m_2} (n-m_1-m_2+1-i) = \sum_{j=1}^{n-m_1-m_2+1} j$$

$$= \frac{(n-m_2-m_1+1)(n-m_2-m_1+2)}{2} = \binom{n-m_1-m_2+2}{n-m_1-m_2}.$$

And since $m = m_1 + m_2$

$$(3) \qquad N^{(2)}_{n,m} = \binom{n-m+2}{n-m}.$$

Suppose the result is true for k parties and consider the case for $k+1$ parties. The order of the parties in the row is given. Again the total number of arrangements is found by holding the leftmost party fixed and summing the ways of seating the k remaining parties as the party on the left is moved successively one seat at a time to the right until no vacant seats remain between parties. If the leftmost party occupies m_{k+1} seats and the remaining k parties occupy m_k seats, the total number of occupied seats is $m = m_{k+1} + k_m$.

$$(4) \quad N^{(k+1)}_{n,m} = N^{(k+1)}_{n,m_k+m_{k+1}} = N^{(k)}_{n-m_{k+1},m_k} + N^{(k)}_{n-m_{k+1}-1,m_k} + \cdots + N^{(k)}_{m_k,m_k}$$

$$= \sum_{i=0}^{n-m_{k+1}-m_k} N^{(k)}_{n-m_{k+1}-i,m_k} = \sum_{i=0}^{n-m_{k+1}-m_k} \binom{n-m_{k+1}-m_k-i+k}{n-m_{k+1}-m_k-i}$$

$$= \sum_{i=0}^{n-m} \binom{n-m-i+k}{n-m-i} = \sum_{j=0}^{n-m} \binom{k+j}{j} = \sum_{j=0}^{n-m} \binom{k+j}{k}$$

$$= \binom{n-m+k+1}{n-m}.$$

The last step follows from the sum of the first two terms

$$(5) \qquad \binom{k+0}{k} + \binom{k+1}{k} = \binom{k+2}{k+1},$$

and the generalization

$$(6) \qquad \binom{k+j}{k+1} + \binom{k+j}{k} = \binom{k+j+1}{k+1}.$$

If at least one unoccupied seat occurs between parties, this seat can be viewed as occurring to the right of each of the first $k-1$ parties and the number of vacant seats remaining is $n-m-k+1$. In this case the number of arrangements will be the number of ways of choosing $n-m-k+1$ things from $n-m+1$ things. The answer is $\binom{n-m+1}{n-m-k+1}$. This result also depends on the total number of occupied seats and the number of parties but not on the number in each party.

The cases considered by Wiggins [1] in which the parties are all of the same size are readily solved by application of the general formula. If there are k parties of size 2 and n seats with $1 \leqq k \leqq n/2$, the number of arrangements of occupied and unoccupied seats is $\binom{n-k}{k}$. This formula is derived immediately from the general formula where the number of unoccupied seats is $n-2k$ and the number of parties is k. For the case in which there are k parties each consisting of m individuals with $1 \leqq k \leqq n/m$, the number of arrangements is $\binom{n-(m-1)k}{k}$. This can be written down immediately by use of the general formula in which the number of parties is k and the number of unoccupied seats is $n-mk$. Since these two cases involve runs which are all of the same length and are indistinguishable as far as arrangements of occupied and unoccupied cells are concerned, the result is valid without specifying the order of the runs. In the general problem, the runs are presumed to be of varied length and the result is valid only for a specified order of the runs.

If the order of the parties is permitted to vary and a vacant seat need not occur between parties, the number of arrangements of occupied and unoccupied seats cannot be found readily unless the individual parties are identified. For example, if all possible orders of the parties are considered, the number of arrangements of occupied and unoccupied seats is reduced if two parties are adjacent without a vacant seat between them, and the number of arrangements is further reduced whenever the number in two or more adjacent parties without vacant seats between them is equal to the number of occupied seats in another group of such parties. However, as long as the linear order of the parties is specified, the general formulas above apply.

References

1. A. D. Wiggins, An occupancy problem involving placement of pairs of balls, this MAGAZINE, 45 (1972) 82–85.

2. William Feller, An Introduction to Probability Theory and Its Applications, vol. I, 3rd. ed., Wiley, New York, 1968.

AN INTEGER PROGRAMMING HANDICAP SYSTEM IN A "WRITE RING TOSSING GAME"

EUGENE F. SCHUSTER, The University of Texas at El Paso

1. The write ring tossing game. During my recent tour in the army, I worked as an operation research (OR) analyst with an OR team working mainly on Department of Defense level projects. One project we inherited was a large simulation model (called INTLOC, an acronym for INTerdiction of Lines of Communication) which was used to provide an empirical technique for evaluating the impact of various aerial interdiction strategies on the throughput capacity of a logistics network. INTLOC rapidly became a computer time gobbling monstrosity. A single INTLOC run on a 65K UNIVAC-1108 took 3–4 hours. We were forced to beg, borrow and steal computer time wherever and whenever we could find it. One place we did find it was at the Cameron Station Military Installation in Washington, D.C. There we

The cases considered by Wiggins [1] in which the parties are all of the same size are readily solved by application of the general formula. If there are k parties of size 2 and n seats with $1 \leqq k \leqq n/2$, the number of arrangements of occupied and unoccupied seats is $\binom{n-k}{k}$. This formula is derived immediately from the general formula where the number of unoccupied seats is $n-2k$ and the number of parties is k. For the case in which there are k parties each consisting of m individuals with $1 \leqq k \leqq n/m$, the number of arrangements is $\binom{n-(m-1)k}{k}$. This can be written down immediately by use of the general formula in which the number of parties is k and the number of unoccupied seats is $n-mk$. Since these two cases involve runs which are all of the same length and are indistinguishable as far as arrangements of occupied and unoccupied cells are concerned, the result is valid without specifying the order of the runs. In the general problem, the runs are presumed to be of varied length and the result is valid only for a specified order of the runs.

If the order of the parties is permitted to vary and a vacant seat need not occur between parties, the number of arrangements of occupied and unoccupied seats cannot be found readily unless the individual parties are identified. For example, if all possible orders of the parties are considered, the number of arrangements of occupied and unoccupied seats is reduced if two parties are adjacent without a vacant seat between them, and the number of arrangements is further reduced whenever the number in two or more adjacent parties without vacant seats between them is equal to the number of occupied seats in another group of such parties. However, as long as the linear order of the parties is specified, the general formulas above apply.

References

1. A. D. Wiggins, An occupancy problem involving placement of pairs of balls, this MAGAZINE, 45 (1972) 82–85.

2. William Feller, An Introduction to Probability Theory and Its Applications, vol. I, 3rd. ed., Wiley, New York, 1968.

AN INTEGER PROGRAMMING HANDICAP SYSTEM IN A "WRITE RING TOSSING GAME"

EUGENE F. SCHUSTER, The University of Texas at El Paso

1. The write ring tossing game. During my recent tour in the army, I worked as an operation research (OR) analyst with an OR team working mainly on Department of Defense level projects. One project we inherited was a large simulation model (called INTLOC, an acronym for INTerdiction of Lines of Communication) which was used to provide an empirical technique for evaluating the impact of various aerial interdiction strategies on the throughput capacity of a logistics network. INTLOC rapidly became a computer time gobbling monstrosity. A single INTLOC run on a 65K UNIVAC-1108 took 3–4 hours. We were forced to beg, borrow and steal computer time wherever and whenever we could find it. One place we did find it was at the Cameron Station Military Installation in Washington, D.C. There we got a chance to put in a little army overtime making a run or two of INTLOC each Sunday night. Working this shift was not the most exciting thing in the world. One would carry in the large box of input cards and data tapes, start the run and wait four hours till the run was over. A quick analysis of the output of the first run would usually spot the input card which was out of order (or the keypunch error) and the second run would begin. It was here in the Univac computer room at Cameron Station in late October of 1969 that I witnessed my first "write ring tossing game." A write ring is a plastic ring that fits into a round groove molded in a magnetic tape reel. When the ring is in place, either reading or writing can occur on the tape during a computer run. When the ring is removed, writing is suppressed and only reading can take place; thus the file is protected from accidental erasure. Now the computer room had an ample supply of these write rings and in order to pass the time between INTLOC runs, my two army buddies devised the following game. Each player would begin with n write rings. A flip of the coin would determine which player began. The first player would toss his n rings at the hose arm of the fire extinguisher on the wall, trying to score as many ringers are possible. When the first player had completed his n tosses, the second player would begin. And so the game would proceed in stages, at each stage a player would retire as many rings as he had made on that stage. The first player to make n ringers was the winner.

After a number of Sunday evening games it became evident that my two buddies, A and B, were not evenly matched. Player A made ringers on about one tenth of his tosses and player B made ringers on about two tenths of his tosses. I was asked to determine a handicapping system which would make this a fair game, that is, to determine how many rings each player should begin with in order that each player had an equal chance of winning. The only answer I could give them at the time was to simulate the game and empirically determine the solution. Of course I was not satisfied with that solution and I have thought about the game from time to time since then. Shortly after I got out of the army I realized I could consider the play of each player as a Markov chain. The transition matrix of this Markov chain has some rather interesting properties which I will now take up.

2. The stochastic analysis of one player's play. Suppose the player begins with r rings. Let the random variable X_k record the number of rings the player has remaining after k stages of the game. Then (assuming that the player's ability does not change from toss to toss) the probability that j rings remain after $k+1$ stages given that i rings remain after k stages is

$$(2.1)\qquad P(X_{k+1} = j \mid X_k = i) = p_{ij} = \begin{cases} \binom{i}{j} q^j(1-q)^{i-j} & \text{if } 0 \leqq j \leqq i \leqq r \\ 0 & \text{otherwise,} \end{cases}$$

where q is the probability of failure on a single toss of the ring. It is then evident that we have a sequence of trials $X_1, X_2, \cdots$ whose outcomes satisfy the following two properties:

(i) Each outcome of each trial belongs to a finite set of outcomes $\{a_0, a_1, a_2, \cdots, a_r\}$

called the state space. If the outcome of the kth trial is i, then i rings remain and we say that the system is in the state a_i at time k.

(ii) The outcome of any trial depends at most upon the outcome of the immediately preceding trial and not upon any other previous outcome; with each pair of states (a_i, a_j) there is the probability p_{ij} that a_j occurs immediately after a_i occurs.

Hence, the sequence $X_1, X_2, \cdots$ is a stochastic process called a (time-homogeneous finite) Markov chain. The matrix $P = (p_{ij})$ is called the transition matrix. Thus, for each i, j and n, the probability that the system changes from state a_i to state a_j in n steps, denoted $p_{ij}^{(n)}$, is given by the n-step transition matrix $P^{(n)} = (p_{ij}^{(n)})$ where $P^{(n)} = P^n$. In our particular case it is interesting that our triangular n-step transition matrix $P^{(n)}$ can be written in closed form. This form is given in the following theorem, the proof of which follows by inducting on n.

THEOREM 1. *Let $P = (p_{ij})$ where p_{ij} is given by* (2.1). *Then $P^n = P^{(n)} = (p_{ij}^{(n)})$ where*

$$p_{ij}^{(n)} = \begin{cases} \binom{i}{j}(q^n)^j(1-q^n)^{i-j} & \text{when } 0 \leqq j \leqq i \leqq r \\ 0 & \text{otherwise.} \end{cases}$$

Note that $P^{(n)}$ is a lower triangular matrix in which the nonzero elements in the ith row are the terms in the binomial expansion of $\{(1-q^n)+q^n\}^i$.

Most questions regarding the play of one player can be answered using the above transition matrix. For example, let the random variable N record the stage at which the player tosses his rth ringer. Since $p_{r,0}^{(n)}$ is the probability that no rings will remain after n stages, the density of N is given by

$$P(N = n) = p_{r,0}^{(n)} - p_{r\,0}^{(n-1)} = (1-q^n)^r - (1-q^{n-1})^r. \tag{2.2}$$

The expected number of stages is given in

THEOREM 2. $E(N) = \sum_{i=1}^{r} \binom{r}{i}(-1)^i(q^i-1)^{-1}$.

Proof. An outline of the proof is as follows:

$$\begin{aligned} E(N) &= \sum_{n=1}^{\infty} n\{(1-q^n)^r - (1-q^{n-1})^r\} \\ &= \sum_{n=1}^{\infty} n \sum_{i=1}^{r} \binom{r}{i}(-1)^i(q^{ni}-q^{ni-i}) \\ &= \sum_{i=1}^{r} \binom{r}{i}(-1)^i \sum_{n=1}^{\infty} n(q^{ni}-q^{ni-i}) \\ &= \sum_{i=1}^{r} \binom{r}{i}(-1)^i\{(q^i)/(1-q^i)^2 - 1/(1-q^i)^2\} \\ &= \sum_{i=1}^{r} \binom{r}{i}(-1)^i(q^i-1)^{-1}. \end{aligned}$$

3. Theoretical fair game. Let $a > 0$ and $b > 0$ be the probabilities that player A and B, respectively, fail to make a ringer on a single toss. Suppose A starts with r rings and B starts with s rings. If AW and BW denote the events "A wins" and "B wins" then we have:

THEOREM 3. $f(r,s;a,b) = P(AW) - P(BW) = \sum_{n=2}^{\infty}\{(1-a^{n-1})^r(1-b^n)^s - (1-a^n)^r(1-b^{n-1})^s\} = 1 + \sum_{i=1}^{r} \sum_{j=0}^{s}\binom{r}{i}\binom{s}{j}(-1)^{i+j}b^j(1-a^{2i})/(1-a^ib^j)$.

Proof. Let AS, BS and $AW(n)$ be the events "A starts," "B starts," and "A wins at the nth stage," respectively. Then

$$\begin{aligned} P(AW) &= P(AW \cap AS) + P(AW \cap BS) \\ &= \sum_{n=1}^{\infty}\{P(AW(n) \cap AS) + P(AW(n) \cap BS)\}. \end{aligned}$$

Let $\{X_n\}$ and $\{Y_n\}$ be the Markov chains which record the number of rings remaining after each stage by player A and B respectively. Then using (2.2) we have

$$\begin{aligned} P(AW(n) \cap AS) &= P(X_n = 0, X_{n-1} > 0,\ Y_{n-1} > 0, AS) \\ &= \{(1-a^n)^r - (1-a^{n-1})^r\}\{1-(1-b^{n-1})^s\}(1/2) \end{aligned}$$

and

$$\begin{aligned} P(AW(n) \cap BS) &= P(X_n = 0, X_{n-1} > 0,\ Y_n > 0, BS) \\ &= \{(1-a^n)^r - (1-a^{n-1})^r\}\{1-(1-b^n)^s\}(1/2). \end{aligned}$$

Thus one can easily show that

$$\begin{aligned} P(AW) - P(BW) &= 2P(AW) - 1 \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} \{(1-a^{n-1})^r(1-b^n)^s - (1-a^n)^r(1-b^{n-1})^s\} \\ &= \lim_{k\to\infty} \left\{ \sum_{n=1}^{k} (1-a^{n-1})^r(1-b^n)^s - \sum_{n=1}^{k-1} (1-a^{n+1})^r(1-b^n)^s \right\} \\ &= \lim_{k\to\infty} \left\{ \sum_{n=1}^{k} (1-b^n)^s[(1-a^{n-1})^r - (1-a^{n+1})^r] + (1-a^{k-1})^r(1-b^k)^s \right\} \\ &= 1 + \lim_{k\to\infty} \sum_{n=1}^{k} (1-b^n)^s \sum_{i=1}^{r} \binom{r}{i}(-1)^i(a^{ni-i} - a^{ni+i}) \\ &= 1 + \lim_{k\to\infty} \sum_{i=1}^{r} \binom{r}{i}(-1)^i \sum_{n=1}^{k} (1-b^n)^s(a^{ni-i} - a^{ni+i}) \\ &= 1 + \lim_{k\to\infty} \sum_{i=1}^{r} \binom{r}{i}(-1)^i \sum_{n=1}^{k} \sum_{j=0}^{s} \binom{s}{j}(-1)^j b^{nj}(a^{ni-i} - a^{ni+i}) \\ &= 1 + \lim_{k\to\infty} \sum_{i=1}^{r} \binom{r}{i}(-1)^i \sum_{j=0}^{s} \binom{s}{j}(-1)^j \sum_{n=1}^{k} b^{nj}(a^{ni-i} - a^{ni+i}) \end{aligned}$$

$$= 1 + \sum_{i=1}^{r} \sum_{j=0}^{s} \binom{r}{i} \binom{s}{j} (-1)^{i+j} \left\{ \frac{b^j}{1-a^i b^j} - \frac{b^j a^{2i}}{1-a^i b^j} \right\}$$

$$= 1 + \sum_{i=1}^{r} \sum_{j=0}^{s} \binom{r}{i} \binom{s}{j} (-1)^{i+j} b^j (1-a^{2i})/(1-a^i b^j).$$

In view of Theorem 3, the game is fair if and only if $f(r,s;a,b) = 0$, i.e., if and only if

$$1 + \sum_{i=1}^{r} \sum_{j=0}^{s} \binom{r}{i} \binom{s}{j} (-1)^{i+j} b^j (1-a^{2i})/(1-a^i b^j) = 0.$$

So far I have not been able to find any exact nontrivial solutions (when $a \neq b$) of this equation. However, in any practical problem there would be a limit, say N, on the total number of rings available to the two players, i.e., $r + s \leqq N$. In this case one might be interested in an efficient solution to the following problems:

(i) Find the fairest game, i.e., find an (r,s) which minimizes $|f(r,s;a,b)|$ subject to $r + s \leqq N$.

(ii) Find an ε-fair game, i.e., given $\varepsilon > 0$, find an (r,s) with $|f(r,s;a,b)| \leqq \varepsilon$.

4. An integer programming solution. The solution to problem (i) is an integer pair contained in the triangle $T_N = \{(r,s)\colon r \geqq 1, s \geqq 1, r + s \leqq N, r \text{ and } s \text{ integers}\}$. Let P_N be the oriented (integer) path in T_N which begins at $(1,1)$ and moves in T_N according to the following recursive rule: At the integer pair (x,y), compute $f = f(x,y;a,b)$. If $f = 0$ or $x + y = N$, then the path ends at (x,y). If $f > 0$ step right one unit to $(x+1,y)$, otherwise step up one unit to $(x, y+1)$. The following Theorems 4 and 5 can then be used to produce algorithms which give efficient solutions to problems (i) and (ii). A numerical illustration is given in Section 5.

THEOREM 4. *The minimum of* $|f(r,s;a,b)|$ *in* T_N *is attained on* P_N *and is not attained in* $T_N - P_N$.

THEOREM 5. *For any* $\varepsilon > 0$ *there exists a positive integer* N *such that* $|f| \leqq \varepsilon$ *at the endpoint of the path* P_N *and* $|f| > \varepsilon$ *for all integer pairs in* T_{N-1}.

We will need the following Lemma 6 in the proof of Theorem 4 and Lemmas 7 and 8 in the proof of Theorem 5. For simplicity let $P = \lim_{N\to\infty} P_N$, $f(r,s) = f(r,s;a,b)$, and let Z^+ denote the set of positive integers. Then

LEMMA 6. *For fixed* a *and* b, $0 < a$, $b < 1$, *the following hold*:

(i) *For fixed* $r \in Z^+$, $f(r,s)$ *is a strictly increasing function of* s *with* $\lim_{s\to\infty} f(r,s) = 1$.

(ii) *For fixed* $s \in Z^+$, $f(r,s)$ *is a strictly decreasing function of* r *with* $\lim_{r\to\infty} f(r,s) = -1$.

Proof. Although one can prove this lemma by working directly with the expressions for the function f as given in Theorem 3, we shall not do so here. Instead, we give the intuitive argument which suggests the lemma and gives some insight into our algorithm. If r is fixed and we increase s, we are giving player B additional rings, thereby increasing player A's chance of winning. Hence, $f(r,s)$ is an increasing function of s. If B has an infinite number of rings then A is certain to win so that $\lim_{s\to\infty} f(r,s)$ should equal 1 for each r, so that (i) should hold. In a similar fashion one can argue that (ii) should hold. Note that f will be positive above P and negative below.

Proof of Theorem 4. The case $N = 1$ is trivial. Suppose then that the theorem holds for $N = 1,2,\cdots,k$ and let (r_{k+1}, s_{k+1}) be the end point of P_{k+1}. Let us first consider the case $f(r_{k+1}, s_{k+1}) > 0$. By the induction hypothesis it suffices to show that the minimum of $|f|$ in T_{k+1} is not attained in $\{(r,s): r+s=k+1\} - \{(r_{k+1}, s_{k+1})\}$. Lemma 6 can be used to show that f is decreasing along the diagonal $r + s = k+1$ so that $0 < f(r_{k+1}, s_{k+1}) < f(r, k+1-r)$ for $r = 1,2,\cdots, r_{k+1}-1$. We claim that $f(r_{k+1}+1, s_{k+1}-1) < 0$. Suppose not. If it were positive then Lemma 6 indicates that $f(r, s_{k+1}-1) > 0$ for $r = 1,2,\cdots, r_{k+1}+1$. But this would say that the path P_{k+1} cannot reach (r_{k+1}, s_{k+1}) (recall, + causes P to move to the right). Hence $f(r_{k+1}+1, s_{k+1}-1) < 0$ and since f is decreasing along the diagonal $f(r_{k+1}+s, s_{k+1}-s) < f(r_{k+1}+1, s_{k+1}-1) < 0$ for $s = 2,3,\cdots, s_{k+1}-1$. There must exist an $(r, s_{k+1}-1)$ on P_k satisfying $f(r_{k+1}+1, s_{k+1}-1) < f(r, s_{k+1}-1) < 0$. This follows from Lemma 6 and the fact that only a negative value of f can cause the path to move upward. Hence, in this case one can easily see that the minimum of $|f|$ is not attained in $T_{k+1} - P_{k+1}$. The case $f(r_{k+1}, s_{k+1}) < 0$ follows in a similar fashion. If $f(r_{k+1}, s_{k+1}) = 0$, P ends at (r_{k+1}, s_{k+1}) and Lemma 6 can be used to show that f can have at most one zero in T_{k+1}. Hence, the theorem holds in this case also.

LEMMA 7. *Fix a, b in $(0,1)$. If $f(r,s;a,b)$ does not vanish on $T = \lim_{N\to\infty} T_N$ then there is a sequence of integer pairs $\{(r_n, s_n)\}_{n=0}^{\infty}$ on P satisfying*:

(i) $f(r_n - 1, s_n) > 0$ *and* $f(r_n, s_n) < 0$, $n \geqq 1$.

(ii) $r_n \geqq r_{n-1} + 1$ *and* $s_n \geqq s_{n-1} + 1$, $n \geqq 1$.

Proof. Let $r_0 = s_0 = 1$. If $n = 1$, one can use Lemma 6 and the definition of the path P to show that there exists a point (r_1, s_1) satisfying (i) and (ii). Let us then assume that we have found (r_n, s_n) on P satisfying (i) and (ii) for each $n = 1,2,\cdots,k$. In this case, $f(r_k, s_k) < 0$. Using Lemma 6 one can conclude that there must exist an s^* with $f(r_k, s^*-1) < 0$ and $f(r_k, s^*) > 0$. Clearly (r_k, s^*) lies on P. Moving right along the path P (from (r_k, s^*)) we can then use Lemma 6 again to conclude that there must exist an r^* with $f(r^*-1, s^*) > 0$ and $f(r^*, s^*) < 0$. Then $(r_{k+1}, s_{k+1}) = (r^*, s^*)$ satisfies (i) and (ii) and the proof is complete.

LEMMA 8. *For each $\varepsilon > 0$ there exists an $R = R(\varepsilon, a)$ such that $r \geqq R$ implies that*

$$0 < f(r,s) - f(r+1,s) < \varepsilon.$$

Proof. Let $\varepsilon > 0$ be given. Then there exists a $K = K(\varepsilon, a)$ such that $0 < \sum_{n=K+1}^{\infty} a^n < \varepsilon/2$. Since $(1-a^K)^r$ tends to zero as r tends to infinity, there must

0.594						
0.585	0.277					
0.576	0.263	0.028				
0.566	0.249	0.011	−0.169			
0.555	0.233	−0.006	−0.187	−0.326		
0.544	0.216	−0.025	−0.206	−0.344	−0.451	
0.531	0.198	−0.045	−0.226	−0.363	−0.469	−0.552
0.517	0.178	−0.067	−0.247	−0.383	−0.488	−0.570
0.502	0.157	−0.090	−0.270	−0.405	−0.508	−0.588
0.485	0.133	−0.115	−0.294	−0.428	−0.529	−0.606
0.466	0.108	−0.142	−0.320	−0.452	−0.551	−0.626
0.445	0.080	−0.171	−0.349	−0.478	−0.574	−0.647
0.421	0.048	−0.203	−0.379	−0.505	−0.599	−0.669
0.394	0.013	−0.239	−0.412	−0.535	−0.625	−0.692
0.362	−0.027	−0.279	−0.448	−0.567	−0.653	−0.717
0.325	−0.073	−0.323	−0.489	−0.603	−0.684	−0.744
0.279	−0.127	−0.374	−0.533	−0.641	−0.717	−0.772
0.222	−0.191	−0.432	−0.583	−0.683	−0.753	−0.802
0.149	−0.269	−0.501	−0.641	−0.731	−0.792	−0.835
0.049	−0.367	−0.583	−0.707	−0.784	−0.835	−0.870
−0.101	−0.498	−0.684	−0.785	−0.845	−0.883	−0.909
−0.357	−0.686	−0.816	−0.880	−0.915	−0.937	−0.952

Fig. 1

exist an $R = R(\varepsilon, a)$ such that $r \geqq R$ implies that $0 < (1-a^K)^r < \varepsilon(1-a)/2$. Using Theorem 3, one can easily see that for $r \geqq R$

$$f(r,s) - f(r+1,s) = \sum_{n=2}^{\infty} \{a^{n-1}(1-a^{n-1})^r(1-b^n)^s - a^n(1-a^n)^r(1-b^{n-1})^s\}$$

$$= \sum_{n=1}^{\infty} a^n(1-a^n)^r\{(1-b^{n+1})^s - (1-b^{n-1})^s\}$$

$$\leqq (1-a^K)^r/(1-a) + \sum_{n=K+1}^{\infty} a^n$$

$$< \varepsilon$$

and the proof is complete.

Proof of Theorem 5. Let $\varepsilon > 0$ be given. We then claim that there must exist an (r,s) on P with $|f(r,s)| < \varepsilon$. If f has a zero then Theorem 4 implies that one zero must lie on the path P. Suppose then that f has no zeroes in $T = \lim_{N\to\infty} T_N$ and let R be as in Lemma 8. Lemma 7 then indicates that we can find an (r,s) on P such that $r \geqq R$, $f(r-1,s) > 0$ and $f(r,s) < 0$. Since $r \geqq R$ Lemma 8 indicates that $0 < f(r-1,s) - f(r,s) < \varepsilon$. But then $-f(r,s) = |f(r,s)| < \varepsilon$, i.e., there exists a point (r,s) on P with $|f(r,s)| < \varepsilon$. Let (r^*, s^*) be the first point on the path P for which $|f(r,s)| < \varepsilon$ and let $N = r^* + s^*$. Then, using Theorem 4, one can easily see that N satisfies the conditions of our present theorem.

5. Numerical example. We now give the handicaps in the original game of Section 1 when $a = 0.9$ and $b = 0.8$ (i.e., players A and B fail on 0.9 and 0.8 of their tosses, respectively) subject to the restriction that the rotal number of rings is at most 23. In this case the fairest game occurs when $r = 3$ and $s = 18$ with $f(3,18;a,b) \cong -0.006$. The three digit approximations to the function $f(r,s;a,b)$ at the points of the triangle T_{23} where $r \leqq 7$ are given in the table of Figure 1. The numbers in boxes correspond to the points (r,s) on the path P_{23}. The horizontal axis is the r axis. Note the f is positive above the path P_{23} and negative below.

6. Other applications. In reliability theory an active redundant configuration is a system of, say r, components connected in parallel as shown in the block diagram of Figure 2. Whenever the system is activated the components are all simultaneously subjected to operation unless they have previously failed. The system functions as long as k or more of the components function.

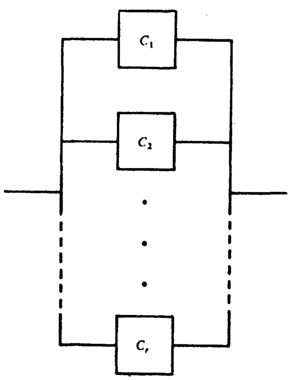

FIG. 2.

Suppose the time to failure of each of the components is exponentially distributed with mean $1/\lambda$ and suppose the system operates for exactly T hours each day. Now, a component whose lifetime follows the exponential distribution shows no aging, i.e., the probability that the component survives day $n+1$ given that it survived day n is the same as the probability that a new component will survive day one. Hence the stochastic analysis of this (active redundant) system is equivalent to the analysis in Section 2 providing we identify

(i) the number of rings r with the number of components, and

(ii) the (constant) probability of failure q on a single toss of a ring with the (constant) probability $e^{-\lambda T}$ that a component will survive day one ($= T$ hours)

The play in the ring tossing game can then be identified with the comparison between competing systems built from two different types of components.

ON APPLICATIONS OF VAN DER WAERDEN'S THEOREM

JOHN R. RABUNG, Randolph-Macon College

1. Equivalent versions of the theorem. In [1] B. L. van der Waerden relates how Artin, Schreier, and he were able to find the proof of the following conjecture of Baudet:

(A) *If the set of positive integers is partitioned in any way into two classes, then for any positive integer l at least one class contains a set of l consecutive members of an arithmetic progression. (Henceforth we shall use the phrase "arithmetic progression of length l" to mean a set of l consecutive members of an arithmetic progression.)*

Aside from the ingenuity of the proof which finally arose, one of the most intriguing aspects of the paper is the manner in which these men were able to manipulate Baudet's conjecture into more manageable, yet equivalent forms. The first such manipulative step was to consider the following statement suggested by Schreier:

(B) *For any positive integer l there exists a positive integer $N(l)$ such that if the set $\{1,2,\cdots,N(l)\}$ is partitioned into two classes, then at least one class contains an arithmetic progression of length l.*

This so-called "finite version" of (A) clearly implies (A) and the converse implication is shown in [1] using a Cantor diagonal approach.

From here it was an easy step for Artin to show that (B) is equivalent to:

(C) *For any positive integers k and l there exists a positive integer $N(k,l)$ such that if the set $\{1,2,\cdots,N(k,l)\}$ is partitioned into k classes, some class contains an arithmetic progression of length l.*

This is the statement which van der Waerden proved. (See [1] or [2].)

Since the appearance of the proof several applications of the theorem have been published. (We shall discuss A. Brauer's application to power residues in another section of this paper.) However, it has become apparent that the real potential for

Suppose the time to failure of each of the components is exponentially distributed with mean $1/\lambda$ and suppose the system operates for exactly T hours each day. Now, a component whose lifetime follows the exponential distribution shows no aging, i.e., the probability that the component survives day $n+1$ given that it survived day n is the same as the probability that a new component will survive day one. Hence the stochastic analysis of this (active redundant) system is equivalent to the analysis in Section 2 providing we identify

(i) the number of rings r with the number of components, and

(ii) the (constant) probability of failure q on a single toss of a ring with the (constant) probability $e^{-\lambda T}$ that a component will survive day one ($=T$ hours)

The play in the ring tossing game can then be identified with the comparison between competing systems built from two different types of components.

ON APPLICATIONS OF VAN DER WAERDEN'S THEOREM

JOHN R. RABUNG, Randolph-Macon College

1. Equivalent versions of the theorem. In [1] B. L. van der Waerden relates how Artin, Schreier, and he were able to find the proof of the following conjecture of Baudet:

(A) *If the set of positive integers is partitioned in any way into two classes, then for any positive integer l at least one class contains a set of l consecutive members of an arithmetic progression. (Henceforth we shall use the phrase "arithmetic progression of length l" to mean a set of l consecutive members of an arithmetic progression.)*

Aside from the ingenuity of the proof which finally arose, one of the most intriguing aspects of the paper is the manner in which these men were able to manipulate Baudet's conjecture into more manageable, yet equivalent forms. The first such manipulative step was to consider the following statement suggested by Schreier:

(B) *For any positive integer l there exists a positive integer $N(l)$ such that if the set $\{1,2,\cdots,N(l)\}$ is partitioned into two classes, then at least one class contains an arithmetic progression of length l.*

This so-called "finite version" of (A) clearly implies (A) and the converse implication is shown in [1] using a Cantor diagonal approach.

From here it was an easy step for Artin to show that (B) is equivalent to:

(C) *For any positive integers k and l there exists a positive integer $N(k,l)$ such that if the set $\{1,2,\cdots,N(k,l)\}$ is partitioned into k classes, some class contains an arithmetic progression of length l.*

This is the statement which van der Waerden proved. (See [1] or [2].)

Since the appearance of the proof several applications of the theorem have been published. (We shall discuss A. Brauer's application to power residues in another section of this paper.) However, it has become apparent that the real potential for application of (C) may lie in the size of the numbers $N(k,l)$. In his proof of (C) van der Waerden constructs numbers $N(k,l)$ which suffice, but it is thought that these constructed numbers are terrifically loose. For example, van der Waerden constructs $N(2,3) = 67$ whereas any $N(2,3) \geqq 9$ will work. And just a glance at his general construction of these numbers suggests that their growth rate is much greater than it need be. As Erdös points out in [3], tightening of the numbers $N(k,l)$ may lead to settling the question of the existence of arbitrarily long strings of prime numbers which are consecutive members of some arithmetic progression.

But the refinement of the numbers $N(k,l)$ will not be achieved without an essentially new proof of (C). Because none have yet been found, one is led back to the reasoning of Artin, Schreier, and van der Waerden that perhaps another version of the statement would be more manageable. Several equivalent forms of (C) have appeared since the proof was published. In this section we present some other equivalent versions of (C). These versions seem more explicitly related to the problem of primes in arithmetic progression than does the statement of (C). We begin with:

(D) *Let $S = \{a_i\}_{i=1}^{\infty}$ be any strictly increasing sequence of positive integers. If there exists a positive integer M such that $a_{i+1} - a_i \leqq M$, for $i = 1,2,\cdots$, then there exist among the members of S arithmetic progressions of arbitrary length.*

This is quickly seen to be a consequence of (C), for consider the following partition of the set of positive integers into M classes:

$$\begin{aligned}
K_0 &= \{a_i\colon a_i \in S\} = S\\
K_1 &= \{a_i + 1\colon a_i \in S\} \cap K_0{}'\\
K_2 &= \{a_i + 2\colon a_i \in S\} \cap (K_0 \cup K_1)'\\
&\vdots \qquad\quad \vdots \qquad\quad \vdots\\
K_j &= \{a_i + j\colon a_i \in S\} \cap (\textstyle\bigcup_{h=1}^{j-1} K_h)'\\
&\vdots \qquad\quad \vdots \qquad\quad \vdots\\
K_{M-1} &= \{a_i + (M-1)\colon a_i \in S\} \cap (\textstyle\bigcup_{h=1}^{M-1} K_h)'.
\end{aligned}$$

It is clear from the nature of S and the construction of the classes, K_j, that this is indeed a partition of the set of positive integers into M classes. Now by (C) there must be an arithmetic progression of length l (arbitrary) in some class, say K_n. Suppose this arithmetic progression is $\{b, b+d, b+2d, \cdots, b+(l-1)d\}$. Since $b \in K_n$, we have $b = a + n$ for some $a \in S$. Similarly each member of this progression is n greater than some member of S. Thus, $\{a, a+d, a+2d, \cdots, a+(l-1)d\} \subset S$, and the result is obtained.

Now (D) immediately yields:

(E) *If the set of positive integers is partitioned into two classes, then at least one of the following holds*:

(1) *One class contains arbitrarily long strings of consecutive integers.*

(2) *Both classes contain arithmetic progressions of arbitrary length.*

This is clear since if (1) does not occur, then (D) applies in both classes. And since (E) readily implies the Baudet conjecture (A), we see that (D) and (E) are each equivalent to van der Waerden's theorem (C).

Now (D) and (E) both yield "finite versions":

(D′) *For any M and l there is a positive integer $N_d(M,l)$ such that any strictly increasing finite sequence $\{a_i\}_{i=1}^{m}$ of positive integers with differences bounded by M (i.e., $a_{i+1} - a_i \leqq M$) and with $a_m - a_1 \geqq N_d(M,l)$ will contain an arithmetic progression of length l.*

(E′) *For any M and l there is a positive integer $N_e(M,l)$ such that whenever the set $\{1,2,\cdots,N_e(M,l)\}$ is partitioned into two classes at least one of the following holds:*

(1) *One class contains M consecutive numbers.*

(2) *Both classes contain arithmetic progressions of length l.*

From either of the statements (D′) or (E′) one sees the connection between van der Waerden's theorem and the problem of primes forming consecutive members of an arithmetic progression. If one can sharpen the number $N_d(M,l)$ enough and observe a relationship between this number and the rate of growth of gaps between consecutive prime numbers, one may be able to settle the question.

We have not been able to generally sharpen an estimate for $N_d(M,l)$, but we have found best possible values of $N_d(M,l)$ for some values of M and l. These are presented in section 3 of this paper.

2. An application of Witt's generalization of van der Waerden's theorem. Very shortly after van der Waerden published his proof in [**2**], A. Brauer [**4**] showed, among other things, that for any positive integers k and l there is a number $Z(k,l)$ such that for any prime $p > Z$ with $p \equiv 1 \pmod{k}$ the reduced residue system $\{1,2,\cdots,p-1\}$ modulo p contains l consecutive numbers, each of which is a kth power residue modulo p. Several authors since have found uniform upper bounds on this string of consecutive kth power residues for fixed k and l. (See for example, [**5**]–[**10**].)

J. H. Jordan [**11**], without the assurance of a general theorem like Brauer's, stepped into the domain of Gaussian integers $\mathbb{Z}[i]$ and proceeded to find several uniform upper bounds for what he called "consecutive" Gaussian integers which are all kth power residues of a prime of sufficiently large norm. The question arises, then, as to whether there is a theorem like Brauer's for the Gaussian domain. In this section we show that there is such a theorem.

It is Ernst Witt's generalization [**12**] of the van der Waerden theorem which allows one to use an approach analogous to Brauer's in not only the Gaussian integers, but in some other domains as well. Witt's theorem may be stated as follows:

Let $\Gamma = \{\gamma_1,\gamma_2,\cdots,\gamma_l\}$ be a fixed set of Gaussian integers. For any positive integer k there is a positive integer $N(k,l)$ such that if the set

$$\Delta = \left\{\sum_{j=1}^{l} a_j\gamma_j : a_j \in Z,\ a_j \geqq 0,\ \sum_{j=1}^{l} a_j = N(k,l)\right\}$$

is partitioned into k classes, some class will contain a homothetic image, Γ', of Γ.

(Here we shall say that Γ' is homothetic to Γ if $\Gamma' = \lambda\Gamma + \alpha = \{\lambda\gamma + \alpha : \gamma \in \Gamma\}$ where λ is a positive integer and α is an arbitrary Gaussian integer.) Now if in the above statement $\gamma_J \in \Gamma$ is such that $|\gamma_J| \geqq |\gamma_i|$, $1 \leqq i \leqq l$, then we see

$$\left| \sum_{j=1}^{l} a_j\gamma_j \right| \leqq \sum_{j=1}^{l} (|a_j|\,|\gamma_j|) \leqq |\gamma_J| \sum_{j=1}^{l} |a_j| = |\gamma_J| N(k,l)$$

since $a_j \geqq 0$. Thus, we may also say that if the set of all Gaussian integers with norm not greater than $(|\gamma_J| N(k,l))^2$ is partitioned into k classes, then some class will contain a homothetic image of Γ. Let $N(k,\Gamma) = (\max\{|\gamma_j| : \gamma_j \in \Gamma\})N(k,l)$. Now we can prove:

THEOREM 1. *Given any finite set of Gaussian integers, say* $\Gamma = \{\gamma_1, \gamma_2, \cdots, \gamma_l\}$, *and any sufficiently large Gaussian prime,* π, *there is a set* $P = \{\rho_1, \rho_2, \cdots, \rho_l\}$ *of quadratic residues modulo* π *such that* P *is a translation of* Γ.

Proof. Let $D(\Gamma) = \max\{|\gamma_j - \gamma_i| : \gamma_i, \gamma_j \in \Gamma\}$ be called the diameter of Γ. By choosing our prime π with large enough norm we can imbed any set of Gaussian integers having finite diameter in a reduced residue system modulo π. (See Hardman and Jordan [13].) Hence, we see that for π with sufficiently large norm there is either a translated image of Γ close to the origin consisting entirely of quadratic residues or there is a quadratic nonresidue, ν, modulo π such that $1 \leqq |\nu| \leqq D(\Gamma) + 1$. Let R represent any finite array of Gaussian integers which is large enough to contain some translation of Γ and $\nu\Gamma = \{\nu\gamma_j : \gamma_j \in \Gamma\}$.

Now consider π to be a Gaussian prime such that $|\pi| > 2N(2,R) + 1$, thus assuring that a reduced residue system (of the Hardman-Jordan type) modulo π, when broken into two classes will contain a homothetic image of R in one of the classes. In particular, either the class of quadratic residues or the class of quadratic nonresidues will contain such an image, say $R' = \lambda R + \alpha$ for some nonnegative integer λ and α a Gaussian integer. Now if λ and the elements of R' have the same quadratic nature modulo π, then multiplication by λ^{-1} modulo π yields $R'' = R + \lambda^{-1}\alpha$ consisting entirely of quadratic residues modulo π.

If λ and the elements of R' are of differing quadratic nature modulo π, then $R'' = R + \lambda^{-1}\alpha$ consists of quadratic nonresidues. But now $R + \lambda^{-1}\alpha$ contains as a subset some translation of $\nu\Gamma$, say $\nu\Gamma + \alpha'$. Since ν is a quadratic nonresidue, $\Gamma + \nu^{-1}\alpha'$ is made up of quadratic residues modulo π, and we are done.

Following Brauer's argument in similar fashion, one can establish

THEOREM 2. *Given any finite set* Γ *of Gaussian integers and any positive integer* k, *there exists a translation of* Γ *consisting entirely of kth power residues modulo any sufficiently large Gaussian prime* π *with* $N(\pi) \equiv 1 \pmod{k}$.

In the interest of moving toward applications of Witt's Theorem similar to those mentioned in the first section of this paper, we note that this theorem has the following equivalent versions analogous to (D) and (E) of the preceding section:

(A) *If* $\Sigma = \{\sigma_i\}_{i=1}^{\infty}$ *is any sequence of Gaussian integers for which there exists a finite set* $H = \{\eta_1, \eta_2, \cdots, \eta_M\}$ *of Gaussian integers such that*

$$\mathbf{Z}[i] \subset \Sigma \cup (\Sigma + \eta_1) \cup (\Sigma + \eta_2) \cup \cdots \cup (\Sigma + \eta_M)$$

then for any $\Gamma = \{\gamma_1, \gamma_2, \cdots, \gamma_l\} \subset \mathbf{Z}[i]$, Σ *contains a homothetic image of* Γ.

This statement is proved in the same way as statement (D) of the preceding section was proved. And from (A) we get:

(B) *For any partition of the set of Gaussian integers into two classes one of the following must occur:*

(1) *One class contains a translation of any finite set of Gaussian integers.*

(2) *Both classes contain a homothetic image of any finite set of Gaussian integers.*

To establish (B) one needs only to observe that in such a partition if we view each class as a sequence of Gaussian integers, either the condition of (A) will apply to both sequences (and, hence, (2) holds) or that condition will fail to hold for one of the sequences. In the latter case the class K where the condition fails will have arbitrarily large "holes" in it in the sense that for any positive integer M we could find $\xi \in \mathbf{Z}[i]$ such that

$$\{\zeta \in \mathbf{Z}[i] : |\zeta - \xi| \leqq M\} \cap K = \varnothing.$$

(For, if not, we could use $H = \{\zeta \in \mathbf{Z}[i] : |\zeta| \leqq M\}$ in the statement of (A).)

So (B), like (E) of the preceding section, brings forth the question of "patterns" of prime Gaussian integers. Of course, just as there are arbitrarily large gaps in the set of rational primes, so there are arbitrarily large holes in the set of Gaussian primes. But still one is led to study a "finite version" of (B) as weighed against the growth rate of holes in the set of Gaussian primes. Here, as in the rational case, such studies have so far yielded little fruit because of the unwieldy size of the constants involved in finite versions of Witt's Theorem. This is not unexpected since Witt's proof is essentially the same as that of van der Waerden in the rational case.

3. Some numerical results. Let $\{a_i\}_{i=1}^m$ be a strictly increasing sequence of positive integers such that for some fixed positive integer M we have $a_{i+1} - a_i \leqq M$ for $i = 1, 2, \cdots, m-1$. From statement (D′) of the first section of this paper we know of the existence of a number $N(M,l)$ such that if $a_m - a_1 \geqq N(M,l)$, then among the members of the given sequence there is an arithmetic progression of length l. We direct our attention to the number $N(M,l)$. Clearly, once such a number is found, any larger number would serve the same purpose. Let $N^*(M,l) = \min\{N(M,l)\}$. Under this definition, displaying a value of $N^*(M,l)$ for some M and l implies the existence of a sequence $\{a_i\}_{i=1}^m$ with differences between successive members bounded by M such that $a_m - a_1 = N^*(M,l) - 1$, and such that the sequence contains no arithmetic progression of length l. In presenting our numerical results we shall also present such sequences which show our constants to be correct. Actually we shall give the sequence of differences associated with the original sequence; that is, if $\{a_i\}_{i=1}^m$ is a sequence to be displayed, we shall instead display the sequence $\{d_i\}_{i=1}^{m-1}$ where $d_i = a_{i+1} - a_i$ for $i = 1, 2, \cdots, m-1$. We shall also impose on our sequences the condition that $d_i + d_{i+1} > M$ for $i = 1, 2, \cdots, m-2$. One easily sees that this condition in no way alters the generality in computations of $N^*(M,l)$.

To give an easy example of how the computations were made, let us consider the calculation of $N^*(2,3)$. That is, we consider all sequences of differences d_i with each $d_i = 1$ or 2 and with $d_i + d_{i+1} > 2$. Suppose $d_1 = 1$. Then $d_2 = 2$ and $d_3 = 1$ or 2. If, however, $d_3 = 2 = d_2$, then two consecutive differences are alike which, of course, means three members of the original sequence are in arithmetic progression. So we consider the case when $d_3 = 1$. This means $d_4 = 2$, and again since $d_4 + d_3 = d_2 + d_1$, we have an arithmetic progression of length $l = 3$ in the original sequence. This exhausts all cases with $d_1 = 1$. Similar argument shows that with $d_1 = 2$ one gets the sequence of differences $\{2,1,2\}$ before exhausting all possibilities. A corresponding original sequence might be $\{1,3,4,6\}$. This is, in one sense, the longest such sequence with no three terms in arithmetic progression. Since here, in the notation of the preceding paragraph, $a_m - a_1 = 5$, we get $N^*(2,3) = 6$.

The following table displays some values of $N^*(M,l)$ which we have found using essentially the above technique and the CDC 3800 computer at the Research Computation Center, Naval Research Laboratory, Washington, D.C.

$N^*(M,l)$	Sequences of differences $\{d_i\}_{i=1}^{m-1}$ of maximal length
$N^*(2,3) = 6$	$\{2,1,2\}$
$N^*(3,3) = 18$	$\{3,2,3,1,3,2,3\}$
$N^*(4,3) = 27$	$\{1,4,3,4,2,4,3,4,1\}$
$N^*(5,3) = 64$	$\{5,4,5,3,5,4,5,1,5,4,5,3,5,4,5\}$
$N^*(6,3) = 102$	$\{5,6,4,6,2,6,4,5,6,5,3,5,6,5,4,6,2,6,4,6,5\}$
$N^*(2,4) = 15$	$\{2,2,1,2,2,1,2,2\}$
$N^*(3,4) = 57$	$\{3,3,2,2,3,3,1,3,3,2,3,3,1,3,3,2,3,3,1,3,1,3,2\}$
$N^*(2,5) = 29$	$\{1,2,1,2,2,2,1,2,2,2,1,2,2,2,1,2,1\}$
$N^*(2,6) = 57$	$\{2,2,1,2,1,2,2,2,1,2,2,2,2,1,2,1,2,1,2,1,2,2,2,2,1,2,2,2,1,$ $2,1,2,2\}$

We also found the partial results $N^*(4,4) \geqq 160$ and $N^*(2,7) \geqq 193$.

References

1. B. L. van der Waerden, Wie der Beweis der Vermutung von Baudet gefunden wurde, Abh. Math. Sem. Univ. Hamburg, 28 (1965) 6–15.

2. ——— Beweis einer Baudet'schen Vermutung, Nieuw Arch. Wisk., 15 (1927) 212–216.

3. P. Erdos, Some recent advances and current problems in number theory, Lectures on Modern Mathematics, vol. 3, T. L. Saaty (editor), Wiley, New York, 1965, pp. 196–244.

4. A. Brauer, Uber Sequenzen von Potenzresten, S.-B. Preuss. Akad. Wiss., (1928) 9–16.

5. M. Dunton, Bounds for pairs of cubic residues, Proc. Amer. Math. Soc., 16 (1965) 330–332.

6. D. Lehmer, E. Lehmer, W. Mills, and J. Selfridge, Machine proof of a theorem on cubic residues, Math. Comp., 16 (1962) 407–415.

7. D. Lehmer and E. Lehmer, On runs of residues, Proc. Amer. Math. Soc., 13 (1962) 102–106.

8. D. Lehmer, E. Lehmer, and W. Mills, Pairs of consecutive power residues, Canad. J. Math., 15 (1963) 172–177.

9. W. Mills and R. Bierstedt, On the bound for a pair of consecutive quartic residues modulo a prime p, Proc. Amer. Math. Soc., 14 (1963) 628–632.
10. J. Rabung and J. Jordan, Consecutive power residues or nonresidues, Math. Comp., 24 (1970) 737–740.
11. J. Jordan, Consecutive residues or nonresidues in the Gaussian integers.
12. E. Witt, Ein kombinatorischer Satz der Elementargeometrie, Math. Nachr., 6 (1951) 261–262.
13. N. Hardman and J. Jordan, A minimum problem connected with complete residue systems in the Gaussian integers, Amer. Math. Monthly, 74 (1967) 559–561.

ELEMENTARY EVALUATION OF $\zeta(2n)$

BRUCE C. BERNDT, The Institute for Advanced Study and University of Illinois

1. Introduction. Let B_j denote the jth Bernoulli number (defined below in Section 2). Euler's formula

$$\zeta(2n) \equiv \sum_{k=1}^{\infty} k^{-2n} = \frac{(-1)^{n-1}(2\pi)^{2n}B_{2n}}{2(2n)!} \qquad (n \geqq 1) \tag{1.1}$$

is one of the most beautiful results of elementary analysis. Perhaps the three most common methods of proving (1.1) are by the use of the Fourier series for the Bernoulli polynomials [**4**, p. 524], by the use of the calculus of residues in conjunction with the Laurent expansion of $\cot x$ (given below) in terms of Bernoulli numbers [**10**, pp. 141–143], and by the method of Euler, described in [**1**], for example. T. M. Apostol [**1**] recently gave a proof of (1.1) that uses knowledge of symmetric functions and one of Newton's formulas. The idea for Apostol's proof can be found in the Yagloms' book [**14**, pp. 131–133], although they only establish (1.1) for $n = 1$ and $n = 2$. I. Skau and E. Selmer [**11**] use a similar method, but they do not explicitly evaluate $\zeta(2n)$ in terms of Bernoulli numbers.

Apostol's paper [**1**] contains a survey of "elementary" methods used to establish (1.1). An even more recent paper of E. L. Stark [**12**] contains a lengthy bibliography of papers on the evaluation of $\zeta(2)$ and $\zeta(2n)$. To the references cited in the two aforementioned papers, one can add the paper of R. Hovstad [**3**] and H. Rademacher's book [**9**, pp. 121–124].

In this paper, two new proofs of (1.1) are given. The proofs use only elementary calculus. The first proof, especially, is suitable for presentation in an ordinary calculus class.

2. Properties of Bernoulli numbers and polynomials. The Bernoulli polynomials $B_n(x)$, $0 \leqq n < \infty$, are defined by

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x)t^n/n! \qquad (|t| < 2\pi). \tag{2.1}$$

9. W. Mills and R. Bierstedt, On the bound for a pair of consecutive quartic residues modulo a prime p, Proc. Amer. Math. Soc., 14 (1963) 628–632.

10. J. Rabung and J. Jordan, Consecutive power residues or nonresidues, Math. Comp., 24 (1970) 737–740.

11. J. Jordan, Consecutive residues or nonresidues in the Gaussian integers.

12. E. Witt, Ein kombinatorischer Satz der Elementargeometrie, Math. Nachr., 6 (1951) 261–262.

13. N. Hardman and J. Jordan, A minimum problem connected with complete residue systems in the Gaussian integers, Amer. Math. Monthly, 74 (1967) 559–561.

ELEMENTARY EVALUATION OF $\zeta(2n)$

BRUCE C. BERNDT, The Institute for Advanced Study and University of Illinois

1. Introduction. Let B_j denote the jth Bernoulli number (defined below in Section 2). Euler's formula

$$\zeta(2n) \equiv \sum_{k=1}^{\infty} k^{-2n} = \frac{(-1)^{n-1}(2\pi)^{2n}B_{2n}}{2(2n)!} \qquad (n \geqq 1) \tag{1.1}$$

is one of the most beautiful results of elementary analysis. Perhaps the three most common methods of proving (1.1) are by the use of the Fourier series for the Bernoulli polynomials [**4**, p. 524], by the use of the calculus of residues in conjunction with the Laurent expansion of $\cot x$ (given below) in terms of Bernoulli numbers [**10**, pp. 141–143], and by the method of Euler, described in [**1**], for example. T. M. Apostol [**1**] recently gave a proof of (1.1) that uses knowledge of symmetric functions and one of Newton's formulas. The idea for Apostol's proof can be found in the Yagloms' book [**14**, pp. 131–133], although they only establish (1.1) for $n = 1$ and $n = 2$. I. Skau and E. Selmer [**11**] use a similar method, but they do not explicitly evaluate $\zeta(2n)$ in terms of Bernoulli numbers.

Apostol's paper [**1**] contains a survey of "elementary" methods used to establish (1.1). An even more recent paper of E. L. Stark [**12**] contains a lengthy bibliography of papers on the evaluation of $\zeta(2)$ and $\zeta(2n)$. To the references cited in the two aforementioned papers, one can add the paper of R. Hovstad [3] and H. Rademacher's book [**9**, pp. 121–124].

In this paper, two new proofs of (1.1) are given. The proofs use only elementary calculus. The first proof, especially, is suitable for presentation in an ordinary calculus class.

2. Properties of Bernoulli numbers and polynomials. The Bernoulli polynomials $B_n(x)$, $0 \leqq n < \infty$, are defined by

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x)t^n/n! \qquad (|t| < 2\pi). \tag{2.1}$$

The Bernoulli numbers B_n, $0 \leqq n < \infty$, are defined by $B_n = B_n(0)$. Formally differentiating both sides of (2.1) with respect to x, we easily deduce that

$$B_n'(x) = nB_{n-1}(x). \tag{2.2}$$

A more rigorous proof may be easily derived from the formula

$$B_n(x) = \sum_{k=0}^{n} \binom{n}{k} B_k x^{n-k},$$

which is easily proved by writing the left side of (2.1) as $\{t/(e^t - 1)\}e^{xt}$, expanding $t/(e^t - 1)$ and e^{xt} into power series about $t = 0$, and then equating coefficients of like powers of t on each side of (2.1).

If $n \geqq 1$,

$$B_{2n+1} = 0. \tag{2.3}$$

This is easily shown upon observing that the left side of

$$\frac{t}{e^t - 1} + \frac{1}{2}t = 1 + \sum_{n=2}^{\infty} B_n t^n/n!$$

is an even function of t. Similarly, since the left side of

$$\frac{te^{t/2}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(1/2)t^n/n!$$

is an even function of t, we deduce that for $n \geqq 0$

$$B_{2n+1}(1/2) = 0. \tag{2.4}$$

Setting $x = 0$ and replacing t by $-iy$ in (2.1), we find that for $|y| < 2\pi$

$$(y/2)\cot(y/2) = -iy/2 + \frac{-iy}{e^{-iy} - 1} \tag{2.5}$$

$$= -iy/2 + \sum_{n=0}^{\infty} B_n(-iy)^n/n! = \sum_{m=0}^{\infty} B_{2m}(-1)^m y^{2m}/(2m)!,$$

by (2.3).

3. First proof of Euler's formula. The germ for the idea behind this proof came from a problem in A. Ostrowski's text [7, p. 196]. We first need the following very weak version of the Riemann-Lebesgue lemma.

LEMMA 1. *Let $f(x)$ be twice continuously differentiable on $[0, a]$, $0 < a < \pi$, and suppose that $f(0) = 0$. Then*

$$\lim_{N\to\infty} \int_0^a f(x)\frac{\sin(Nx)}{\sin x}dx = 0.$$

Proof. Put $g(x) = f(x)/\sin x$ and integrate by parts to obtain

$$(3.1) \qquad \int_{0+}^{a} g(x)\sin(Nx)dx = -\frac{g(x)\cos(Nx)}{N}\Big|_{0+}^{a} + \frac{1}{N}\int_{0+}^{a} g'(x)\cos(Nx)dx.$$

By L'Hospital's rule, $\lim_{x\to 0} g(x) = f'(0)$. Thus, the integrated term on the right side of (3.1) tends to 0 as N tends to ∞. Also, by L'Hospital's rule, $\lim_{x\to 0} g'(x) = f''(0)/2$. Thus, the integrand on the right side of (3.1) is bounded on $[0, a]$. It follows that the last expression on the right side of (3.1) tends to 0 as N tends to ∞, and the proof is complete.

First proof of (1.1). Let k and n be positive integers. Integrating by parts twice with the aid of (2.2), we obtain, if $n > 1$,

$$\begin{aligned} I_n(k) &\equiv \int_0^{1/2} \{B_{2n}(x) - B_{2n}\}\cos(2\pi kx)dx \\ &= -\frac{2n}{2\pi k}\int_0^{1/2} B_{2n-1}(x)\sin(2\pi kx)dx \\ &= -\frac{2n(2n-1)}{(2\pi k)^2}\int_0^{1/2} B_{2n-2}(x)\cos(2\pi kx)dx, \end{aligned}$$

where in the last integration by parts we used (2.3) and (2.4). If we integrate by parts a total of $2n-1$ times, we find upon the continued use of (2.2), (2.3) and (2.4) that

$$(3.2) \qquad \begin{aligned} I_n(k) &= \frac{(-1)^n(2n)!}{(2\pi k)^{2n-1}}\int_0^{1/2} B_1(x)\sin(2\pi kx)dx \\ &= \frac{(-1)^n(2n)!}{(2\pi k)^{2n-1}}\left\{-\frac{B_1(x)\cos(2\pi kx)}{2\pi k}\Big|_0^{1/2}\right\} \\ &= \frac{(-1)^{n-1}(2n)!}{2(2\pi k)^{2n}}, \end{aligned}$$

by (2.4) and the fact that $B_1 = -1/2$.

Now sum both sides of (3.2) from $k = 1$ to $k = N$ to get

$$(3.3) \qquad \begin{aligned} \frac{(-1)^{n-1}(2n)!}{2(2\pi)^{2n}}\sum_{k=1}^{N} k^{-2n} &= \int_0^{1/2}\{B_{2n}(x) - B_{2n}\}\sum_{k=1}^{N}\cos(2\pi kx)dx \\ &= \int_0^{1/2}\{B_{2n}(x) - B_{2n}\}\left\{\frac{\sin\{(2N+1)\pi x\}}{2\sin(\pi x)} - \frac{1}{2}\right\}dx. \end{aligned}$$

(The given expression for the sum of $\cos(2\pi kx)$, $1 \leqq k \leqq N$, is easily proved by induction on N. Alternatively, the result can be obtained by summing the geometric series $e^{2\pi ikx}$, $0 \leqq k \leqq N$, and then taking the real part.) Now apply Lemma 1 with $f(\pi x) = \{B_{2n}(x) - B_{2n}\}/2$. Upon letting N tend to ∞ in (3.3) we obtain

$$(3.4)\qquad \frac{(-1)^{n-1}(2n)!}{2(2\pi)^{2n}}\zeta(2n) = -\frac{1}{2}\int_0^{1/2}\{B_{2n}(x) - B_{2n}\}dx$$

$$= -\frac{1}{2(2n+1)}\int_0^{1/2} B'_{2n+1}(x)dx + \frac{1}{4}B_{2n}$$

$$= B_{2n}/4,$$

where we have again employed (2.2), (2.3), and (2.4). Equation (3.4) is clearly equivalent to (1.1), and the proof is finished.

It is worthwhile to note that in the proof above we are, in fact, calculating the Fourier coefficients of $B_{2n}(x)$.

4. Second proof of Euler's formula. In our next proof we generalize the method of the Yagloms [**14**], Holme [**2**], Skau and Selmer [**11**], Papadimitriou [**8**], and Apostol [**1**]. However, we need to generalize only the case $n = 1$.

The simplest, most elementary proof of the next lemma is due to E. H. Neville [**5**]. We shall give a more detailed recapitulation of his proof.

LEMMA 2. *For real, nonintegral x, we have*

$$(4.1)\qquad \pi^2\csc^2(\pi x) = \sum_{k=-\infty}^{\infty}(k+x)^{-2}.$$

Proof. As in [**8**], it easily follows from DeMoivre's theorem that for each positive integer m,

$$\sin(2m+1)\theta = \sin^{2m+1}\theta P_m(\cot^2\theta),$$

where $P_m(y)$ is a polynomial of degree m with constant term $2m+1$. Put $\theta = \phi - \pi k/(2m+1)$, where k is an integer such that $-m \leqq k \leqq m$. It follows that

$$\sin^2(2m+1)\theta = \sin^2(2m+1)\phi = \{\sin^{2m+1}\phi P_m(\cot^2\phi)\}^2.$$

But $y^mP_m((1-y)/y)$ is a polynomial in y of degree m, and so $Q(y) \equiv y\{y^mP_m((1-y)/y)\}^2$ is a polynominal in y of degree $2m+1$. Hence, the equation above states that the polynomial

$$R(y) \equiv Q(y) - \sin^2(2m+1)\theta$$

has zeros at the points $y = \sin^2\phi = \sin^2\{\theta + \pi k/(2m+1)\}$, $-m \leqq k \leqq m$. The coefficient of y and the constant term of the polynomial $R(y)$ are $(2m+1)^2$ and $-\sin^2(2m+1)\theta$, respectively. Hence, if we let $\theta = \pi x/(2m+1)$ and form the sum of the reciprocals of the roots of $R(y)$, we deduce that

$$(4.2)\qquad \sum_{k=-m}^{m}\csc^2\frac{\pi(k+x)}{2m+1} = \sum_{k=-m}^{m}\cot^2\frac{\pi(k+x)}{2m+1} + (2m+1)$$

$$= (2m+1)^2\csc^2(\pi x).$$

Without loss of generality, we shall now assume $0 < |x| < 1/2$, since each side of (4.1) obviously has period 1.

Since $\sin y < y < \tan y$ for $0 < y < \pi/2$, we have

$$\cot^2 y < y^{-2} < 1 + \cot^2 y,$$

for $0 < |y| < \pi/2$. Let $y = (k+x)\pi/(2m+1)$ and sum on k, $-m \leqq k \leqq m$, (note that for each such y, $0 < |y| < \pi/2$) to obtain

$$\sum_{k=-m}^{m} \cot^2 \frac{(k+x)\pi}{2m+1} < \frac{(2m+1)^2}{\pi^2} \sum_{k=-m}^{m} (k+x)^{-2}$$

$$< 2m+1 + \sum_{k=-m}^{m} \cot^2 \frac{(k+x)\pi}{2m+1}.$$

Multiplying both sides by $\pi^2/(2m+1)^2$, using (4.2), and letting m tend to ∞, we deduce that

$$\sum_{k=-\infty}^{\infty} (k+x)^{-2} = \lim_{m\to\infty} \frac{\pi^2}{(2m+1)^2}\{4m^2\csc^2(\pi x) + (4m+1)\cot^2(\pi x) + 2m\}$$

$$= \pi^2\csc^2(\pi x).$$

To prove Euler's formula (1.1), we shall need the following recursion relation for Bernoulli numbers [**6**, p. 66].

LEMMA 3. *For* $n \geqq 2$,

$$-(2n-1)B_{2n} = \sum_{k=0}^{n} \binom{2n}{2k} B_{2k}B_{2n-2k}. \tag{4.3}$$

Proof. An easy proof can be given by induction, but we shall proceed directly. A slight rearrangement of (2.1) gives

$$1/(e^t-1) + 1/2 - t/4 = 1/t - t/6 + \sum_{n=2}^{\infty} B_{2n}t^{2n-1}/(2n)! \tag{4.4}$$

Differentiating both sides of (4.4), we obtain

$$-t^2\frac{d}{dt}\left\{\frac{1}{e^t-1} + \frac{1}{2} - \frac{t}{4}\right\} = -t^2\left\{\frac{-e^t}{(e^t-1)^2} - \frac{1}{4}\right\} \tag{4.5}$$

$$= \left\{\frac{t}{e^t-1} + \frac{t}{2}\right\}^2 = 1 + t^2/6 - \sum_{n=2}^{\infty}(2n-1)B_{2n}t^{2n}/(2n)!$$

On the other hand, from (2.1)

$$\left\{\frac{t}{e^t-1} + \frac{t}{2}\right\}^2 = \left\{\sum_{n=0}^{\infty} B_{2n}t^{2n}/(2n)!\right\}^2. \tag{4.6}$$

Comparing coefficients on the right sides of (4.5) and (4.6), we find that for $n \geqq 2$

$$-\frac{(2n-1)B_{2n}}{(2n)!}=\sum_{k=0}^{n}\frac{B_{2k}B_{2n-2k}}{(2k)!(2n-2k)!},$$

which is equivalent to (4.3).

Second proof of (1.1). For $0<|x|<1$, we find from (4.1) and (2.5) that

$$\begin{aligned}\sum_{\substack{k=-\infty\\k\neq 0}}^{\infty}(k+x)^{-2}&=\pi^2\csc^2(\pi x)-1/x^2=\pi^2\{1+\cot^2(\pi x)\}-1/x^2\\&=\pi^2+\frac{1}{x^2}\left\{\sum_{m=0}^{\infty}B_{2m}(-1)^m(2\pi x)^{2m}/(2m)!\right\}^2-1/x^2\\&=\pi^2+\frac{1}{x^2}\sum_{k=0}^{\infty}(-1)^k\left(\sum_{j=0}^{k}\frac{B_{2j}B_{2k-2j}}{(2j)!(2k-2j)!}\right)(2\pi x)^{2k}-1/x^2\\&=\pi^2-B_2(2\pi)^2-\sum_{k=2}^{\infty}(-1)^k(2\pi)^{2k}(2k-1)B_{2k}x^{2k-2}/(2k)!,\end{aligned}\tag{4.7}$$

where we have employed Lemma 3. On the interval $|x|\leqq 1/2$, the left side of (4.7) converges uniformly and thus represents a continuous function there. Hence, we may let x tend to 0 on both sides of (4.7) and find that $2\zeta(2)=\pi^2/3$, which is, of course, (1.1) with $n=1$.

Now differentiate both sides of (4.7) $2n-2$ times with respect to x, where x is not a nonzero integer and $n\geqq 2$. We then obtain

$$(2n-1)!\sum_{\substack{k=-\infty\\k\neq 0}}^{\infty}(k+x)^{-2n}=-\sum_{k=n}^{\infty}(-1)^k(2\pi)^{2k}B_{2k}\frac{x^{2k-2n}}{2k(2k-2n)!}.\tag{4.8}$$

Setting $x=0$ in (4.8), we deduce that for $n\geqq 2$

$$2(2n-1)!\zeta(2n)=(-1)^{n-1}(2\pi)^{2n}B_{2n}/2n,$$

which is equivalent to (1.1).

There exist several recursion formulae similar to (4.3) with products of two Bernoulli numbers on the right side. See Nielsen's book [**6**] for many such relations. From (1.1) a corresponding recursion relation for $\zeta(2n)$ may then be established. Without recourse to Bernoulli numbers, several authors have found such recursion relations for $\zeta(2n)$. One can then usually establish (1.1) by an easy induction.

5. Concluding remarks. Finally, another method for evaluating $\zeta(2n)$ can be based on the infinite product expansion for $\sin x$,

$$\frac{\sin x}{x}=\prod_{k=1}^{\infty}\left(1-\frac{x^2}{k^2\pi^2}\right)\qquad(-\infty<x<\infty).\tag{5.1}$$

Neville [5] has given a very simple, elementary proof of (5.1) that depends upon (4.2). This method of evaluating $\zeta(2n)$ essentially goes back to Euler. The determination of $\zeta(2)$ and $\zeta(4)$ from (5.1) is a problem in Titchmarsh's book [13, p. 34]. See also a paper of Zerr [15]. To find $\zeta(2n)$, raise each side of (5.1) to the nth power, expand each side into a Maclaurin series, and then equate coefficients of x^{2n}.

The author is very grateful to Tom M. Apostol for suggestions which substantially improved the quality of this paper.

Supported in part by National Science Foundation grant GP-36418X1.

References

1. Tom M. Apostol, Another elementary proof of Euler's formula for $\zeta(2n)$, Amer. Math. Monthly, 80 (1973) 425–431.

2. Finn Holme, En enkel beregning av $\sum_{k=1}^{\infty} 1/k^2$, Nordisk Mat. Tidskr., 18 (1970) 91–92.

3. Rolf Hovstad, The series $\sum_1^{\infty} 1/k^{2p}$, the area of the unit circle and Leibniz' formula, Nordisk Mat. Tidskr., 20 (1972) 92–98.

4. Konrad Knopp, Theory and Application of Infinite Series, Blackie, London, 1951.

→ E. H. Neville, A trigonometrical inequality, Proc. Cambridge Philos. Soc., 47 (1951) 629–632.

6. N. Nielsen, Traité Élémentaire des Nombres de Bernoulli, Gauthier-Villars, Paris, 1923.

7. A. Ostrowski, Vorlesungen über Differential-und Integralrechnung, Dritter Band, Verlag Birkhäuser, Basel, 1954.

8. Ioannis Papadimitriou, A simple proof of the formula $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$, Amer. Math. Monthly, 80 (1973) 424–425.

9. H. Rademacher, Topics in Analytic Number Theory, Springer-Verlag, Berlin, 1973.

10. Giovanni Sansone and Johan Gerretsen, Lectures on the Theory of Functions of a Complex Variable, vol. I, P. Noordhoff, Groningen, 1960.

11. Ivar Skau and Ernst S. Selmer, Noen anvendelser av Finn Holmes methode for beregning av $\sum_{k=1}^{\infty} 1/k^2$, Nordisk Mat. Tidskr., 19 (1971) 120–124.

→ E. L. Stark, The series $\sum_{k=1}^{\infty} k^{-s}, s = 2, 3, 4, \ldots$, once more, this MAGAZINE, 47 (1974) 197–202.

13. E. C. Titchmarsh, The Theory of Functions, 2nd ed., Oxford University Press, London, 1939.

14. A. M. Yaglom and I. M. Yaglom, Challenging Mathematical Problems with Elementary Solutions, vol. II, trans. by James McCawley, Jr., Holden Day, San Francisco, 1967.

15. G. B. M. Zerr, Summation of series, Amer. Math. Monthly, 5 (1898) 128–135.

INCONSISTENT AND INCOMPLETE LOGICS

JOHN GRANT, University of Florida, Gainesville

1. Introduction. In propositional logic every formula is either true or false but not both. In this paper we describe 3 logics of propositions and investigate some of their properties. Our first logic, the b-logic, allows inconsistencies in the sense that a formula may be both true and false. Our second logic, the n-logic, allows incompleteness in the sense that a formula may be neither true nor false. Finally, our g-logic allows both inconsistencies and incompleteness. Our approach is semantical throughout this paper.

Neville [5] has given a very simple, elementary proof of (5.1) that depends upon (4.2). This method of evaluating $\zeta(2n)$ essentially goes back to Euler. The determination of $\zeta(2)$ and $\zeta(4)$ from (5.1) is a problem in Titchmarsh's book [13, p. 34]. See also a paper of Zerr [15]. To find $\zeta(2n)$, raise each side of (5.1) to the nth power, expand each side into a Maclaurin series, and then equate coefficients of x^{2n}.

The author is very grateful to Tom M. Apostol for suggestions which substantially improved the quality of this paper.

Supported in part by National Science Foundation grant GP-36418X1.

References

1. Tom M. Apostol, Another elementary proof of Euler's formula for $\zeta(2n)$, Amer. Math. Monthly, 80 (1973) 425–431.

2. Finn Holme, En enkel beregning av $\sum_{k=1}^{\infty} 1/k^2$, Nordisk Mat. Tidskr., 18 (1970) 91–92.

3. Rolf Hovstad, The series $\sum_1^{\infty} 1/k^{2p}$, the area of the unit circle and Leibniz' formula, Nordisk Mat. Tidskr., 20 (1972) 92–98.

4. Konrad Knopp, Theory and Application of Infinite Series, Blackie, London, 1951.

5. E. H. Neville, A trigonometrical inequality, Proc. Cambridge Philos. Soc., 47 (1951) 629–632.

6. N. Nielsen, Traité Élémentaire des Nombres de Bernoulli, Gauthier-Villars, Paris, 1923.

7. A. Ostrowski, Vorlesungen über Differential-und Integralrechnung, Dritter Band, Verlag Birkhäuser, Basel, 1954.

8. Ioannis Papadimitriou, A simple proof of the formula $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$, Amer. Math. Monthly, 80 (1973) 424–425.

9. H. Rademacher, Topics in Analytic Number Theory, Springer-Verlag, Berlin, 1973.

10. Giovanni Sansone and Johan Gerretsen, Lectures on the Theory of Functions of a Complex Variable, vol. I, P. Noordhoff, Groningen, 1960.

11. Ivar Skau and Ernst S. Selmer, Noen anvendelser av Finn Holmes methode for beregning av $\sum_{k=1}^{\infty} 1/k^2$, Nordisk Mat. Tidskr., 19 (1971) 120–124.

12. E. L. Stark, The series $\sum_{k=1}^{\infty} k^{-s}, s = 2, 3, 4, \ldots$, once more, this MAGAZINE, 47 (1974) 197–202.

13. E. C. Titchmarsh, The Theory of Functions, 2nd ed., Oxford University Press, London, 1939.

14. A. M. Yaglom and I. M. Yaglom, Challenging Mathematical Problems with Elementary Solutions, vol. II, trans. by James McCawley, Jr., Holden Day, San Francisco, 1967.

15. G. B. M. Zerr, Summation of series, Amer. Math. Monthly, 5 (1898) 128–135.

INCONSISTENT AND INCOMPLETE LOGICS

JOHN GRANT, University of Florida, Gainesville

1. Introduction. In propositional logic every formula is either true or false but not both. In this paper we describe 3 logics of propositions and investigate some of their properties. Our first logic, the b-logic, allows inconsistencies in the sense that a formula may be both true and false. Our second logic, the n-logic, allows incompleteness in the sense that a formula may be neither true nor false. Finally, our g-logic allows both inconsistencies and incompleteness. Our approach is semantical throughout this paper.

Now we introduce our notations. They differ somewhat from the usual notations but are more appropriate for our presentation.

$P = \{p_1, p_2, p_3, \cdots\}$ is the set of propositional variables.

$L = \{p_1, \sim p_1, p_2, \sim p_2, \cdots\}$ is the set of literals.

F is the set of formulas. F is defined as the smallest class such that $L \subseteq F$ and if $A \in F$ and $B \in F$ then $(A \wedge B) \in F$ and $(A \vee B) \in F$.

We use only the connectives $\sim$ (not), $\wedge$ (and), and $\vee$ (or). We use the letters A and B for formulas.

Notice that we apply $\sim$ only to propositional variables. However, consider the usual definition of formula where negation is allowed to apply to arbitrary formulas. Any such formula may be transformed to an element of F by applying the transformations A for $\sim\sim A$, $\sim A \vee \sim B$ for $\sim(A \wedge B)$, and $\sim A \wedge \sim B$ for $\sim(A \vee B)$ as many times as necessary so that at the end $\sim$ is applied only to propositional variables. So when we write $\sim A$ for a formula we mean the element of F obtained from $\sim A$ by these transformations. When we assert A we mean that A is true, and when we assert $\sim A$ we mean that A is false.

To motivate the need for the logics discussed in this paper, we now give an example of a situation where propositional logic is not the appropriate logic to use. Let us suppose that P stands for a set of scientific statements whose truth or falsity is evaluated by 4 observers. The first observer performs various experiments and calculations on the basis of which he determines for each p_i whether it is true or false (exactly one of these alternatives being allowed), and writes p_i if true, $\sim p_i$ if false. We let O_1 be the set of literals written by the first observer. The second observer performs his own experiments and calculations on the basis of which he determines the truth or falsity of each p_i where again exactly one of these alternatives is allowed. We let O_2 be the set of literals written by the second observer. We may for example have the following,

$O_1 = \{p_1, \sim p_2, \sim p_3, p_4, p_5, \cdots\}$,

$O_2 = \{p_1, p_2, \sim p_3, p_4, \sim p_5, \cdots\}$.

The third and fourth observers are not able to conduct their own experiments and calculations. They are simply given O_1 and O_2. Let us assume that the third observer accepts all the conclusions of both the first and second observers. Then in our example

$$O_3 = \{p_1, p_2, \sim p_2, \sim p_3, p_4, p_5, \sim p_5, \cdots\}.$$

Note that p_2 is both true and false for the third observer, while p_1 is true but not false for him. Let us now assume that the fourth observer accepts the truth or falsity of a proposition only if both the first and second observers agree. Then in our example

$$O_4 = \{p_1, \sim p_3, p_4, \cdots\}.$$

Note that while p_3 is false for the fourth observer, p_2 is neither true nor false for him.

What logic should these observers use? In this paper we try to answer that question. While propositional logic suffices for the first 2 observers, it does not suffice for the last 2. The first 3 observers may use our b-logic which allows for inconsistencies, while the first, second, and fourth observers may use our n-logic which allows for incompleteness. Finally all 4 observers may use our generalized logic, the g-logic.

2. Propositional logic. In this section we review some concepts and results from propositional logic using our notations. For an exposition using standard notations see [3] Pages 9–13.

(1) A subset V of L is called an elementary valuation if for every pair $\{p_i, \sim p_i\}$ exactly one member of the pair is a member of V.

We think of the members of V as the true literals of the elementary valuation. For example O_1 and O_2 are elementary valuations, but O_3 and O_4 are not.

Every elementary valuation V generates a valuation $\bar{V}$, where $\bar{V} \subseteq F$, as follows:

$$(*)\qquad \left\{\begin{array}{ll}\text{(i)} & \text{if } A \text{ is a literal then } A \in \bar{V} \text{ if and only if } A \in V.\\ \text{(ii)} & \text{If } A = B \vee C \text{ then } A \in \bar{V} \text{ if and only if } B \in \bar{V} \text{ or } C \in \bar{V}.\\ \text{(iii)} & \text{If } A = B \wedge C \text{ then } A \in \bar{V} \text{ if and only if } B \in \bar{V} \text{ and } C \in \bar{V}.\end{array}\right\}$$

If $A \in \bar{V}$ then we say that A is true in the valuation $\bar{V}$; if $\sim A \in \bar{V}$ then we say that A is false in the valuation $\bar{V}$.

(2) A formula A is a tautology if $A \in \bar{V}$ for every valuation $\bar{V}$.

(3) A formula A is a contradiction if $A \notin \bar{V}$ for any valuation $\bar{V}$.

(4) We say that A and B are equivalent if for every valuation $\bar{V}$, $A \in \bar{V}$ if and only if $B \in \bar{V}$. The set of equivalence classes of formulas form a Boolean algebra as follows:

$$\|A\| \vee \|B\| = \|A \vee B\|,\ \|A\| \wedge \|B\| = \|A \wedge B\|,\ \|A\|' = \|\sim A\|.$$

The set of contradictions forms one equivalence class, namely the 0 of this Boolean algebra; and the set of tautologies forms another equivalence class, namely the 1 of this Boolean algebra. (See [2] Pages 43–46.)

(5) If $\bar{V}_1$ and $\bar{V}_2$ are 2 different valuations then neither $\bar{V}_1 \cap \bar{V}_2$ nor $\bar{V}_1 \cup \bar{V}_2$ are valuations.

Proof. If $\bar{V}_1 \cap \bar{V}_2 = \bar{V}$ for some valuation $\bar{V}$ then for some i neither $p_i \in V$ nor $\sim p_i \in V$. Thus V is not an elementary valuation. Similarly, if $\bar{V}_1 \cup \bar{V}_2 = \bar{W}$ then for some i both $p_i \in W$ and $\sim p_i \in W$ so that W is not an elementary valuation.

(6) If $\bar{W} \supseteq \bar{V}$ and $\bar{V}$ is a valuation then $\bar{W}$ is a valuation if and only if $\bar{W} = \bar{V}$.

Proof. Assume that $\bar{W} \supseteq \bar{V}$ and $\bar{W}$ is a valuation. Then $\bar{W}$ is generated by some elementary valuation W such that $W \supseteq V$. This implies that $W = V$, so that $\bar{W} = \bar{V}$. The converse is immediate.

3. Inconsistent logic (b-logic). In this section we describe a logic which allows for inconsistencies.

(1) A subset V of L is called an elementary b-valuation if for every pair $\{p_i, \sim p_i\}$ at least one member of the pair is a member of V.

It follows that the following 3 conditions are equivalent:

(i) V is an elementary b-valuation.

(ii) There is an elementary valuation W such that $W \subseteq V \subseteq L$.

(iii) There is a set $\{W_i\}$ of elementary valuations such that $V = \cup W_i$.

We think of the members of V as the true literals of the elementary b-valuation. Then O_1, O_2, and O_3 are elementary b-valuations, but O_4 is not. Every elementary b-valuation V generates a b-valuation $\bar{V}$ according to (*) in Section 2.

(2) A formula A is a b-tautology if $A \in \bar{V}$ for every b-valuation $\bar{V}$.

The set of b-tautologies coincides with the set of tautologies.

Proof. By (1) every b-tautology is a tautology. Now suppose that A is a tautology and let $\bar{V}$ be any b-valuation. $\bar{V}$ is generated by some elementary b-valuation V. By (1)(ii) there is an elementary valuation U such that $U \subseteq V$. Then $A \in \bar{U}$ and $\bar{U} \subseteq \bar{V}$, so $A \in \bar{V}$. Thus A is a b-tautology.

(3) A formula A is a b-contradiction if $A \notin \bar{V}$ for any b-valuation $\bar{V}$.

There are no b-contradictions.

Proof. F is a b-valuation and for every formula A, $A \in F$.

(4) We say that A and B are b-equivalent if for every b-valuation $\bar{V}$, $A \in \bar{V}$ if and only if $B \in \bar{V}$.

If A and B are b-equivalent then they are equivalent but not vice versa. For example, the formulas $p_1 \wedge \sim p_1$ and $p_2 \wedge \sim p_2$ are equivalent but not b-equivalent.

The set of b-equivalence classes of formulas form a distributive lattice with a 1 but no 0 and an operation $'$ such that

$$\|A\| \vee \|A\|' = 1, \ (\|A\| \vee \|B\|)' = \|A\|' \wedge \|B\|',$$
$$(\|A\| \wedge \|B\|)' = \|A\|' \vee \|B\|', \text{ and } \|A\|'' = \|A\|.$$

The proof is similar to the proof of the proposition that the set of equivalence classes of formulas form a Boolean algebra.

(5) If $\bar{V}_1$ and $\bar{V}_2$ are b-valuations then $\bar{V}_1 \cup \bar{V}_2$ is a b-valuation in case $\bar{V}_1 \cup \bar{V}_2 = \overline{(V_1 \cup V_2)}$. However the equation $\bar{V}_1 \cup \bar{V}_2 = \overline{(V_1 \cup V_2)}$ does not hold in general. For example, if $p_1 \in V_1$, $\sim p_1 \notin V_1$, $p_1 \notin V_2$, and $\sim p_1 \in V_2$ then $p_1 \wedge \sim p_1 \in \overline{(V_1 \cup V_2)}$ but $p_1 \wedge \sim p_1 \notin \bar{V}_1 \cup \bar{V}_2$. Similarly, the equation $\bar{V}_1 \cap \bar{V}_2 = \overline{(V_1 \cap V_2)}$ does not hold in general. First of all $V_1 \cap V_2$ need not be an elementary b-valuation. But even if $V_1 \cap V_2$ is an elementary b-valuation the equation need not hold. For example, if $p_1 \in V_1$, $\sim p_1 \in V_1$, $p_2 \in V_1$, $\sim p_2 \notin V_1$ and $p_1 \notin V_2$, $\sim p_1 \in V_2$, $p_2 \in V_2$, $\sim p_2 \in V_2$ then $p_1 \vee \sim p_2 \in \bar{V}_1 \cap \bar{V}_2$ but $p_1 \vee \sim p_2 \notin \overline{(V_1 \cap V_2)}$.

(6) The set of b-valuations is a semilattice with a 1 under inclusion. (See [1] Page 19 for the definition of semilattice.)

Proof. F is the 1 of this semilattice since $\bar{V} \subseteq F$ for all b-valuations $\bar{V}$. It suffices to prove that given b-valuations $\bar{U}$ and $\bar{V}$, the least upper bound (lub) of $\{\bar{U}, \bar{V}\}$ always exists. We claim that lub($\{\bar{U}, \bar{V}\}$) is $\overline{(U \cup V)}$. For certainly $\bar{U} \subseteq \overline{(U \cup V)}$ and $\bar{V} \subseteq \overline{(U \cup V)}$. Suppose now that $\bar{U} \subseteq \bar{W}$ and $\bar{V} \subseteq \bar{W}$. Then $U \subseteq W$ and $V \subseteq W$, so $U \cup V \subseteq W$. Therefore $\overline{(U \cup V)} \subseteq \bar{W}$.

4. Incomplete logic (*n*-logic). In this section we describe a logic which allows for incompleteness.

(1) A subset V of L is called an elementary n-valuation if for every pair $\{p_i, \sim p_i\}$ at most one member of the pair is a member of V.

It follows that the following 3 conditions are equivalent:

(i) V is an elementary n-valuation.

(ii) There is an elementary valuation W such that $V \subseteq W$.

(iii) There is a set $\{W_i\}$ of elementary valuations such that $V = \cap W_i$.

We think of the members of V as the true literals of the elementary n-valuation. Then O_1, O_2, and O_4 are elementary n-valuations, but O_3 is not. Every elementary n-valuation V generates an n-valuation $\bar{V}$ according to (*) in Section 2.

(2) A formula A is an n-tautology if $A \in \bar{V}$ for every n-valuation $\bar{V}$.

There are no n-tautologies.

Proof. ϕ is an n-valuation and there is no formula A such that $A \in \phi$.

(3) A formula A is an n-contradiction if $A \notin \bar{V}$ for any n-valuation $\bar{V}$.

The set of n-contradictions coincides with the set of contradictions.

Proof. By (1) every n-contradiction is a contradiction. Now suppose that A is a contradiction and let $\bar{V}$ be any n-valuation. $\bar{V}$ is generated by some elementary n-valuation V. By (1)(ii) there is an elementary valuation U such that $V \subseteq U$. Then $A \notin \bar{U}$, and $\bar{V} \subseteq \bar{U}$, so $A \notin \bar{V}$. Thus A is an n-contradiction.

(4) We say that A and B are n-equivalent if for every n-valuation $\bar{V}$, $A \in \bar{V}$ if and only if $B \in \bar{V}$.

If A and B are n-equivalent then they are equivalent but not vice versa. For example, the formulas $p_1 \vee \sim p_1$ and $p_2 \vee \sim p_2$ are equivalent but not n-equivalent.

The set of n-equivalence classes of formulas form a distributive lattice with a 0 but no 1 and an operation $'$ such that

$$\|A\| \wedge \|A\|' = 0, \ (\|A\| \vee \|B\|)' = \|A\|' \wedge \|B\|',$$
$$(\|A\| \wedge \|B\|)' = \|A\|' \vee \|B\|', \text{ and } \|A\|'' = \|A\|.$$

The proof is similar to the proof of the proposition that the set of equivalence classes of formulas form a Boolean algebra.

(5) If $\bar{V}_1$ and $\bar{V}_2$ are n-valuations then $\bar{V}_1 \cap \bar{V}_2$ is an n-valuation in case $\bar{V}_1 \cap \bar{V}_2 = \overline{(V_1 \cap V_2)}$. However the equation $\bar{V}_1 \cap \bar{V}_2 = \overline{(V_1 \cap V_2)}$ does not hold in general. For example, if $p_1 \in V_1$ and $\sim p_1 \in V_2$ then $p_1 \vee \sim p_1 \in \bar{V}_1 \cap \bar{V}_2$ but $p_1 \vee \sim p_1 \notin \overline{(V_1 \cap V_2)}$. Similarly, the equation $\bar{V}_1 \cup \bar{V}_2 = \overline{(V_1 \cup V_2)}$ does not hold in general. First of all $V_1 \cup V_2$ need not be an elementary n-valuation. But even if $V_1 \cup V_2$ is an elementary n-valuation, the equation need not hold. For example, if $p_1 \in V_1$, $p_2 \notin V_1$, $p_1 \notin V_2$, and $p_2 \in V_2$ then $p_1 \wedge p_2 \in \overline{(V_1 \cup V_2)}$ but $p_1 \wedge p_2 \notin \bar{V}_1 \cup \bar{V}_2$.

(6) The set of n-valuations is a semilattice with a 0 under inclusion.

Proof. ϕ is the 0 of this semilattice since $\phi \subseteq \bar{V}$ for all n-valuations $\bar{V}$. It suffices to prove that given n-valuations $\bar{U}$ and $\bar{V}$, the greatest lower bound (glb) of $\{\bar{U}, \bar{V}\}$ always exists (see [1] Pages 10 and 19). We claim that glb($\{\bar{U}, \bar{V}\}$) is $\overline{(U \cap V)}$. For certainly $\overline{(U \cap V)} \subseteq \bar{U}$ and $\overline{(U \cap V)} \subseteq \bar{V}$. Suppose now that $\bar{W} \subseteq \bar{U}$ and $\bar{W} \subseteq \bar{V}$. Then $W \subseteq U$ and $W \subseteq V$, so $W \subseteq U \cap V$. Therefore $\bar{W} \subseteq \overline{(U \cap V)}$.

5. Generalized logic (g-logic). In this final section we describe a logic which combines features of the b-logic and the n-logic. We state only the results and give no proofs since the proofs are similar to the ones given in previous sections.

(1) Every subset of L is called an elementary g-valuation. Thus O_1, O_2, O_3, and O_4 are all elementary g-valuations. Every elementary g-valuation V generates a g-valuation $\bar{V}$ according to (*) in Section 2.

(2) A formula A is a g-tautology if $A \in \bar{V}$ for every g-valuation $\bar{V}$.

There are no g-tautologies.

(3) A formula A is a g-contradiction if $A \notin \bar{V}$ for any g-valuation $\bar{V}$.

There are no g-contradictions.

(4) We say that A and B are g-equivalent if for every g-valuation $\bar{V}$, $A \in \bar{V}$ if and only if $B \in \bar{V}$.

If A and B are g-equivalent then they are both b-equivalent and n-equivalent but not vice versa. For example, the formulas $(p_1 \vee \sim p_1) \wedge (p_2 \vee \sim p_2)$ and $(p_1 \vee p_2) \wedge (\sim p_1 \vee \sim p_2)$ are both b-equivalent and n-equivalent but not g-equivalent.

The set of g-equivalence classes of formulas form a distributive lattice with no 0, no 1, and an operation $'$ such that $(\|A\| \vee \|B\|)' = \|A\|' \wedge \|B\|'$,

$$(\|A\| \wedge \|B\|)' = \|A\|' \vee \|B\|', \text{ and } \|A\|'' = \|A\|.$$

(5) If $\bar{V}_1$ and $\bar{V}_2$ are g-valuations then $\bar{V}_1 \cup \bar{V}_2$ is a g-valuation in case $\bar{V}_1 \cup \bar{V}_2 = \overline{(V_1 \cup V_2)}$ and $\bar{V}_1 \cap \bar{V}_2$ is a g-valuation in case $\bar{V}_1 \cap \bar{V}_2 = \overline{(V_1 \cap V_2)}$. The equations $\bar{V}_1 \cup \bar{V}_2 = \overline{(V_1 \cup V_2)}$ and $\bar{V}_1 \cap \bar{V}_2 = \overline{(V_1 \cap V_2)}$ do not hold in general.

(6) The set of g-valuations is a Boolean algebra under inclusion.

Note. The notions presented in this paper for propositional logic have been extended by the author to first-order logic in a paper under preparation.

References

1. G. Grätzer, Universal Algebra, Van Nostrand, Princeton, 1968.
2. P. Halmos, Lectures on Boolean Algebras, Van Nostrand, Princeton, 1963.
3. R. M. Smullyan, First-Order Logic, Springer-Verlag, New York, 1968.

JACOBI'S SOLUTION OF LINEAR DIOPHANTINE EQUATIONS

M. S. WATERMAN, Idaho State University, Pocatello

1. Introduction. C. G. J. Jacobi's 1869 paper *Über die Auflösung der Gleichung*, $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = fu$ [1] is a careful treatment of linear Diophantine equations. Although Jacobi's first solution is exactly that used by modern authors such as Niven and Zuckerman [2, p. 94–98], he introduces the beautiful concept of equivalent systems of variables and uses this concept to establish the validity of his solution.

5. Generalized logic (*g*-logic). In this final section we describe a logic which combines features of the *b*-logic and the *n*-logic. We state only the results and give no proofs since the proofs are similar to the ones given in previous sections.

(1) Every subset of L is called an elementary g-valuation. Thus O_1, O_2, O_3, and O_4 are all elementary g-valuations. Every elementary g-valuation V generates a g-valuation $\bar{V}$ according to (*) in Section 2.

(2) A formula A is a g-tautology if $A \in \bar{V}$ for every g-valuation $\bar{V}$.

There are no g-tautologies.

(3) A formula A is a g-contradiction if $A \notin \bar{V}$ for any g-valuation $\bar{V}$.

There are no g-contradictions.

(4) We say that A and B are g-equivalent if for every g-valuation $\bar{V}$, $A \in \bar{V}$ if and only if $B \in \bar{V}$.

If A and B are g-equivalent then they are both b-equivalent and n-equivalent but not vice versa. For example, the formulas $(p_1 \vee \sim p_1) \wedge (p_2 \vee \sim p_2)$ and $(p_1 \vee p_2) \wedge (\sim p_1 \vee \sim p_2)$ are both b-equivalent and n-equivalent but not g-equivalent.

The set of g-equivalence classes of formulas form a distributive lattice with no 0, no 1, and an operation $'$ such that $(\|A\| \vee \|B\|)' = \|A\|' \wedge \|B\|'$,

$$(\|A\| \wedge \|B\|)' = \|A\|' \vee \|B\|', \text{ and } \|A\|'' = \|A\|.$$

(5) If $\bar{V}_1$ and $\bar{V}_2$ are g-valuations then $\bar{V}_1 \cup \bar{V}_2$ is a g-valuation in case $\bar{V}_1 \cup \bar{V}_2 = \overline{(V_1 \cup V_2)}$ and $\bar{V}_1 \cap \bar{V}_2$ is a g-valuation in case $\bar{V}_1 \cap \bar{V}_2 = \overline{(V_1 \cap V_2)}$. The equations $\bar{V}_1 \cup \bar{V}_2 = \overline{(V_1 \cup V_2)}$ and $\bar{V}_1 \cap \bar{V}_2 = \overline{(V_1 \cap V_2)}$ do not hold in general.

(6) The set of g-valuations is a Boolean algebra under inclusion.

Note. The notions presented in this paper for propositional logic have been extended by the author to first-order logic in a paper under preparation.

References

1. G. Grätzer, Universal Algebra, Van Nostrand, Princeton, 1968.
2. P. Halmos, Lectures on Boolean Algebras, Van Nostrand, Princeton, 1963.
3. R. M. Smullyan, First-Order Logic, Springer-Verlag, New York, 1968.

JACOBI'S SOLUTION OF LINEAR DIOPHANTINE EQUATIONS

M. S. WATERMAN, Idaho State University, Pocatello

1. Introduction. C. G. J. Jacobi's 1869 paper *Über die Auflösung der Gleichung*, $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = fu$ [**1**] is a careful treatment of linear Diophantine equations. Although Jacobi's first solution is exactly that used by modern authors such as Niven and Zuckerman [**2**, p. 94–98], he introduces the beautiful concept of equivalent systems of variables and uses this concept to establish the validity of his solution.

The purpose of this paper is to present the theory of equivalent systems of variables and apply them to linear Diophantine equations. The material on equivalent systems is, we feel, of interest in its own right and, while we try to follow Jacobi as much as possible, some of the work is not to be found in his paper. For example, Proposition 1 and Theorems 2, 3, and 4 do not appear in Jacobi's paper.

This material could compose the basis of an independent study project or take home exam in undergraduate number theory.

2. Equivalent systems of variables. We begin with the basic definition.

DEFINITION 1. *Let* $X = (x_1, x_2, \cdots, x_n)$ *and* $Y = (y_1, y_2, \cdots, y_n)$ *be* $1 \times n$ *vectors of variables. X and Y are said to be equivalent systems of variables if* (1) *each vector is defined in terms of the other by a set of linear equations without constant term and* (2) *the vector X has integer values if and only if Y does. If X and Y are related in this fashion we will write* $X \sim Y$. *If X and Y are not equivalent systems of variables, we will write* $X \nsim Y$.

Of course (1) means that $X = YA$ and $Y = XB$ where A and B are $n \times n$ matrices. It is easy to show that (2) implies these matrices must have integer elements.

Jacobi also has another definition regarding linear systems.

DEFINITION 2. *Suppose X and Y are* $1 \times n$ *vectors of variables and* $X = YA$, $Y = XB$. *Then these two systems of linear equations are called reciprocal systems if* $A = B^{-1}$.

Now we relate the two definitions.

PROPOSITION 1. *Suppose* $X \sim Y$ *with* $X = YA$ *and* $Y = XB$. *Then these equations represent reciprocal systems.*

Proof. We have $X = YA = X(BA)$ so that $X(I-BA) = 0$. Since X is a vector of variables, we fix the index j and let $x_j = 1$ and $x_i = 0$ for all $i \neq j$. This implies the jth row of $I - BA$ is the zero vector, and, since j is arbitrary, $I - BA = 0$. Similarly $I - AB = 0$ so that $A = B^{-1}$.

The next theorem shows that the set of possible matrices in equivalent systems of variables forms the unimodular group (with integer elements).

THEOREM 1. *Let* $X = YA$ *where all* x_i*'s and* y_j*'s are distinct variables and the elements of A are integers. Then* $X \sim Y$ *if and only if* $\det(A) = \pm 1$.

Proof. First assume $X \sim Y$. Then $X = YA$, $Y = XB$, and, by Proposition 1, $A^{-1} = B$. Therefore, A^{-1} must have integer elements. It follows that both $\det A$ and $\det A^{-1}$ must be integers with $(\det A)(\det A^{-1}) = 1$. Therefore, $\det A = \pm 1$.

Next assume $\det A = \pm 1$. Then $Y = XA^{-1}$ where $A^{-1} = (\det A)^{-1}(\operatorname{adj} A) = \pm(\operatorname{adj} A)$ has integer elements.

To see why x_i and y_i must be distinct variables consider

$$(x, t) = (y, t)\begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}.$$

Now $\det\begin{pmatrix}3&1\\2&1\end{pmatrix} = 1$ but the equations cannot hold unless $y = 0$. Therefore, $(x, t) \nsim (y, t)$.

As one would expect $\sim$ is an equivalence relation. The proof is left as an exercise.

THEOREM 2. $\sim$ *is an equivalence relation.*

We now need to define two operations. Let $X = (x_1, x_2, \cdots, x_n)$ and $Z = (z_1, z_2, \cdots, z_n)$. Define $X \vee Z = (x_1, x_2, \cdots, x_n, z_1, z_2, \cdots, z_n)$. Also, let $X \ominus Z$ be the vector of all x variables which are not also Z variables. The next results study $\sim$ under these operations.

THEOREM 3. *If* $X_1 \sim Y_1$ *and* $X_2 \sim Y_2$, *then* $X_1 \vee X_2 \sim Y_1 \vee Y_2$.

Proof. If $X_1 = A_1 Y_1$ and $X_2 = A_2 Y_2$, then

$$X_1 \vee X_2 = (Y_1 \vee Y_2)\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

The other equality follows in the same manner.

THEOREM 4. *Let* $X_1 \sim Y_1$ *and* $X_2 \sim Y_2$ *where* t *is an* X_2 *and a* Y_1 *variable. Then* $(X_1 \vee X_2) \ominus (t) \sim (Y_1 \vee Y_2) \ominus (t)$.

Proof. Using the proof of Theorem 3,

$$X_1 \vee X_2 = (Y_1 \vee Y_2)\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

Take the equation for t and substitute for t into each other occurrence for t in this system of equations. Thus, we obtain $(X_1 \vee X_2) \ominus (t)$ as linear integer combination of $(Y_1 \vee Y_2) \ominus (t)$. To complete the proof, reverse the roles of $X_1 \vee X_2$ and $Y_1 \vee Y_2$.

In Jacobi's paper he solves the equation

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = fu$$

where f is the greatest common divisor of $\alpha_1, \alpha_2, \cdots, \alpha_n$. (That is, $f = (\alpha_1, \alpha_2, \cdots, \alpha_n)$.) He sets

$$u = y_1 = \left(\frac{\alpha_1}{f}\right)x_1 + \left(\frac{\alpha_2}{f}\right)x_2 + \cdots + \left(\frac{\alpha_n}{f}\right)x_n$$

and introduces certain variables $y_2, y_3, \cdots, y_n$ so that $X \sim Y$. Then the problem is solved since $X = YA$ and, with y_1 fixed, we can let $y_2, y_3, \cdots, y_n$ vary over all possible integer values and obtain all possible values of $X = (x_1, x_2, \cdots, x_n)$.

3. **Two variable linear diophantine equations.** In this section we give the usual Euclidean algorithm solution of

(1) $$\alpha_1 x_1 + \alpha_2 x_2 = fu,$$

where α_1, α_2 are fixed integers and $f = (\alpha_1, \alpha_2)$. Of course $(\alpha_1/f, \alpha_2/f) = 1$ and

Euclid's algorithm provides us with integers γ and β such that

$$\gamma\frac{\alpha_1}{f}-\beta\frac{\alpha_2}{f}=1.$$

Then, if z is an arbitrary integer,

$$\alpha_1\left(\gamma u-\frac{\alpha_2}{f}z\right)+\alpha_2\left(-\beta u+\frac{\alpha_1}{f}z\right)=fu.$$

Our solution to (1) is then

$$x_1=\gamma u-\frac{\alpha_2}{f}z,\qquad x_2=-\beta u+\frac{\alpha_1}{f}z, \tag{2}$$

where z is an arbitrary integer.

To see that (2) has all (x_1,x_2) such that (1) holds, write (2) as

$$(x_1,x_2)=(u,z)\begin{bmatrix}\gamma & -\beta\\ \frac{-\alpha_2}{f} & \frac{\alpha_1}{f}\end{bmatrix}=(u,z)A.$$

Note that $\det A=1$. Thus $(x_1,x_2)\sim(u,z)$ and our solution is complete.

4. General linear diophantine equations. As we remarked in the first section, we wish to solve

$$\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_nx_n=fu \tag{3}$$

where $f=(\alpha_1,\alpha_2,\cdots,\alpha_n)$ and $\alpha_1,\alpha_2,\cdots,\alpha_n$ are integer constants. The main task is to find $Y=(y_1,y_2,\cdots,y_n)$ such that $X\sim Y$. To do this some new equations must be introduced.

Let $f_2=(\alpha_1,\alpha_2)$. Then consider

$$\alpha_1x_1+\alpha_2x_2=f_2y_2.$$

By section 3, there exists z_1 such that $(x_1,x_2)\sim(z_1,y_2)$. Then let $f_3=(f_2,\alpha_3)$ and consider

$$f_2y_2+\alpha_3x_3=f_3y_3.$$

Again there exists z_2 such that $(y_2,x_3)\sim(z_2,y_3)$.

Letting $f_i=(f_{i-1},\alpha_i)$, $i=3,\cdots,n$, we obtain the following equations:

$$\begin{aligned}\alpha_1x_1+\alpha_2x_2&=f_2y_2,\\ f_2y_2+\alpha_3x_3&=f_3y_3,\\ f_3y_3+\alpha_4x_4&=f_4y_4,\\ &\cdots\cdots\\ f_{n-1}y_{n-1}+\alpha_nx_n&=f_ny_n.\end{aligned} \tag{4}$$

Of course $y_n = u$ and $f_n = f$.

Repeated applications of section 3 yield

$$(5)\qquad \begin{aligned} (x_1, x_2) &\sim (z_1, y_2), \\ (y_2, y_3) &\sim (z_2, y_3), \\ (y_3, x_4) &\sim (z_3, y_4), \\ &\cdots\cdots \\ (y_{n-1}, x_n) &\sim (z_{n-1}, y_n). \end{aligned}$$

Theorem 4 applied to the first two lines of (5) yields $(x_1, x_2, x_3) \sim (z_1, z_2, y_3)$ This relation and the third line (5) yields $(x_1, x_2, x_3, x_4) \sim (z_1, z_2, z_3, y_4)$. Proceeding in the same manner, we obtain

$$(x_1, x_2, \cdots, x_n) \sim (z_1, z_2, \cdots, z_{n-1}, y_n)$$

or

$$(6)\qquad (x_1, x_2, \cdots, x_n) \sim (z_1, z_2, \cdots, z_{n-1}, u).$$

Of course our solution is obtained in the obvious way from (5). From the first two sets of equations, we eliminate y_2. Then successively, eliminate $y_3, y_4, \cdots, y_{n-1}$. This process was indicated in our passage from (5) to (6). Jacobi actually finds the matrices associated with (6) and therefore solves the linear diophantine equation. He shows the determinant of the two associated matrices is 1 but we do not include his further results here.

This work was partly performed under the auspices of the U. S. Atomic Energy Commission while the author was a faculty participant of the Associated Western Universities at Los Alamos Scientific Laboratory.

References

1. C. G. J. Jacobi, Über die Auflösung der Gleichung $\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n = fu$, J. Reine Angew. Math., 69 (1869) 1–28.

2. Ivan Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, Wiley, New York, London, 1960.

THE CONVERGENCE OF JACOBI AND GAUSS-SEIDEL ITERATION

STEWART VENIT, California State University-Los Angeles

1. Introduction. The Jacobi and Gauss-Seidel methods are perhaps the best known iterative procedures for solving a nonsingular linear system of n equations:

$$(1)\qquad \sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, 2, \cdots, n.$$

We will assume that the equations are ordered so that all the a_{ii} are nonzero.

Of course $y_n = u$ and $f_n = f$.

Repeated applications of section 3 yield

$$
\begin{aligned}
&(x_1, x_2) \sim (z_1, y_2),\\
&(y_2, y_3) \sim (z_2, y_3),\\
&(y_3, x_4) \sim (z_3, y_4),\\
&\quad\cdots\cdots\cdots\cdots\\
&(y_{n-1}, x_n) \sim (z_{n-1}, y_n).
\end{aligned} \tag{5}
$$

Theorem 4 applied to the first two lines of (5) yields $(x_1, x_2, x_3) \sim (z_1, z_2, y_3)$ This relation and the third line (5) yields $(x_1, x_2, x_3, x_4) \sim (z_1, z_2, z_3, y_4)$. Proceeding in the same manner, we obtain

$$(x_1, x_2, \cdots, x_n) \sim (z_1, z_2, \cdots, z_{n-1}, y_n)$$

or

$$(x_1, x_2, \cdots, x_n) \sim (z_1, z_2, \cdots, z_{n-1}, u). \tag{6}$$

Of course our solution is obtained in the obvious way from (5). From the first two sets of equations, we eliminate y_2. Then successively, eliminate $y_3, y_4, \cdots, y_{n-1}$. This process was indicated in our passage from (5) to (6). Jacobi actually finds the matrices associated with (6) and therefore solves the linear diophantine equation. He shows the determinant of the two associated matrices is 1 but we do not include his further results here.

This work was partly performed under the auspices of the U. S. Atomic Energy Commission while the author was a faculty participant of the Associated Western Universities at Los Alamos Scientific Laboratory.

References

1. C. G. J. Jacobi, Über die Auflösung der Gleichung $\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n = fu$, J. Reine Angew. Math., 69 (1869) 1–28.

2. Ivan Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, Wiley, New York, London, 1960.

THE CONVERGENCE OF JACOBI AND GAUSS-SEIDEL ITERATION

STEWART VENIT, California State University-Los Angeles

1. Introduction. The Jacobi and Gauss-Seidel methods are perhaps the best known iterative procedures for solving a nonsingular linear system of n equations:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, 2, \cdots, n. \tag{1}$$

We will assume that the equations are ordered so that all the a_{ii} are nonzero.

To apply either method to this system, first arbitrarily choose an initial approximation, $x_i^{(0)}$, $i = 1, 2, \cdots, n$. Then, generate successive iterates, $x_i^{(k)}$, $k = 1, 2, 3, \cdots$, by using the formulas:

Jacobi:

$$x_i^{(k+1)} = \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j^{(k)}\right) \Big/ a_{ii} \qquad (i = 1, 2, \cdots, n).$$

Gauss-Seidel:

$$x_i^{(k+1)} = \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)}\right) \Big/ a_{ii} \qquad (i = 1, 2, \cdots, n).$$

Under certain conditions (one of which is given later), the approximations, $x_i^{(k)}$, will converge to the solution of (1) as $k \to \infty$.

Notice that in obtaining iterate $(k+1)$ from iterate k, the Gauss-Seidel scheme makes use of "more recent information" than does the Jacobi scheme. Hence, it is natural to assume that the former method will converge whenever the latter one does, and at a faster rate. It is known, however, that this is not always the case. Several authors (see, for example, Fox [**1**, p. 194]) have given examples to show that when $n = 3$ it is possible for either method to converge while the other diverges. We will prove this to be true for all $n \geqq 3$. We will also show for all $n \geqq 3$ that when both methods converge, the rate of convergence of Jacobi iteration may be less than, equal to, or greater than that of Gauss-Seidel iteration.

2. Matrix formulation of the problem. The system (1) can be written in matrix form as

$$Ax = b, \tag{1'}$$

where A is a nonsingular square matrix of order n with nonzero diagonal elements, and $\boldsymbol{x}$ and $\boldsymbol{b}$ are n-vectors. We can generate a class of iterative schemes to solve (1′) by "splitting" A as $A = B - C$, where B is nonsingular, and then rewriting (1′) as $\boldsymbol{x} = (B^{-1}C)\boldsymbol{x} + B^{-1}\boldsymbol{b}$. This motivates the definition of a general class of iterative methods. Denoting the kth iterate by $\boldsymbol{x}^{(k)}$, we choose $\boldsymbol{x}^{(0)}$ arbitrarily and define

$$\boldsymbol{x}^{(k+1)} = (B^{-1}C)\boldsymbol{x}^{(k)} + B^{-1}\boldsymbol{b}, \text{ for } k = 0, 1, 2, \cdots. \tag{2}$$

For simplicity of notation, let $M = B^{-1}C$. Then (2) becomes

$$\boldsymbol{x}^{(k+1)} = M\boldsymbol{x}^{(k)} + B^{-1}\boldsymbol{b}, \quad \text{for } k = 0, 1, 2, \cdots. \tag{2'}$$

DEFINITION. The *spectral radius of a matrix* N, *denoted* $\rho(N)$, *is given by*

$$\rho(N) = \max\{|\lambda_i| : \lambda_i \textit{ is an eigenvalue of } N\}.$$

DEFINITION. *The iteration scheme* (2′) *converges if for all initial vectors* $\boldsymbol{x}^{(0)}$, $\lim \boldsymbol{x}^{(k)} = \boldsymbol{x}$ *(component-wise) as* $k \to \infty$, *where* $\boldsymbol{x}$ *is the solution of* (1′).

The proof of the following theorem, as well as a more complete discussion of iterative schemes, can be found in Isaacson and Keller [**2**, chapter 2, section 4].

THEOREM 1. *The iteration scheme* (2′) *converges if and only if* $\rho(M) < 1$.

The Jacobi and Gauss-Seidel methods are special cases of the general scheme (2). For Jacobi iteration, we have $b_{ij} = a_{ij}$ if $i = j$, and $b_{ij} = 0$ otherwise. The Gauss-Seidel scheme is obtained by taking $b_{ij} = a_{ij}$ if $i \geqq j$, and $b_{ij} = 0$ otherwise. For both schemes, the matrix C is given by $C = B - A$, and since A has nonzero diagonal elements, B is nonsingular.

3. **Relative convergence theorems.** In the following theorems, the matrix A is the coefficient matrix of the linear system (1). It is assumed to be nonsingular with nonzero diagonal elements, and we denote its order by n.

THEOREM 2. *If* $n = 2$, *Jacobi iteration converges if and only if Gauss-Seidel iteration converges.*

Proof. Let the elements of A be denoted by a_{ij} where i, $j = 1,2$. Then direct computation yields $\rho(M) = \sqrt{|a_{21}a_{12}|/|a_{11}a_{22}|}$ for Jacobi and $\rho(M) = |a_{21}a_{12}|/|a_{11}a_{22}|$ for Gauss-Seidel. The theorem then follows by applying Theorem 1.

THEOREM 3. *For any* $n > 2$, *it is possible for Jacobi iteration to converge while Gauss-Seidel iteration diverges, and conversely.*

Proof. Let a and b be parameters, and consider

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & a \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ b & 0 & 0 & \cdots & 1 & 1 \end{bmatrix}.$$

For this matrix A, we explicitly compute $M = B^{-1}C$ for the two methods, and obtain the characteristic equation, $\det(\lambda I - M) = 0$, for each. They are

$$\lambda^n - ab\lambda^{n-2} + (-1)^{n+1}a = 0 \text{ for Jacobi, and} \tag{3}$$

$$\lambda^{n-1}(\lambda - a(b + (-1)^n)) \text{ for Gauss-Seidel.} \tag{4}$$

(i) Take $a \geqq 1$ and $b = (-1)^{n+1}$. Then, for Gauss-Seidel, $\rho(M) = 0$, while for Jacobi, $\rho(M) \geqq 1$. The last inequality follows from the fact that the absolute value of the product of the roots of (3) is equal to $a \geqq 1$, and consequently, there is at least one eigenvalue, λ, with $|\lambda| \geqq 1$.

(ii) Take $a = \frac{1}{2}(-1)^{n+1}$ and $b = (-1)^n$. Then, for Gauss-Seidel $\rho(M) = 1$. Equation (3) becomes $\lambda^n + \frac{1}{2}\lambda^{n-2} + \frac{1}{2} = 0$. We will use Rouché's theorem to show that all roots of this equation lie within the unit circle.

Let $f(\lambda) = \lambda^n + \frac{1}{2}$ and $g(\lambda) = \frac{1}{2}\lambda^{n-2}$. Since $f(\lambda)$ has n zeroes within the unit circle, we would like to show that $f(\lambda)$ and $f(\lambda) + g(\lambda)$ have the same number of zeroes in this domain. To do this, let $\delta > 0$ and define C_δ to be the boundary of the region

$$\{\lambda \mid |\lambda| \leqq 1\} \cup \{\lambda \mid |\lambda + 1| \leqq \delta\}.$$

For each $\delta > 0$, on C_δ we have $|f(\lambda)| > |g(\lambda)|$, and hence $f(\lambda)$ and $f(\lambda) + g(\lambda)$ have the same number of zeroes within C_δ. Moreover, $\lambda = -1$ is not a zero of $f(\lambda) + g(\lambda)$. Consequently, $f(\lambda)$ and $f(\lambda) + g(\lambda)$ have the same number of zeroes within the unit circle, as desired.

DEFINITION. *The (asymptotic) rate of convergence, R, of a convergent iteration scheme of the form* (2′) *is given by* $R = -\log_{10}(\rho(M))$, *if* $\rho(M) \neq 0$. *If* $\rho(M) = 0$, $R = \infty$.

Notice that the smaller the spectral radius, the greater the value of R. This will result in faster convergence.

THEOREM 4. *Let R_1 and R_2 denote the respective rates of convergence of the Jacobi and Gauss-Seidel schemes. Assume that both schemes converge. Then*:
(i) *if* $n = 2$, *then* $R_1 \leqq R_2$;
(ii) *if* $n > 2$, *then R_1 may be less than, equal to, or greater than R_2.*

Proof. (i) An inspection of the expressions for $\rho(M)$ in the proof of Theorem 2 shows this to be the case.

(ii) Let A be as in the proof of Theorem 3.

If $b = 0$ and $0 < a < 1$, then $R_1 < R_2$.

If $a = 0$, then $R_1 = R_2 = \infty$.

Finally, as we have seen in the proof of Theorem 3, when $a = \frac{1}{2}(-1)^{n+1}$ and $b = (-1)^n$, for Jacobi iteration $\rho(M) \equiv \rho_J < 1$, while for Gauss-Seidel iteration $\rho(M) \equiv \rho_G = 1$. Since the roots of an equation are continuous functions of the coefficients, fixing $b = (-1)^n$, there is a value of a with $|a| < \frac{1}{2}$, for which $\rho_J < \rho_G < 1$. In this case $R_1 > R_2$.

At several points in this note, we have obtained $\rho(M) = 0$, and consequently $R = \infty$. The following theorem points out the significance of this value of R:

THEOREM 5. *In the iteration scheme* (2′), *suppose* $\rho(M) = 0$. *Then,* $\boldsymbol{x}^{(k)} = \boldsymbol{x}$ *for all* $k \geqq n$, *where $\boldsymbol{x}$ is the solution of* (1′).

Proof. Let $\boldsymbol{e}^{(k)} = \boldsymbol{x}^{(k)} - \boldsymbol{x}$. Then, $\boldsymbol{e}^{(k)}$ satisfies $\boldsymbol{e}^{(k)} = M^k \boldsymbol{e}^{(0)}$. Since $\rho(M) = 0$, M is similar to a strictly upper triangular matrix, U. (That is, $u_{ij} = 0$ if $i \geqq j$.) Hence, $U^k = 0$ for all $k \geqq n$. But, $M^k = P^{-1}U^kP$ for some nonsingular matrix P, and hence $M^k = 0$ for all $k \geqq n$. This gives $\boldsymbol{e}^{(k)} \equiv \boldsymbol{x}^{(k)} - \boldsymbol{x} = 0$ for all $k \geqq n$, as desired.

It should be noted in closing that in many cases of practical interest, the Gauss-Seidel method *is* better than the Jacobi one. For a detailed discussion of this aspect see Varga [**3**, chapter 3].

References

1. L. Fox, An Introduction to Numerical Linear Algebra, Oxford University Press, New York, 1965.
2. E. Isaacson and H. Keller, Analysis of Numerical Methods, Wiley, New York, 1966.
3. R. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, 1962.

A GENERALIZATION OF KRASNOSELSKI'S THEOREM ON THE REAL LINE

BRUCE P. HILLAM, California State Polytechnic University

Recently Bailey [1] gave a proof of Krasnoselski's Theorem [2] restricted to to the real line. Krasnoselski's Theorem is the following result:

THEOREM 1. *If K is a convex, bounded subset of a uniformly convex Banach space and if f is a mapping of K into a compact subset of K such that* $\|f(x)-f(y)\| \leqq \|x-y\|$, *then the sequence obtained by choosing x_1 in K and defining* $x_{n+1} = \frac{1}{2}[x_n + f(x_n)]$ *converges to some z in K and $f(z) = z$.*

Bailey gave a proof for the case in which K is a closed interval of the real line. Bailey's result is

THEOREM 2. *If f takes $[a,b]$ into itself and $|f(x)-f(y)| \leqq |x-y|$, then the sequence obtained by choosing x_1 in $[a,b]$ and defining $x_{n+1} = \frac{1}{2}[x_n + f(x_n)]$ converges to some z in $[a,b]$ and $f(z) = z$.*

Recall that a function $f\colon [a,b] \to [a,b]$ satisfies a Lipschitz condition with constant L if for all x and y in $[a,b]$, $|f(x)-f(y)| \leqq L|x-y|$. A function that satisfies a Lipschitz condition is clearly continuous. Finally, a sequence $\{x_n\}$ is said to converge monotonically if either $x_n \leqq x_{n+1}$ or $x_n \geqq x_{n+1}$ for all integers n.

Using the fact that the real line is totally ordered, the following more general theorem with a much more elementary proof is possible. Clearly Theorem 2 follows directly from Theorem 3.

THEOREM 3. *Let $f\colon [a,b] \to [a,b]$ be a function that satisfies a Lipschitz condition with constant L. Let x_1 in $[a,b]$ be arbitrary and define $x_{n+1} = (1-\lambda)x_n + \lambda f(x_n)$ where $\lambda = 1/(L+1)$. If $\{x_n\}$ denotes the resulting sequence, then $\{x_n\}$ converges monotonically to a point z in $[a,b]$ where $f(z) = z$.*

Proof. Without loss of generality we can assume $f(x_n) \neq x_n$ for all n. Suppose $f(x_1) > x_1$ and let p be the first point greater than x_1 such that $f(p) = p$. Since $f(x_1) > x_1$ and $f(b) \leqq b$, the continuity of f implies there is such a point.

Claim. If $x_1 < x_2 < \cdots < x_n < p$ and $f(x_i) > x_i$ for $i = 1,2,\cdots,n$, then $f(x_{n+1}) > x_{n+1}$ and $x_{n+1} < p$. Suppose $p < x_{n+1}$, then $x_n < p < x_{n+1}$ thus $0 < p - x_n < x_{n+1} - x_n = \lambda(f(x_n) - x_n)$. Hence it follows

References

1. L. Fox, An Introduction to Numerical Linear Algebra, Oxford University Press, New York, 1965.
2. E. Isaacson and H. Keller, Analysis of Numerical Methods, Wiley, New York, 1966.
3. R. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, 1962.

A GENERALIZATION OF KRASNOSELSKI'S THEOREM ON THE REAL LINE

BRUCE P. HILLAM, California State Polytechnic University

Recently Bailey [1] gave a proof of Krasnoselski's Theorem [2] restricted to to the real line. Krasnoselski's Theorem is the following result:

THEOREM 1. *If K is a convex, bounded subset of a uniformly convex Banach space and if f is a mapping of K into a compact subset of K such that* $\|f(x)-f(y)\| \leqq \|x-y\|$, *then the sequence obtained by choosing x_1 in K and defining* $x_{n+1} = \frac{1}{2}[x_n + f(x_n)]$ *converges to some z in K and* $f(z) = z$.

Bailey gave a proof for the case in which K is a closed interval of the real line. Bailey's result is

THEOREM 2. *If f takes $[a,b]$ into itself and* $|f(x) - f(y)| \leqq |x-y|$, *then the sequence obtained by choosing x_1 in $[a,b]$ and defining* $x_{n+1} = \frac{1}{2}[x_n + f(x_n)]$ *converges to some z in $[a,b]$ and $f(z) = z$.*

Recall that a function f: $[a,b] \to [a,b]$ satisfies a Lipschitz condition with constant L if for all x and y in $[a,b]$, $|f(x)-f(y)| \leqq L|x-y|$. A function that satisfies a Lipschitz condition is clearly continuous. Finally, a sequence $\{x_n\}$ is said to converge monotonically if either $x_n \leqq x_{n+1}$ or $x_n \geqq x_{n+1}$ for all integers n.

Using the fact that the real line is totally ordered, the following more general theorem with a much more elementary proof is possible. Clearly Theorem 2 follows directly from Theorem 3.

THEOREM 3. *Let f: $[a,b] \to [a,b]$ be a function that satisfies a Lipschitz condition with constant L. Let x_1 in $[a,b]$ be arbitrary and define* $x_{n+1} = (1-\lambda)x_n + \lambda f(x_n)$ *where* $\lambda = 1/(L+1)$. *If $\{x_n\}$ denotes the resulting sequence, then $\{x_n\}$ converges monotonically to a point z in $[a,b]$ where $f(z) = z$.*

Proof. Without loss of generality we can assume $f(x_n) \neq x_n$ for all n. Suppose $f(x_1) > x_1$ and let p be the first point greater than x_1 such that $f(p) = p$. Since $f(x_1) > x_1$ and $f(b) \leqq b$, the continuity of f implies there is such a point.

Claim. If $x_1 < x_2 < \cdots < x_n < p$ and $f(x_i) > x_i$ for $i = 1,2,\cdots,n$, then $f(x_{n+1}) > x_{n+1}$ and $x_{n+1} < p$. Suppose $p < x_{n+1}$, then $x_n < p < x_{n+1}$ thus $0 < p - x_n < x_{n+1} - x_n = \lambda(f(x_n) - x_n)$. Hence it follows

$$0 < \frac{1}{\lambda}|x_n - p| = (L+1)|x_n - p| < |f(x_n) - x_n|$$
$$\leqq |f(x_n) - f(p)| + |p - x_n|$$

or

$$L|x_n - p| < |f(x_n) - f(p)|,$$

which is a contradiction of the fact that f satisfies a Lipschitz condition. Thus $x_{n+1} < p$ and $f(x_{n+1}) > x_{n+1}$ by the choice of p, and the claim is proved.

Using the induction hypothesis it follows that $x_n < x_{n+1} < p$ for all integers n. Since a bounded monotonic sequence converges, $\{x_n\}$ converges to some point z. By the triangle inequality it follows that

$$|z - f(z)| \leqq |z - x_n| + |x_n - f(x_n)| + |f(x_n) - f(z)|$$
$$= |z - x_n| + \frac{1}{\lambda}|x_{n+1} - x_n| + |f(x_n) - f(z)|.$$

Clearly the right hand side tends to 0 as $n \to \infty$. Thus $f(z) = z$. If $f(x_1) < x_1$ a similar argument holds; q.e.d.

Using a somewhat more sophisticated argument, one can allow λ to be any number less than $2/(L+1)$ but the resulting sequence $\{x_n\}$ need not converge monotonically. The following example shows this last result is best possible.

Let $f: [0,1] \to [0,1]$ be defined by

$$f(x) = \begin{cases} 1, & 0 \leqq x < \dfrac{L-1}{2L} \\ -Lx + \frac{1}{2}(L+1), & \dfrac{L-1}{2L} \leqq x \leqq \dfrac{L+1}{2L} \\ 0, & \dfrac{L+1}{2L} < x < 1, \end{cases}$$

where $L > 1$ is arbitrary. Note f satisfies a Lipschitz condition with constant L. Let $\lambda = 2/(L+1)$ and let $x_1 = (L-1)/2L$. Then $x_2 = (1-\lambda)x_1 + \lambda f(x_1) = (L+1)/2L$, $x_3 = (1-\lambda)x_2 + \lambda f(x_2) = (L-1)/2L$, etc.

Theorem 3 appears to be new even for the real line. It would be very interesting if it could be extended to higher dimensions, but this would require a new proof that does not rely on the total ordering of the space.

References

1. D. F. Bailey, Krasnoselski's theorem on the real line, Amer. Math. Monthly, 81 (1974) 506–507.

2. M. A. Krasnoselski, Two remarks on the method of successive approximations, Uspehi Math. Nauk (N. S.), (1) 10 (1955) 123–127.

ON ALMOST RELATIVELY PRIME INTEGERS

ALAN H. STEIN, University of Connecticut

The probability that two integers are relatively prime is $1/\zeta(2)$, where ζ is the Riemann Zeta Function and $\zeta(2) = \pi^2/6$. One might then ask about the probability that two integers are "almost" relatively prime in some sense. More precisely, given a set P, what is the probability that the greatest common divisor of two integers is in P? It will be shown that this probability is that proportion of $\zeta(2)$ contributed by integers in P.

THEOREM 1. *Let P be a set of positive integers, let $C(P) = (\sum_{n \in P} 1/n^2)/\zeta(2)$ and let $P(x)$ = the number of pairs of positive integers both less than or equal to x whose greatest common divisor is in P. Then*

$$P(x) = C(P)x^2 + O(x \log^2 x). \tag{1}$$

To prove this we will need the special case $P = P_0 = \{1\}$, where we can obtain a stronger result:

(2) LEMMA. $P_0(x) = 1/\zeta(2)x^2 + O(x \log x)$.

Notation. The following notation will be used:
(3) $\nu(n)$ = the number of distinct prime divisors of n.
(4) $\mu(n)$ = the Mobius function.
(5) (m, n) = the greatest common divisor of m and n.
(6) $[x]$ = the greatest integer less than or equal to x.

Proof of the Lemma.

$$\begin{aligned} P_0(x) &= \sum_{m \leqq x} \sum_{\substack{n \leqq x \\ (m,n)=1}} 1 = \sum_{m \leqq x} \sum_{n \leqq x} \sum_{d \mid (m,n)} \mu(d) \\ &= \sum_{m \leqq x} \sum_{d \mid m} \mu(d)[x/d] = \sum_{m \leqq x} (x \sum_{d \mid m} \mu(d)/d + O(\sum_{d \mid m} 1)) \\ &= x \sum_{m \leqq x} \sum_{d \mid m} \mu(d)/d + O(\sum_{d \leqq x} x/d) \\ &= x \sum_{d \leqq x} (\mu(d)/d)[x/d] + O(x \log x). \end{aligned} \tag{7}$$

Since

$$\sum_{d \leqq x} (\mu(d)/d)[x/d] = x \sum_{d \leqq x} \mu(d)/d^2 + O(\log x) = x/\zeta(2) + O(\log x), \tag{8}$$

this yields (2).

Proof of Theorem 1. The key observation is that

$$P(x) = \sum_{\substack{d \in P \\ d \leqq x}} P_0(x/d). \tag{9}$$

This is true because if $(m,n) \in P$ then $m = m'(m,n)$, $n = n'(m,n)$ with $(m',n') = 1 \in P_0$. We can thus bootstrap our way up from the case of relatively prime integers. The lemma combined with (9) yields

$$P(x) = \sum_{\substack{d \in P \\ d \leqq x}} (x/d)^2/\zeta(2) + O\Big(\sum_{\substack{d \in P \\ d \leqq x}} (x/d)\log(x/d)\Big). \tag{10}$$

The first term is

$$(x^2/\zeta(2)) \sum_{\substack{d \in P \\ d \leqq x}} 1/d^2 = (x^2/\zeta(2))\big(\sum_{d \in P} 1/d^2 + O(1/x)\big) = C(P)x^2 + O(x). \tag{11}$$

The second term is

$$O\Big(x \log x \sum_{\substack{d \in P \\ d \leqq x}} 1/d\Big) = O(x \log^2 x). \tag{12}$$

Putting (10), (11) and (12) together yields (1). Q.E.D.

For certain sets P, far better errors can be obtained. For example, if P is finite then $\sum_{d \in P, d \leqq x} 1/d = O(1)$ so the error is $O(x \log x)$. Among infinite P, consider $P_e = \{d\colon d \text{ is a positive integer and } \nu(d) = e\}$ for some positive integer e. In this case, Theorem 427 of [1] yields

$$\sum_{\substack{d \in P_e \\ d \leqq x}} 1/d \leqq \Big(\sum_{\substack{\nu(d)=1 \\ d \leqq x}} 1/d\Big)^e = O(\log^e \log x) \tag{13}$$

so the error shrinks to $O(x(\log x)(\log\log x)^e)$.

This problem may also be tackled using analytic methods by considering the generating function $\sum_{(m,n) \in P} 1/(m^s n^t)$. Surprisingly, however, the best errors the author has been able to obtain using analytic methods have been much larger than the ones obtained very easily using elementary methods.

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford, London, 1968.

2. T. Nagell, Introduction to Number Theory, Chelsea, New York, 1964.

A GROUP THEORETIC PRESENTATION OF THE ALTERNATING GROUP ON FIVE SYMBOLS, A_5

EUGENE SCHENKMAN, Purdue University

Let $G = \langle a, b \mid a^5 = 1 = b^2 = (ab)^3 \rangle$. This means that G is the group generated by elements a, b subject to the relations to the right of the vertical bar. An expression for G in this form is called an abstract definition or presentation of G. Every group

This is true because if $(m,n) \in P$ then $m = m'(m,n)$, $n = n'(m,n)$ with $(m',n') = 1 \in P_0$. We can thus bootstrap our way up from the case of relatively prime integers. The lemma combined with (9) yields

$$P(x) = \sum_{\substack{d \in P \\ d \leq x}} (x/d)^2/\zeta(2) + O\Big(\sum_{\substack{d \in P \\ d \leq x}} (x/d)\log(x/d)\Big). \tag{10}$$

The first term is

$$(x^2/\zeta(2)) \sum_{\substack{d \in P \\ d \leq x}} 1/d^2 = (x^2/\zeta(2))\Big(\sum_{d \in P} 1/d^2 + O(1/x)\Big) = C(P)x^2 + O(x). \tag{11}$$

The second term is

$$O\Big(x \log x \sum_{\substack{d \in P \\ d \leq x}} 1/d\Big) = O(x \log^2 x). \tag{12}$$

Putting (10), (11) and (12) together yields (1). Q.E.D.

For certain sets P, far better errors can be obtained. For example, if P is finite then $\sum_{d \in P, d \leq x} 1/d = O(1)$ so the error is $O(x \log x)$. Among infinite P, consider $P_e = \{d : d \text{ is a positive integer and } \nu(d) = e\}$ for some positive integer e. In this case, Theorem 427 of [1] yields

$$\sum_{\substack{d \in P_e \\ d \leq x}} 1/d \leq \Big(\sum_{\substack{\nu(d)=1 \\ d \leq x}} 1/d\Big)^e = O(\log^e \log x) \tag{13}$$

so the error shrinks to $O(x(\log x)(\log\log x)^e)$.

This problem may also be tackled using analytic methods by considering the generating function $\sum_{(m,n) \in P} 1/(m^s n^t)$. Surprisingly, however, the best errors the author has been able to obtain using analytic methods have been much larger than the ones obtained very easily using elementary methods.

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford, London, 1968.

2. T. Nagell, Introduction to Number Theory, Chelsea, New York, 1964.

A GROUP THEORETIC PRESENTATION OF THE ALTERNATING GROUP ON FIVE SYMBOLS, A_5

EUGENE SCHENKMAN, Purdue University

Let $G = \langle a, b \mid a^5 = 1 = b^2 = (ab)^3 \rangle$. This means that G is the group generated by elements a, b subject to the relations to the right of the vertical bar. An expression for G in this form is called an abstract definition or presentation of G. Every group has a presentation, actually many presentations, and two groups are isomorphic if they have identical presentations. Also two finite groups are isomorphic if one is a homomorphic image of the other and their orders are identical. Our object is to show that G is isomorphic to the alternating group A_5 and also to the group of two by two matrices over the finite prime field Z_p modulo $\pm\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$.

We begin by observing that in A_5 if we consider $A = (12345)$ of order 5 and $B = (12)(34)$ of order 2, then $AB = (245)$ has order 3. Thus A_5 satisfies the defining relations of G and hence A_5 is a homomorphic image of G since A, B generate A_5. The latter follows from the fact that $H = \langle A, B\rangle$ contains subgroups of orders $5(K = \langle A\rangle)$, $3(L = \langle AB\rangle)$ and $4(M = \langle BA^2BA^3BA^2, B\rangle)$ which implies that H has to have order at least 60. We also note that if $\alpha = \left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)$ and $\beta = \left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$ then $\alpha, \beta, \alpha\beta$ have orders 5, 2 and 3 respectively. Since α and β generate the group of two by two matrices over Z_5 modulo $\pm\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$, this group also satisfies the defining relations of G and hence is also a homomorphic image of G.

If we prove that G has at most sixty elements we can conclude (since A_5 and the above matrix group have exactly sixty elements) that G has exactly sixty elements and is therefore isomorphic to the above groups. We prove that G has at most sixty elements as follows.

Let $d = a^b$, the conjugate of a by b. Then $d = bab$ has order 5 (being a conjugate of a) and $dad = b$ has order 2. Consequently $b^d = ad^2$ has order 2 and $ad^2ad^2 = 1$ or $d^2ad^2 = a^{-1}$. Then $a^{d^2} = d^3ad^2 = da^{-1}$ and $(da^{-1})^b = d^b(a^{-1})^b = ad^{-1}$ since $a^b = d$ and $d^b = a$. Thus conjugation by b inverts $da^{-1} = a^{d^2}$. It follows that a is inverted when conjugated by $d^2bd^3 = ba^2ba^3b$ (an element of order 2 since it is a conjugate of b). So a and ba^2ba^3b generate a dihedral group any of whose elements not a power of a have order 2. In particular $h = ba^2ba^3ba^2$ has order 2 and inverts a.

Consider now $a^3d^2a^3 = a^3ba^2ba^3$. Since $(ad^2)^2 = 1$ and $a^5 = 1$, $a^3d^2a^3a^3d^2a^3 = 1$. Hence $a^3ba^2ba^3$ has order 2 and is equal to its inverse $a^2ba^3ba^2$. Since h, b and bh have order 2, h and b generate a four-group and h commutes with b.

We now consider the set S of elements a^i, a^ih, a^iba^j, a^ihba^j for $i,j = 0,1,2,3,4$. There are at most sixty such elements. We shall show that $Sa \subseteq S$ and $Sb \subseteq S$ and hence that $G \subseteq S$ since G is generated by a and b. By the form of the elements of S and the fact that $aha = h$ (from $h^{-1}ah = a^{-1}$ and $h = h^{-1}$), it is clear that $Sa \subseteq S$. To show that $Sb \subseteq S$ the only problems come from elements of the form a^iba^jb and a^ihba^jb. First we show that $ba^jb \in S$. Since $(ab)^3 = 1$, $bab = a^{-1}ba^{-1} = a^4ba^4 \in S$ and $ba^{-1}b = ba^4b = aba \in S$. Since $h^2 = 1$, $h = h^{-1} = a^3ba^2ba^3b$ from which it follows that $ba^2b = a^2hba^2 \in S$. From $h = ba^2ba^3ba^2$, it follows that $ba^3b = a^3bha^3 = a^3hba^3 \in S$. So we have $ba^jb \in S$. Hence $a^iba^jb \in S$ since clearly $a^iS \subseteq S$. Similarly we have $a^ihba^jb \in S$ since $a^ihS \subseteq S$. The latter follows from the facts that, $hb = bh$, $h = h^{-1}$, $ha^i = a^{-i}h$ and $ha^ih = a^{-i}$. Consequently $Sb \subseteq S$ and so we have $G \subseteq S$. Hence we have shown that G has at most sixty elements.

The author is indebted to Johnny L. Houston for help in rewriting the final draft.

ON THE NUMBER OF SUBGROUPS OF INDEX TWO — AN APPLICATION OF GOURSAT'S THEOREM FOR GROUPS

R. R. CRAWFORD and K. D. WALLACE, Western Kentucky University

Bruckheimer, Bryan, and Muir [1] studied groups that could be expressed as the union of three proper subgroups ("3-groups") and thereby provided a paper delightfully appropriate and accessible to the student enrolled in an introductory course in group theory. The characterization obtained in [1] for 3-groups was: A group G is a 3-group if and only if there exists a homomorphism from G onto the Klein Four Group. The following observations are easy consequences of this characterization:

(1) G is a 3-group if and only if G contains more than one subgroup of index two.

(2) If A and B are distinct subgroups of index two in the group G then there exists a unique proper subgroup C of G such that $G = A \cup B \cup C$. Moreover $C = (A \cap B) \cup (G - (A \cup B))$ and C has index two in G.

Let $I_2(G)$ denote the number of distinct subgroups of index two in the group G. The above observations show that the subgroups of index two of a group G for which $I_2(G) > 1$ and $I_2(G)$ is finite form a Steiner triple system and consequently that $I_2(G)$ is congruent to either 1 or 3 modulo 6.

The purpose of this note is to provide an elementary application of Goursat's Theorem for Groups in determining which nonnegative integers occur as the number, $I_2(G)$, of distinct subgroups of index two in some group G. It should be noted that this application of Goursat's Theorem can easily be modified to resolve the analogous problem for normal subgroups of prime index p.

The following lemma allows us to restrict consideration to finite Abelian groups and hence provides a significant reduction to the problem.

LEMMA 1. *Let G be a group such that $I_2(G)$ is finite. Then there exists a finite Abelian group A having order a power of two such that $I_2(G) = I_2(A)$.*

Proof. Let $H_1, \cdots, H_n$ be the distinct subgroups of index two in G and let $H = H_1 \cap \cdots \cap H_n$. Since each H_i is normal in G and contains the commutator subgroup of G, H enjoys these same properties. Thus G/H is Abelian and $|G/H| = [G\colon H] = [G\colon H_1 \cap \cdots \cap H_n] = 2^n$. Finally, the correspondence established by the fundamental theorem on group homomorphisms yields a bijection between the subgroups of index two in G/H and the subgroups of index two in G, so we may take $A = G/H$.

Since any finite Abelian group may be expressed as a direct product of cyclic groups, it is clear that a result giving a relationship between the subgroups of a product $G \times H$ and the subgroups of the factors G and H would be very useful. Goursat's Theorem provides precisely such a relationship. Although this theorem is well known (see Lambek [2]), it has perhaps been neglected by authors at an elementary level.

THEOREM 2 (Goursat's Theorem for Groups). *Let G and H be groups. Then there is a bijection between the set S of all subgroups of $G \times H$ and the set T of all 5-tuples (A, B, C, D, f) where $B \trianglelefteq A \leqq G$, $D \trianglelefteq C \leqq H$, and $f: A/B \to C/D$ is an isomorphism (here $\leqq$ denotes subgroup and $\trianglelefteq$ denotes normal subgroup).*

Proof. We content ourselves with defining the bijection and its inverse, leaving the details of the verification, which are all routine, to the reader. Given a subgroup $K \leqq G \times H$, we let

$$A_K = \{g \in G \mid (g,h) \in K \text{ for some } h \in H\},$$

$$B_K = \{g \in G \mid (g,1) \in K\},$$

$$C_K = \{h \in H \mid (g,h \in K \text{ for some } g \in G\},$$

$$D_K = \{h \in H \mid (1,h) \in K\},$$

and define $_K$: $A_K/B_K \to C_K/D_K$ by $f_K(gB_K) = hD_K$ if $(g,h) \in K$. Then we may obtain a map $\theta: S \to T$ by setting $\theta(K) = (A_K, B_K, C_K, D_K, f_K)$. Finally, define $\Gamma: T \to S$ by $\Gamma(A,B,C,D,f) = \{(g,h) \in A \times C \mid f(gB) = hD\}$.

We retain the notation of Theorem 2 for the remainder of the paper.

COROLLARY 3. *Let G and H be finite groups. Then* $|A_K| \cdot |D_K| = |K| = |B_K| \cdot |C_K|$.

Proof. Given any $g \in A_K$ and $h \in C_K$, we have $(g,h) \in K$ if and only if $f_K(gB_K) = hD_K$ which is in turn equivalent to $h \in f_K(gB_K)$. Thus each element of A_K is paired, in K, with $|f_K(gB_K)| = |D_K|$ elements. There being $|A_K|$ distinct choices for g, we have $|K| = |A_K| \cdot |D_K|$. The second equality follows by symmetry.

COROLLARY 4. *Let G and H be finite groups. Then* $[G : A_K] \cdot [H : D_K] = [G \times H : K] = [G : B_K] \cdot [H : C_K]$.

COROLLARY 5. *Let G and H be finite groups. Then* $I_2(G \times H) = I_2(G) + I_2(H) + I_2(G)I_2(H)$.

Proof. By Corollary 4, K has index two in $G \times H$ if and only if $[G: A_K] \cdot [H: D_K] = 2 = [G: B_K] \cdot [H: C_K]$. Note that the factor groups A_K/B_K and C_K/D_K must have order either 1 or 2 and consequently there is a unique isomorphism $f: A_K/B_K \to C_K/D_K$. We are thus led to consideration of the following three cases:

Case (I) $[G: A_K] = 2$ and $[A_K: B_K] = 1$. In this case $C_K = D_K = H$. There being $I_2(G)$ choices for A_K, we obtain $I_2(G)$ distinct 5-tuples (A_K, A_K, H, H, f) and consequently $I_2(G)$ distinct subgroups of index two in $G \times H$.

Case (II) $[G: A_K] = 1$ and $[A_K: B_K] = 1$. It follows that $C_K = D_K$ is a subgroup of index two in H and $I_2(H)$ subgroups of index two in $G \times H$ are obtained.

Case (III) $[G:A_K] = 1$ and $[A_K:B_K] = 2$. In this case $[C_K:D_K] = 2$ and $C_K = H$. Consequently, there are $I_2(G)$ choices for B_K, $I_2(H)$ choices for D_K, and hence $I_2(G)I_2(H)$ subgroups of index two in $G \times H$ determined by this case.

Since each cyclic group of even order contains a unique subgroup of index two, we have:

COROLLARY 6. *Let n be a nonnegative integer and G the direct product of n cyclic groups each of even order. Then* $I_2(G) = 2^n - 1$.

Combining Lemma 1 and Corollary 6, we obtain:

THEOREM 7. *Let n be a nonnegative integer. There exists a group G such that* $I_2(G) = n$ *if and only if* $n = 2^k - 1$ *for some nonnegative integer k.*

Remarks. For p a prime, let $N_p(G)$ denote the number of distinct normal subgroups of G having index p. Apparent modifications in the preceding development yield:
(1) $N_p(G \times H) = N_p(G) + N_p(H) + (p-1)N_p(G)N_p(H)$; and (2) if $N_p(G)$ is finite, then $N_p(G) = (p^k - 1)/(p-1)$ for some nonnegative integer k.

References

1. M. Bruckheimer, A. C. Bryan and A. Muir, Groups which are the union of three subgroups, Amer. Math. Monthly, 77 (1970) 52–57.
2. J. Lambek, Goursat's theorem and the Zassenhaus lemma, Canad. J. Math., 10 (1958) 45–56.

ON THE REPRESENTATION OF A POSSIBLE SOLUTION SET OF FERMAT'S LAST THEOREM

CHARLES J. MIFSUD, The Mitre Corporation

Introduction. It is my intention in this article to report a rather interesting fact about the solution of

$$a^n + b^n = c^n \tag{1}$$

which might have a special appeal to readers of this journal. The presentation does not represent any advance or improvement in the status of the solution of Fermat's Last Theorem.

Instead of concentrating our attention on proving that for certain values of n equation (1) is impossible, as has been done in past research, we will direct our attention to the possible range of values of the number triplet (a, b, c) that might satisfy (1). Thus, a computer programmer planning to make a test of a particular triplet for a given value of n would be concerned with determining the maximum number of digits of each member of the triplet. Thus, if c is a 10 digit number and $n = 100$, then the programmer must have at his disposal a multiple-precision package of order 100 which can operate on 1000 digit operands. The facts to be unfolded here are quite easy to understand, and all information needed in the analysis is self-contained.

We will determine a lower bound for the number of digits required to represent

Since each cyclic group of even order contains a unique subgroup of index two, we have:

COROLLARY 6. *Let n be a nonnegative integer and G the direct product of n cyclic groups each of even order. Then* $I_2(G) = 2^n - 1$.

Combining Lemma 1 and Corollary 6, we obtain:

THEOREM 7. *Let n be a nonnegative integer. There exists a group G such that* $I_2(G) = n$ *if and only if* $n = 2^k - 1$ *for some nonnegative integer k.*

Remarks. For p a prime, let $N_p(G)$ denote the number of distinct normal subgroups of G having index p. Apparent modifications in the preceding development yield:

(1) $N_p(G \times H) = N_p(G) + N_p(H) + (p-1)N_p(G)N_p(H)$; and (2) if $N_p(G)$ is finite, then $N_p(G) = (p^k - 1)/(p-1)$ for some nonnegative integer k.

References

1. M. Bruckheimer, A. C. Bryan and A. Muir, Groups which are the union of three subgroups, Amer. Math. Monthly, 77 (1970) 52–57.

2. J. Lambek, Goursat's theorem and the Zassenhaus lemma, Canad. J. Math., 10 (1958) 45–56.

ON THE REPRESENTATION OF A POSSIBLE SOLUTION SET OF FERMAT'S LAST THEOREM

CHARLES J. MIFSUD, The Mitre Corporation

Introduction. It is my intention in this article to report a rather interesting fact about the solution of

$$a^n + b^n = c^n \tag{1}$$

which might have a special appeal to readers of this journal. The presentation does not represent any advance or improvement in the status of the solution of Fermat's Last Theorem.

Instead of concentrating our attention on proving that for certain values of n equation (1) is impossible, as has been done in past research, we will direct our attention to the possible range of values of the number triplet (a, b, c) that might satisfy (1). Thus, a computer programmer planning to make a test of a particular triplet for a given value of n would be concerned with determining the maximum number of digits of each member of the triplet. Thus, if c is a 10 digit number and $n = 100$, then the programmer must have at his disposal a multiple-precision package of order 100 which can operate on 1000 digit operands. The facts to be unfolded here are quite easy to understand, and all information needed in the analysis is self-contained.

We will determine a lower bound for the number of digits required to represent at least one member of the number triplet that might satisfy (1) since this will dictate the minimum precision requirements.

Necessary background. Our analysis is based on the following three Lemmas:

LEMMA 1. *First we note that any two members of* (1) *are relatively prime since if any two had a factor in common then this factor would also divide the third member. Thus, a factor common to all three members could be removed without affecting equality* (1).

LEMMA 2. *We next note that if* (1) *holds, one of the number triplets must be even. If more than one was even, then Lemma* 1 *which states that any two are relatively prime would be contradicted. All three members cannot be odd since the sum or difference of any two odd members is an even number, and hence, the third member must be even.*

LEMMA 3. *If n is odd and if A and B are odd, then expressions*

$$(A^n + B^n)/(A + B) \text{ and } (A^n - B^n)/(A - B) \text{ are odd.}$$

$$(A^n \pm B^n)/(A \pm B) = A^{n-1} \mp A^{n-2}B + \cdots \mp AB^{n-2} + B^{n-1}.$$

If A and B are odd, then the expression on the right contains an odd number of terms and can be represented as

$$\sum_{i=1}^{2l+1} (2k_i + 1) = 2 \sum_{i=1}^{2l+1} k_i + 2l + 1,$$

which is odd.

Concluding Results. If we place the even term of (1) on one side and the remaining odd terms on the other side and then apply Lemma 3, one of the three equations below will follow:

$$c - b = 2^n r, \; r \geqq 1, \tag{2}$$

$$c - a = 2^n s, \; s \geqq 1, \tag{3}$$

$$a + b = 2^n t, \; t \geqq 1. \tag{4}$$

We first observe that $c > 2^n$ in either equations (2) or (3). And from equation (4) we conclude that either $a \geqq 2^{n-1}$ or $b \geqq 2^{n-1}$ since if $a < 2^{n-1}$ and $b < 2^{n-1}$ then $a + b < 2^n$ contrary to (4).

The results of equations (2), (3) and (4) are credited to Rowe [**1**].

As a result of this analysis we must conclude that one of the number triplets must be as great as 2^{n-1} if n satisfies (1). Next we take note that it has been reported in [**2**] that equation (1) is impossible for values $n \leqq 25{,}000$. Therefore, in order to test the validity of (1) with meaningful quantities, we must be able to manipulate numbers in the computer that lie in the range of $(2^{24{,}999})^{25{,}000} = 2^{624{,}975{,}000}$. If we assume 3.33 binary bits $\cong$ 1 decimal digit in precision, then we must work with decimal numbers in the range of $10^{187{,}680{,}180}$. This means we must deal with numbers having almost 188 million digits of precision. It is highly unlikely that a solution set of Fermat's Last Theorem will ever be discovered with the use of a computer with such huge number representations (storage) and with such high order precision (computer time) required before adequate testing can be performed.

References

1. Abstract, Bull. Amer. Math. Soc., 20 (1913) 68–69.
2. Selfridge and Pollock, Notices Amer. Math. Soc., January 1964, p. 97.

NOTE ON NON-EUCLIDEAN PRINCIPAL IDEAL DOMAINS

KENNETH S. WILLIAMS, Carleton University, Ottawa

It is well known that every Euclidean domain is a principal ideal domain. In [4] Wilson's object is to show, in a manner accessible to students in an undergraduate algebra class, that the ring of integers of the field $Q(\sqrt{-19})$ is a principal ideal domain which is not Euclidean. The proof that it is a principal ideal domain is based on a theorem of Dedekind and Hasse and the proof that it is not Euclidean is based upon the work of Motzkin [2]. However, as much of what Wilson does in the latter proof is unnecessary for the required purpose, it is our purpose to give a simpler treatment.

If D is an integral domain, we let $\tilde{D}$ denote the collection of units of D together with 0, so that $D - \tilde{D} = \emptyset$ if and only if D is a field. An element $u \in D - \tilde{D}$ is called a universal side divisor if for any $x \in D$ there exists some $z \in \tilde{D}$ such that $u \mid x - z$.

THEOREM. *Let D be an integral domain which is not a field (so that $D - \tilde{D} \neq \emptyset$) and which has no universal side divisors. Then D is not Euclidean.*

Proof. Suppose that D is a Euclidean domain, with Euclidean function d, which has no universal side divisors. Consider the nonempty subset $S = \{d(v): v \in D - \tilde{D}\}$ of the nonnegative integers. It possesses a least element, say $d(u)$, $u \in D - \tilde{D}$. For any $x \in D$ there exists $y, z \in D$ such that $x = uy + z$, where either (i) $z = 0$ or (ii) $z \neq 0$ and $d(z) < d(u)$. If (i) holds then $u \mid x$. If (ii) holds, by the minimality of $d(u)$, z must be a unit. Thus in both cases $u \mid x - z$ for some $z \in \tilde{D}$, and so u is a universal side divisor which is impossible.

COROLLARY. *The rings of integers of $Q(\sqrt{-19})$, $Q(\sqrt{-43})$, $Q(\sqrt{-67})$, $Q(\sqrt{-163})$ are not Euclidean.*

Proof. Let $D = 19, 43, 67$ or 163 and suppose that the ring

$$R = \{a + b(1 + \sqrt{-D})/2:\ a, b \text{ integers}\}$$

of integers of $Q(\sqrt{-D})$ contains a universal side divisor u. As the only units of R are ± 1, u must be a nonunit divisor of 2 or 3. In R, 2 and 3 are irreducible and therefore the only possible universal side divisors are 2, -2, 3, and -3.

numbers having almost 188 million digits of precision. It is highly unlikely that a solution set of Fermat's Last Theorem will ever be discovered with the use of a computer with such huge number representations (storage) and with such high order precision (computer time) required before adequate testing can be performed.

References

1. Abstract, Bull. Amer. Math. Soc., 20 (1913) 68–69.
2. Selfridge and Pollock, Notices Amer. Math. Soc., January 1964, p. 97.

NOTE ON NON-EUCLIDEAN PRINCIPAL IDEAL DOMAINS

KENNETH S. WILLIAMS, Carleton University, Ottawa

It is well known that every Euclidean domain is a principal ideal domain. In [**4**] Wilson's object is to show, in a manner accessible to students in an undergraduate algebra class, that the ring of integers of the field $Q(\sqrt{-19})$ is a principal ideal domain which is not Euclidean. The proof that it is a principal ideal domain is based on a theorem of Dedekind and Hasse and the proof that it is not Euclidean is based upon the work of Motzkin [**2**]. However, as much of what Wilson does in the latter proof is unnecessary for the required purpose, it is our purpose to give a simpler treatment.

If D is an integral domain, we let $\tilde{D}$ denote the collection of units of D together with 0, so that $D-\tilde{D}=\varnothing$ if and only if D is a field. An element $u\in D-\tilde{D}$ is called a universal side divisor if for any $x\in D$ there exists some $z\in\tilde{D}$ such that $u\,|\,x-z$.

THEOREM. *Let D be an integral domain which is not a field (so that $D-\tilde{D}\neq\varnothing$) and which has no universal side divisors. Then D is not Euclidean.*

Proof. Suppose that D is a Euclidean domain, with Euclidean function d, which has no universal side divisors. Consider the nonempty subset $S=\{d(v):v\in D-\tilde{D}\}$ of the nonnegative integers. It possesses a least element, say $d(u)$, $u\in D-\tilde{D}$. For any $x\in D$ there exists $y,z\in D$ such that $x=uy+z$, where either (i) $z=0$ or (ii) $z\neq 0$ and $d(z)<d(u)$. If (i) holds then $u\,|\,x$. If (ii) holds, by the minimality of $d(u)$, z must be a unit. Thus in both cases $u\,|\,x-z$ for some $z\in\tilde{D}$, and so u is a universal side divisor which is impossible.

COROLLARY. *The rings of integers of $Q(\sqrt{-19})$, $Q(\sqrt{-43})$, $Q(\sqrt{-67})$, $Q(\sqrt{-163})$ are not Euclidean.*

Proof. Let $D=19, 43, 67$ or 163 and suppose that the ring

$$R=\{a+b(1+\sqrt{-D})/2:\ a,b \text{ integers}\}$$

of integers of $Q(\sqrt{-D})$ contains a universal side divisor u. As the only units of R are ± 1, u must be a nonunit divisor of 2 or 3. In R, 2 and 3 are irreducible and therefore the only possible universal side divisors are 2, -2, 3, and -3.

However, none of these divides any of the integers

$$\tfrac{1}{2}(1+\sqrt{-D}),\ \tfrac{1}{2}(3+\sqrt{-D}),\ \tfrac{1}{2}(-1+\sqrt{-D}),$$

so that no such universal side divisor u can exist. Hence by the theorem, R is not Euclidean.

Recently Stark [3] has shown that the only complex quadratic fields $Q(\sqrt{-D})$ whose rings of integers are principal ideal domains are given by $D = 1, 2, 3, 7, 11, 19, 43, 67, 163$ and since it is well known [1] that the first five of these are Euclidean (with respect to the norm), the above corollary gives all the complex quadratic fields whose rings of integers are non-Euclidean principal ideal domains.

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford, New York, 1962, Theorem 246, p. 213.

2. Th. Motzkin, The Euclidean algorithm, Bull. Amer. Math. Soc., 55 (1949) 1142–1146.

3. H. M. Stark, A complete determination of the complex quadratic fields of class-number one, Michigan Math. J., 14 (1967) 1–27.

4. J. C. Wilson, A principal ideal ring that is not a Euclidean ring, this MAGAZINE, 46 (1973) 34–38.

NOTES AND COMMENTS

David Kullman notes that the article *5-con triangles* by Richard G. Pawley in the Mathematics Teacher, May 1967, vol. 60, pp. 438–443, includes results equivalent to several of those in the article *Almost congruent triangles* by Bruce B. Peterson in our September 1974 issue.

From Aaron R. Todd regarding *Almost congruent triangles*, this MAGAZINE, vol. 47, Sept. 1974, No. 4 by Robert T. Jones and Bruce Peterson: The authors should mention their use of continuity in at least one of the several places they need the concept, especially as they ban it so forcibly in their proof of the existence of almost congruent triangles. For that matter, a principle of continuity such as the following is useful in filling a gap in traditional constructions of triangles: *If a point on the arc of a circle lies on one side of a straight line and another point of the arc lies on the other side of the line, then there is a point of the arc lying on the line.*

Comment by Graham Lord, Temple University, Philadelphia, Pa. on *A note on Mersenne numbers* by Steve Ligh and Larry Neal, this MAGAZINE, vol. 47, September 1974, No. 4, pp. 231–233. Theorem 1 of the note is just a special case of the result: If n is a natural number > 1, then $2^n - 1$ is not the mth power of a natural number $m > 1$. This result is quoted in *Elementary theory of numbers* by W. Sierpinski who gives as a reference C. G. Gerono, Nouv. Ann. Math. (2) 9(1870) pp. 469–471, 10 (1871) pp. 204–206.

However, none of these divides any of the integers

$$\tfrac{1}{2}(1+\sqrt{-D}),\ \tfrac{1}{2}(3+\sqrt{-D}),\ \tfrac{1}{2}(-1+\sqrt{-D}),$$

so that no such universal side divisor u can exist. Hence by the theorem, R is not Euclidean.

Recently Stark [3] has shown that the only complex quadratic fields $Q(\sqrt{-D})$ whose rings of integers are principal ideal domains are given by $D = 1, 2, 3, 7, 11, 19, 43, 67, 163$ and since it is well known [1] that the first five of these are Euclidean (with respect to the norm), the above corollary gives all the complex quadratic fields whose rings of integers are non-Euclidean principal ideal domains.

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford, New York, 1962, Theorem 246, p. 213.

2. Th. Motzkin, The Euclidean algorithm, Bull. Amer. Math. Soc., 55 (1949) 1142–1146.

3. H. M. Stark, A complete determination of the complex quadratic fields of class-number one, Michigan Math. J., 14 (1967) 1–27.

4. J. C. Wilson, A principal ideal ring that is not a Euclidean ring, this MAGAZINE, 46 (1973) 34–38.

NOTES AND COMMENTS

David Kullman notes that the article *5-con triangles* by Richard G. Pawley in the Mathematics Teacher, May 1967, vol. 60, pp. 438–443, includes results equivalent to several of those in the article *Almost congruent triangles* by Bruce B. Peterson in our September 1974 issue.

From Aaron R. Todd regarding *Almost congruent triangles*, this MAGAZINE, vol. 47, Sept. 1974, No. 4 by Robert T. Jones and Bruce Peterson: The authors should mention their use of continuity in at least one of the several places they need the concept, especially as they ban it so forcibly in their proof of the existence of almost congruent triangles. For that matter, a principle of continuity such as the following is useful in filling a gap in traditional constructions of triangles: *If a point on the arc of a circle lies on one side of a straight line and another point of the arc lies on the other side of the line, then there is a point of the arc lying on the line.*

Comment by Graham Lord, Temple University, Philadelphia, Pa. on *A note on Mersenne numbers* by Steve Ligh and Larry Neal, this MAGAZINE, vol. 47, September 1974, No. 4, pp. 231–233. Theorem 1 of the note is just a special case of the result: If n is a natural number > 1, then $2^n - 1$ is not the mth power of a natural number $m > 1$. This result is quoted in *Elementary theory of numbers* by W. Sierpinski who gives as a reference C. G. Gerono, Nouv. Ann. Math. (2) 9(1870) pp. 469–471, 10 (1871) pp. 204–206.

From E. T. Steller: Zalman Usiskin in his article in this MAGAZINE, 46 (1973) pages 203–208 suggests at the end "possible explorations," the second one of which reads: "Do the perfect nth power patterns hold for the standard generalizations of the Pascal triangle?"

It seems to me that the answer to this question must be negative for the following reason:

Looking at the first appearance of a prime number p in the Pascal triangle it is obvious that this happens in the pth line, second column and that all elements of this line except the first and last have p as a factor. Furthermore p is a factor of all elements in the following line excepting of course the first two and the last two. Looking at line no. $(p+2)$ all elements except six have p as a factor. This of course follows immediately from the relation

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

If I may be so bold I would like to suggest a change in the notation $\binom{n}{k}$ for the coefficient of x^k in the expansion of $(1+x)^n$, so that it can be used in the expansion of $(1+x+x^2+\cdots+x^l)^n$ as well. The change is the addition of a subscript indicating the number of terms in the polynomial.

Equation (1) would then be written

$$\binom{n+1}{k}_2 = \binom{n}{k-1}_2 + \binom{n}{k}_2 = \sum_{j=0}^{1} \binom{n}{k-j}_2. \tag{1a}$$

In general we would have

$$\binom{n+1}{k}_l = \sum_{j=0}^{l-1} \binom{n}{k-j}_l. \tag{2}$$

The first lines of Pascal's triangle for $l = 2, 3, 4$ are as follows:

$l = 2$

1	1				
1	2	1			
1	3	3	1		
1	4	6	4	1	
1	5	10	10	5	1

$l = 3$

1	1	1								
1	2	3	2	1						
1	3	6	7	6	3	1				
1	4	10	$\overline{16}$	19	16	10	4	1		
1	5	15	30	$\overline{45}$	$\underline{51}$	45	30	15	5	1

$l = 4$

1	1	1	1												
1	2	3	4	3	2	1									
1	3	6	10	12	12	10	6	3	1						
1	4	10	20	31	40	44	40	31	20	10	4	1			
1	5	15	35	$\overline{65}$	$\underline{101}$	135	$\underline{155}$	$\underline{155}$	135	101	65	35	15	5	1

Now looking at the generalization with $l = 3$, we find that 7 already appears in line 3, 19 in line 4. But because we do not have relation (1), 7 appears only once in that line and of course not at all in the next line and only 4 times in line 6.

Again the generalization with $l = 4$ produces 31 in the fourth line and 101 in the fifth line. So it would seem that to even find one perfect square pattern, the pattern must span many more than three lines, but doing this increases the likelihood of finding larger and unmatched primes in the lower lines.

Perhaps it is defeatist to consider "Exploration 2" closed, but because of equation (2) and its implications I do not hold out much hope for a simple and general nth power pattern even for $n = 2$.

BOOK REVIEWS

EDITED BY ADA PELUSO AND WILLIAM WOOTON

*Materials intended for review should be sent to: Professor Ada Peluso, Department of Mathematics, Hunter College of CUNY, 695 Park Avenue, New York, New York 10021, or to Professor William Wooton, 1495 La Linda Drive, Lake San Marcos, California 92069. A boldface capital **C** in the margin indicates that a review is based in part on classroom use.*

Discrete Mathematical Structures and Their Applications. By Harold S. Stone. Science Research Associates, Inc., Chicago, 1973. 387 + xii pp. $11.95.

This book is a well-written introduction to the parts of mathematics frequently needed by students in computer engineering and rarely found together in one course. The chapters cover elementary material on sets, relations and functions, groups, Polya's theory of counting, coding theory, semigroups, finite state machines, rings and fields, linear finite state machines, and switching functions. The book contains good exercises, a good bibliography, and a complete index and index of notation. The material is nicely presented and, in addition to the technical details, the history and applications of the material are discussed.

The author could have made the book more versatile by including material applicable to the needs of mathematically oriented computer science (as opposed to computer engineering) students. In particular, material on graphs, recurrence relations and growth rates, generating functions, permutations and combinations, computability, and self-reproduction of machines would be more relevant than the material on semigroups, rings, fields, etc.

E. M. REINGOLD, University of Illinois at Urbana-Champaign

Now looking at the generalization with $l = 3$, we find that 7 already appears in line 3, 19 in line 4. But because we do not have relation (1), 7 appears only once in that line and of course not at all in the next line and only 4 times in line 6.

Again the generalization with $l = 4$ produces 31 in the fourth line and 101 in the fifth line. So it would seem that to even find one perfect square pattern, the pattern must span many more than three lines, but doing this increases the likelihood of finding larger and unmatched primes in the lower lines.

Perhaps it is defeatist to consider "Exploration 2" closed, but because of equation (2) and its implications I do not hold out much hope for a simple and general nth power pattern even for $n = 2$.

BOOK REVIEWS

EDITED BY ADA PELUSO AND WILLIAM WOOTON

Materials intended for review should be sent to: Professor Ada Peluso, Department of Mathematics, Hunter College of CUNY, 695 Park Avenue, New York, New York 10021, or to Professor William Wooton, 1495 La Linda Drive, Lake San Marcos, California 92069. A boldface capital **C** *in the margin indicates that a review is based in part on classroom use.*

Discrete Mathematical Structures and Their Applications. By Harold S. Stone. Science Research Associates, Inc., Chicago, 1973. 387 + xii pp. $11.95.

This book is a well-written introduction to the parts of mathematics frequently needed by students in computer engineering and rarely found together in one course. The chapters cover elementary material on sets, relations and functions, groups, Polya's theory of counting, coding theory, semigroups, finite state machines, rings and fields, linear finite state machines, and switching functions. The book contains good exercises, a good bibliography, and a complete index and index of notation. The material is nicely presented and, in addition to the technical details, the history and applications of the material are discussed.

The author could have made the book more versatile by including material applicable to the needs of mathematically oriented computer science (as opposed to computer engineering) students. In particular, material on graphs, recurrence relations and growth rates, generating functions, permutations and combinations, computability, and self-reproduction of machines would be more relevant than the material on semigroups, rings, fields, etc.

E. M. REINGOLD, University of Illinois at Urbana-Champaign

PROBLEMS AND SOLUTIONS

EDITED BY DAN EUSTICE, The Ohio State University

ASSOCIATE EDITOR, J. S. FRAME, Michigan State University; Assistant Editors: DON BONAR, Denison University, WILLIAM MCWORTER, JR. and L. F. MEYERS, The Ohio State University

Readers of this department are invited to submit for solution problems believed to be new that may arise in study, in research, or in extra-academic situations. Problems may be submitted from any branch of mathematics and ranging in subject content from that accessible to the talented high school student to problems challenging to the professional mathematician. Proposals should be accompanied by solutions, when available, and by any information that will assist the editor. Ordinarily, problems in well-known textbooks should not be submitted.

The asterisk (*) will be placed by the problem number to indicate that the proposer did not supply a solution. Readers' solution are solicited for all problems. Proposers' solutions may not be "best possible" and solutions by others will be given preference.

Send all communications for this department to Dan Eustice, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.

To be considered for publication, solutions should be mailed before December 1, 1975.

PROPOSALS

937. *Proposed by Norman Schaumberger, Bronx Community College.*

Which is greater: e^{π} or $(e^e \cdot \pi^e \cdot \pi^{\pi})^{1/3}$?

938. *Proposed by S. C. Geller and W. C. Waterhouse, Cornell University.*

Let Σa_n be an infinite series, and set $s_n = a_1 + \cdots + a_n$. A familiar theorem of Abel says that if the a_n are positive and Σa_n diverges, then $\Sigma(a_n/s_n)$ also diverges. If we allow arbitrary signs, can we make Σa_n diverge to $+\infty$ while $\Sigma(a_n/s_n)$ converges?

939. *Proposed by Richard A. Gibbs, Fort Lewis College.*

Consider an $n \times n \times n$ cube consisting of n^3 unit cubes. Using only the unit cubes, determine, in terms of n: (1) the number of possible sizes of rectangular parallelepipeds "imbedded" in the cube, (2) the number of cubes of all sizes "imbedded" in the cube, and (3) the number of rectangular parallelepipeds of all sizes "imbedded" in the cube.

940. *Proposed by Edwin P. McCravy, Midlands Technical Education Center.*

To elaborate on an old problem of Dudeney [1], let us suppose a clock has minute and hour hands of the same length and indistinguishable. [Ignore the fact that on all clocks the hour hand is the one nearer the clock face.] Of the set of all instants in a 12-hour period, consider the partition:

A = set of all instants when the clock reading would be ambiguous;

B = set of all instants when the reading would not be ambiguous.

Which, if either, of these sets is finite?

---

1. H. E. Dudeney, 536 Puzzles and Curious Problems, edited by M. Gardner, Charles Scribner's Sons, New York, 1967, p. 14.

941. *Proposed by Stanley Rabinowitz, Maynard, Mass.*

Show that each of the following expressions is equal to the nth Legendre polynomial.

$$\text{(i)}\ \frac{1}{n!}\begin{vmatrix} x & 1 & 0 & 0 & & \cdots & 0 \\ 1 & 3x & 2 & 0 & & \cdots & 0 \\ 0 & 2 & 5x & 3 & & \cdots & 0 \\ 0 & 0 & 3 & & & & \vdots \\ \vdots & \vdots & \vdots & & & & n-1 \\ 0 & 0 & 0 & \cdots & & n-1 & (2n-1)x \end{vmatrix},\ \text{(ii)}\ \frac{1}{n!}\begin{vmatrix} x & 1 & 0 & 0 & & \cdots & 0 \\ 1 & 3x & 1 & 0 & & \cdots & 0 \\ 0 & 4 & 5x & 1 & & \cdots & 0 \\ 0 & 0 & 9 & \ddots & & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & & 1 \\ 0 & 0 & 0 & \cdots & & (n-1)^2 & (2n-1)x \end{vmatrix}.$$

942. *Proposed by M. S. Klamkin, University of Waterloo.*

Determine the maximum value of

$$S = \sum_{1 \leqq i < j \leqq n} \left(\frac{x_i x_j}{1 - x_i} + \frac{x_i x_j}{1 - x_j} \right)$$

where $x_i \geqq 0$ and $x_1 + x_2 + \cdots + x_n = 1$.

943. *Proposed by Charles W. Trigg, San Diego, California.*

Early in his reign as Emperor of the West, Charlemagne ordered a pentagonal fort to be built at a strategic point of his domain. As good luck charms, he had a third order magic square with all prime elements engraved on each wall. The five magic squares were different from each other, but they had the same magic constant—the year in which the fort was completed. The fort proved its ability to resist attack midway through his reign.

On this evidence, reconstruct the magic squares.

944. *Proposed by Richard Johnsonbaugh, Chicago State University and R. Rangarajan, Tata Institute of Fundamental Research, independently.*

Compute the total number of distinct auctions in contract bridge.

QUICKIES

From time to time this department will publish problems which may be solved by laborious methods, but which with the proper insight may be disposed of with dispatch. Readers are urged to submit their favorite problems of this type, together with the elegant solution and the source, if known.

Q620. Let S be the set of the first n positive integers, r be an integer and $T = S \cup \{r\}$. There exists an integer in T such that its removal results in a set in which the sum of its elements is divisible by n.

[*Submitted by Sidney Penner*]

Q621. Show that in a square pyramid with all edges equal, a dihedral angle formed by two triangular faces is twice a dihedral angle at the base.

[*Submitted by Charles W. Trigg*]

Q622. If G, F are integrable, $a > 0$, $G(x) \geqq F(x) \geqq 0$, and $\int_0^1 xF(x)\,dx = \int_0^a xG(x)\,dx$, show that $\int_0^1 F(x)\,dx \leqq \int_0^a G(x)\,dx$.

[*Submitted by M. S. Klamkin.*]

Q623. It is known that the range of a nonconstant polynomial function with integral coefficients cannot consist wholly of primes. The range of the polynomial $2x-1$, however, contains all the odd primes. Is there a polynomial of degree greater than 1 whose range contains all the primes?

[*Submitted by Erwin Just and Norman Schaumberger*]

Q624. Think of a square matrix as placed on a checkered board, so that the leading diagonal consists entirely of white squares. Then if the signs of all the entries on black squares are changed the eigenvalues are unchanged.

[*Submitted by I. J. Good*]

(Answers on page 186)

SOLUTIONS

Irrational Sides

901. [May, 1974] *Proposed by Leon Bankoff, Los Angeles, California.*

The sides of any triangle are rational (or integral) only if the ratio of the inradius to the circumradius is rational. Is the converse true?

I. *Solution by Vaclav Konecny, Kostelec-Stipa, Czechoslovakia.*

The converse is not true because in the equilateral triangle with any side (rational or irrational) the ratio of the inradius to the circumradius is $1/2$.

II. *Solution by Michael Goldberg, Washington, D.C.*

The problem could have been stated as follows: If the sides a, b, c of a triangle have rational ratios, then the ratio of the inradius r to the circumradius R is rational. If K is the area of the triangle and s is its semiperimeter, then $R = abc/K$ and $r = K/s$. From this, we obtain $R/r = 4K^2/abcs$. Since $K^2 = s(s-a)(s-b)(s-c)$, $R/r = 4(s-a)(s-b)(s-c)$ and is rational.

But, given the fact that R/r is rational, it does not follow that the ratios of the sides of the triangle are always rational. Consider the triangle ABC whose sides are 3, 4, and 5. Then $R = 5/2$ and $r = 1$. Now construct the isosceles triangle EFG

with the same R and r. The distance d between the centers of the circles is given by $d^2 = R^2 - 2rR$. Then, $(e/2)^2 = R^2 - (d-r)^2$ and $f^2 = (e/2)^2 + (R - d + r)^2$. Substituting in the values of r and R, we find that $e/f = (30 + 10\sqrt{5})^{1/2}/5$, which is not rational.

Also solved by Thomas E. Elsner, J. R. Hanna, Ann D. Holley, Joseph D. E. Konhauser, Dave Logothetti, F. G. Schmitt, Jr., Michael I. Shamos, E. P. Starke, Charles W. Trigg, Alan Wayne, and Kenneth M. Wilke.

Antigens

902. [May, 1974] *Proposed by J. Michael McVoy and Anton Glaser, Pennsylvania State University.*

When only the two antigens A and B were known, there were four blood types, corresponding to the four subsets of these antigens. Now that antigen Rh has been brought to light, there are eight blood types. Emergency blood donations are subject to the rule that the donor's set of antigens must be a subset of the recipient's set of antigens. Thus any pair of blood types falls into one of three categories: (i) each owner may donate to the other, the two types being the same; (ii) only one owner may donate to the other; and (iii) neither owner may donate to the other.

Let n be the number of antigens on which blood typing and above rule are based. Category (iii) has 0, 1, and 9 pairs for $n = 1$, 2, and 3 respectively. How many pairs are in (iii) for $n = k$?

Solution by R. A. Gibbs, Fort Lewis College.

We want to find $f(k)$, the number of pairs of subsets of a k-element set where neither member of the pair is a subset of the other. Choose a j-element subset of a k-element set. There are $2^j - 1$ proper subsets and $2^{k-j} - 1$ proper subsets of this set and therefore $2^k - 2^j - 2^{k-j} + 1$ subsets which can be paired with the given set in the desired manner. Since there are $\binom{k}{j}$ j-element subsets from which to select and since the ordering of the pair of subsets is immaterial, it follows that $f(k) = \frac{1}{2} \sum_{j=0}^{k} \binom{k}{j}(2^k - 2^j - 2^{k-j} + 1)$, which is readily shown to be $2^{k-1}(2^k + 1) - 3^k$.

Also solved by Eric Brosius, Gerald Corrigan, Stephen C. Currier, Jr., Stanley C. Eisenstat & Michael I. Shamos, Thomas E. Elsner, Michael Goldberg, Martin Charles Golumbic, C. T. Haskell, N. J. Kuenzi, Kay P. Litchfield, Thomas C. Lominac, Graham Lord, Joseph V. Michalowicz, Armand Orensztein, G. W. Valk, and the proposers.

Unique Cryptarithm

903. [May, 1974] *Proposed by Alan Wayne, Holiday, Florida.*

Solve the equation

$$TWO \times ZERO = NOTHING$$

in which each letter represents a denary digit, with different letters representing different digits, and such that GOO, ROO, TOO, WOO and ZOO are primes. Is the information about the primes essential for a unique solution?

I. *Solution by Thomas E. Elsner, General Motors Institute.*

The five required primes must end in 11, 33, 77, or 99. Direct investigation shows $\{T, W, Z, R\} = \{2, 5, 6, 8\}$ since we must have $O = 7$ and hence $G = 9$. Elimination yields the following solution:

$$(867)(2057) = 1783419.$$

II. *Independently, John Beidler, University of Scranton, Sidney Kravitz, Dover, N. J., and Kay P. Litchfield, Provo, Utah.*

The above used computers to show the solution above is unique without the restrictions to primes.

Also solved by Joseph V. Michalowicz, Anne Marie Ross & Dale Woods, John M. Samoyolo, Kenneth M. Wilke, and the proposer.

Tangent Lines to a Cubic

905. [May, 1974] *Proposed by Marlow Sholander, Case Western Reserve University.*

Let Γ be the graph of $y = f(x) = ax^3 + bx^2 + cx + d$. Given only Γ, how does one construct through a point on Γ lines tangent to Γ?

Solution by the proposer.

We call DEF a *trace* of Γ if these three points are collinear and on Γ. We allow E and F to coincide and call DEE a trace if line DE is tangent to Γ at E. The following construction can readily be shown to be valid.

Choose F on Γ and draw trace DEF. Construct M and N so that M is the midpoint of DE and NF. Construct similarly M' and N' collinear with a trace $D'E'F$. Then MM' and NN' are parallel to the y-axis. Let MM' and NN' meet Γ at G and H, respectively. The traces FGG and HFF are the tangents sought.

The above construction fails when F is the point of inflection of Γ. Then F, M and N coincide. In this case, construct trace AFC parallel to the x-axis, and construct trace HCC as above. Construct line L parallel to the y-axis so that segment FC is divided in the ratio of 2 to 1. Then, with line L meeting HC at T, FT is tangent to Γ at F.

Partial solutions by M. T. Bird, Ragnar Dybvik (Norway), and Michael Goldberg.

A Product

906. [May, 1974] *Proposed by Robert Guy, Framingham, Massachusetts.*

Prove that $\prod_{m=1}^{n} m^{2[n/m]} - \mathscr{I}(m) = 1$, where [ ] is the bracket function and $\mathscr{I}(m)$ is the number of divisors of m.

Note. In the original statement above, the factor $-\mathscr{I}(m)$ should be in the exponent.

Solution by M. G. Greening, The University of New South Wales.

This is equivalent to proving that

$$\text{(i)} \qquad \sum_{m=1}^{n} \{2[n/m] - 2\mathscr{I}(m)\} \log m = 0.$$

$$\sum_{m=1}^{n} 2[n/m] \log m = \sum_{x=1}^{n} [n/x] \log x + \sum_{y=1}^{n} [n/y] \log y$$

$$\text{(ii)} \qquad = \sum_{x=1}^{n} \log x \sum_{y=1}^{[n/x]} 1 + \sum_{y=1}^{n} \log y \sum_{y=1}^{[n/y]} 1$$

$$\text{(iii)} \qquad = \sum_{m=1}^{n} \sum_{xy=m} (\log x + \log y)$$

$$= \sum_{m=1}^{n} \log m \sum_{d|m} 1$$

$$= \sum_{m=1}^{n} 2\mathscr{I}(m) \log m$$

which establishes (i).

We can interpret (ii) geometrically as the sum of $(\log x + \log y)$ over the lattice points (x, y) for which $0 < x \leqq n$, $0 < y \leqq n$, so that (iii) represents the summation over the subsets of lattice points lying on the hyperbolas $xy = m$.

Also solved by Vaclav Konecny (Czechoslovakia), Graham Lord, Kumar Murty & Ram Murty, Bob Prielipp, and the proposer.

Nowhere Continuous

907. [May, 1974] *Proposed by Warren Page, New York City Community College.*

The characteristic function of the rationals, although discontinuous at every point of the real line R, is equal almost everywhere to a continuous function on R. Is it possible to construct a function discontinuous at every point of R which is not equal almost everywhere to a measurable (Lebesgue) function?

Solution by M. B. Gregory, University of North Dakota.

Since a function which is equal almost everywhere to a measurable function must itself be measurable [Royden, *Real Analysis*, 2nd edition, p. 68], the proposer requests a nonmeasurable, nowhere continuous function. The essential fact is that there exist discontinuous linear functions (functions such that $f(x+y) = f(x) + f(y)$ for all real x and y). It is easy to show that such functions are nonmeasurable and nowhere continuous. See Hewitt and Stromberg, *Real and Abstract Analysis*, p. 49, for a proof of the existence of discontinuous linear functions.

Also solved by the proposer.

ANSWERS

A 620. Let w be the sum of the $n+1$ elements of T. Since S constitutes a complete residue system mod n, there exists $j \in S$ such that $w \equiv j \bmod n$. Clearly $T - \{j\}$ is a set with the desired property.

Editor's comment. Note the generalization to $T = S \cup R$, where R is an arbitrary finite set of integers.

A 621. Two such pyramids placed base to base form a regular octahedron in which all dihedral angles are equal. The plane of the bases bisects four of these dihedral angles.

A 622. Since

$$\int_a^1 xF(x)dx = r\int_a^1 F(x)dx = \int_0^a x\{G(x) - F(x)\}dx = s\int_0^a \{G(x)-F(x)\}dx$$

where $1 \geqq r \geqq a$ and $a \geqq s \geqq 0$, we have that $\int_a^1 F(x)dx \leqq \int_0^a \{G(x) - F(x)\}\,dx$ which is equivalent to the desired result.

Remarks. The problem arose in showing that the time of vertical ascent of a particle subject to gravity and air resistance is less than the time of descent. One can give another proof by showing that the speed of ascent is greater than the speed of descent at corresponding heights.

A 623. If f is of degree $\geqq 2$, then it is easily shown that $\Sigma[1/f(i)]$ converges where summation extends over $\{i : f(i) \neq 0\}$. It is known, however, that the sum of the reciprocals of the primes is divergent. This contradiction requires that the answer to the question be given in the negative.

A 624. If $Ax = \lambda x$, with $x = (x_1, x_2, \cdots, x_n)$, then we must have $A^{\#}x^{\#} = \lambda x^{\#}$, where $A^{\#}$ denotes the matrix A with the signs of the 'black squares' changed and $x^{\#} = (x_1, -x_2, \cdots, (-1)^{n+1}x_n)$. Thus the eigenvalues are the same.

(Quickies on page 181)