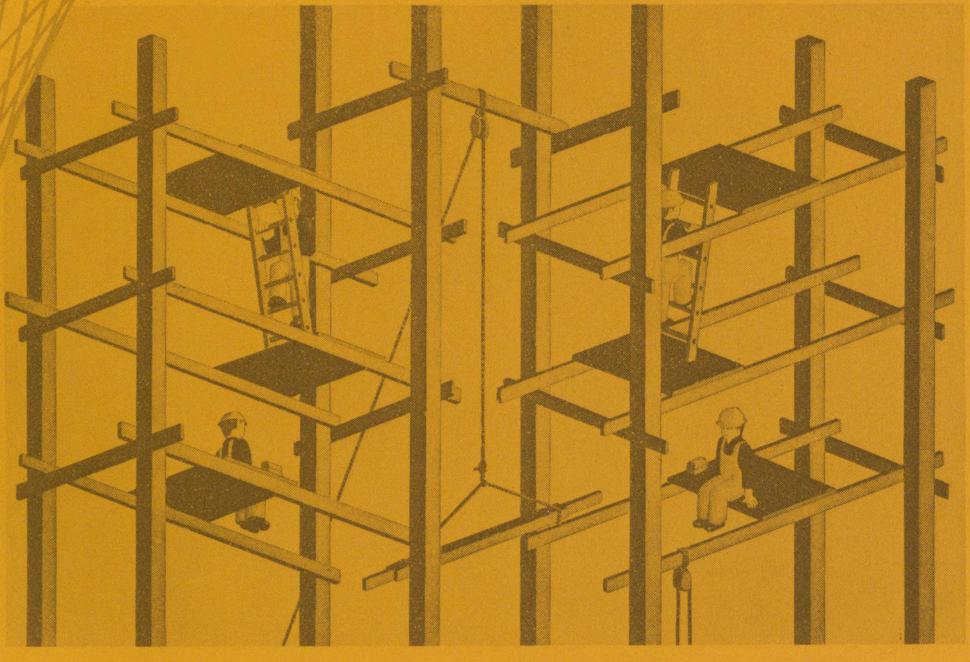
MATHEMATICS MAGAZINE



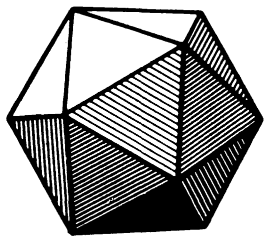
Vol. 51, No. 2
March, 1978
CODEN: MAMGA8

PERIODIC BEHAVIOR • TRISECTING LINES

POLYNOMIAL ROOTS • SAFETY OF ROPES

MATHEMATICS MAGAZINE

Volume 51, Number 2, March 1978



COVER: This original etching "Scaffolding 2" by Jonathan Talbot (1971) presents a paradox of representation of 3-dimensional symmetry on a plane page. Further information may be found in the News and Letters section. To refocus your eyes, see Möbius on page 105.

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

EDITORIAL POLICY

Mathematics Magazine is a journal of collegiate mathematics designed to enrich undergraduate study of the mathematical sciences. The *Magazine* should be an inviting, informal journal emphasizing good mathematical exposition of interest to undergraduate students. Manuscripts accepted for publication in the *Magazine* should be written in a clear and lively expository style. The *Magazine* is not a research journal, so papers written in the terse "theorem-proof-corollary-remark" style will ordinarily be unsuitable for publication. Articles printed in the *Magazine* should be of a quality and level that makes it realistic for teachers to use them to supplement their regular courses. The editors especially invite manuscripts that provide insight into applications and history of mathematics. We welcome other informal contributions, for example, brief notes, mathematical games, graphics and humor.

Editorial correspondence should be sent to: Mathematics Magazine, Department of Mathematics, St. Olaf College, Northfield, Minnesota 55057. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine*. They should be typewritten and double spaced on $8\frac{1}{2}$ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added; the printers will insert printed letters on the illustration in the appropriate locations.

Authors planning to submit manuscripts may find it helpful to obtain the more detailed statement of guidelines available from the editorial office.

ABOUT OUR AUTHORS

Ralph P. Boas ("Estimating Remainders") grew up in and around English departments and thus received an early introduction to writing and editing; he has served as an editor on many occasions including, currently, for the *Monthly*. His interest in estimating integrals goes back to his work as a student with D. V. Widder where he became convinced of the value of estimation as a tool in analysis. The present article developed out of an interest in harmonic series, stimulated by one of Martin Gardner's columns, combined with a desire to show how to extract as much as possible from the simple method of comparing a series with an integral.

Maurice Shrader-Frechette ("Complementary Rational Numbers") taught high school for four years before beginning his graduate study. He received his doctorate from the University of Notre Dame in 1973, and now teaches at Spalding College. The idea for his article arose when, while playing with a pocket calculator, he noticed that some rationals have the complementing property for the two halves of their period and some do not.

Philip D. Straffin, Jr. ("Periodic Points of Continuous Functions") is Associate Professor of Mathematics at Beloit College. He holds a B.A. from Harvard, an M.A. from Cambridge, and a Ph.D. from Berkeley. Although his graduate training was in algebraic topology, he has abiding interests in the methods of global analysis and mathematical ecology. These two interests combine in the present article.

BUSINESS INFORMATION. Mathematics Magazine is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. Ordinary subscriptions are $12 per year. Members of the Mathematical Association of America or of Mu Alpha Theta may subscribe at special reduced rates. Colleges and university mathematics departments may purchase bulk subscriptions (5 or more copies to a single address) for distribution to undergraduate students.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, Suite 310, 1225 Connecticut Avenue, N.W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.

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

ARTICLES

Estimating Remainders

Arguments using a sawtooth function produce efficient integral formulas for remainders.

RALPH P. BOAS
Northwestern University
Evanston, IL 60201

1. Introduction

How do you find the remainder R_n after the nth term of a convergent series? If you interpret "find" to mean "find exactly", you most probably can't do it, but sometimes you can find quite accurate approximations to the remainder. For a series that happens to be the Taylor series of a fairly simple function, there is in fact a useable formula for the remainder; otherwise most textbooks provide little help. Let us suppose that our series has the form $\sum_{k=1}^{\infty} f(k)$, where f is a positive function that tends rather regularly to zero, say with $f(x)$ decreasing to zero and $|f'(x)|$ also decreasing to zero (so that the graph of f is, in particular, convex downward). In that case you could easily compare the area under the curve from $x = 1$ to $x = n$ with $\sum_{k=1}^{n} f(k)$, and you might well come up with formulas like

$$\int_n^\infty f(t)dt - \tfrac{1}{2}f(n) \leqq R_n \leqq \int_n^\infty f(t)dt - \tfrac{1}{2}f(n+1) \tag{1.1}$$

(cf. [5]).

For a divergent series there are no remainders but it is well known (and easy to verify) that, under the same assumptions on f, the difference

$$D_n = \sum_{k=1}^{n} f(k) - \int_1^n f(t)dt$$

approaches a limit γ (which is called Euler's constant when $f(k) = 1/k$). We can then ask how fast D_n approaches its limit. You might find

$$\tfrac{1}{2}f(n+1) \leqq D_n - \gamma \leqq \tfrac{1}{2}f(n-1) \tag{1.2}$$

(cf. [7]).

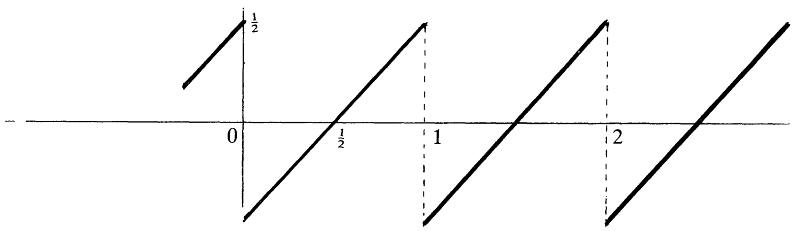
The one kind of series for which every textbook gives some help with remainders is a convergent alternating series $\sum_{k=1}^{\infty}(-1)^{k-1}f(k)$, with $f(x)$ decreasing to zero. We all learn that the remainder after the nth term has the sign of the first term neglected and absolute value less than the absolute value of that term. But is this the most that we can say?

The point of this article is that there is a systematic method — now some 200 years old — for getting approximations for the remainders in convergent series and for the differences $D_n - \gamma$ for divergent series; moreover, it is elementary enough to be given in a first course in calculus; and it produces, very easily, more accurate results than you will get by *ad hoc* methods unless you are unusually lucky. (In particular, it will improve the estimates cited above.) This method uses the very simplest versions of the so-called Euler-Maclaurin formula (see, for example, [4], pp. 520 ff.), namely the following identities, in which $P_1(t) = t - [t] - \frac{1}{2}$, and $[t]$ is the integral part of t:

$$\sum_{k=n+1}^{m} f(k) = \int_{n+1}^{m} f(t)dt + \tfrac{1}{2}\{f(n+1)+f(m)\} + \int_{n+1}^{m} P_1(t)f'(t)dt, \tag{1.3}$$

$$\sum_{k=n+1}^{m} f(k) = \int_{n+1/2}^{m+1/2} f(t)dt + \int_{n+1/2}^{m+1/2} P_1(t)f'(t)dt. \tag{1.4}$$

Either formula (once discovered) is readily verified by breaking the second integral into a sum of integrals between the discontinuities of P_1 (at the integers) and integrating by parts. (It would be harder to think up the formulas in the first place.) Note for future reference that P_1 is periodic with period 1, and its integral over a period is 0. Its graph is a sawtooth:



Our standard hypotheses will be that $f(t) > 0$, and f and f' are monotonic, and approach 0 as $t \to +\infty$. (Then f decreases but f' is negative and increases, so that $|f'|$ decreases.) Then the argument commonly used to show that $\int_0^\infty t^{-1}\sin t\,dt$ converges shows that

$$\int_n^\infty P_1(t)f'(t)dt \tag{1.5}$$

converges. Indeed, we can write (1.5) as

$$\int_n^{n+1/2} + \int_{n+1/2}^{n+1} + \int_{n+1}^{n+3/2} + \cdots .$$

The first integral is positive since $P_1(t) \leqq 0$ on the first half-period and $f'(t) < 0$; the second integral is negative; and so on. Thus we have converted the integral (1.5) into an alternating series. Furthermore, the graph of $P_1(t)$ on $(n, n+1)$ is symmetric about the point $(n+\frac{1}{2}, 0)$, whereas $|f'(t)|$ decreases; consequently the magnitude of $\int_n^{n+\frac{1}{2}}$ exceeds the magnitude of $\int_{n+1/2}^{n+1}$ and thus the alternating series has terms whose absolute values decrease. Finally, by the mean-value theorem for integrals we have

$$\left|\int_n^{n+1/2} P_1(t)f'(t)dt\right| \leqq \max|P_1(t)| \int_n^{n+1/2} |f'(t)|\,dt$$

$$\leqq \tfrac{1}{2}\int_n^{n+1/2} |f'(t)|\,dt = \tfrac{1}{2}\{f(n) - f(n+\tfrac{1}{2})\} \to 0.$$

Hence the alternating series converges, so (1.5) converges.

If we now look at (1.3) and (1.4), we see that $\Sigma_n^\infty f(k)$ and $\int_n^\infty f(t)dt$ converge or diverge together. (This is the content of the "integral test" for convergence of series, which can of course be established without any hypothesis on f'.)

2. Convergent series

Assuming convergence, we can let $m \to \infty$ in (1.3) and (1.4) and obtain

$$R_n = \sum_{k=n+1}^{\infty} f(k) = \int_{n+1}^{\infty} f(t)dt + \tfrac{1}{2}f(n+1) + \int_{n+1}^{\infty} P_1(t)f'(t)dt, \tag{2.1}$$

$$R_n = \sum_{k=n+1}^{\infty} f(k) = \int_{n+1/2}^{\infty} f(t)dt + \int_{n+1/2}^{\infty} P_1(t)f'(t)dt. \tag{2.2}$$

These are the basic formulas for remainders. They are exact, but not very convenient for calculation because of the second integral on the right. We now proceed to estimate that integral.

It should be clear from the graph of P_1 that

$$\int_{n+1}^{\infty} f'(t)P_1(t)dt > 0, \tag{2.3}$$

since $f'(t) < 0$ and $|f'(t)|$ is larger over any interval $(m, m+\frac{1}{2})$, where $P_1(t) < 0$, than over the adjacent interval $(m+\frac{1}{2}, m+1)$, where $P_1(t) > 0$. Hence we read off from (2.1) that

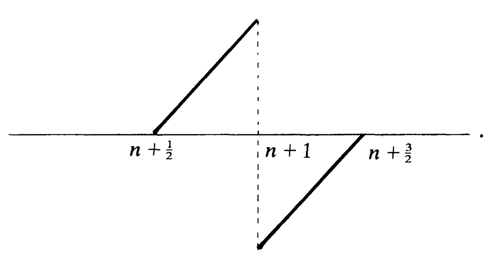
$$R_n > \int_{n+1}^{\infty} f(t)dt + \tfrac{1}{2}f(n+1). \tag{2.4}$$

This inequality, which is derived differently in [6], follows (as we have just seen) with very little work from (2.1). To see that (2.4) actually gives a larger lower bound for R_n than the left-hand side of (1.1) does, we can write the difference of the two estimates in the form

$$\int_{n+1}^{\infty} f(t)dt + \tfrac{1}{2}f(n+1) - \left\{ \int_{n}^{\infty} f(t)dt - \tfrac{1}{2}f(n) \right\} = -\int_{n}^{n+1} f(t)dt + \tfrac{1}{2}\{f(n) + f(n+1)\}.$$

Since $|f'(t)|$ decreases, the curve $y = f(x)$ is convex downward, and the area under the chord (which is the second term on the right) exceeds the area under the curve (the integral in the first term on the right).

To obtain an upper bound for R_n we start from (2.2). Since a period of $P_1(t)$ starting at $n+\frac{1}{2}$ looks like



We see, by the same argument used for (2.3), that

$$\int_{n+1/2}^{\infty} P_1(t)f'(t)dt < 0, \tag{2.5}$$

since now $P_1(t)$ is positive in the intervals where $|f'(t)|$ is larger. Hence (2.2) tells us that

$$R_n < \int_{n+1/2}^{\infty} f(t)dt \tag{2.6}$$

which is somewhat more informative than the right-hand side of (1.1) (the upper bound given in [**4**]). To see this, rewrite the right-hand side of (1.1) as follows:

$$\int_n^\infty f(t)dt - \tfrac{1}{2}f(n+1) = \int_{n+1/2}^\infty f(t)dt + \int_n^{n+1/2} f(t)dt - \tfrac{1}{2}f(n+1).$$

Since f decreases,

$$\int_n^{n+1/2} f(t)dt > \tfrac{1}{2}f(n+\tfrac{1}{2}),$$

and so

$$\int_n^{n+1/2} f(t)dt - \tfrac{1}{2}f(n+1) > \tfrac{1}{2}\{f(n+\tfrac{1}{2}) - f(n+1)\} > 0.$$

In fact, (2.6) is even better than the more complicated upper bound

$$R_n < \int_n^\infty f(t)dt - \tfrac{1}{2}f(n+1) - \tfrac{1}{4}|f'(n+1)| \tag{2.7}$$

given in [**6**].

Let us verify the last statement. To show that

$$\int_{n+1/2}^\infty f(t)dt < \int_n^\infty f(t)dt - \tfrac{1}{2}f(n+1) - \tfrac{1}{4}|f'(n+1)|$$

we need to show that

$$\tfrac{1}{2}f(n+1) - \int_n^{n+1/2} f(t)dt < \tfrac{1}{4}f'(n+1).$$

Now

$$\int_n^{n+1/2} f(t)dt > \tfrac{1}{2}f(n+\tfrac{1}{2})$$

since f decreases, so

$$\tfrac{1}{2}f(n+1) - \int_n^{n+1/2} f(t)dt < \tfrac{1}{2}\{f(n+1) - f(n+\tfrac{1}{2})\} = \tfrac{1}{4}f'(x),$$

for some x such that $n+\tfrac{1}{2} < x < n+1$, by the mean-value theorem for derivatives. But f' is negative and increasing, so $\tfrac{1}{4}f'(x) < \tfrac{1}{4}f'(n+1)$, wherever x may be in the interval $(n+\tfrac{1}{2}, n+1)$.

Some numerical results are instructive. Take $f(t) = t^{-2}$ and $n = 100$; write $R = R_{100}$. Then according to (1.1),

$$0.009950000 < R < 0.009950986.$$

According to (2.4) and (2.6),

$$0.009950049 < R < 0.009950249,$$

whereas according to (2.7),

$$R < 0.009950500.$$

We have now seen that the systematic approach produces results as good as (and sometimes better than) the more haphazard methods, and on the whole with less work.

However, the observant reader will have noticed that we obtained the strongest upper bound from (2.2) and the strongest lower bound from (2.1). To be honest, I have to admit to having tried both formulas in both cases and presented only the more satisfactory results.

However, it is possible to proceed the other way if we are willing to have more complicated bounds. We shall show that

$$R_n < \int_{n+1}^{\infty} f(t)dt + \tfrac{1}{2}f(n+1) + \tfrac{1}{8}|f'(n+1)|. \tag{2.8}$$

This is slightly better than (2.6) (and hence better than the equally complicated (2.7)), as we shall show presently. To derive (2.8), start from (2.1) and write

$$\int_{n+1}^{\infty} f'(t)P_1(t)dt = \int_{n+1}^{n+3/2} f'(t)P_1(t)dt + \int_{n+3/2}^{\infty} f'(t)P_1(t)dt. \tag{2.9}$$

As in deriving (2.5), we see that the second integral on the right of (2.9) is negative. Hence we obtain

$$R_n < \int_{n+1}^{\infty} f(t)dt + \tfrac{1}{2}f(n+1) + \int_{n+1}^{n+3/2} f'(t)P_1(t)dt.$$

In $\int_{n+1}^{n+3/2}$, the largest value of $|f'(t)|$ is at $t = n+1$, so by the mean-value theorem,

$$R_n < \int_{n+1}^{\infty} f(t)dt + \tfrac{1}{2}f(n+1) + |f'(n+1)| + \int_{n+1}^{n+3/2} P_1(t)dt$$

(since $P_1(t) \leqq 0$ in the range of integration, and $f'(t) < 0$). Hence (2.8) follows.

To see that (2.8) is an improvement of (2.6) is not quite easy. We want to show that

$$\int_{n+1}^{\infty} f(t)dt + \tfrac{1}{2}f(n+1) + \tfrac{1}{8}|f'(n+1)| < \int_{n+1/2}^{\infty} f(t)dt, \tag{2.10}$$

that is,

$$\tfrac{1}{2}f(n+1) + \tfrac{1}{8}|f'(n+1)| < \int_{n+1/2}^{n+1} f(t)dt,$$

or simply

$$\tfrac{1}{8}|f'(n+1)| < \int_{n+1/2}^{n+1} \{f(t) - f(n+1)\}dt. \tag{2.11}$$

Now

$$\begin{aligned}\int_{n+1/2}^{n+1} \{f(t) - f(n+1)\}dt &= -\int_{n+1/2}^{n+1} dt \int_{t}^{n+1} f'(u)du \\ &= -\int_{n+1/2}^{n+1} f'(u)du \int_{n+1/2}^{u} dt \\ &= -\int_{n+1/2}^{n+1} f'(u)(u - [u] - \tfrac{1}{2})du \\ &= \int_{n+1/2}^{n+1} \{-f'(u)\}P_1(u)du \\ &> -f'(n+1)\int_{n+1/2}^{n+1} P_1(u)du = \tfrac{1}{8}|f'(n+1)|.\end{aligned}$$

Hence we have (2.11) and therefore (2.10).

When $f(t)=t^{-2}$ and $n=100$, the right-hand side of (2.8) is 0.0099502475, so that the improvement is only in the 9th decimal place.

3. Divergent series

We have now done essentially all the work necessary to obtain better bounds than (1.2) for the difference $D_n-\gamma$. If we rewrite (1.3) with $n=0$ and then let $m\to\infty$, the left-hand side approaches γ, so

$$\gamma=\tfrac{1}{2}f(1)+\int_1^\infty P_1(t)f'(t)dt.$$

Hence

$$D_n-\gamma=\sum_{k=1}^n f(k)-\int_1^n f(t)dt-\gamma$$

$$=\tfrac{1}{2}\{f(1)+f(n)\}+\int_1^n P_1(t)f'(t)dt-\tfrac{1}{2}f(1)-\int_1^\infty P_1(t)f'(t)dt$$

(3.1) $$=\tfrac{1}{2}f(n)-\int_n^\infty f'(t)P_1(t)dt.$$

Similarly, (1.4) leads to

(3.2) $$D_n-\gamma=\int_n^{n+1/2} f(t)dt-\int_{n+1/2}^\infty f'(t)P_1(t)dt.$$

The integral in (3.1) is just the integral in (2.1), with its sign changed. We saw in (2.3) that

$$\int_n^\infty f'(t)P_1(t)dt>0,$$

and so (3.1) leads to

(3.3) $$D_n-\gamma<\tfrac{1}{2}f(n).$$

In the same way, we saw in (2.5) that the integral in (3.2) is positive, so

(3.4) $$D_n-\gamma>\tfrac{1}{2}f(n+\tfrac{1}{2}).$$

Inequalities (3.3) and (3.4) are improvements over (1.2) (and the proofs are shorter than the cited proof of (1.2)).

4. Alternating series

Let the series be $\Sigma_1^\infty(-1)^{n-1}f(n)$ with f decreasing to 0. Then the remainder after the sum of the first m terms is

(4.1) $$R_m=(-1)^m f(m+1)+(-1)^{m+1}f(m+2)+\cdots$$

$$=\tfrac{1}{2}(-1)^m f(m)+\tfrac{1}{2}(-1)^m\{[f(m+1)-f(m)]-[f(m+2)-f(m+1)]+\cdots\}$$

$$=\tfrac{1}{2}(-1)^m f(m)+\tfrac{1}{2}(-1)^m\left\{\int_m^{m+1} f'(t)dt-\int_m^{m+1} f'(t)dt+\cdots\right\}.$$

If $T(t)$ denotes the "square wave" equal to $+1$ on $(m,m+1)$ and to -1 on $(m+1,m+2)$, and so on, we can write the preceding formula as

$$R_m = \tfrac{1}{2}(-1)^m f(m) + \tfrac{1}{2}(-1)^{m-1}\int_m^\infty \{-f'(t)\}T(t)\,dt,$$

or as

$$(4.2)\qquad (-1)^{m-1}R_m = -\tfrac{1}{2}f(m) + \tfrac{1}{2}\int_m^\infty \{-f'(t)\}T(t)\,dt,$$

and also (since $T(t) = 1$ on $(m, m+1)$) as

$$(-1)^{m-1}R_m = -\tfrac{1}{2}f(m) + \tfrac{1}{2}\int_m^{m+1} \{-f'(t)\}dt + \tfrac{1}{2}\int_{m+1}^\infty \{-f'(t)\}T(t)\,dt,$$

that is, as

$$(4.3)\qquad (-1)^{m-1}R_m = -\tfrac{1}{2}f(m+1) + \tfrac{1}{2}\int_{m+1}^\infty \{-f'(t)\}T(t)\,dt.$$

By using the same kind of reasoning that we used in §2, we see that, since $-f'(t)$ is positive and decreasing, the integral in (4.2) is positive whereas the integral in (4.3) is negative. Consequently the remainder after the mth term satisfies

$$(4.4)\qquad -\tfrac{1}{2}f(m) < (-1)^{m-1}R_m < -\tfrac{1}{2}f(m+1)$$

Therefore, whether m is odd or even, the remainder R_m has its sign opposite to that of the last term retained, and $|R_m|$ is between $\tfrac{1}{2}f(m)$ and $\tfrac{1}{2}f(m+1)$, i.e., the absolute value of the remainder is between half that of the last term retained and half that of the first term neglected. This is only slightly more than half the usual estimate of the remainder (of course, under somewhat more restrictive hypotheses). It seems that the left side of (4.4) was never noticed explicitly until 1962 **[2]**; **[1]** (p. 59, 1st ed.; p. 65, 2d ed.) gives an inequality equivalent to $|R_m| < f(m) - \tfrac{1}{2}f(m+1) < \tfrac{1}{2}\{f(m) + |f'(m)|\}$, and also gives the right-hand side of (4.4). Most textbooks seem to have remained indifferent to this information.

The inequalities (4.4) account for the rule of thumb that when you break off an alternating series you should subtract half the last term (or add half the next term) to improve the approximation. This can be made more precise by estimating the integral in (4.3) or (4.4); and still more precise by noticing that the series in (4.1) is itself an alternating series of terms of decreasing absolute value, and applying the reasoning of the present section to it in its turn; and so on. For examples of what can be accomplished by this kind of reasoning in special cases, see **[3]**.

Finally, we observed in §1 that the second integrals in the formulas (2.1) and (2.2) for the remainders in series of positive terms can be written as alternating series of integrals. If you apply (4.4) to these series, the estimates in (2.4), (2.6) and (2.8) can be improved still further.

References

[1] T. J. I'A. Bromwich, An Introduction to the Theory of Infinite Series, 1st and 2nd ed., Macmillan, London, 1908 and 1947.

[2] P. Calabrese, A note on alternating series, Amer. Math. Monthly, 69 (1962) 215–217; repr. in Selected Papers in Calculus, 1968, pp. 352–353.

[3] C. Goldsmith, Calculation of ln 2 and π, Math. Gaz., 55 (1971) 434–436.

[4] K. Knopp, Theory and Application of Infinite Series, Blackie, London and Glasgow, 1928.

[5] R. K. Morley, The remainder in computing by series, Amer. Math. Monthly, 57 (1950) 550–551; repr. in Selected Papers in Calculus, 1968, pp. 324–325.

[6] R. K. Morley, Further note on the remainder in computing by series, Amer. Math. Monthly, 58 (1951) 410–412.

[7] S. R. Tims, J. A. Tyrrell, Approximate evaluation of Euler's constant, Math. Gaz., 55 (1971) 65–67.

Complementary Rational Numbers

A survey and history of intriguing patterns found in repeating decimals.

MAURICE SHRADER-FRECHETTE
Spalding College
Louisville, KY 40203

The repeating decimal representations of some fractions exhibit a fascinating pattern. Consider, for example, $1/7 = .\overline{142857}$:

— When the first *three* digits of the period are added to the second three, the sum is $142 + 857 = 999$; *three* 9's.

— When successive *pairs* of digits are added, the sum is $14 + 28 + 57 = 99$; *two* 9's.

— And when successive *single* digits are added, the sum is $1 + 4 + 2 + 8 + 5 + 7 = 27$. In the previous cases, the sum consisted of three 9's, then two 9's, so in this case we would naturally like to see one 9. In fact, if the high order digit 2 is added to the 7 we do get a single 9!

One can check the decimal representations of other rational numbers and find that this phenomenon occurs often enough to be named. The accompanying table of reciprocals (TABLE 1) provides many examples. Let us call a rational number **complementary of order *n*, *r*** if and only if, beginning with *any* digit of its period, the sum of r successive blocks of n digits is represented by n 9's, stipulating that if the sum consists of more than n digits, the number formed by the additional ("high order") digits on the left should be added to the n-digit number on the right. Roughly speaking, the extra digits on the left "carry" to appropriate columns on the right.

Note that although our example 1/7 had a numerator of 1, the definition applies to other fractions as well. $5/7 = .\overline{714285}$ is complementary of orders 1,6 and 2,3 and 3,2. Note also that we do not limit our digits to a single occurrence of the period. However, the product nr must be some multiple of the period length. For example, 1/3 is complementary of order 1,3.

It is natural to wonder why complementarity of a given order is exhibited by some numbers and not by others. In this article therefore, we shall first explain this phenomenon. Then, after looking at some applications of complementarity, we shall see that in fact every rational exhibits complementarity of many different orders. The investigation of the precise orders to which a given rational belongs will lead us to four rules relating orders to the prime powers that are factors of the denominator. These rules in turn will help us to determine when complementarity will be exhibited *within* the period. Finally we shall look at the history of the concept of complementarity.

A necessary and sufficient condition

Our first task is to explain why it is that for some rationals, adding r successive blocks of n digits from the decimal representation gives a sum represented by n 9's. In order to add r successive blocks of n digits, we could multiply our rational sucessively by $1, 10^n, (10^n)^2, \ldots, (10^n)^{r-1}$. The resulting r numbers would have the successive blocks in the same position after the decimal point. The r numbers could be added, and the sum of the blocks would automatically be calculated. High order digits would even carry as we stipulated.

For example, to add three successive pairs of digits in the representation of 1/14 we could write

$$1/14 = .0\overline{714285}$$

$$10^2 \cdot 1/14 = 7.1\overline{428571}$$

$$10^4 \cdot 1/14 = 714.2\overline{857142}$$

The successive blocks 71, 42, and 85 are positioned in columns such that they are ready to be added. Likewise, 14, 28, and 57 are also positioned above each other. No matter what three successive blocks one considers, they are positioned properly in the above arrangement. The sum $1/14 + 10^2 \cdot 1/14 + 10^4 \cdot 1/14 = 721.4999999$ has a period consisting entirely of 9's, so the sum of successive blocks is represented by 9's, and 1/14 is therefore complementary of order 2,3. Note that the extra high order digit in the sum $71 + 42 + 85$ is carried.

Similarly, any rational z/x is complementary of order n, r if and only if the sum

$$z/x + 10^n \cdot z/x + (10^n)^2 \cdot z/x + \cdots + (10^n)^{r-1} \cdot z/x$$

x	$1/x$	x	$1/x$
1	$.\overline{9}$	26	$.03\overline{84615}$
2	$.4\overline{9}$	27	$.\overline{037}$
3	$.\overline{3}$	28	$.03\overline{571428}$
4	$.24\overline{9}$	29	$.\overline{0344827586206896551724137931}$
5	$.1\overline{9}$	30	$.0\overline{3}$
6	$.1\overline{6}$	31	$.\overline{032258064516129}$
7	$.\overline{142857}$	32	$.03124\overline{9}$
8	$.124\overline{9}$	33	$.\overline{03}$
9	$.\overline{1}$	34	$.0\overline{2941176470588235}$
10	$.0\overline{9}$	35	$.0\overline{285714}$
11	$.\overline{09}$	36	$.02\overline{7}$
12	$.08\overline{3}$	37	$.\overline{027}$
13	$.\overline{076923}$	38	$.0\overline{263157894736842105}$
14	$.0\overline{714285}$	39	$.\overline{025641}$
15	$.0\overline{6}$	40	$.024\overline{9}$
16	$.0624\overline{9}$	41	$.\overline{02439}$
17	$.\overline{0588235294117647}$	42	$.0\overline{238095}$
18	$.0\overline{5}$	43	$.\overline{023255813953488372093}$
19	$.\overline{052631578947368421}$	44	$.02\overline{27}$
20	$.04\overline{9}$	45	$.0\overline{2}$
21	$.\overline{047619}$	46	$.0\overline{2173913043478260869565}$
22	$.0\overline{45}$	47	$.\overline{0212765957446808510638297872340425531914893617}$
23	$.\overline{0434782608695652173913}$	48	$0208\overline{3}$
24	$.041\overline{6}$	49	$.\overline{020408163265306122448979591836734693877551}$
25	$.03\overline{9}$	50	$.01\overline{9}$

TABLE 1

has a period consisting of 9's, i.e., this sum is a terminating decimal. From now on, we represent the sum

$$1 + 10^n + (10^n)^2 + \cdots + (10^n)^{r-1}$$

by the symbol $t_{n,r}$.

Terminating decimals are rationals whose denominator is a power of 2 times a power of 5. Thus z/x is complementary of order n, r if and only if for some f, g, and y, $(z/x)t_{n,r} = y/(2^f 5^g)$. Clearing fractions we get $z \cdot 2^f 5^g \cdot t_{n,r} = xy$. Since x divides the right side, it divides the left. And since we may presume that z/x is reduced to lowest terms, x must divide $2^f 5^g \cdot t_{n,r}$. Thus we have the following explanation for complementarity in the form of a necessary and sufficient condition: *Let z/x be a rational in lowest terms. Then z/x is complementary of order n, r if and only if the denominator x is some power of 2, times a power of 5, times a factor of $t_{n,r}$.*

Notice that complementarity depends only on the denominator. Thus for example, since $3 \cdot 37 = 111 = t_{1,3}$, we know that any fraction with a denominator of 3, 37, or 111 is complementary of order 1,3. And since $7 \cdot 11 \cdot 13$ is 1001, or $t_{3,2}$, we know that any fraction with a denominator of 7, 11, 13 or 77, 91, 143 or 1001 is complementary of order 3,2.

Further applications

Rationals which belong to an order $n, 2$ are particularly interesting because of the ease with which one can convert the fractional form to the decimal and vice versa. Consider some examples in which the period begins immediately after the decimal point.

57/73 belongs to order 4,2 because the denominator 73 is a factor of $t_{4,2}$: $73 \cdot 137 = 10001$. Thus if we could find the first four digits in the decimal representation, the second four would be their complement. The first four may of course be found by division, but there is an even quicker way to calculate them. Multiply the numerator 57 by the other factor of 10001, that is 137, and subtract one: $57 \cdot 137 - 1 = 7808$, so $57/73 = .\overline{78082191}$.

The explanation for this shortcut is that when we multiply the decimal representation of 57/73 by 10^4, the first four digits in the period become a whole number p, and the second four digits move into the first four places after the decimal point. Since the two halves of the period are complementary, $(57/73) \cdot 10^4 + 57/73$ is p plus $.\overline{9999}$. But $.\overline{9999}$ is the same as 1, so $(57/73) \cdot 10001 = p + 1$, i.e., $57 \cdot 137 - 1 = p$.

On the other hand, it is also easy to convert a decimal to a fraction when the number belongs to order $n, 2$. Suppose we have the decimal $.\overline{72932706}$ and we note that it is complementary of order 4,2. Letting its fractional representation be z/x and letting p be the number represented by the first four digits of the period, we see from the equation $(z/x) \cdot 10^4 + z/x = p + 1$ that $z/x = (p+1)/10001$, i.e., $.\overline{72932706} = 7294/10001$.

Perhaps there is a way that computers can be designed to do arithmetic with numbers belonging to some order $n, 2$. Only the first n digits in the complementing block would have to be stored in memory since the second n would be the complement of the first n digits.

Every rational is complementary of some order

One might wonder whether for a given rational one can always add up successive blocks of n digits and eventually have a sum represented by n nines. To find the answer let us consider an arbitrary denominator x and an arbitrary natural number n. Then x is some power of 2, times a power of 5, times a number x' which is relatively prime to 2 and 5. Because complementarity is equivalent to the condition that x' divides $t_{n,r}$ for some r, our question is whether there is an r such that this is so.

The trick to employ when determining whether a number x' divides a number of a certain form is to look at $x' + 1$ numbers of that form. Consider the numbers $t_{n,1}, t_{n,2}, \ldots, t_{n,x'+1}$. At least two of them must be in the same residue class mod x', say $t_{n,i}$ and $t_{n,j}$, where $i < j$. If we subtract them, the difference will be divisible by x' and will look like one of the t's followed by a few zeros: $t_{n,j} - t_{n,i} = t_{n,j-i} \cdot (10^n)^i$. Since x' is relatively prime to 10, it must divide $t_{n,j-i}$. Letting $j - i$ be r, we can conclude that *for each rational number and integer n there is some number r such that the rational is complementary of order n, r.* In fact there is an r less than or equal to x'.

The discovery that every rational belongs to *some* order leads us naturally to search for *all* of the orders to which a rational belongs. Thus for a given rational z/x and a given natural number n, we are interested in all values of r for which x' divides $t_{n,r}$.

Four Rules

Four of the results recently published by Hayashi in this MAGAZINE **[10]** enable us to determine these values. We continue to use our doubly subscripted notation $t_{n,r}$ instead of his singly subscripted notation.

R1. Let x' be relatively prime to 10, let n be any natural number, and let r be the smallest number such that x' divides $t_{n,r}$. Then for any s, x' divides $t_{n,s}$ if and only if s is a multiple of r.

Thus it is enough to search for the minimum r and then take all of its multiples.

R2. If x' and x'' are relatively prime to 10 and to each other, then for a given n, the least r for which the product $x'x''$ divides $t_{n,r}$ is the least common multiple of the least r' and r'' for x' and x'' respectively.

So we would search for minimum r's for the prime power factors of x', since these will be relatively prime to each other. For example, $1/(7\cdot 31)$ belongs to order 3,10 because 1/7 belongs to order 3,2 and 1/31 to order 3,5. 10 is the 1cm of 2 and 5.

R3. If p is a prime other than 2 or 5 such that p divides 10^n-1 and if p^a is any power of p, then the smallest value of r for which p^a divides $t_{n,r}$ is $r=p^a$.

For example, 7 divides 10^6-1, so 7^1 divides $t_{6,7}$ and 7^2 divides $t_{6,49}$ etc.

R4. If p is a prime other than 2 or 5 and n is a natural number, then the sequence of minimum values of r for which $p, p^2, p^3, \dots$ divide $t_{n,r}$ is at first constant and eventually becomes geometric with multiplier p.

Thus for a particular prime p and a particular natural number n, let e be the smallest value such that p divides $t_{n,e}$: then p^2 *may* also divide $t_{n,e}$; p^3 may also. And so on until for some number c, p^c divides $t_{n,e}$ but p^{c+1} does not. However, $R4$ then assures us that p^{c+1} does divide $t_{n,ep}$ and that p^{c+2} divides t_{n,ep^2} etc. For example, 7 divides $t_{1,6}$ but 7^2 does not divide it. Thus $R4$ tells us that 7^2 divides $t_{1,6\cdot 7}$ and 7^3 divides $t_{1,6\cdot 7^2}$ etc.

The sequence $6, 6\cdot 7, 6\cdot 7^2, \dots$ in the example becomes geometric immediately, but there are examples in which this does not happen. William Shanks **[19]** discovered that the prime 487 divides $t_{1,486}$ and 487^2 also divides $t_{1,486}$.

Complementarity within the period: Part one

Before applying the rules to the search for all orders for which z/x is complementary, let us first note that they give us some information about complementarity *within* the period.

For $1/7=.\overline{142857}$ we need only look at one occurrence of the period to see that when successive single digits are added the sum is 9, when successive pairs are added the sum is 99, and when successive triples are added the sum is 999. However if we wish to add successive groups of four digits, it takes *two* occurrences of the period before the sum reaches 9999: $1428+5714+2857$.

We wonder whether there is any way of telling from the denominator whether a number will exhibit complementarity within the period. As most number theory books explain, to find the length of the period of any rational number, factor the denominator into $2^f 5^g x'$, where x' is relatively prime to 2 and 5. The length of the period is then the smallest number y such that x' divides 10^y-1. See **[20]**, for example.

It turns out that if the factor x' is a power of 3, complementarity will never be exhibited *within* the period. Let f, g, and a be any natural numbers; the length of the period of $z/(2^f 5^g 3^a)$ is the smallest y such that 3^a divides 10^y-1, which may be written $9\cdot t_{1,y}$. If a is 1 or 2, the period length is 1 since both 3 and 9 divide the first factor. And if a is greater than 2, the period length is the least value of y for which 3^{a-2} divides the second factor, $t_{1,y}$. R3 tells us that y is 3^{a-2}, so for $a>2$, the period length of

$z/(2^f5^g3^a)$ is 3^{a-2}. (For example, 1/27 has three digits in its period and 1/81 has nine.) No matter what natural number n we consider, $R3$ tells us that the minimal r for which $z/(2^f5^g3^a)$ is complementary of order n, r is 3^a. Since 3^a is more than the period length, it always takes more digits than are in the period to exhibit complementarity in a fraction whose denominator has its factor x' equal to 3^a.

We turn next to any prime $p > 5$ and wonder whether complementarity is possible within the period when $x' = p^a$, where a is some natural number. Once we know the period length we will investigate orders n, r such that the product nr equals the period length. This is necessary because complementarity within the period means that the period can be broken into r blocks of n digits, and the sum of these blocks is represented by n 9's. Because we will always need to add together more than one block to get 9's, we will only consider values of r greater than 1.

If $x' = p^a$, the period length of $z/(2^f5^gx')$ is the smallest number y such that p^a divides $9 \cdot t_{1,y}$. Since p is relatively prime to 9, y must be the smallest number such that p^a divides the second factor, $t_{1,y}$. By $R4$, the sequence of minimum values of y for which $p, p^2, p^3, \ldots$ divide $t_{1,y}$ is at first constant and then it is geometric with multiplier p. Thus the period length of $z/(2^f5^gp^a)$ will be term number a in this sequence.

If we let c be the number of constant terms in the sequence and let e be the constant, i.e., the period length of any fraction with denominator p, then the following table illustrates that y, the period length of $z/(2^f5^gp^a)$, equals e for values of a up to and including c, and it equals ep^{a-c} for values of a above c:

$$x' \text{ factor:} \quad p^1, \ldots, p^c, p^{c+1}, p^{c+2}, \ldots, p^a, \ldots,$$

$$\text{period length:} \quad e, \ldots, e, ep, ep^2, \ldots, ep^{a-c}, \ldots .$$

So suppose we have a fraction $z/(2^f5^gp^a)$ and we choose n and r such that nr does equal the period length y and $r > 1$. Then complementarity of order n, r will be exhibited within the period if and only if p^a divides $t_{n,r}$. Because y is the period length, we know that p^a does divide $t_{1,y}$, which may be factored $t_{1,n} \cdot t_{n,r}$.

In case y equals e, then n is too small for p to divide the first factor, so in this case p^a does divide $t_{n,r}$. In other words, any fraction $z/(2^f5^gp^a)$ whose period length is e, the same as that of $1/p$, exhibits complementarity of any order n, r where $nr = e$. For example, 1/487 and $1/487^2$ each have a period length of 486 and so they exhibit complementarity of orders 2,243 and 243,2 and 3,162 and 162,3 etc.

In case the period length of $z/(2^f5^gp^a)$ is a composite, $y = ep^{a-c}$, we also want to know the conditions under which p^a divides the second factor of $t_{1,y} = t_{1,n} \cdot t_{n,r}$. If p does not divide the first factor, then p^a does divide the second. And if p does divide the first factor, then p^a does not divide the second; otherwise p^{a+1} would divide $t_{1,y}$ and y is too small for that. Thus the condition is that p does not divide $t_{1,n}$. By $R1$ this is equivalent to the condition that e does not divide n. In other words, any fraction $z/(2^f5^gp^a)$ whose period length is $ep^{a-c} = nr$ exhibits complementarity of order n, r within its period if and only if e does not divide n.

For example, 1/7 has a period length of 6 and 1/49 has a period length of 42. The period of 1/49 does not exhibit complementarity of order 6, 7 although it does exhibit complementarity of all other possible orders: 1,42 and 2,21 and 3,14 and 7,6 and 14,3 and 21,2.

In the next section we shall see that when the factor x' of the denominator is a product of one or more prime powers, complementarity within the period depends on the period lengths of the prime power reciprocals in an interesting way.

Complementarity within the period: Part two

Let $p_1^{a_1} \cdots p_k^{a_k}$ be the prime power factorization of x', and for each prime p_i, let e_i be the period length of the reciprocal $1/p_i$. It turns out that when we restrict our attention to values of r which are not divisible by any of the primes p_i, the period exhibits complementarity of order n, r if and only if there is some one power of r which divides each e_i, but no higher power divides any of them. For example, when we multiply 1/7, which has period length 6, by 1/19, which has period length 18, the period length of the product is the least common multiple, 18. Both 6 and 18 are divisible by the same power of 2, so the period of $1/(7 \cdot 19)$ does exhibit complementarity of order 9,2. However, the highest power of 3 which divides 6 is not the same as the power which divides 18, so the period of $1/(7\cdot 19)$ does not exhibit complementarity of order 6,3.

To understand why this condition works, remember that nr, the period length of z/x, is the lcm of the period lengths of $1/p_1^{a_1},\dots,1/p_k^{a_k}$. We have seen that for each i, the period length of $1/p_i^{a_i}$ is e_i times some power of p_i, possibly the zeroth power. Thus the period length of z/x is the lcm of the e_i's times some powers of $p_1,\dots,p_k$.

First assume that the same power of r divides each e_i and no higher power of r divides any of them. If the period length of z/x is divided by r, the quotient n is not divisible by any of the numbers e_i since the highest power of r which divides n is too low; here we are using the condition that r is not divisible by any of the primes. Since n is not divisible by any e_i, none of the primes divides 10^n-1. But each prime power divides $10^{nr}-1=(10^n-1)t_{n,r}$, since nr is the period length. So each divides the second factor. Hence their product x' divides $t_{n,r}$. We therefore conclude that the period of z/x is complementary of order n,r.

Conversely suppose that the period of z/x exhibits complementarity of order n,r where r is relatively prime to x'. We wish to show that the same power of r divides each e_i and no higher power of r divides any e_i. Suppose, on the contrary, that a certain e_j were divisible by a higher power of r than the one that divides a certain e_h; then e_h would divide n. Consequently p_h would divide 10^n-1. By $R1$ and $R3$, the values of r' for which $1/p_h^{a_h}$ would be complementary of order n,r' would be multiples of $p_h^{a_h}$. But since $p_h^{a_h}$ divides x', which divides $t_{n,r}$, we know that $1/p_h^{a_h}$ is complementary of order n,r. Then r would have to be a multiple of $p_h^{a_h}$. But this would contradict the assumption that r is relatively prime to x'. Therefore the same power of r divides each e_i, and no higher power divides any e_i.

The condition which we have just proved required that r be relatively prime to x'. If we drop this condition, we can still say that the period of z/x exhibits complementarity of order n,r if and only if x' divides $t_{n,r}$. This is true if and only if each prime power factor of x' divides $t_{n,r}$. For those prime powers whose reciprocals have period length nr, we have shown in the previous section how to determine whether the condition holds. For the others, the methods of the next section will enable us to determine whether their reciprocals also belong to order n,r.

Finding all the orders of complementarity

We have seen that for a particular rational z/x and a particular natural number n, there are many values of r such that z/x is complementary of order n,r. We have also seen that for a particular n, $R1$ assures us that it is enough to find only the minimum r since all other values are multiples.

And $R2$ permits us to factor x' into prime powers p^a, find the orders n,r to which each $1/p^a$ belongs, and take the lcm of these r's.

In case p is 3, $R3$ enables us to conclude that $1/3^a$ is complementary of order $n,3^a$ but not complementary of order n,r if $r<3^a$.

For the other primes there are three cases. First suppose that $p>5$ and n is a *multiple* of e, the period length of $1/p$. 10^n-1 may be factored $(10^e-1)t_{e,n/e}$, and by the fact that e is the period length, p divides the first factor. Thus p divides 10^n-1. Using $R3$ we may conclude that $1/p^a$ is complementary of order n,p^a, but it is not complementary of order n,r if $r<p^a$.

For example, let p be 7 and n be 6. 6 is a multiple of 6, the period length of 1/7. Thus 1/7 must be complementary of order 6, 7 but not of order 6, 6 or any lower order. And 1/49 must be complementary of order 6,49 but not 6,48 or any lower order.

For the second case, suppose that $p>5$ and n is a *factor* of e, the period length of $1/p$. Since the case $n=e$ has just been covered, presume that $n<e$. We write $10^e-1=(10^n-1)t_{n,e/n}$ and note that while p does divide 10^e-1, it cannot divide the first factor since n is smaller than the period length. Hence p divides $t_{n,e/n}$, and $1/p$ must be complementary of order $n,e/n$ but not complementary of order n,r if $r<e/n$.

In this case, e/n is the first term of the sequence described by $R4$. The rest of the sequence may be found by successively dividing $p^2,p^3,\dots$ into $t_{n,e/n}$ and noting whether the remainder is 0. The sequence will have the constant value e/n until one of the powers does not divide evenly; then it will rise geometrically with the multiplier p. The term in the sequence corresponding to p^a is therefore the minimum r for which $1/p^a$ belongs to order n, r.

For example, take $p = 487$ and $n = 3$. 3 is a factor of 486, the period length of 1/487. From a previous example we know that both 1/487 and $1/487^2$ are complementary of order 3,162. To find whether $1/487^3$ is also complementary of order 3,162 we divide 487^3 into $t_{3,162}$. This one will be left as an exercise for the reader's computer.

For the final case, suppose that $p > 5$ but n is *neither* a factor nor a multiple of e. Let m be the lcm of n and e. Then $(10^n - 1)t_{n,m/n} = 10^m - 1$ is divisible by p since $10^m - 1$ may also be written $(10^e - 1)t_{e,m/e}$ and p divides the first factor. However, p does not divide $10^n - 1$ since n is not a multiple of e. Hence it divides the factor $t_{n,m/n}$ and $1/p$ is complementary of order $n, m/n$. m/n is the *smallest* possible value since m is the *least* common multiple. The rest of the sequence is found as in the previous case.

For example, let us find the orders for which $1/63 = .\overline{015873}$ is complementary. $63 = 3^2 \cdot 7$, so we investigate $1/3^2$ and 1/7.

Let $n = 1$. 3^2 divides $t_{1,9}$ and 7 divides $t_{1,6}$. Hence the minimum r is 18, the lcm of 9 and 6. 1/63 belongs to orders 1,18 and 1,36 and 1,54, etc.

Let $n = 2$. 3^2 divides $t_{2,9}$ and 7 divides $t_{2,3}$. Hence the minimum r is 9. 1/63 belongs to orders 2,9 and 2,18 and 2,27, etc.

Let $n = 3$. 3^2 divides $t_{3,9}$ and 7 divides $t_{3,2}$. Hence the minimum r is 18.

Let $n = 4$. 3^2 divides $t_{4,9}$ and 7 divides $t_{4,3}$. Hence the minimum r is 9.

Etc.

Historical Development

The notion of complementarity which we have presented has several facets. As we review the history of the concept, we see that there seems to have been no time at which all of its components were considered together. Let us recall six facts about complementarity and then look at the historical development of the concept:

— It applies to any rational number, not only to reciprocals of whole numbers and not only to rationals whose denominators are prime or are relatively prime to the base of the numeration system, e.g., 10.

— One or more repetitions of the period are involved.

— These repetitions of the period are broken into two or more blocks of n digits each.

— The blocks, not just corresponding digits, are added.

— The extra high order digits of this sum are added to the n low order digits.

— The resulting sum must be represented by n copies of the digit which is one less than the base, for example 9.

Usually mathematicians looked for complementarity within only one occurrence of the period. And usually they broke the period into two parts. Frequently they considered only reciprocals of whole numbers; sometimes they even restricted their attention to reciprocals of primes. Often the factor 3, and even more often the factors 2 and 5 were excluded from the denominator. Sometimes researchers added single digits instead of whole blocks. The high order digits were hardly ever carried. But the aim was always to get a sum having to do with 9 or whatever digit was one less than the base.

The first person to publish on complementarity, according to Dickson **[6]**, was H. Goodwyn **[8]**. He noted that for primes $\geqq 7$, corresponding digits in the two half periods of the reciprocal add up to 9. Of course this is true only if the period has an even number of digits.

Someone noticed in 1829 that when the digits of the period of a rational number are added one by one and the high order digits of the sum are then added to the low order digit, the final result is 9. This phenomenon is what we have called complementarity of order $1, r$ where r is the length of the period. However, we have seen that the property does not *always* hold. For example, if the denominator is a power of 3, it does not hold. The observation appeared anonymously in **[1]** and was repeated in **[2]**.

A few years later, according to L. E. Dickson [**6**], E. Midy published a pamphlet containing his discoveries about periodic decimals [**16**]. It occurred to him to break the period of a repeating decimal into blocks of equal length, then add the corresponding digits in each block. For example, breaking the period of $1/7 = .142857$ into three blocks of two digits each he would add the first digits, $1+2+5=8$, and the second digits, $4+8+7=19$. He had formulas relating these sums to the remainders produced during the process of dividing the denominator of the fraction into the numerator.

Midy noticed that when the period is broken into two blocks of n digits each, corresponding digits in the two blocks will add to 9 in case the denominator of the fraction is relatively prime to 10 and to $10^n - 1$. This implies the condition which we proved earlier, namely that the factor of the denominator which is relatively prime to 10 should divide $10^n + 1 = t_{n,2}$.

P. Lafitte [**13**] proved that if the denominator of a fraction is a prime and the period of its decimal representation contains an even number of digits, then the two halves will sum to $99\ldots9$; thus he proved Midy's discovery in the special case of a prime denominator.

M. Jenkins [**12**] knew that for the two halves of the period of $1/x$ to be complementary, it is both necessary and sufficient that the factors of x, other than 2 and 5, divide $10^n + 1$, where the period length is $2n$. Using this fact he proved another condition to be necessary and sufficient: aside from 2 and 5, the reciprocals of all the other prime factors of x must have even numbers of digits in their periods, and these period lengths must be divisible by the same power of 2.

In the second section on complementarity within the period, we extended Jenkins' discovery, as well as his method of proof, to the case in which the period is broken into any number of parts, not just two. In his proof, Jenkins used an incorrect formula for the length of the period of the reciprocal of a prime power, but fortunately his reasoning did not actually depend upon the mistaken portion of the formula. He also did not seem to realize that his condition applies to any fraction z/x and not just to a reciprocal $1/x$.

P. Mansion [**15**] applied some theorems of E. Catalan [**5**] concerning the remainders one gets when converting z/x to a decimal by division. He found that if the length of the period is $x-1$, then the two halves will add up to $99\ldots9$. He was forced to restrict himself to the case in which the period had maximal length because his proof uses the fact that all of the numbers $1,\ldots,x-1$ are remainders in the division of x into z.

In 1880, O. Schlömilch [**18**] stated that for corresponding digits in the two halves of the period of a reciprocal to add up to 9, it is necessary and sufficient that the denominator divide $10^n + 1$ for some n. He left the proof to the reader and suggested the use of the formula for the sum of a geometric progression. He did not seem to realize that this condition only applies to that factor of the denominator which is relatively prime to 10.

V. Bouniakowsky [**4**] read Schlömilch's note and called attention to the fact that complementarity is not only exhibited by reciprocals. However he did not seem to realize that the property is completely independent of the numerator, for he only suggested that in addition to $1/x$, one should examine the fractions $(x-1)/x$ and $[(x-1)/2]/x$.

Bouniakowsky pointed out the interesting fact that decimals arising in a certain way (from Fibonacci sequences) exhibit complementarity in the two halves of their periods. For example, he thought that the two half periods of .112358437189 887641562819 were complementary in the sense defined by Schlömilch. The digits of this decimal are generated by beginning with two 1's. Thereafter, every digit is the sum of the previous two digits, except that whenever the sum is a two-digit number, the digits are added to form a single digit.

Though he does not say so explicitly, he must have noticed that this example, as well as his others, are not really complementary in Schlömilch's strict sense, namely that *each pair* of corresponding digits in the two halves of the period add up to 9. His twelfth and twenty-fourth digits add up to 18. Bouniakowsky did formulate a different notion of complementarity, namely that corresponding digits in the two half periods add up to a *multiple* of 9. But this is not equivalent to the divisibility condition which Midy, Jenkins, and Schlömilch knew about.

To make his examples truly complementary he should have replaced the second (or first) appearance of a 9 by a 0. With this change his examples also exhibit within their periods complementarity of orders 1,24 and 2,12 and 3,8 and 4,6 and 6,4 but not 8,3. Order 8,3 is left out because the fractional representations of these rationals have the denominator $t_{12,2}$, and this is not a factor of $t_{8,3}$.

In 1912, E. B. Escott [7] used the problem section of the *American Mathematical Monthly* to ask whether anyone knew when the two halves of the period of a fraction will add up to 99 . . . 9. He and others gave several proofs that fractions having denominators which divide $10^n + 1$ for some n have this property. Like Schlömilch they did not seem to realize that this condition is merely sufficient.

Rademacher and Toeplitz [17] used the same reasoning as Lafitte to show that when a fraction with a prime denominator has a period with an even number of digits, the first half plus the second is 99 . . . 9.

Although R. E. Green [9] considered only the reciprocals of primes, he seems to be the first person since Midy to examine the notion of breaking the period into several blocks. He improved upon Midy by adding whole blocks and getting a sum consisting of 9's once the extra high order digits are added to the low order digits. He found that complementarity is exhibited by the period if and only if the prime denominator divides a number of the type which we have denoted $t_{n,r}$.

W. G. Leavitt [14] worked with composite denominators, but he broke the period into only two parts. He did not seem to realize that the divisibility condition is necessary as well as sufficient. His proof is interesting in that it uses the remainders which occur in the division process.

Beck, Bleicher and Crowe [3] also used remainders and broke the period into two parts. Their denominators are prime.

Ross Honsberger [11] treated fractions with prime denominators and broke the period in half. His proof used sums of geometric series and Fermat's Little Theorem.

There are still some unanswered questions: Can Jenkins' condition be extended still further than we have taken it? Do Bouniakowsky's Fibonacci decimals hold further secrets? Are complementary decimals adaptable to computer arithmetic? And if each order n, r is considered to be a point in a lattice, is the resulting structure significant or interesting?

References

[1] Anonymous, Mech. Mag., 313 (1828 or 1829) 411.

[2] Anonymous, Ueber die Eigenschaften der Zahlen, ed. J. G. Dingler, Polytechnisches J., (1829) 68.

[3] Anatole Beck, Michael Bleicher, and Donald Crowe, Excursions into Mathematics, Worth, New York, 1969, pp. 465–470.

[4] V. Bouniakowsky, Quelques remarques sur les propriétés d'une classe particulière des fractions décimales périodiques, Bull. Acad. Impériale Sci. Saint-Pétersbourg, 27 (1881) 362–369.

[5] E. Catalan, Sur les fractions décimales périodiques, Nouvelles Ann. Math., 1 (1842) 457–470.

[6] L. E. Dickson, History of the Theory of Numbers, vol. 1, Stechert, New York, 1934, pp. 161–172.

[7] E. B. Escott et al., Solutions of problems, Amer. Math. Monthly, 19 (1912) 130–132.

[8] H. Goodwyn, J. Natural Philosophy, Chemistry, and the Arts, 1 (1802) 314–316.

[9] R. E. Green, Primes and recurring decimals, Math. Gaz., 47 (1963) 25–33.

[10] E. K. Hayashi, Factoring integers whose digits are all ones, this MAGAZINE, 49 (1976) 19–22.

[11] Ross Honsberger, A property of some repeating decimals, Ingenuity in Mathematics, Random House, New York, 1970, 147–156.

[12] M. Jenkins, Mathematical Questions and Solutions, 7 (1867) 31–32.

[13] P. Lafitte, Théorème sur les fractions périodiques, Nouvelles Ann. Math., 5 (1846) 397–399.

[14] W. G. Leavitt, A Theorem on Repeating Decimals, Amer. Math. Monthly, 74 (1967) 669–673.

[15] Paul Mansion, Sur quelques propriétés des fractions périodiques, Nouvelle Correspondence Math., 1 (1874–75) 8–12.

[16] E. Midy, De Quelques Propriétés des Nombres et des Fractions Décimales Périodiques, Nantes, 1836, 21pp.

[17] Hans Rademacher and Otto Toeplitz, The Enjoyment of Mathematics: Selections from Mathematics for the Amateur, Princeton University Press, Princeton, 1957, pp. 158–160.

[18] O. Schlömilch, Notiz über gewisse periodische Decimalbrüche, Z. Math. Physik, 25 (1880) 416.

[19] William Shanks, Remarks chiefly on $487^2 \equiv 486$, Proc. Roy. Soc. London, 25 (1877) 551–553.

[20] B. M. Stewart, Theory of Numbers, 2nd ed., Macmillan, New York, 1964, pp. 339–340.

Periodic Points of Continuous Functions

Analysis of the cyclic behavior of points under repeated application of a function yields insights into population patterns.

PHILIP D. STRAFFIN, JR.
Beloit College
Beloit, WI 53511

Let $f\colon R \to R$ be a continuous function, and denote the nth iterate of f by f^n: $f^1(x) = f(x)$ and $f^n(x) = f(f^{n-1}(x))$ for $n > 1$. A number x is said to be a point of period k for f if $f^k(x) = x$ and $f^i(x) \neq x$ for all $0 < i < k$. Suppose a continuous function f has points of period k; must it also have points of other periods $l \neq k$? The obvious answer would seem to be "no": why should there be any connection between points of period k and points of period l? Yet a little thought will show that there should be at least some results along these lines.

For instance, if a continuous function f has a point a of period $k > 1$, then it must also have a fixed point, that is, a point of period 1. To see this, suppose $f(a) > a$ and consider the sequence of points $a, f(a), f^2(a), f^3(a), \ldots, f^{k-1}(a), f^k(a) = a$. Then there must be a point $b = f^i(a)$ in the sequence such that $f(b) < b$, or else the sequence would constantly increase and could not return to a. But then the Intermediate Value Theorem (applied to $f(x) - x$) implies that there must be a number c between a and b such that $f(c) = c$. A symmetric argument works if $f(a) < a$.

Hence we have

$$\text{Period } k \Rightarrow \text{Period } 1 \text{ for all } k.$$

In 1975, Li and Yorke published (in [**2**]) a surprising theorem on this question: *If a continuous function $f\colon R \to R$ has a point of period* 3, *then it has points of all periods.* In other words,

$$\text{Period } 3 \Rightarrow \text{Period } l \text{ for all } l.$$

Clearly the periodic behavior of a function with points of all periods is extremely complex. In fact, considering the periodic behavior of physical and biological systems which can be modeled using such functions, Li and Yorke called such behavior "chaos", and titled their paper "Period Three Implies Chaos". Li and Yorke also produced a counterexample to show that Period 5 does not imply Period 3.

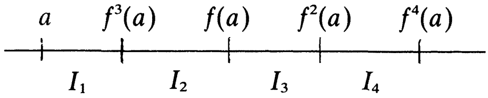
The purpose of this paper is to discuss general techniques for using a point of period k to deduce the existence of points of other periods. We will do this first by constructing an appropriate directed graph that will enable us to generalize Li and Yorke's result by showing, in section two below, that whenever a continuous function f has points of odd period k greater than one, then f has points of every period $l \geqq k - 1$. In other words,

$$\text{Period } k \Rightarrow \text{Period } l \text{ for all odd } k > 1 \text{ and all } l \geqq k - 1.$$

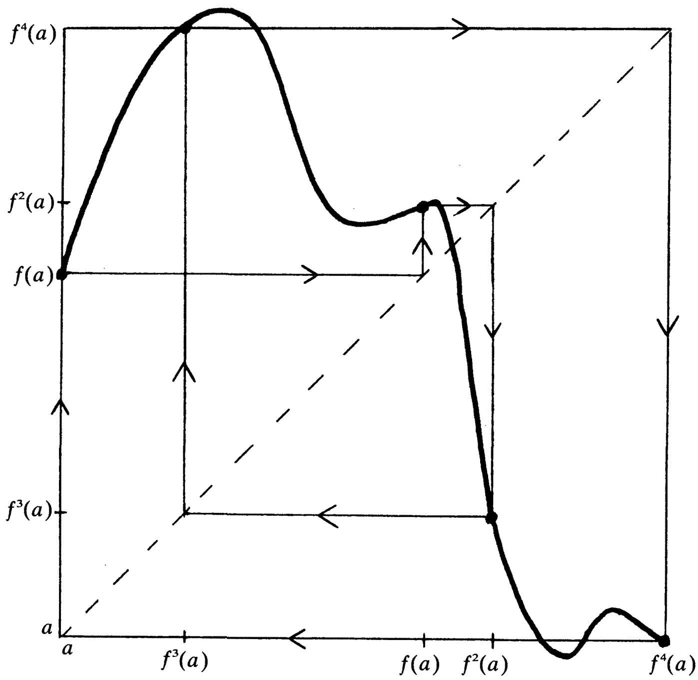
Unknown to Li and Yorke in 1975, and unknown to me when I gave the talk on which this article is based, the Russian mathematician A. N. Sharkovsky had already answered (in 1964) the general question of when period k implies period l. In section three we present Sharkovsky's theorem and show how the method of directed graphs gives a partial proof of his result. Finally, in section four we discuss the relevance of the material of the first three sections to recent work in population ecology.

1. Constructing periodic point digraphs

Consider a continuous function f which has a point x of period k. Among x and its iterates under f, let a be the smallest number. Then starting with a, the iterates of a are spread out along the real line to the right of a, and determine $k-1$ finite closed intervals. Call these intervals, from left to right, $I_1, I_2, \ldots, I_{k-1}$. For example, the function in FIGURE 1 has a point of period five, and gives rise to the intervals



Now consider $f(I_1)$. In our example, I_1 has endpoints a and $f^3(a)$, which get mapped to $f(a)$ and $f^4(a)$. Hence by the Intermediate Value Theorem, $f(I_1)$ must include all points between $f(a)$ and $f^4(a)$. In other words, $f(I_1) \supset I_3 \cup I_4$. Similarly, we get $f(I_2) \supset I_4$, $f(I_3) \supset I_2 \cup I_3$, and $f(I_4) \supset I_1$.



A continuous function with a point of period five.

FIGURE 1

One convenient way to organize this information is in a directed graph, usually called, for short, a digraph. Label the vertices of the digraph $I_1, I_2, \ldots, I_{k-1}$ and draw a directed arc from I_i to I_j if $f(I_i) \supset I_j$. (For our example, we get the digraph in FIGURE 2.) This construction starting with the periodic point and ending with the digraph is perfectly general: it works for any continuous function f with a point of period k, for any k. We will call a digraph arising from this construction a ***k*-periodic point digraph.** The interesting thing is that one can read from the digraph information about other periodic points which f must have.

THEOREM A. *If a k-periodic point digraph associated to f has a non-repetitive cycle of length l, then f must have a point of period l.*

Cycles, as we interpret them, are allowed to use a vertex or edge more than once; by a "non-repetitive cycle" we simply mean one that does not consist *entirely* of a cycle of smaller length traced several times. In the digraph of FIGURE 2, $I_1I_3I_2I_4I_1I_3I_2I_4I_1$ is a repetitive cycle of length eight (illegal), while $I_1I_3I_3I_3I_3I_3I_2I_4I_1$ is a non-repetitive cycle of length eight (legal).

The proof of Theorem A is modeled on Li and Yorke's proof. It uses two lemmas which are standard in analysis courses:

LEMMA 1. *Suppose I and J are closed intervals, f continuous, and $J \subset f(I)$. Then there is a closed interval $Q \subset I$ such that $f(Q) = J$.*

LEMMA 2. *Suppose I is a closed interval, f continuous, and $I \subset f(I)$. Then f has a fixed point in I.*

To prove the theorem, suppose we are given a non-repetitive sequence of closed intervals $I^0, I^1, \ldots, I^l = I^0$ (the superscript refers merely to the position in the sequence) such that $f(I^i) \supset I^{i+1}$. We wish to show that f has a point of period l. Consider the diagram:

$$\begin{array}{ccc} I^0 & \xrightarrow{f} & f(I^0) \\ & & \cup \\ & & I^1 \end{array} .$$

(The arrow in this and subsequent diagrams represents an onto map.) Using Lemma 1, we can find a closed interval Q_1 to fill in the diagram:

$$\begin{array}{ccc} I^0 & \xrightarrow{f} & f(I^0) \\ \cup & & \cup \\ Q_1 & \xrightarrow{f} & I^1 \end{array} .$$

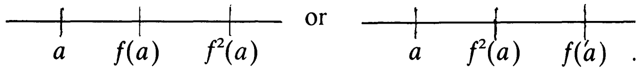
In fact, using Lemma 1 repeatedly, we can construct the following diagram:

$$\begin{array}{lllll} I^0 \xrightarrow{f} f(I^0) & & & & \\ \cup \qquad\quad \cup & & & & \\ Q_1 \xrightarrow{f} I^1 & \xrightarrow{f} f(I^1) & & & \\ \cup & \qquad \cup & & & \\ Q_2 \xrightarrow{\qquad f^2 \qquad} & I^2 & \xrightarrow{f} f(I^2) & & \\ \cup & & \qquad \cup & & \\ \vdots & & \qquad \ddots & & \\ \cup & & & & \\ Q_{l-1} \xrightarrow{\qquad\qquad f^{l-1} \qquad\qquad} & & & I^{l-1} \xrightarrow{f} & f(I^{l-1}) \\ \cup & & & & \cup \\ Q_l \xrightarrow{\qquad\qquad\qquad f^{l} \qquad\qquad\qquad} & & & & I^l = I^0 \end{array} .$$

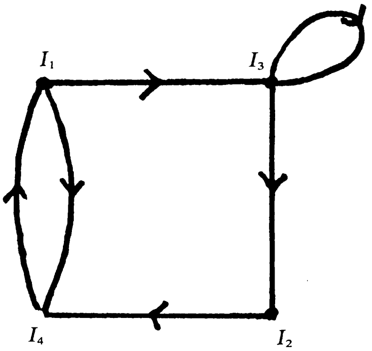
Now reading up the left hand column and across the bottom row, we see that $Q_l \subset I^0 = f^l(Q_l)$. Hence by Lemma 2, f^l has a fixed point x in Q_l. Since the sequence of intervals is non-repetitive, one can show that x is indeed a point of period l for f (rather than a point of period $< l$).

2. Analying periodic point digraphs

For $k = 3$, there are two possible orderings of points of period three:

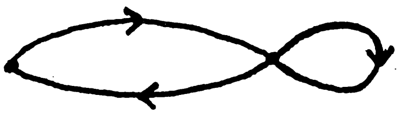


These orderings give rise to the same digraph, illustrated in FIGURE 3; we call this the Li-Yorke digraph. Since it clearly has non-repetitive cycles of all lengths, we recover the Li-Yorke theorem by applying Theorem A.



The periodic point digraph of the function in FIGURE 1.

FIGURE 2



The Li–Yorke digraph.

FIGURE 3

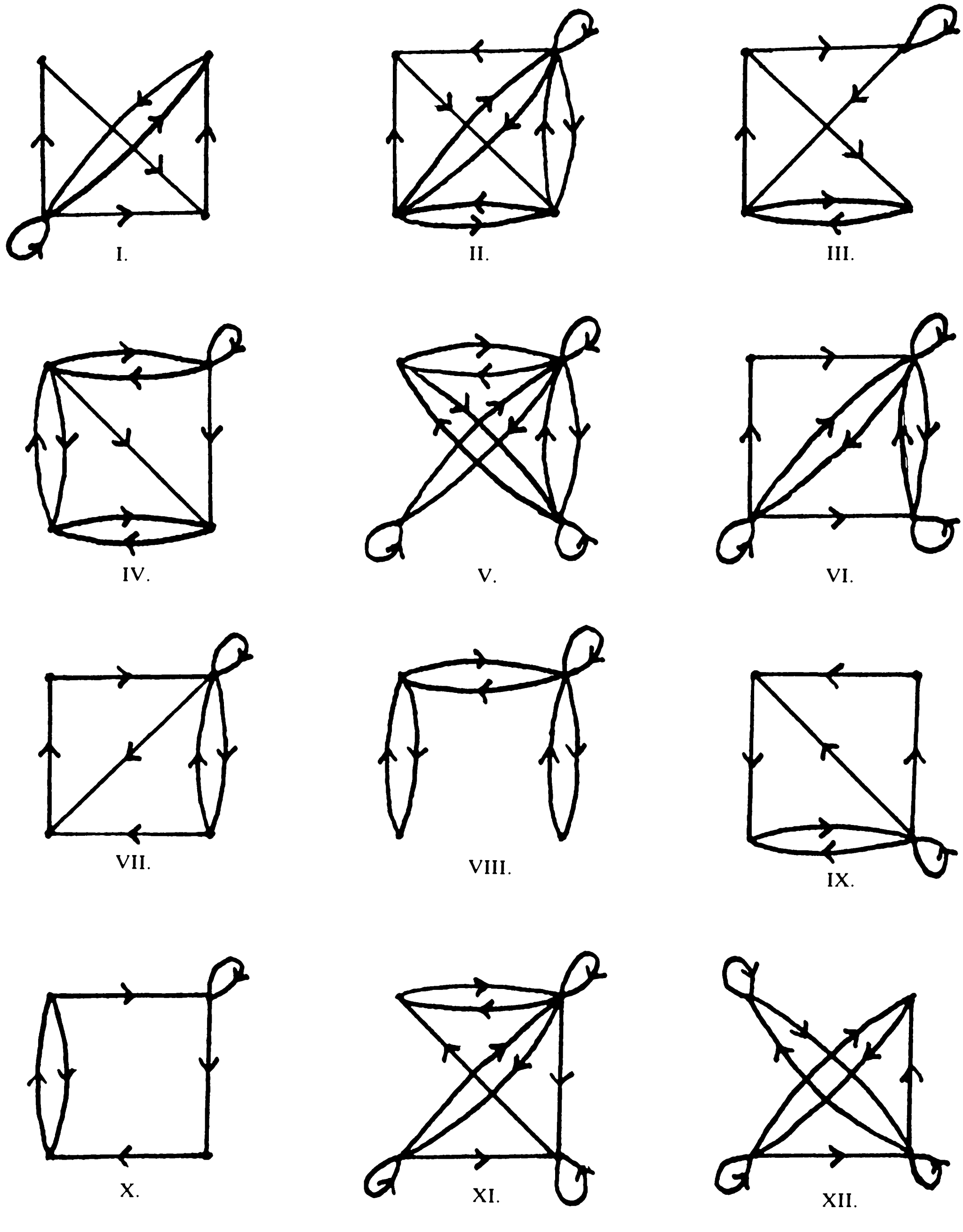
For $k = 5$, there are twelve possible digraphs, all shown in FIGURE 4. All except III and X contain the Li-Yorke digraph as a subgraph, and hence correspond to functions with points of all periods. You can check that III also has non-repetitive cycles of all lengths. X, our first example, has cycles of all lengths except three, and it corresponds to Li and Yorke's counterexample that Period 5 does not imply Period 3. However, we can conclude that

$$\text{Period } 5 \Rightarrow \text{Period } l \text{ for all } l \neq 3.$$

In general, the number of k-periodic point digraphs increases factorially with k, so that exhaustive analysis quickly becomes unmanageable. However, we can use an inductive technique to prove a general result:

THEOREM B. *If a continuous function f: $R \to R$ has a point of odd period $k > 1$, then it must have periodic points of all periods greater than or equal to $k - 1$.*

Proof. We first show that for all $k > 1$, a k-periodic point digraph has a cycle of length k. To do this, recall the notation of section one. Set $I^0 = I_1$, which has a as one endpoint. Define I^n for $n \geqq 1$ to be the interval I_i which has $f^n(a)$ as one endpoint and is contained in $f(I^{n-1})$. This construction gives $I^k = I^0$, so that we obtain a cycle of length k. (For instance, in the example of section one, I^1 would be the interval which has $f(a)$ as one endpoint and is contained in $f(I^0) = f(I_1)$. This is I_3. You can check that the construction gives $I_1I_3I_3I_2I_4I_1$ as the cycle of length five.) Since there are only $k - 1$ vertices in a k-periodic point digraph, some vertex must be repeated in this k-cycle, and at this vertex, the k-cycle decomposes into two cycles of smaller length. (For instance, $I_1I_3I_3I_2I_4I_1$ decomposes into the 4-cycle $I_1I_3I_2I_4I_1$ and the loop at I_3.) Since the repeated interval has only two endpoints, it can appear only twice in the original k-cycle. Hence it can appear only once in each of the smaller cycles, and both of these smaller cycles must be non-repetitive. Finally, when k is odd, one of the smaller cycles must be of odd length.



All 5-periodic point digraphs.

FIGURE 4

The theorem can now be proved by induction on odd $k > 1$. We already have the result for $k = 3$ or 5. Suppose it is true for all odd numbers between 1 and k. If f has a point of period k, then we know its periodic point digraph has a cycle of length k, and this cycle decomposes into two smaller cycles, one of which is of odd length and non-repetitive. If this length is not 1, we are done by Theorem A and the induction hypothesis. If it is of length 1, (i.e., is a loop) then we can get non-repetitive cycles of all lengths greater than or equal to $k - 1$ by traveling the loop as often as we need, and then completing the complementary $k - 1$ cycle. Again, Theorem A immediately gives the desired result.

3. Sharkovsky's theorem

In 1964 A. N. Sharkovsky completely answered the question of when period k implies period l. His work [7] was published in the *Ukranian Journal of Mathematics* in Russian, and has not been translated. His result is remarkable, and deserves to be better known in the West than it has been.

THEOREM. *A continuous function f: $R \to R$ which has a point of period k, must also have a point of period l precisely when k precedes l in the following ordering*

$$3, 5, 7, 9, \ldots, 3 \cdot 2, 5 \cdot 2, \ldots, 3 \cdot 2^2, 5 \cdot 2^2, \ldots, \ldots, ., ., 2^3, 2^2, 2, 1$$

of all positive integers.

Thus our observation that period k implies period 1 for all k, just corresponds to 1 being at the end of Sharkovsky's ordering, and Li and Yorke's result just corresponds to 3 being at the beginning. Theorem B also clearly follows from Sharkovsky's Theorem. You can read off other interesting results, for instance that a continuous function f: $R \to R$ which has a point of period not equal to a power of 2, must have an infinite number of periodic points.

Sharkovsky's Theorem can be derived from three basic results:

(i) Period $k \Rightarrow$ Period 2, for all $k > 1$.

(ii) Any odd period $> 1 \Rightarrow$ all higher odd periods.

(iii) Any odd period $> 1 \Rightarrow$ all even periods.

Perhaps you can see, and Sharkovsky shows in his article, how to combine (i), (ii) and (iii) with careful consideration of f^{2^n} to obtain the general result. Of the three basic results, (i) is fairly easy to show directly. However, (ii) and (iii) are not obvious, and Sharkovsky's proof of them is long and very complicated. He constructs so many sequences of points that eight complex figures and most of the letters of the Greek alphabet are necessary to keep track of them. Sharkovsky's Theorem is an example of a common occurrence in mathematics — an elegant result whose first proof is extremely inelegant.

Theorem B comes close to providing a more elegant proof of Sharkovsky's Theorem for it embodies (ii) and most of (iii). To complete the proof of Sharkovsky's Theorem, we would only have to show that any odd period $k > 1$ implies all even periods between 2 and $k - 1$. Could the reader fill this gap using the method of k-periodic point digraphs?

Finally, we should note that it is possible to approach these questions using completely different techniques. For instance, John Guckenheimer [1] has recently succeeded in proving Sharkovsky's Theorem for a certain class of functions using the methods of symbolic dynamics.

4. Population ecology

Recent concern about the pure mathematical question of the nature of periodic points of continuous functions was generated by ecologists, who in turn had been stimulated by earlier work of a meteorologist. References can be found in [2], [3], [4], and [5]. The ecological problem is to describe the behavior over time of, say, an insect population with discrete generations, which might behave according to the equation

$$x_{t+1} = rx_t(1 - x_t).$$

Here x_t is the size of the tth generation, r is the "intrinsic rate of increase", and $(1 - x_t)$ is a damping term due to environmental limitations. In this model, periodic points of the function $f(x) = rx(1 - x)$ would correspond to cyclical behavior of the insect population.

What happens for this function is that as r increases, more and more periodic points begin to appear. Points of period 2 appear as soon as $r > 3.00$, period 4 comes in at $r \approx 3.45$, then periods 8, 16, etc. All periods of the form 2^n are in by $r \approx 3.57$, when points of other periods begin to appear. The first points of odd period appear when $r \approx 3.68$, and period 3 finally comes in at $r \approx 3.83$. As you can see, the first appearance of points of different periods appears to run exactly backwards through

Sharkovsky's ordering, as r increases between 3.00 and 3.83. If this is true, and no one has yet produced a formal proof that it is, we would have one family of counterexamples, much simpler than the counterexamples Sharkovsky originally constructed in [7], to show that the Sharkovsky result is strict: no other implications are possible. Ecologically, an infinite number of periodic points by $r = 3.57$, and points of all periods by $r = 3.83$, means that for r this large, periodic behavior becomes chaotic. An excellent survey of these kinds of results from an ecological point of view has recently appeared in [6].

References

[1] John Guckenheimer, On the bifurcation of maps of the interval, preprint.

[2] T.-Y. Li and James Yorke, Period three implies chaos, Amer. Math. Monthly, 82 (1975) 985–992.

[3] E. N. Lorenz, The problem of deducing the climate from the governing equations, Tellus, 16 (1964) 1–11.

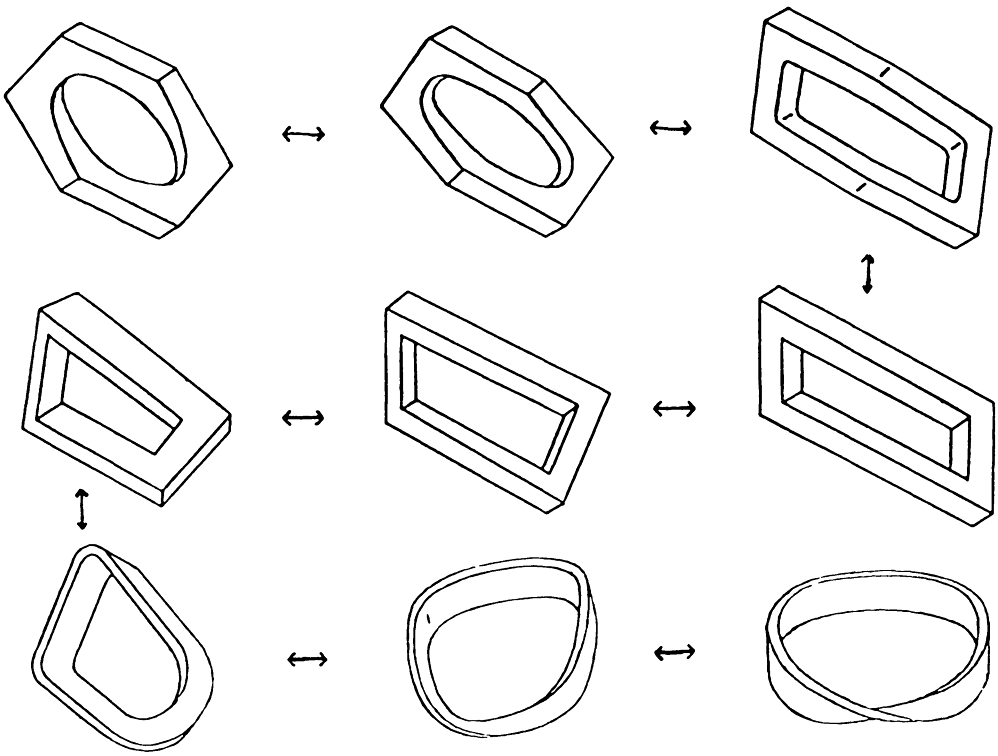
[4] Robert May, Biological populations with non-overlapping generations: stable points, stable cycles, and chaos, Science, 186 (1974) 645–647.

[5] ———, Biological populations obeying difference equations: stable points, stable cycles, and chaos, J. Theoret. Biol., 51 (1975) 511–524.

[6] ———, Simple mathematical models with very complicated dynamics, Nature, 261 (1976) 459–467.

[7] A. N. Sharkovsky, Co-existence of the cycles of a continuous mapping of the line into itself, Ukranian Math. J., 16 (1964) 61–71 (Russian).

Möbius



CLIFF LONG
Bowling Green State Univ.
Bowling Green, Ohio

NOTES

Calculator Algorithms

JAMES C. KROPA
Clayton Junior College
Morrow, GA 30260

Increased emphasis on numerical calculations in elementary mathematics courses coupled with the popularity of pocket calculators makes it appropriate to examine the mathematical algorithms used in these calculators. Even though these algorithms can be found in the literature, this literature is not readily available to the mathematician and is filled with initially confusing terminology such as "pseudo-division", "pseudo-rotation", and "cross-addition" and with special computer oriented concepts. In this paper, we present the calculator algorithms in a manner readily understandable to a mathematician and his student. We will begin with a trigonometric algorithm, which serves as a paradigm for exponential and logarithmic computations, and shall show, at the end of this note, how its *schema* can unify even the multiplication and division computations.

The trigonometric algorithms of most pocket calculators use the CORDIC (Coordinate Rotation Digital Computer) technique which is based on the 1624 work of Henry Briggs in *Arithmetica Logarithmica.* In actual practice there are several versions of the same basic algorithm; we will

$$R_1^2 = X_1^2 + Y_1^2 \qquad w_1 = \tan\alpha_1 = \frac{Y_1}{X_1} \qquad X_2 = X_1 - w_2 Y_1 \qquad R_2 = R_1\sqrt{1+w_2^2}$$

$$R_2^2 = X_2^2 + Y_2^2 \qquad w_2 = \tan\alpha_2 \qquad Y_2 = Y_1 + w_2 X_1 \qquad \quad = R_0\sqrt{1+w_1^2}\sqrt{1+w_2^2}$$

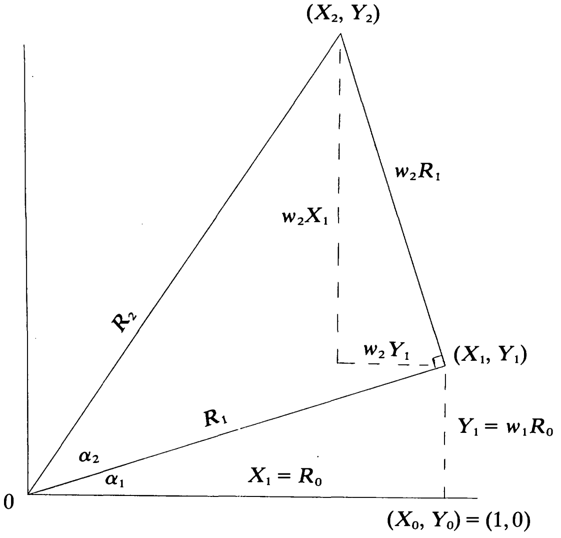


FIGURE 1

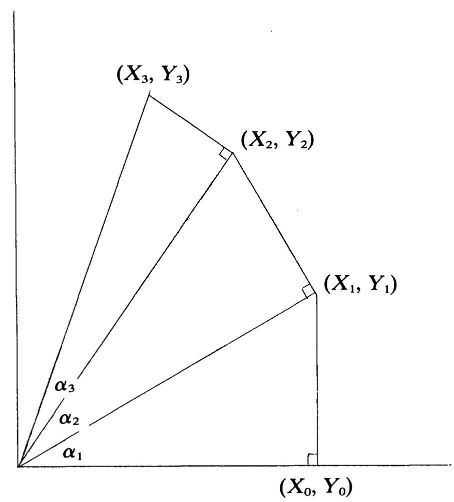


FIGURE 2

indicate alternate versions within parentheses where appropriate. It is also difficult to find out which version is used in which calculator. The emphasis in this paper will be on the conceptional understanding of the algorithms used in calculators rather than on supplying complete detail which can be found in the references. For trigonometric computations the fundamental idea of the CORDIC algorithm is based on the relations illustrated in FIGURE 1.

Suppose one wants to find the trigonometric functions of a positive angle θ of less than 360°. The algorithm finds acute angles $\alpha_1, \alpha_2, \ldots, \alpha_n$ so that $\theta \approx \alpha_1 + \alpha_2 + \cdots + \alpha_n$, sets $w_i = \tan \alpha_i$, and then performs the following calculations $X_0 = 1$, $Y_0 = 0$, $X_i = X_{i-1} - w_i Y_{i-1}$, $Y_i = Y_{i-1} + w_i X_{i-1}$, for $i = 1, 2, \ldots, n$. (The situation for $n = 3$ is illustrated in FIGURE 2.) When this calculation is completed, the trigonometric functions of θ can be calculated using the usual trigonometric formulas applied to the coordinates (X_n, Y_n). Note that (X_n, Y_n) is probably not on the unit circle. (In certain versions of the algorithm the length from the origin to (X_n, Y_n) given by $\sqrt{1+w_1^2}\sqrt{1+w_2^2}\cdots\sqrt{1+w_n^2}$ which equals $(1/\cos\alpha_1)(1/\cos\alpha_2)\cdots(1/\cos\alpha_n)$ is used instead of some trigonometric formulas.)

The computational power of the algorithm is enhanced by choosing $\alpha_1, \alpha_2, \ldots, \alpha_n$ so that $w_1, w_2, \ldots, w_n$ are simple one significant digit numbers. Since most calculators use decimal digits (which are binary coded), each w_i will be 10^{-k} for some nonnegative integer k. For example if $X_{i-1} = .4214$ and $Y_{i-1} = .6582$ and $w_i = .0010$, then the computations necessary to compute X_i and Y_i involve multiplication by .0010 which can be considered a shift operation and addition or subtraction. If each w_i is 10^{-k} for some k, then α_i must be $\tan^{-1} 10^{-k}$. (In a binary computer each w_i is 2^{-k} and α_i is $\tan^{-1} 2^{-k}$ for some k.) The approximate values for the numbers $\tan^{-1} 10^{-k}$ for $k = 0, 1, 2, \ldots, , m$ are stored in the calculator where m is chosen with consideration to the number of digits of the calculator.

To choose particular α_i for each given θ, the following procedure is often used. Find the largest nonnegative integer q_0 so that $q_0 \tan^{-1} 1 \leqq \theta$. If q_0 is greater than or equal to 1, then let $\alpha_1 = \alpha_2 = \cdots = \alpha_{q_0} = 45°\ (= \tan^{-1} 1)$. Now find the largest nonnegative integer q_1 so that $q_0 \tan^{-1} 1 + q_1 \tan^{-1} .1 \leqq \theta$. If q_1 is greater than or equal to 1, then let all the α_i from $q_0 + 1$ to $q_0 + q_1$ be $\tan^{-1} .1$ which is approximately 5.71°. This process is continued until θ is approximately equal to $\Sigma_{j=0}^{k} q_j \tan^{-1} 10^{-j}$.

The following example should help illustrate the algorithm. First we tabulate basic values of α_i:

k	0	1	2	3	4
$\tan^{-1} 10^{-k}$	45°	5.711°	.573°	.057°	.006°

We apply the algorithm to compute tan 51°. Since $51° \approx 45° + 5.711° + 0(.573°) + 5(.057°) + 0(.006°)$, we have $q_0 = 1$, $q_1 = 1$, $q_2 = 0$, $q_3 = 5$, $q_4 = 0$. The data for this computation is given in TABLE 1. From it we can see that $\tan 51° \approx X_7/Y_7 = 1.1045/.8945 \approx 1.2348$. (Of course most calculators use ten or more decimal places rather than five.)

(It might be noted that certain calculators using CORDIC for θ less than or equal to 90° have a certain predetermined set of angles $\beta_1, \beta_2, \ldots, \beta_m$, namely: $\beta_1 = \tan^{-1} 1$, $\beta_2 = \beta_3 = \cdots = \beta_{10} = \tan^{-1} .1$, $\beta_{11} = \cdots = \beta_{19} = \tan^{-1} .01$, $\beta_{20} = \cdots = \beta_{28} = \tan^{-1} .001$, etc. For an angle θ the α_i are chosen so that $\theta \approx \alpha_1 + \alpha_2 + \cdots + \alpha_m$ and for each i, α_i is plus or minus β_i. This method, although it is longer for certain angles, has the advantage that the vector (X_m, Y_m) for any angle θ will have a length given by

i	α_i	$\Sigma\alpha_i$	w_i	X_i	Y_i
0				1.0000	.0000
1	45°	45.000°	1	1.0000	1.0000
2	5.711°	50.711°	.1	.9000	1.1000
3	.057°	50.768°	.001	.8989	1.1009
4	.057°	50.825°	.001	.8978	1.1018
5	.057°	50.882°	.001	.8967	1.1027
6	.057°	50.939°	.001	.8956	1.1036
7	.057°	50.996°	.001	.8945	1.1045

TABLE 1

$K_m = (1/\cos\beta_1)(1/\cos\beta_2)\cdots(1/\cos\beta_m)$ that can be stored in the calculator. Note that for each i, $\cos\alpha_i = \cos\beta_i$.)

For angles greater than 360° or less than 0°, an appropriate (positive or negative) multiple of $2\pi = 360°$ is added to the angle to bring it into the range of 0° to 360°. Even though the algorithm has been described using degree measure, many calculators and computers also compute with radian measure and many use only radian measure internally. It turns out that many calculators do not accurately compute the trigonometric functions of large numbers, greater than 10^{10}, because the value of π used in the calculator to reduce the given number to between 0 and 2π is only given to 10 to 13 significant figures.

Turning now to inverse trigonometric functions, suppose one of $\sin^{-1}w, \cos^{-1}w$, or $\tan^{-1}w$ is to be calculated, where there are restrictions on w for the first two. For simplicity assume w is greater than 0. In each case there are nonnegative numbers x and y so that $x^2+y^2=1$ and so that either $y = w$, $x = w$, or $y/x = w$, respectively. In each case the measure of the angle θ whose tangent is y/x needs to be found.

One version of the algorithm does the following. Let $(X_0, Y_0) = (x, y)$, and calculate $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ using the previous formulas where the α_i are chosen from $\{-\tan^{-1}1, -\tan^{-1}.1, -\tan^{-1}.01, \ldots\}$ so that Y_i gets small but stays nonnegative. For example, if $(x, y) = (X_0, Y_0) = (.6, .8)$, $\alpha_1 = -45°$, $(X_1, Y_1) = (1.4, .2)$, $\alpha_2 = -\tan^{-1}.1 \approx -5.7°$, $(X_2, Y_2) = (1.42, .06)$, $\alpha_3 = -\tan^{-1}.01$, etc. When the process is complete, $\theta \approx -(\alpha_1 + \cdots + \alpha_n)$. Note that in this process nonnegative integers q_i can be found so that $\theta \approx \sum_{j=0}^{m} q_j \tan^{-1}10^{-j}$.

(Another version of the algorithm starts out with $(X_0, Y_0) = (1, 0)$ and computes (X_i, Y_i) with a predetermined set of angles, except for sign, where the sign of each angle is chosen by reference to the original coordinates (x, y). Knowing the length of each (X_i, Y_i) helps in the process.)

For some other transcendental functions two different approaches are frequently used. First an algorithm for hyperbolic trigonometric functions can be generated in a similar manner to regular trigonometric functions. After hyperbolic trigonometric functions are computed, exponentials and then logarithms can be generated using the usual mathematical relationships. For details of this approach consult **[3]**. The second approach uses algorithms for transcendental functions found in **[2]**. These algorithms have some surprising mathematical similarity to the trigonometric algorithms that have been illustrated.

To obtain a similar algorithm for the natural logarithm, $\ln(1+(y/x))$ is computed. (To obtain $\ln z$, find $\ln(1+(y/x))$ where $x = 1$ and $y = z-1$ when $z \geqq 1$, and find $-\ln(1+(y/x))$ where $x = z$ and $y = 1-z$ when $z < 1$.) To compute $\ln(1+(y/x))$ let $X_0 = x$, $Y_0 = y$, $X_i = X_{i-1} + w_i X_{i-1}$, and $Y_i = Y_{i-1} - w_i X_{i-1}$, for $i = 1, 2, \ldots, n$, where each $w_i = 10^{-j}$ for some nonnegative integer j. In a manner similar to that used for the inverse trigonometric functions, the largest values of 10^{-j} are used to reduce the values of the Y_i's to 0, keeping them positive. For each 10^{-j}, a nonnegative integer q_j is found indicating how many times 10^{-j} is used. It turns out that

$$\ln(1+(y/x)) \approx \sum_{j=0}^{m} q_j \ln(1+10^{-j})$$

where the values of $\ln(1+10^{-j})$ for $j = 0, 1, \ldots, m$ are stored in the calculator.

To obtain a similar algorithm for the exponential function, we compute $x(e^y-1)$: To do this, let $X_0 = x$, $Y_0 = y$, $X_i = X_{i-1} + w_i X_{i-1}$, and $Y_i = Y_{i-1} + w_i X_{i-1}$, for $i = 1, 2, \ldots, n$, where the $w_i = 10^{-j}$ for some j are chosen in the following manner. For each j the largest nonnegative integer q_j is found so that $y \approx \sum_{j=0}^{m} q_j \ln(1+10^{-j})$. (The values $\ln(1+10^{-j})$ for $j = 0, 1, \ldots, m$ are, as usual, stored in the calculator.) The w_i are then chosen in the same manner as for the tangent, namely

$$\text{if } q_0 \geqq 1,\ w_1 = w_2 = \cdots = w_{q_0} = 10^{-0} = 1,$$

$$\text{if } q_1 \geqq 1,\ w_{q_0+1} = \cdots = w_{q_0+q_1} = 10^{-1}, \text{ etc.}$$

Under these conditions the sequence Y_n converges to $x(e^y-1)$.

To help the reader begin to understand the validity of the last two algorithms, we outline briefly a special procedure for calculating e^z. Suppose numbers $w_1, w_2, \ldots, w_n$ can be found so that $z = \sum_{i=1}^n \ln(1+w_i)$; then $e^z = \prod_{i=1}^n (1+w_i)$. Now let $Z_0 = 1$ and $Z_i = Z_{i-1} + w_i Z_{i-1}$, for $i = 1, 2, \ldots, n$. Then $Z_1 = 1 + w_1$, $Z_2 = (1+w_1) + w_2(1+w_1) = (1+w_1)(1+w_2), \ldots, Z_n = \prod_{i=1}^n (1+w_i) = e^z$.

One has to be impressed by the mathematical similarity of the four algorithms for trigonometric, inverse trigonometric, logarithmic, and exponential functions that we have presented. Their similarity enables them to be microprogrammed into one master program in the calculator. To indicate further why these algorithms are so adaptable to calculator use, we will conclude with a brief discussion of how calculators can multiply and divide.

When 524 is multiplied by 3.71, $524 \times 3.71 = 524 \times 3 + 524 \times .7 + 524 \times .01 = 3(524 \times 1) + 7(524 \times .1) + 1(524 \times .01)$. So set, $X_0 = 524$, $Y_0 = 0$, $X_i = X_{i-1} + 0\,Y_{i-1}$, and $Y_i = Y_{i-1} + w_i X_{i-1}$, for $i = 1, 2, \ldots, 11$, where $w_1 = w_2 = w_3 = 1$, $w_4 = w_5 = \cdots = w_{10} = .1$, $w_{11} = .01$. It turns out that $Y_{11} = 524 \times 3.71$. Note that the number of w_i's which are 1 (q_0 in earlier algorithms) is 3, the number of w_i's which are .1 (q_1 in earlier algorithms) is 7, and the number of w_i's which are .01 (q_2 in earlier algorithms) is 1. The numbers 3, 7, 1 are the digits of 3.71.

Turning now to division, assume y/x is less than 10. Let q_0 be the largest nonnegative integer so that $xq_0 \leqq y$; then $q_0 < 10$. Now let q_1 be the largest nonnegative integer so that $xq_0 + xq_1(.1) \leqq y$. Continuing this process, nonnegative integers $q_0, q_1, \ldots, q_m$ all less than 10 are found so that $y \approx x \sum_{j=0}^m q_j 10^{-j}$. It is clear that y/x approximately equals $q_0 . q_1 q_2 \cdots q_m$.

In each of the algorithms for transcendental functions, part of the calculation was similar to the multiplication just presented, and part was similar to the division just presented. The use of such terminology as pseudo multiplication and pseudo division in the references should now be apparent.

References

[1] D. S. Cochran, Algorithms and accuracy in the *HP*−35, Hewlett-Packard Journal, 23, 10, (June 1972) 10–11.

[2] J. E. Meggitt, Pseudo division and pseudo multiplication processes, IBM J., (April 1962) 210–226.

[3] H. Schmid and A. Bogacki, Use decimal CORDIC for generation of many transcendental functions, EDN Magazine, February 20, 1973, 64–73.

[4] J. E. Volder, The CORDIC trigonometric technique, IRE Transactions on Electronic Computers, September 1959, 330–334.

Rope Strength under Dynamic Loads: The Mountain Climber's Surprise

JOEL ZEITLIN

California State University,
Northridge, CA 91374

After taking up rock climbing, I became quite interested in the strength of ropes. A rope must be strong enough not only to hold the climber's "dead" weight but also to support the climber in case of a fall. In this note we develop formulas for the static or breaking strength of a rope and for the force exerted on a rope by the fall of a given weight from a given height. We then use these formulas to determine the strength of some standard ropes and, finally, the strength of a system involving two ropes of different size tied together in series.

The static strength of a rope is the (dead) weight it takes to break the rope. For a rope with uniform cross sectional area, static strength B is given by a constant times the cross sectional area. Thus for a rope with diameter d the static strength is given by $B(d) = cd^2$, where c is some constant. This is, in fact, the force needed to break the rope.

If we drop a weight on the rope from a given height, then we will generate greater force at the moment when the rope is stretched to its maximum displacement. We can analyze this situation by assuming that the rope behaves according to Hooke's Law (which is usually applied to springs) that force equals a constant multiple of displacement. (Manufacturers' charts of elongation as a function of load show that the relationship is not exactly linear, but that calculations based on Hooke's Law will give a good approximation.) Hence we will assume that $F = kx$ where x is the displacement (or stretch) of the rope, and $k = EA/L$ is Hooke's constant that depends on the cross sectional area $A = \pi d^2/4$, the length L, and the material constant E of the rope. According to Hooke's Law, maximum force occurs at maximum displacement.

We write x as a function of time, $x(t)$, where $t = 0$ at the time the stretching begins and t_1 is the time of maximum displacement. Thus $x(0) = 0$ while $x(t_1)$ equals maximum displacement $x_{\max}$. Moreover, $\dot{x}(0)$ is the initial velocity v_0 at the moment the rope begins to stretch, while $\dot{x}(t_1) = 0$ at the moment the rope has stretched as far as it will and begins to spring back. We are really interested in $F_{\max}$, the maximum force which will be produced at t_1 in order to determine when $F_{\max} = B(d)$.

The mass on the end of the rope will be subject to a downward force of $w = mg$ and an upward force of kx. The algebraic sum of these forces gives $m\ddot{x}$. Since we are taking x positive in the downward direction, we have $m\ddot{x} = -kx + w$. Multiplying both sides by $\dot{x}$ and integrating from $t = 0$ to $t = t_1$ we obtain

$$-\frac{1}{2} mv_0^2 = -\frac{1}{2} kx_{\max}^2 + wx_{\max},$$

where $x_{\max} = x(t_1)$. Solving for $x_{\max}$ yields

$$x_{\max} = \frac{w \pm \sqrt{w^2 + kmv_0^2}}{k}.$$

The physical nature of our problem indicates that we should choose the plus sign.

Consider now a body starting from rest and falling for time t a distance s until the rope is taut, stretching is about to begin, and the velocity is v_0. Then $s = gt^2/2$ and $v_0 = gt$, so $v_0^2 = 2sg$. Since $k = EA/L$ (where A is the cross-sectional area and L is the length of the rope) we have

$$x_{\max} = \frac{w + \sqrt{w^2 + 2EA(s/L)mg}}{k} = \frac{w + \sqrt{w^2 + 2EAfw}}{k},$$

where $f = s/L$ is called the fall factor and $w = mg$ is the weight. The maximum force is given by

$$F_{\max} = kx_{\max} = w + \sqrt{w^2 + 2EAfw}. \tag{1}$$

Thus for any given weight, the maximum force produced depends on the fall factor f, i.e., on the ratio of distance fallen to length of rope. So the maximum force produced by a 10 meter fall on a 20 meter rope is the same as that produced by a 10 cm fall on a 20 cm rope. (To make this plausible, you may think of longer ropes as having more cushion.)

We want the critical (or maximum) weight, w_m, which the rope will sustain. That is, we wish to find w_m, so that $F_{\max} = B$, or

$$w_m + \sqrt{w_m^2 + 2EAfw_m} = B.$$

Solving for w_m we obtain

$$(2) \qquad w_m = \frac{B^2}{(2EAf+2B)}.$$

We now must obtain the proportionality constant c in the equation for static strength B and the material constant E in Hooke's Law. The *Edelrid Guide to Mountaineering Ropes* [**1**] gives the static or breaking strength of their 11 mm nylon rope to be about 2175 kg. Since $B(11) = c(11)^2 = 2175$ we obtain $c = 2175/121 = 17.975 \approx 18$ kg/mm^2. Thus $B = 18d^2$. The *Edelrid Guide* also reports [**1**, pp. 35] that an impact force of 970 kg was created by dropping a weight of 80 kg with a fall factor of 1.78. Substituting these values into (1) and solving yields $EA = 2780$ and so $E = 2780/\pi(5.5)^2 = 29.25$ kg/mm^2.

The fall factor f usually varies from 0 to 2. The value $f = 2$ is the worst possible situation which occurs when one climbs the full length of rope above a point of protection where the rope is fastened and then falls past the point of protection the full length of the rope. We obtain the critical weight by setting $f = 2$ and substituting into equation (2):

$$(3) \qquad w_m = \frac{(18d^2)^2}{(2)(29.25)(\pi)(d/2)^2(2)+(2)18d^2} = 2.53d^2,$$

where w_m is the weight in kilograms and d is the diameter in millimeters. For $d = 11$ we get $w_m = 306.13$ kg, so a climber should be very safe with this size rope.

My curiosity and fears were aroused anew when I realized that in an actual climbing situation the sturdy 11 mm rope is often connected to a point of security (such as a "nut" carefully wedged into a narrowing granite crack) by a thinner rope. What we are dealing with, then, is two different ropes tied together. It then seemed appropriate to consider ropes of smaller diameter. The thinnest rope used is a 5 mm rope. For $d = 5$, (3) yields $w_m = 63.25$ kg (or 139.15 lbs.). *I weigh more than this*!

But the 5 mm rope is not used alone. To be complete we need to analyze the strength of a rope which is made up of two different sized ropes tied together. In practice, a thin rope is attached to a secure point and then a second, thicker rope is tied to the thin rope. Assuming all connections are ideal (i.e., the knots don't weaken the structure) and that Hooke's Law holds, we obtain $F = k_c x$ where k_c is the constant for the combined length of rope, and x is the amount of stretch. But $x = x_1 + x_2$ where x_i is the amount of stretch in the ith rope. Let k_i be the matching Hooke's constant and assume that the force at any instant is constant throughout the full length of rope. Hence $F = k_1 x_1$ and $F = k_2 x_2$. Combining these we see that

$$(4) \qquad k_c = \frac{F}{x_1+x_2} = \frac{F}{F/k_1+F/k_2} = \frac{1}{1/k_1+1/k_2}.$$

We now consider a concrete situation. Assume we have 1 meter of 5 mm diameter rope and 3 meters of 11 mm diameter rope. Since $k = EA/L$, we obtain $k_1 = 574$, $k_2 = 927$ and so $k_c = 354$. Replacing EA by kL in (2) yields the formula $w_m = B^2/(2ks+2B)$. We have $B = B(5) = (18)(5^2) = 450$, $s = 8$, and $k = 354$ which yields $W = 30.85$ kg $= 68$ lbs. This is a surprise — the combination of a 5 mm rope and an 11 mm rope is not nearly as strong as a 5 mm rope by itself! Indeed we can say that *the chain is weaker than its weakest link.*

This result follows from the "series resistance law" (4) and is independent of physical measurement. A possible explanation of this result is that as the rope stretches, the thick part is more resistant to stretching than the thin part, and so more stretching will take place in the thin part of the rope. Thus the fall factor has effectively been enlarged. Another model could be constructed by assuming that all the stretching must occur in the thinnest (or most easily stretched) part of the rope. This would give a fall factor of 8 in our example and would produce an even lower critical weight of about 20 kg.

Reference

[1] A Guide to Mountaineering Ropes, Edelrid Works, Isny (Allgäu), West Germany.

A Continued Fraction Algorithm for Approximating All Real Polynomial Roots

DAVID ROSEN
Swarthmore College
Swarthmore, PA 19081

JEFFREY SHALLIT, student
Princeton University
Princeton, NJ 08540

It is almost presumptuous to write another paper on approximating the real roots of a polynomial with real coefficients because so much is known, theoretically and practically, about the subject. However, the advent of computers makes it worthwhile to look again at some long forgotten theorems that can now be efficiently implemented on a computer to find all the real roots (and their multiplicities) of a real polynomial. Since a real number can be accommodated explicitly on the computer only by rational approximations, the generality of our method is restricted by the computer to polynomials with rational coefficients. Indeed, all the examples we tested had integral coefficients because a polynomial with rational coefficients is easily made equivalent to a polynomial with integral coefficients. For all practical purposes, therefore, our polynomials have integral coefficients.

The theorem that lies at the heart of our algorithm was proved in 1836 by Vincent ([**6**], [**7**]) and uses a continued fraction algorithm to separate the roots. Once a root is separated out, we continue with a continued fraction algorithm that approximates the root to any desired accuracy. An interesting feature of the program is that there is no problem with round-off error because not only are the coefficients integers, but integer arithmetic is used and all computations are exact. We programmed in APL because it handles polynomials so efficiently.

Our algorithm for finding all real roots first tests for rational roots, using the rational root theorem, and then factors out the linear factors, leaving a polynomial whose real roots are all irrational. The multiplicity of each rational root is also noted. The algorithm then tests the new polynomial for multiple irrational roots using an elegant procedure described in Uspensky [**6**, p. 65]. Let X_1 denote the product of all linear factors corresponding to simple roots; let X_2 denote the product of all linear factors corresponding to double roots, etc. If there are no roots of multiplicity k, we set $X_k = 1$. If, for example, $P(x) = (x-1)(x-2)(x-3)^2(x-4)^2(x-5)^3$, then $P(x)$ can be written in the form $P(x) = X_1X_2^2X_3^3$, where $X_1 = (x-1)(x-2)$, $X_2 = (x-3)(x-4)$, and $X_3 = (x-5)$. In general a polynomial $P(x)$ with r multiple roots can be written in the form $P(x) = a_0X_1X_2^2\cdots X_r^r$ where a_0 is a constant and $X_i = (x-b_1)\cdots(x-b_j)$ is a polynomial whose simple roots are all the roots of multiplicity i of P. The X_i are obtained as follows:

Let $D_1 = \gcd\{P, P'\} \equiv (P, P')$, where P' is the derivative of P. Generate a sequence of polynomials $D_2 = (D_1, D_1')$, $D_3 = (D_2, D_2')$, etc. An easy calculation shows that

$$D_1 = X_2X_3^2X_4^3\cdots X_m^{m-1}, \quad D_2 = X_3X_4^2\cdots X_m^{m-2}, \quad D_3 = X_4\cdots X_m^{m-3},$$

and so on. The sequence of common divisors $D_1, D_2, D_3, \ldots,$ ends with a constant term D_{m-1} which says that there are no roots of multiplicity $> m$. The polynomials

$$P_1 = \frac{P}{D_1} = X_1X_2\cdots X_m, \quad P_2 = \frac{D_1}{D_2} = X_2\cdots X_m, \quad P_m = \frac{D_{m-1}}{D_m} = X_m,$$

are now used to find an explicit polynomial for each X_i. Indeed $X_1 = P_1/P_2$, $X_2 = P_2/P_3, \ldots,$ and $X_m = P_m$. Observe that as the roots of each X_i are found, we know that the multiplicity of each root is

i. Our APL program determines the X_i. (To do polynomial algebra algorithmically, the reader may find Knuth [**3**] and Uspensky [**6**] useful and helpful.)

We have now reduced our problem to finding the roots of a polynomial having only simple roots. The second portion of our algorithm separates out each real root. This too is described completely by Uspensky [**6**, p. 127], using a separation theorem published by Vincent in the first volume of Liouville's Journal [**7**]. Uspensky's version and proof of this remarkable theorem is given in his appendix [**6**, p. 298].

Before continuing with Vincent's theorem, we need to discuss a few elementary notions about continued fractions. (References [**1**], [**2**] (pp. 129–151) and [**5**] are excellent sources, as are many books on number theory which treat continued frations.) A (simple regular) continued fraction is the expression

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}},$$

where the a_i are positive integers and a_0 may be any integer. As a space saver we write the continued fraction as

$$a_0 + \frac{1}{a_1+}\,\frac{1}{a_2+}\cdots\frac{1}{a_n+}\cdots;$$

the a_i are called partial quotients. The finite section

$$R_n = a_0 + \frac{1}{a_1+}\,\frac{1}{a_2+}\cdots\frac{1}{a_n}$$

is called the $(n+1)$th convergent. If we define two sequences of numbers by the recursion formulae

$$P_{-2} = 0, \quad P_{-1} = 1, \quad P_m = a_m P_{m-1} + P_{m-2},$$
$$Q_{-2} = 1, \quad Q_{-1} = 0, \quad Q_m = a_m Q_{m-1} + Q_{m-2},$$

then it can be shown inductively that $R_m = P_m/Q_m$, for $m = 0, 1, \ldots$.

Although most theorems that can be proved using continued fractions can be derived otherwise, there is in the theory of continued fractions a neat elegance that can be useful in many ways. The part of the theory that makes continued fraction representation of polynomial roots worthwhile lies in the rather precise estimation of an irrational number by an infinite set of certain rationals. Every rational number has a finite continued fraction representation. (For example, $\frac{23}{16} = 1 + \frac{1}{2+}\frac{1}{3+}\frac{1}{2}$.) Conversely every continued fraction of finite length denotes a rational number, and the representation is unique if we require that the last term is greater than 1. (Observe that $n = n - 1 + \frac{1}{1}$; hence $\frac{23}{16} = 1 + \frac{1}{2+}\frac{1}{3+}\frac{1}{1+}\frac{1}{1}$ as well.) More important, however, is the theorem that every infinite continued fraction represents a real number in the sense that its convergents $R_n = P_n/Q_n$ form a sequence of rational numbers that converge to some real number. Conversely every real number has a unique fraction representation which may be obtained by using a greatest integer-type algorithm.

The convergence of the sequence of convergents of a real number α guarantees that for every $\varepsilon > 0$, $|\alpha - (P_n/Q_n)| < \varepsilon$, for n sufficiently large. But the theory yields much more, namely, that

$$\left|\alpha - \frac{P_n}{Q_n}\right| < \frac{1}{Q_n Q_{n+1}} < \frac{1}{Q_n^2} \text{ for all } n. \tag{1}$$

So to find an adequate rational approximation to α, any algorithm that produces the continued fraction of the real number can be terminated at that n for which $1/Q_n^2$ is less than the desired accuracy. It is this inequality that we exploit in determining how far our real root algorithm should be carried out.

We are now ready to return to Vincent's theorem which we state informally, as does Uspensky. *Let $a, b, c, \ldots,$ be an arbitrary sequence of positive integers and use them to transform a polynomial equation without multiple roots by a series of successive substitutions*

$$x = a + \frac{1}{y}, \quad y = b + \frac{1}{z}, \quad z = c + \frac{1}{w}, \ldots, .$$

Then regardless of the choice of the integers $a, b, c, \ldots,$ we will come eventually to a transformed equation with not more than 1 *variation of sign.* The reference in the theorem to variations in sign refers to the Descartes rule of sign, ([**6**], p. 121) which states that the number of positive real roots of an equation with real coefficients is never greater than the number of variations of sign in the sequence of its coefficients, and if less, then always by an even number. The rule is effective when the number of variations is either 0 (the polynomial has no positive roots) or 1 (the polynomial has exactly one positive root).

Vincent's theorem is implemented by substituting $x = 1 + y$, $y > 0$ to capture those positive roots greater than 1 or $x = 1/(1 + y)$ to get those less than 1. (The negative real roots are obtained, as usual, by using the transformation $x' = -x$ in the original polynomial and then going through the whole procedure for positive real roots.) If the transformed equations have no variations or just one, we are done; i.e., if after substituting $x = 1 + y$ the equation has no variation of sign, then the original equation has no roots greater than 1. If it has one variation, then there is exactly one root greater than 1. A similar explanation accompanies the substitution $x = 1/(1 + y)$. If one or both of the transformed equations have more than one variation, then we transform them by substituting $y = 1 + z$ and $y = 1/(1 + z)$ and continue. Observe that a sequence of transformations $x = 1 + y$, $y = 1 + z, \ldots,$ followed by one of the type $u = 1/(1 + v)$ is equivalent to two transformations, one of the type $x = a + 1/y$ followed by $y = 1 + z$. Hence implementation of this process results in a Vincent sequence of transformations of the form

$$x = a_1 + \frac{1}{y}, \quad y = a_2 + \frac{1}{z}, \quad \ldots, \quad u = a_j + \frac{1}{v}, \quad v = \frac{1}{w},$$

where the a_i are all positive. (The substitution $v = 1/w$ does not change the number of variations in sign, as one can easily see by actually substituting $x = 1/w$ in a polynomial.)

The interval in which the root lies can be determined by combining all the transformations into a continued fraction and expressing it in terms of its convergents. The theory of continued fractions provides the useful formula

$$a_0 + \frac{1}{a_1 +} \cdots \frac{1}{+ a_m + 1/\xi} = \frac{(P_m \xi + P_{m-1})}{(Q_m \xi + Q_{m-1})}, \tag{2}$$

where P_{m-1}/Q_{m-1} and P_m/Q_m are consecutive convergents of the continued fraction

$$a_0 + \frac{1}{a_1 +} \cdots \frac{1}{+ a_m}.$$

Since $0 < \xi < \infty$, the interval in which the root lies has as boundaries P_{m-1}/Q_{m-1} and P_m/Q_m obtained by taking $\xi = 0$, and $\xi = \infty$ in (2).

Now that the polynomial X_i has been transformed so that each positive real root is the root of a real polynomial with only one variation in sign, we proceed to the third part of the algorithm — finding the continued fraction expansion for the one positive root of this polynomial. Suppose $F(x) = a_0 + a_1 x + \cdots + a_r x^r$ has only one variation of sign. We begin by locating the root between consecutive integers. The smaller of the two integers where the change of sign occurs is the greatest integer of the root which we denote by r_0 ($r_0 \geqq 0$). The substitution $X = r_0 + (1/T_1)$ gives a transformed polynomial in T_1, where we notice that T_1 is the tail of the continued fraction of the root.

The resulting equation in T_1, which we denote by $F_1(T_1)$, cannot have two positive roots. For suppose $F_1(t)=0$ and $F_1(t')=0$; then $F(r_0+(1/t))=0=F(r_0+(1/t'))$, which contradicts the assertion that $F(x)$ has only one positive root. We next test $F_1(T_1)=0$ for a change of sign between consecutive integers, the smaller being the greatest integer in T_1. Making the substitution $T_1=r_1+(1/T_2)$ yields an equation $F_2(T_2)=0$ which, similarly, has only one positive root, say t_2. From (2) we deduce that the exact root X of F is given by

$$r_0+\frac{1}{r_1+1/t_2}=(P_1t_2+P_0)/(Q_1t_2+Q_0);$$

therefore X lies between P_1/Q_1 and P_0/Q_0. Continuing in this way we obtain the mth convergent of the continued fraction expansion of the positive root. The exact value of the root lies between P_{m-1}/Q_{m-1} and P_m/Q_m. The value of m is set using (1): Decide on the desired accuracy and stop the calculations when $1/Q_m^2$ is less than it.

Our algorithm carries out the procedure by using Newton's method at each step to find a rough approximation to the greatest integer of the root. Newton's method with an initial approximation of 1 quickly determines the root roughly to $\pm.5$. This approximation is an integer, say, t. We then test $t-1, t, t+1$ to see which is really the greatest integer of the root. In this way round-off error in Newton's method is avoided. The reason for using Newton is that we quickly get three integers to test for a change of sign in the values of the polynomial. Without Newton the algorithm would have to test all integers from 1 to r_m+1, which could be expensive if r_m is very large.

The transformed polynomial is obtained using Lang and Trotter's technique [**4**, p. 117] which is computationally more efficient than the brute force method described above. Briefly, given a polynomial $P_n(x)$ of degree d with positive leading coefficient and a unique simple irrational root $y_n>1$ the new polynomial P_{n+1} is constructed by finding $r_n=[y_n]$, the largest integer such that $P_n(r_n)<0$. Define $B_n(x)=P_n(x+r_n)$ and $P_{n+1}(x)=-x^dB_n(x^{-1})=-x^dP_n(r_n+x^{-1})$ which has only positive powers of x. Having started with a polynomial with exactly one change of sign, and positive leading coefficient, the sign of the constant term is minus. The substitution $x+r_n$ translates the root to be between 0 and 1 so the number of variation of signs is unchanged and the constant term again is minus since the leading coefficient remains positive. Multiplying $B_n(x^{-1})$ by $-x^d$ forces the leading coefficient to be positive.

As an example, consider the polynomial $P_1(x)=3x^3+2x^2+x-7$. The polynomial changes sign between 1 and 2 so $r_1=1$. Thus

$$\begin{aligned}B_1(x)=P_1(x+r_1)&=3(x+1)^3+2(x+1)^2+(x+1)-7\\&=3(x^3+3x^2+3x+1)+2x^2+4x+2+x+1-7\\&=3x^3+11x^2+14x-5.\end{aligned}$$

Substituting $x=y^{-1}$, we get $B_1(y^{-1})=3y^{-3}+11y^{-2}+14y^{-1}-5$. Hence

$$-y^3B_1(y^{-1})=-3-11y^2-14y+3y^3=P_2(y).$$

Because of the special infinite precision integer arithmetic subroutines developed by the second author, our algorithm gives the continued fraction of each root to any desired length. For convenience, we summarize the algorithm here in its entirety.

1. Test for rational roots and their multiplicity using the rational root theorem; factor them out, leaving a polynomial whose real roots are irrational.
2. Test for multiple roots using Uspensky's algorithm. That is, find polynomials whose roots are all simple but have the same multiplicity in the original polynomial.
3. Use Vincent's theorem to separate the roots; for each root find a polynomial having that root as the only positive root.
4. Find the continued fraction approximation of the root using Newton's method to get a first approximation for each partial quotient.

5. Stop the algorithm at P_n/Q_n when $1/Q_n^2$ is less than the desired accuracy.

6. Find negative roots by replacing x by $-x$ in the original equation and start at the beginning.

We have not tried to compare the speed of our algorithm with other algorithms, but we do assert that ours is quick and accurate, provided multiple precision routines are developed. Greater accuracy of the approximation does slow up the computations. To test our algorithm we tried some of the polynomials suggested by Vincent [7]. The equation $x^6-12x^4-2x^3+37x^2+10x-10=0$ has roots $-1\pm\sqrt{2}$, $1\pm\sqrt{3}$, $\pm\sqrt{5}$; two roots in the interval $(2,3)$, two roots in the interval $(-3,-2)$, one in $(0,1)$ and one in $(-1,0)$. Our algorithm neatly separated out the roots and gave the continued fraction representation for each. In his paper, Vincent shows that there is exactly one root in $(0,1)$ and one in $(-1,0)$ for $x^6-6x^5+40x^3+60x^2-x-1=0$. Our algorithm gives the continued fractions as

$$\frac{1}{7+}\frac{1}{1+}\frac{1}{1+}\frac{1}{2+}\frac{1}{5+}\cdots \quad\text{and}\quad -1+\frac{1}{1+}\frac{1}{6+}\frac{1}{1+}\frac{1}{13+}\frac{1}{2+}\cdots.$$

One might wonder why Vincent and Uspensky didn't continue the separation algorithm to produce the continued fraction of the root. Undoubtedly, they were worn out from the tedious hand computation required to separate the roots.

This paper was presented at the Conference on Computations in Algebra and Number Theory, University of New Brunswick, Canada, August, 1975. Some of this work was done while the first author was a Fulbright-Hays Professor at University College Cork, Ireland, and he acknowledges with thanks the help of Mr. R. Studdert at UCC who programmed the continued fraction algorithm for cubic irrationals on the IBM 1130 using commercial subroutines for handling huge integers (100 digits). The second author did all the programming in APL and acknowledges with thanks permission of Dr. K. Iverson and Mr. A. Falkoff to use the facilities at IBM's Philadelphia System Development Division. We wish to thank Professor H.W. Brinkmann for remembering Vincent's theorem and steering us to Uspensky, the only place where it seems to be discussed outside of the original paper.

References

[1] G. Chrystal, Algebra, A. & C. Black, London, 1926.
[2] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Clarendon Press, Oxford, 1938.
[3] D. Knuth, The Art of Computer Programming, vol 2 (Seminumerical Algorithms), Addison-Wesley, 1969.
[4] S. Lang and H. Trotter, Continued fractions for some algebraic numbers, J. Reine Angew. Math. (1972) 112–134.
[5] C. D. Olds, Continued Fractions, Random House, New York, 1963.
[6] J. V. Uspensky, Theory of Equations, McGraw-Hill, New York, 1948.
[7] A. J. H. Vincent, Note Sur la Résolution des Équations Numériques, J. Math. Pures Appl., 1 (1836) 341–372.

High Card Point Counts

JAMES G. WENDEL
University of Michigan
Ann Arbor, MI 48109

It is an old tradition in the teaching of elementary probability and statistics to illustrate various concepts through consideration of games of chance involving coins, dice, roulette wheels or playing cards; indeed, history tells us that the subject originated in such considerations. In this note we offer some examples based on the distribution of certain cards in a bridge deal, in which an entire 52 card deck is partitioned into four 13-card hands. The value of the face cards in such hands offers a rich source of examples of elementary combinatorics, the use of linearity and symmetry in the calculation of expectations, correlations and their relation to angle cosines, conditional probability, and generating functions. Background on these basic concepts may be found in such books as Dwass [**2**] and Feller [**3**]. Books on bridge probabilities and their application to over-the-table bidding and play include [**5**] and the classic [**1**].

Throughout this note we consider a standard deck of 52 cards, in which the value of each card is defined as 4, 3, 2, or 1, respectively, for an ace, king, queen, or jack; all other cards have value 0. For a bridge hand of 13 cards, the sum of the values of the cards in the hand is known as the high card point count of the hand. For simplicity, we will call this sum the **value** of the hand. The total value of the deck is, of course, $4(4+3+2+1)=40$. We regard all deals as equally likely, each with probability $(13!)^4(52!)^{-1}=1.864\times10^{-29}$. Then a given hand of 13 cards has probability

$$h=\frac{39!/13!^3}{52!/13!^4}=\binom{52}{13}^{-1}=1.575\times10^{-12},$$

so that all hands are equally likely, each with probability h. We let the random variable X represent the value of a player's hand, Y that of his partner, and U, V those of their opponents.

It is clear that the possible values of these random variables are the integers $0,1,\ldots,37$; for example, $X=0$ means that the given player's hand contains no face cards, while $X=37$ represents the strongest possible hand (in our sense), having four each of aces, kings, and queens, and one jack. Next, since hands may be permuted without affecting probabilities, it is clear that the probability distributions of X, Y, U, V, are identical. More generally, the random variables X, Y, U, V are symmetrically dependent or **exchangeable** in the sense that $P=\{X=x,\ Y=y,\ U=u,\ V=v\}=P\{X=x',\ Y=y',\ U=u',\ V=v'\}$ for any permutation (x',y',u',v') of (x,y,u,v). But they are certainly not independent. For instance, if one player has a lot of high-card points then the chances are diminished for another player to have many. This lack of independence is, of course, illustrated further by the identity

$$X+Y+U+V=40, \tag{1}$$

the deck's total value.

Can we determine the probability distribution $f(x)=P\{X=x\}$? A few values are easily obtained, e.g., $f(0)=\binom{36}{13}h=3.639\times10^{-3}$, $f(1)=4\binom{36}{12}h=7.884\times10^{-3}$, $f(2)=[4\binom{36}{12}+\binom{4}{2}\binom{36}{11}]h=0.01356$ (because 2 points can come from one queen or two jacks) and $f(37)=4h=6.229\times10^{-12}$. But $f(10)$ seems almost hopeless, because there are so many combinations of face cards that add up to 10 points. Later we will find the generating function of X and use it to tabulate the probabilities $f(x)$. Meanwhile, we will investigate ways to find the expectation and variance of X without direct appeal to their definitions

$$\mu=E(X)=\sum_{x=0}^{37}xf(x)$$

$$\sigma^2=\mathrm{Var}(X)=E((X-\mu)^2)=\sum_{x=0}^{37}x^2f(x)-\mu^2.$$

Similarly, we will determine the covariance and correlation of the values of two hands without explicit reference to the joint distribution $P\{X=x,\ Y=y\}$. The linearity of expectation and the exchangeability of X, Y, U, V make this possible.

To begin we calculate $E(X)$, actually in two different ways. First, think of X as the sum of the values of the individual cards in the hand:

$$X=X_1+X_2+\cdots+X_{13}. \tag{2}$$

The random variables X_i are exchangeable; their common distribution is given by $c(x) = P\{X_i = x\}$, with values $c(4) = c(3) = c(2) = c(1) = 1/13$, $c(0) = 9/13$. This gives $E(X_i) = \Sigma_{x=0}^{4} x\, c(x) = 10/13$. Then from (2), $E(X) = 13(10/13) = 10$, one-fourth of the total value of the deck. This highly intuitive result can also be obtained by taking the expectation of both sides of (1). We obtain, by linearity, $E(X) + E(Y) + E(U) + E(V) = E(40) = 40$. By exchangeability, the four expectations on the left are identical; hence $E(X) = 10$.

Next we calculate the variance by means of the equation $\text{Var}(X) = E(X^2) - [E(X)]^2 = E(X^2) - 100$. From (2) we have

$$X^2 = \sum_{i=1}^{13} X_i^2 + \sum_{i \neq j} X_i X_j$$

and therefore by linearity and exchangeability $E(X^2) = 13E(X_1^2) + (13)(12)E(X_1X_2)$. Using the distribution $c(x)$ we can readily compute that $E(X_1^2) = 30/13$, but the term $E(X_1X_2)$ requires the use of the joint distribution of X_1 and X_2:

$X_2 \backslash X_1$	0	1	2	3	4
0	$315k$	$36k$	$36k$	$36k$	$36k$
1	$36k$	$3k$	$4k$	$4k$	$4k$
2	$36k$	$4k$	$3k$	$4k$	$4k$
3	$36k$	$4k$	$4k$	$3k$	$4k$
4	$36k$	$4k$	$4k$	$4k$	$3k$

where $k = [3 \cdot 13 \cdot 17]^{-1} = 1/663$. Then $E(X_1X_2) = 360/663$, $E(X^2) = 1990/17$, and $\text{Var}(X) = 290/17 = 17\frac{1}{17}$. We give a simpler derivation of the variance later.

Similar but more strenuous difficulties will attend the calculation of the covariance of the values of two hands; therefore, the determination of their corrrelation coefficient would seem to involve much messy arithmetic. But, as we shall now show, finding the correlation is almost trivial, and has some interesting byproducts. We call on (1) again; since the covariance of a random variable with a constant is zero, we have $0 = \text{Cov}(X, 40) = \text{Cov}(X, X + Y + U + V) = \text{Var}(X) + 3\,\text{Cov}(X, Y)$, where we have again used linearity and exchangeability. From the last equation we find immediately that the correlation coefficient $r_{X,Y}$ is $\text{Cov}(X, Y)/[\text{Var}(X)\text{Var}(Y)]^{1/2} = \text{Cov}(X, Y)/\text{Var}(X) = -1/3$.

The argument just used, and its result, provides a concrete illustration of correlation coefficients interpreted as cosines of angles. Consider a regular tetrahedron $ABCD$ with centroid at the origin O; the problem is to determine the central angle, $\theta = \angle AOB$. Think of A, B, C, D as unit vectors from the origin. Since the centroid of the configuration is at the origin we have $A + B + C + D = 0$. Taking the dot-product of both sides of this equation with the vector A we get $1 + 3\cos\theta = 0$, hence $\cos\theta = -1/3$. (Readers familiar with organic chemistry can think of θ as the angle between $C-H$ bonds in the methane molecule CH_4.)

Returning to the probabilistic setting we note several generalizations of the correlation computation. First, it is clear that the correlation $-1/3$ depends neither on the size of the deck nor on the actual point-values assigned to cards (provided they're not all the same); from the method of calculation one sees that the same result would be obtained for a deck of size $4k$ divided into 4 hands each of size k, with any non-constant prior assignment of points to cards. A further generalization replaces 4 by any positive integer m: for a deck of size mk divided into m hands of k cards each, the correlation between point-totals for hands equals $-1/(m-1)$. As a particular case take $m = 52$ and $k = 1$. We find that the correlation between the high card points of any two distinct cards is $r_{X_1,X_2} = -1/51$, which of course can also be obtained from the joint probability distribution of X_1 and X_2.

With the last result we are in position to go back to the variance calculation. Recalling (2) we have

$$(3)\ \mathrm{Var}(X)=\sum_{i=1}^{13}\mathrm{Var}(X_i)+\sum_{i\neq j}\sum\mathrm{Cov}(X_i,X_j)=13\,\mathrm{Var}(X_1)+156\,\mathrm{Cov}(X_1,X_2)=(13-156/51)\mathrm{Var}(X_1)$$

since $\mathrm{Cov}(X_1,X_2)=r_{X_1,X_2}\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}=r_{X_1,X_2}\mathrm{Var}(X_1)$. Since $E(X_1^2)=30/13$ and $E(X_1)^2=(10/13)^2$ we get $\mathrm{Var}(X_1)=290/169$, which on substitution in (3) yields again $\mathrm{Var}(X)=290/17$.

Let us conclude this exploration by considering the following ridiculous method for dealing a bridge hand. We take cards from the deck one at a time; with each draw we toss a fair coin, and if it comes up heads we take the card into the hand; otherwise we discard it. Say that the size of the hand is the random variable N. Then N has a binomial distribution, and the probability that $N=13$ is pretty small, namely $p=2^{-52}\binom{52}{13}=1.410\times10^{-4}$. But *conditional on the event* $N=13$, the distribution of points in the hand will be the same as that of X when the hand is dealt in the customary way.

Let Z be the total value obtained by the coin-tossing method. We can write Z in the form $Z=Z_1+Z_2+\cdots+Z_{52}$ where Z_i is 0 if card i is not taken into the hand, and is the value of the card if it is used. At the same time we write N as $N=I_1+I_2+\cdots+I_{52}$ where I_j is 1 or 0 according as card j is or is not taken; in other words, I_j is the indicator of the event that card j goes into the hand. The successive pairs $(I_1,Z_1),(I_2,Z_2),\ldots,(I_{52},Z_{52})$ are independent. Therefore, for the bivariate generating function of N and Z (cf. [**3**, p. 279]) we have

$$G(s,t)=E(s^Nt^Z)=\prod_{i=1}^{52}E(s^{I_i}t^{Z_i})=2^{-52}(1+st^4)^4(1+st^3)^4(1+st^2)^4(1+st)^4(1+s)^{36},$$

because a card with value k contributes a factor $(1+st^k)/2$, and there are four cases each with $k=4,3,2,1$, and thirty-six with $k=0$.

The joint generating function can also be written in the form

$$G(s,t)=\sum_{n=0}^{52}s^nE(t^Z\,|\,N=n)P\{N=n\}.$$

It follows that the generating function for X, namely $E(t^X)=E(t^Z\,|\,N=13)$, is p^{-1} times the coefficient of s^{13} in the polynomial $G(s,t)$. This is simply h times the coefficient of s^{13} in the polynomial $H(s,t)=(1+st^4)^4(1+st^3)^4(1+st^2)^4(1+st)^4(1+s)^{36}$. There is no neat way to write the final result, but since $E(t^X)=\sum_{x=0}^{37}f(x)t^x$, a bit of clerical labor yields the values $f(x)$; they are the coefficients of $s^{13}t^x$ in $hH(s,t)$, and are tabulated in TABLE 1.

I am obliged to the referee for directing me to [**4**], which contains extensive tabulations of high card point counts and other bridge probabilities. In particular, TABLE XX of [**4**] gives values for $f(x)$,

x	$f(x)$	x	$f(x)$	x	$f(x)$
0	3.639×10^{-3}	13	6.914×10^{-2}	26	1.167×10^{-4}
1	7.884×10^{-3}	14	5.693×10^{-2}	27	4.907×10^{-5}
2	1.356×10^{-2}	15	4.424×10^{-2}	28	1.857×10^{-5}
3	2.462×10^{-2}	16	3.311×10^{-2}	29	6.672×10^{-6}
4	3.845×10^{-2}	17	2.362×10^{-2}	30	2.198×10^{-6}
5	5.186×10^{-2}	18	1.605×10^{-2}	31	6.113×10^{-7}
6	6.554×10^{-2}	19	1.036×10^{-2}	32	1.719×10^{-7}
7	8.028×10^{-2}	20	6.435×10^{-3}	33	3.521×10^{-8}
8	8.892×10^{-2}	21	3.779×10^{-3}	34	7.061×10^{-9}
9	9.356×10^{-2}	22	2.100×10^{-3}	35	9.827×10^{-10}
10	9.405×10^{-2}	23	1.119×10^{-3}	36	9.449×10^{-11}
11	8.945×10^{-2}	24	5.590×10^{-4}	37	6.299×10^{-12}
12	8.027×10^{-2}	25	2.643×10^{-4}		

The probability distribution of high card point count in a bridge hand.

TABLE 1

$0 \leqq x \leqq 30$. I also want to thank Jeff Rubens of Pace University, co-editor of the *Bridge World* magazine, for a number of helpful comments. The table of high card point count probabilities in [**5**, p. 289] is due to him.

References

[1] Émile Borel and André Chéron, Théorie Mathématique du Bridge, Gauthier-Villars, Paris, 1955.
[2] Meyer Dwass, Probability, Theory and Applications, W. A. Benjamin, New York, 1970.
[3] William Feller, An Introduction to Probability Theory and its Applications, 3rd ed., vol. I, Wiley, New York, 1968.
[4] F. H. Frost, Bridge Odds Complete, George Coffin, Waltham, Massachusetts, 1971.
[5] C. H. Goren, Go with the Odds, Macmillan, New York, 1969.

Mean and Variance for Covering Sets of Congruences

JOSEPH H. SILVERMAN, student
Brown University
Providence, RI 02912

A covering set of congruences, as introduced by Paul Erdös in a paper in 1952 [**1**], is a system of congruences of the form $x \equiv b_i\,(n_i)$, $1 < n_1 < n_2 < \cdots < n_k$, such that every integer satisfies at least one of the congruences. For example, it is easy to see that $x \equiv 0(2)$, $x \equiv 0(3)$, $x \equiv 1(4)$, $x \equiv 1(6)$, $x \equiv 11(12)$, is a covering set of congruences. Erdös made two conjectures in his paper, both of them still unsolved. The first is that for every integer n there exists a covering whose moduli $n_1, n_2, \ldots, n_k$ are all greater than n. The second is that there does not exist a covering all of whose moduli are odd. In view of these conjectures, it is natural that previous work has focused primarily on the set of moduli, either specifying necessary conditions on this set (e.g., [**2**], [**3**]) or producing specific examples of coverings whose moduli satisfy certain interesting properties (e.g., [**4**], [**5**], [**6**]). The major purpose of this note is to present some elementary results dealing with such systems of congruences for a *fixed* set of moduli $n_1, n_2, \ldots, n_k$, as $b_1, b_2, \ldots, b_k$ are allowed to vary over all possible congruence classes.

With this in mind, it is convenient to define a covering in a slightly different, but fully equivalent way. First we set some notation. Fix a set $N = \{n_1, n_2, \ldots, n_k\}$ of distinct integers greater than 1. Let $\beta = Z/n_1Z \times Z/n_2Z \times \cdots \times Z/n_kZ$. An element $B \in \beta$ will often be written as $(b_1, b_2, \ldots, b_k)$, where each b_i is a representative integer of an equivalence class in Z/n_iZ. Taken together with N, each $B = (b_1, b_2, \ldots, b_k) \in \beta$ determines a subset of Z given by: $\langle N, B\rangle = \{x \in Z\colon x \equiv b_i\,(n_i) \text{ for some } i\}$. Using this notation we may restate Erdös' definition: A pair (N, B) as above is a **covering of Z** if and only if $\langle N, B\rangle = Z$. N is called the set of moduli of the covering, and B is called the k-tuple of congruence classes.

Even if (N, B) is not a covering, it is still of interest to have some way of measuring "how large" $\langle N, B\rangle$ is in relation to all of Z. This prompts the following definition: Let $l = \mathrm{LCM}(N)$, the least common multiple of the set of moduli. The **density** of a pair (N, B) is

$$\mu(N, B) = \frac{1}{l}\,|\langle N, B\rangle \cap \{1, 2, \ldots, l\}|.$$

From the definition, $\mu(N, B)$ appears only to measure what fraction of $\{1, 2, \ldots, l\}$ is in $\langle N, B\rangle$. However, it actually corresponds quite well to the intuitive idea of (size of $\langle N, B\rangle$)/(size of Z), because

$x \in \langle N, B\rangle$ implies $x \pm l \in \langle N, B\rangle$. Thus all blocks of l consecutive integers intersect $\langle N, B\rangle$ in essentially the same way. It is easy to see that a pair (N, B) is a covering of Z if and only if $\mu(N, B) = 1$.

A first step in the analysis of μ is to show that coverings in $\{(N, B)\colon B \in \beta\}$ come in groups of l. If $B = (b_1, b_2, \ldots, b_k) \in \beta$ and $z \in Z$, we let $B + z$ be an abbreviation for $(b_1 + z, b_2 + z, \ldots, b_k + z) \in \beta$. Then since $x \in \langle N, B\rangle$ if and only if $x + z \in \langle N, B + z\rangle$, $\mu(N, B) = \mu(N, B + z)$. Call $B_1, B_2 \in \beta$ related if $B_1 = B_2 + z$ for some integer z. This is an equivalence relation, so it divides β into disjoint equivalence classes. Each class has exactly l elements, because for all $B \in \beta$, $B = B + z$ if and only if $l \mid z$. Since $\mu(N, B) = \mu(N, B + z)$, the number of coverings in $\{(N, B)\colon B \in \beta\}$ is divisible by $l = LCM(N)$.

It would be ideal, for a given N, to have some simple method of determining which B in β maximizes $\mu(N, B)$. For example, if $n_1, n_2, \ldots, n_k$ are pairwise relatively prime, then it follows from the Chinese remainder theorem that $\mu(N, B)$ is constant as a function of B. (This will also follow from Proposition 2 below.) As another example, let $N = \{n, n^2, \ldots, n^k\}$ for some $n > 1$. Then an optimum B is $(1, n, n^2, \ldots, n^{k-1})$, for which $\mu(N, B) = (1 - n^{-k})/(n - 1)$.

A lower bound on this maximum may be found by computing $\bar{\mu}(N)$, the arithemetic mean of the set $\{\mu(N, B)\colon B \in \beta\}$. The following proposition gives a simple formula for $\bar{\mu}(N)$.

PROPOSITION 1. $\bar{\mu}(N) = 1 - \Pi_{n \in N}(1 - 1/n)$.

Proof. For each $B = (b_1, b_2, \ldots, b_k) \in \beta$ and each integer x, define a covering function $s_B(x)$ to be 0 if $x \in \langle N, B\rangle$ and 1 if $x \notin \langle N, B\rangle$. For a fixed integer x, we compute for future reference

$$\begin{aligned}\sum_{B \in \beta} s_B(x) &= |\{B \in \beta\colon s_B(x) = 1\}| \\ &= |\{(b_1, b_2, \ldots, b_k)\colon 0 \leqq b_i < n_i \text{ and } b_i \not\equiv x(n_i)\}| \\ &= (n_1 - 1)(n_2 - 1)\cdots(n_k - 1) \\ &= \prod_{n \in N}(n - 1).\end{aligned}$$

Use of the covering function gives a simple expression for the density: $\mu(N, B) = (1/l)\Sigma_{x=1}^{l}(1 - s_B(x)) = 1 - (1/l)\Sigma_{x=1}^{l} s_B(x)$. For convenience, let $m = |\beta| = n_1 n_2 \cdots n_k$. We are now ready to compute $\bar{\mu}(N)$:

$$\begin{aligned}\bar{\mu}(N) &= \frac{1}{|\beta|}\sum_{B \in \beta}\mu(N, B) \\ &= 1 - \frac{1}{m}\sum_{B \in \beta}\frac{1}{l}\sum_{x=1}^{l} s_B(x) \\ &= 1 - \frac{1}{l}\sum_{x=1}^{l}\frac{1}{m}\sum_{B \in \beta} s_B(x) \\ &= 1 - \frac{1}{l}\sum_{x=1}^{l}\frac{1}{m}\prod_{n \in N}(n - 1) \\ &= 1 - \prod_{n \in N}\left(1 - \frac{1}{n}\right).\end{aligned}$$

One example of interest is given by the modulus set $N_k = \{3, 5, 7, 9, \ldots, 2k + 1\}$, for which $\bar{\mu}(N_k) = 1 - (2^k \cdot k!)^2/(2k + 1)!$. If k is sufficiently large, one can use Stirling's formula to obtain the approximation $\bar{\mu}(N_k) \approx 1 - \sqrt{\pi k}/(2k + 1)$. Since Erdös' conjecture is that there does not exist a covering using entirely odd moduli, the fact that the average density using N_k approaches 1 as $k \to \infty$

may be significant. This fact (i.e., $\text{Lim}_{k\to\infty}\bar{\mu}(N_k)=1$) is stronger than $\text{Lim}_{k\to\infty}\max_{B\in\beta}\mu(N_k,B)=1$; the latter limit follows directly by letting P_k denote the set of primes in N_k and noting that

$$\max_{B\in\beta}\mu(N_k,B)\geqq\max_{B\in\beta}\mu(P_k,B)=1-\prod_{p\in P_k}\left(1-\frac{1}{p}\right).$$

That this last expression approaches 1 as $k\to\infty$ is a theorem of Gauss.

Proceeding in a statistical vein, one might be tempted to compute the standard deviation σ of $\{\mu(N,B)\colon B\in\beta\}$. Proposition 2 gives a formula for σ^2. This formula can be derived in a manner similar to that used in Proposition 1. However, since the ensuing algebra is rather lengthy, only the result will be presented and the proof will be left for the interested reader.

The following conventions are used in the statement of Proposition 2. As before, let $m=n_1n_2\cdots n_k=|\beta|$. The sum over $\theta\subset N$ is a sum over all subsets θ of N. $N\backslash\theta$ is the set difference, everything in N that is not also in θ. If $\theta=\{x,y,\ldots,z\}\subset N$, then $r_\theta=(xy\cdots z)/\text{LCM}(x,y,\ldots,z)$. It might be noted that r_θ measures the pairwise relative primality of the elements of θ. Thus $r_\theta=1$ if and only if the elements of θ are pairwise relatively prime. Finally, the empty product and r_Φ are both assigned the value 1.

PROPOSITION 2. *The variance σ^2 of $\{\mu(N,B)\colon B\in\beta\}$ is given by*

$$\sigma^2=\frac{1}{m^2}\sum_{\theta\subset N}\left[\prod_{n\in N\backslash\theta}n(n-2)\right](r_\theta-1).$$

By combining all of the results of this note, we can obtain a very weak necessary condition on the moduli of a covering. If there is any covering using N as the modulus set, then, as we observed above, there are l of them; call them $B_1,B_2,\ldots,B_l\in\beta$ with $\mu(N,B_i)=1$. Combined with Proposition 1, this serves to give a lower bound of

$$\frac{1}{m}\sum_{i=1}^{l}(\mu(N,B_i)-\bar{\mu}(N))^2=\frac{l}{m}\prod_{n\in N}\left(1-\frac{1}{n}\right)^2$$

on σ^2. This bound can be compared with the expression in Proposition 2 to prove that certain modulus sets do not admit coverings. Sharper and more useful results should be possible, provided two things can be done. First we must find more efficient ways of distinguishing between those distributions $\{\mu(N,B)\colon B\in\beta\}$ which do contain a 1 and those which do not. Second we need a method to estimate upper and lower bounds for σ^2 that is both fairly accurate and mathematically tractable. For the expression for σ^2 in Proposition 2, which involves a sum over the power set of N, is both unwieldy to manipulate and beyond the range of modern computers to calculate even for N of moderate size. (For example, see **[5]** for a covering using 124 moduli.)

In particular, if $N_k=\{3,5,7,9,\ldots,2k+1\}$ as before, we can use the estimate $\bar{\mu}(P_k)=\bar{\mu}(\{\text{primes in } N_k\})\leqq\mu(N_k,B)$ for all $B\in\beta$ to find the upper bound $\sigma(N_k)\leqq\Pi_{p\in P_k}(1-1/p)$ for all k. This allows us to conclude that $\text{Lim}_{k\to\infty}\sigma^2(N_k)=0$. However, more careful analysis of the rate at which $\sigma^2(N_k)\to 0$ might help shed further light on Erdös' second conjecture, a task which I will leave as a challenge for the interested reader.

References

[1] P. Erdös, On a problem concerning congruence systems, Math. Lapok, 3 (1952) 122–128.
[2] M. Billik and H. M. Edgar, Covering sets of congruences, this MAGAZINE, 46 (1973) 265–270.
[3] S. K. Stein, Unions of arithmetic sequences, Math. Ann., 134 (1958) 289–294.
[4] J. Jordan, Covering classes of residues, Canad. J. Math., 19 (1967) 514–519.
[5] R. F. Churchhouse, Covering sets and systems of conguences, Computers in Mathematical Research, 1968, pp. 20–36.
[6] J. D. Swift, Sets of covering congruences, Bull. Amer. Math. Soc., 60 (1954) 390.

Joint vs. Individual Normality

RICHARD A. VITALE
Claremont Graduate School
Claremont, CA 91711

Saying that random variables X and Y are *jointly* normal is, of course, a much stronger statement than claiming X and Y are *individually* normal. The gap between the two concepts is a crucial one but often puzzling to the student. By way of illustration, Feller [**1**, pp. 99–100], for instance, provides examples which depend on density function constructions. We give here an elementary example based on the probability integral transformation $U = F(X)$ where X is a random variable with continuous cumulative distribution function F.

We begin with a standard normal variable X with distribution function N, and define a transformation T (see FIGURE 1) by the formula:

$$T(x) = \begin{cases} x, & 0 \leqq x \leqq 1/2, \\ \frac{3}{2} - x, & 1/2 < x \leqq 1. \end{cases}$$

Then $U = N(X)$ is uniformly distributed on $[0, 1]$ (see, for instance, the recent note [**2**] by Schuster in this MAGAZINE), as is $V = T(U)$. Therefore $Y = N^{-1}(V) = N^{-1}(T(N(X))$ is also standard normal.

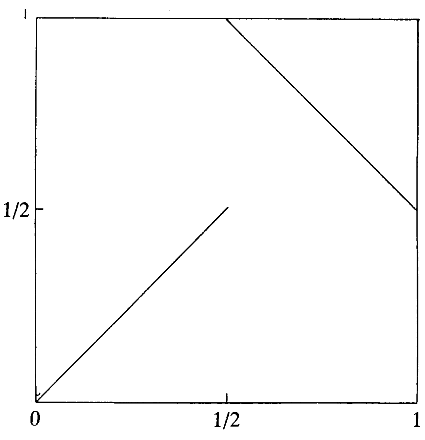


FIGURE 1

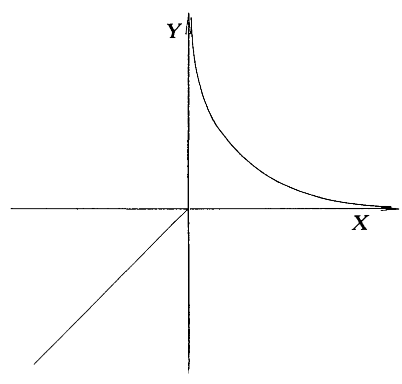


FIGURE 2

Now there are various ways to argue against the joint normality of X and Y. One approach is to plot the possible values of (X, Y) (see FIGURE 2). It is evident (by rotating the picture through $-45°$) that $W = (X + Y)/\sqrt{2}$ cannot achieve values on some interval $[0, \alpha]$ yet is not concentrated at a single point. This implies that W is not normally distributed (not even degenerately so). Yet if X and Y were jointly normal, $X + Y$ and hence W would be normal.

References

[1] W. Feller, An Introduction to Probability Theory and Its Applications, vol. II, 2nd ed., Wiley, New York, 1971.

[2] E. F. Schuster, The probability integral transformation: a simple proof, this MAGAZINE, 49 (1976) 242.

Segment Trisection in Absolute Geometry

HUBERT J. LUDWIG
Ball State University
Muncie, IN 47306

Absolute geometry consists of the deductive consequences of a postulational basis for Euclidean geometry with the axiom which leads to uniqueness of parallelism deleted. It thus involves those results which are common to Euclidean and hyperbolic geometry. There are procedures and results which are often presented in the context of Euclidean geometry which do not depend upon the parallelism assumption and hence are actually results in absolute geometry. One of these is a common segment bisection procedure. In this note we shall consider an extension of this procedure which gives a method for segment trisection in Euclidean geometry but which fails to trisect segments in absolute geometry.

In absolute geometry a segment AB may be bisected by (1) setting up rays $\overrightarrow{AX}$ and $\overrightarrow{BY}$ on opposite sides of line AB such that $\angle BAX \cong \angle ABY$, and (2) locating points A' on $\overrightarrow{AX}$ and B' on $\overrightarrow{BY}$ such that $AA' \cong BB'$. Segment $A'B'$ will then cut AB at its midpoint [2, p. 134]. In Euclidean geometry a segment AB may be trisected by performing step (1) of the segment bisection procedure from absolute geometry and then locating points A' and A'' on $\overrightarrow{AX}$, B' and B'' on $\overrightarrow{BY}$ such that $AA' \cong A'A'' \cong BB' \cong B'B''$. Segments $A'B''$ and $A''B'$ will then cut AB at its trisection points [3, p. 233]. It is a straightforward task to show that this segment trisection process, if used in absolute geometry, will produce points M and N on segment AB such that $AM \cong NB$. However, this trisection procedure will not trisect the segment in absolute geometry.

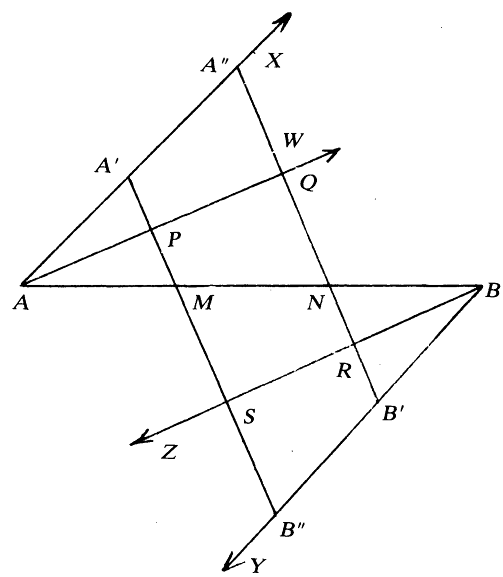


FIGURE 1

Let AB be a segment of length 3. Set up rays $\overrightarrow{AX}$ and $\overrightarrow{BY}$ on opposite sides of line AB such that $\angle BAX \cong \angle ABY$. Locate points A' and A'' on $\overrightarrow{AX}$, B' and B'' on $\overrightarrow{BY}$ such that segments $AA', A'A'', BB', B'B''$ all have unit length; see FIGURE 1. It can then be shown, in absolute geometry that $A'B''$ and $A''B'$ will cut segment AB at M and N respectively [2, p. 133], and that the order A, M, N, B holds.

We now assume that M and N are the trisection points of AB and hence that the segments $AM, MN, NB, AA', A'A'', BB'$ and $B'B''$ all have unit length. Let $\overrightarrow{AW}$ bisect $\angle BAX$ internally and let $\overrightarrow{BZ}$ bisect $\angle ABY$ internally. In absolute geometry, the internal angle bisector at the vertex of an isosceles triangle is the perpendicular bisector of the base. Thus, since triangles MAA' and NAA'' are isosceles, $\overrightarrow{AW}$ will be perpendicular to $A'M$ at P and to $A''N$ at Q, where $P \neq Q$. Similarly, $\overrightarrow{BZ}$ will be perpendicular to $B'N$ at R and to $B''M$ at S, where $R \neq S$. Neither of P, Q is R, S since $\acute{AW}$ and $\overrightarrow{BZ}$ are non-intersecting. Thus $PQRS$ is a quadrilateral having an angle sum of four right angles. Therefore the geometry is Euclidean, since an absolute geometry is Euclidean if and only if there exists a quadrilateral which has an angle sum of four right angles [**1**, p. 13].

References

[1] David Gans, An Introduction to Non-Euclidean Geometry, Academic Press, New York, 1973.

[2] Ellery B. Golos, Foundations of Euclidean and Non-Euclidean Geometry, Holt, Rinehart and Winston, New York, 1968.

[3] A. E. Hallerberg, The geometry of the fixed compass, The Mathematics Teacher, 52 (1959) 230–244.

Imitating the Euclidean Metric

IRA ROSENHOLTZ

University of Wyoming
Laramie, WY 82071

It is not hard to prove that Euclidean space has certain local topological properties, e.g., local connectedness and local compactness. What is required is to show the existence of a basis for the topology so that each basis element has the desired topological property (connectedness, compact closure, etc.). In fact, as everyone notices, the metric balls with respect to the usual Euclidean distance have these properties! This makes things so easy that one is tempted to feel that he is dealing with an over-simple special case. Some of this feeling may be relieved by the following very pretty but not particularly well-known theorem: *A metric space is both connected and locally connected if and only if it has a metric inducing the same topology such that each metric ball is connected* [**2**, pp. 95–96]. In this paper we prove an analogous result for locally compact metric spaces, and demonstrate further similarities between such spaces and Euclidean spaces.

We begin by noting that if a metric space X has a metric D so that the closure of each D-ball is compact, then, in addition to being locally compact, X must be σ-compact (that is, X must be the union of countably many compact sets). To see this observe that $X = \bigcup_{N=1}^{\infty} \mathrm{cl}_X(B_D(x, N))$, where $B_D(z, r)$ denotes the ball $\{p \mid D(p, z) < r\}$ and $\mathrm{cl}_X(A)$ denotes the closure of A in X. This proves the easy half of our main result:

THEOREM. *X is a locally compact and σ-compact metric space if and only if there is a metric D for X such that each D-ball has compact closure.*

Before proving the nontrivial half of this result, notice that the simple example $(0, 1]$ with the usual metric is a locally compact, σ-compact metric space, but not all balls have compact closure, e.g., $B(x, r)$ where $r \geqq x > 0$. The second half of our theorem claims that $(0, 1]$ can be remetrized so that each ball has compact closure. The metric $D(x, y) = |x - y| + |(1/x) - (1/y)|$ suffices, and it is instructive to follow the proof through with this example. Note that the one-point compactification of $(0, 1]$ is $[0, 1]$.

Proof. Assume that X is a locally compact, σ-compact metric space. If X is actually compact, then the result is trivial — any metric for X works — so we may assume that X is not compact. Let $Y = X \cup \{\infty\}$ be the one-point compactification of X. Since X is σ-compact, it follows directly that X is Lindelöf (every open cover has a countable subcover), and thus X is second countable (X has a countable basis). This makes Y in turn second countable, since a countable basis for X together with complements in Y of finite unions of compact neighborhoods in X forms a countable basis for Y. Also Y is Hausdorff since X is locally compact. So Y, being compact and Hausdorff, is regular. The Urysohn metrization theorem ([**3**], p. 75) shows that Y is a metric space.

Let d be a metric for Y (and hence, by restriction, for X). Now in the metric for X, not every d-ball has compact closure, primarily because ∞ will often be in the closure of these balls. We adjust this by defining D on X by:

$$D(x,y) = d(x,y) + \left| \frac{1}{d(x,\infty)} - \frac{1}{d(y,\infty)} \right| .$$

It is easy to see that D is a metric since the triangle inequality happens term-by-term. Also, D induces the same topology as d on X, because $D(x_n, x) \to 0$ if and only if $d(x_n, x) \to 0$.

Finally, we must show that the closure in X of each r-ball (in the D metric) is compact. So suppose $x \in X$ and $r > 0$. If $y \in B_D(x,r)$, then $D(x,y) < r$, so $|(1/d(x,\infty)) - (1/d(y,\infty))| < r$, and we get $d(y,\infty) > [r + (1/d(x,\infty))]^{-1}$. Thus, ∞ is neither a point nor a limit point of $B_D(x,r)$, and so $\mathrm{cl}_X(B_D(x,r)) = \mathrm{cl}_Y(B_D(x,r))$, which is compact. This completes the proof.

This metric D shares several nice properties with the usual Euclidean metric. In particular, a subset of X is compact if and only if it is closed and bounded with respect to D. This follows as in the Euclidean case from the fact that such a set is a closed subset of a ball which has compact closure. Also the metric D is a complete metric for X since any D-Cauchy sequence must, of course, be bounded, lie in some ball, and thus have a limit in the closure of that ball. (Actually, any locally compact metric space has a complete metric, even though it may not be σ-compact, but that's another story — see [**1**], p. 196 and p. 294.) Finally, many of the standard theorems about locally compact, σ-compact metric spaces — for example, that each such space is the union of a sequence $U_1, U_2, \ldots$ of open sets such that, for each n, U_n has compact closure and $\mathrm{cl}_X(U_n) \subseteq U_{n+1}$ — become "one-liners" using the metric D.

Since we now have two slightly different remetrization theorems, it is natural to inquire whether a metric space which satisfies the hypotheses of both theorems will have a metric which does everything. Our final result says "yes".

THEOREM. *If X is a connected locally connected, locally compact, σ-compact metric space, then there is a metric for X such that each metric ball is connected and has compact closure.*

Proof. First, let D be a metric for X such that the closure of each D-ball is compact, as above. Then let D^* be defined as follows: $D^*(x,y) = \inf\{r \mid x$ and y belong to a connected set whose D-diameter is $r\}$. Then D^* is a metric for X having the property that each ball is connected ([**2**], p. 96). Furthermore, since $B_{D^*}(x,r) \subseteq B_D(x,r)$, it follows that each ball has compact closure.

Though the hypothesis of σ-compactness in the previous theorem is necessary, it is superfluous. The reason for this is that a locally compact metric space is locally separable, and a connected, locally separable metric space is separable ([**4**], p. 75). But a separable locally compact metric space is σ-compact.

The author would like to thank the referee and editors for the exceptionally fine job they did on this little paper. Their devotion to improving the exposition was extraordinary, and their suggestions are warmly appreciated.

References

[1] James Dugundji, Topology, Allyn and Bacon, Boston, 1966.
[2] D. N. Hall and G. L. Spencer, Elementary Topology, Wiley, New York, 1967.
[3] J. G. Hocking and G. S. Young, Topology, Addison–Wesley, Reading, Mass., 1961.
[4] M. H. A. Newman, Topology of Plane Sets, Cambridge University Press, Cambridge, 1964.
[5] Stephen Willard, General Topology, Addison–Wesley, Reading, Mass., 1970.

PROBLEMS

Dan Eustice, Editor
Leroy Meyers, Associate Editor
The Ohio State University

Proposals

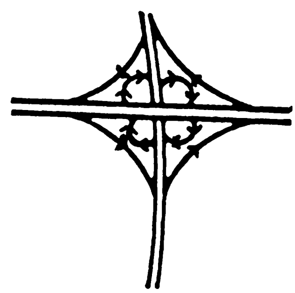
To be considered for publication, solutions should be mailed before November 1, 1978.

1033. For given positive integers $n_1, n_2, \ldots, n_k$, when is

$$\int_0^{2\pi} \cos n_1\theta \cos n_2\theta \cdots \cos n_k\theta \, d\theta$$

different from zero and what is its value? [*H. Kestelman, University College, London.*]

1034. We are familiar with the standard clover-leaf interchange [CLI] which has, inside the four ramps for making right-hand turns, the arrangement whereby left-hand turns are achieved by turning right into lanes which outline the four leaf clover. Your car approaches the CLI from the south. A mechanism has been installed so that at each point where there exists a choice of directions, the car turns to the right with fixed probability r.



a. If $r = 1/2$, what is your chance of emerging from the CLI going west?

b. Find the value of r which maximizes your chance of westward departure.

[*Marlow Sholander, Case Western Reserve University.*]

1035. A is a real $n \times n$ matrix. Do there exist orthogonal matrices B such that $A + B$ is real orthogonal? [*H. Kestelman, University College, London.*]

1036. If $a_1, a_2, \ldots, a_N$ are complex numbers such that $|a_N| > \sum_{k=0}^{N-1}|a_k|$, show that $\sum_{n=0}^{N} a_n \cos n\theta = 0$ has at least $2N$ solutions for $0 \leqslant \theta < 2\pi$. [*Joseph Silverman, Cambridge, Massachusetts.*]

1037. Let n be an integer greater than 2 and let $A_1, A_2, \ldots, A_n$ be non-empty sets of positive integers with the property that $a \in A_1$ and $b \in A_{i+1}$ implies $a+b \in A_{i+2}$, where we identify A_{n+1} as A_1 and A_{n+2} as A_2.

(a) If $1 \in A_1$ and $2 \in A_2$, find an integer that belongs to at least two of the sets.

(b)* Is it possible for $A_1, A_2, \ldots, A_n$ to be pairwise disjoint? [*James Propp, Great Neck, New York.*]

1038. Define the following sequence of square matrices:

$$M(1) = [1], M(2) = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}, M(3) = \begin{bmatrix} 6 & 7 & 8 \\ 9 & 10 & 11 \\ 12 & 13 & 14 \end{bmatrix}, \ldots.$$

Find the sum of the elements on the main diagonal of $M(n)$. [*Douglas Lewan, Antrim, New Hampshire.*]

Quickies

Solutions to Quickies appear at the conclusion of the Problems section.

Q651. Given any triangle ABC. Divide BC (respectively, AC, AB) into n equal segments by means of points A_i (respectively, B_i, C_i), $i = 1, 2, \ldots, n-1$. Prove that

$$\sum_{i=1}^{n-1} \left(|AA_i|^2 + |BB_i|^2 + |CC_i|^2\right) = \frac{(n-1)(5n-1)}{6n}(a^2 + b^2 + c^2).$$

[*Geoffrey Kandall, Hamden, Connecticut.*]

Q652. Show that $\Sigma_{i=1}^{n}(1 + \tan\alpha_i) \leqslant \sqrt{2}\, \Sigma_{i=1}^{n} \sec\alpha_i$ when $\sec\alpha_i > 0$. When does equality hold? (This is a generalization of Q472, March 1970.) [*Murray S. Klamkin, University of Alberta.*]

Solutions

No Polynomials **September 1976**

990. Prove that the identity $f(x+1)/g(x+1) - f(x)/g(x) = h(1/x)$ is not satisfied by any non-constant polynomials f, g, and h. [*Harry W. Hickey, Arlington, Virginia.*]

Solution: Note first that $f(x)/g(x) = k(x)$ cannot be a polynomial since $|k(x+1) - k(x)| \to \infty$ as $x \to 0$. Thus there is at least one complex number r such that $g(r) = 0$ but $f(r) \neq 0$. Let $r_1, \ldots, r_m$ be the complete set of such numbers, arranged in nondecreasing order of their real parts. Then $f(x+1)/g(x+1) - f(x)/g(x)$ has at least two distinct poles, namely $r_1 - 1$ and r_m, while the only pole of $h(1/x)$ is at 0.

G. A. HEUER
Concordia College

Also solved by Steven Alexander, John Briggs & Aron Pinker, Jesse Deutsch, Daniel S. Freed, Donald C. Fuller, Richard A. Groeneveld, Eli L. Isaacson, Jordan Levy, Peter A. Lindstrom, James McKim, R. M. Mathsen, Frank R. Olson, Adam Riese, H. T. Sedinger, V. Srinivas (India), J. M. Stark, and the proposer.

An Identity in a Ring **September 1976**

991. Let a and b be elements of a finite ring such that $ab^2=b$. Prove that $bab=b$. [*F. S. Cater, Portland State University.*]

Solution: From $ab^2=b$, we get $ab^{n+1}=b^n$ for $n\geqslant 1$. Let m be the least positive integer such that $b^m=b^{m+k}$ for some positive integer k (the ring is finite). If $m\geqslant 2$, we have $b^{m-1}=ab^m=ab^{m+k}=b^{m+k-1}$, so m must be 1. Then $(ba)b=(ba)b^{k+1}=bb^k=b$. (Note that only the semigroup structure is used.)

KARL W. HEUER, student
California Institute of Technology
G. A. HEUER
Concordia College

Also solved by Louis I. Alpert, Sydney Bulman-Fleming (Canada), Vasily C. Cateforis, Lee O. Hagglund, Erwin Just, Paul T. Karch & George A. Novacky, Jr., Ronald King & Gerald Thompson, Jerry M. Metzger, Leonard L. Palmer, R. Plagge, Sally Ringland, Maurice Shrader-Frechette, St. Olaf Problem Group, and the proposer.

Exceptional Hexagons **September 1976**

992. Call a vertex of a convex hexagon *ordinary* if it is the intersection of at least three diagonals or sides of different lengths. Otherwise, let the vertex be called *exceptional.*

(a) Prove that at least one vertex of a convex hexagon is ordinary.

(b)* What is the maximum number of exceptional vertices that a convex hexagon can have? [*Kenneth Fogarty, Erwin Just, and Norman Schaumberger, Bronx Community College.*]

Solution: We first show that no two adjacent vertices can be exceptional, then exhibit a hexagon with three exceptional vertices. Thus, three is the desired maximum.

Suppose that two adjacent vertices, v_1 and v_2, are exceptional, with distances to the other vertices a,b and b,c respectively, with b as the distance between v_1 and v_2. As the hexagon lies entirely on one side of the line v_1v_2, the positions of the other four vertices are determined by the respective distances to v_1 and v_2, that is, by the distance pairs (a,b), (a,c), (b,b), and (b,c). There is one point for each pair, in particular, one for the pair $(\min[a,b],\min[b,c])$. (In FIGURE 1, $\min[a,b]=a$ and $\min[b,c]=b$.) But this point is interior to the pentagon determined by the other five vertices, contradicting the convexity of the hexagon. Thus there can be at most three exceptional vertices, and so at least three ordinary vertices.

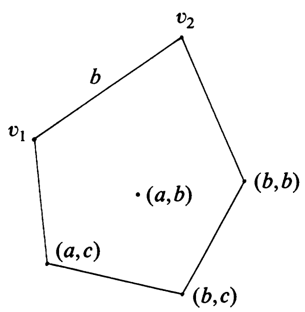


FIGURE 1

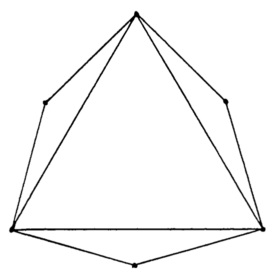


FIGURE 2

Now since no adjacent vertices are exceptional, we could have as many as three exceptional vertices only if they were alternate vertices. We construct one: let the exceptional vertices form an equilateral triangle with side r. Locate each of the remaining vertices at the midpoint of the arc of the circle of radius r centered at one vertex cut off by the other two. (FIGURE 2.)

STEVEN ALEXANDER, student
Princeton University

Also solved by Heiko Harborth (West Germany) and Dinh Thê' Hùng. Partial solutions by Donald C. Fuller, Michael Goldberg, and the proposers.

Fixed Points of Iterates **September 1976**

993. Let g be a continuous function from $[0,1]$ to $[0,1]$ with $g(0)=0$. If for each x in $[0,1]$, there is a positive integer $n(x)$ such that $g^{n(x)}(x)=x$ (the $n(x)$th iterate of g), then show that $g(x)=x$ for all x in $[0,1]$. [*F. David Hammer, University of Illinois at Chicago Circle.*]

Solution: One sees easily and directly that the relation

$$y \sim x \quad \text{iff} \quad y=g^k(x) \quad \text{for some } k \geqslant 0$$

is an equivalence relation, and the condition $g^{n(x)}(x)=x$ guarantees that every equivalence class (orbit) is finite. Suppose that for some x_0, $g(x_0)\neq x_0$; then $n_0=n(x_0)>1$. Let $x_i=g^i(x_0)$, for $i=1,2,\ldots,n_0$, and $x_j=x_i$ if $j\equiv i \pmod{n_0}$. If x_r is the least element in the orbit of x_0, we have $g(x_{r-1})=x_r<x_{r+1}=g(x_r)$, so that $0<x_r<x_{r-1}$, while $g(0)<g(x_{r-1})<g(x_r)$. By continuity, $g(u)=g(x_{r-1})$ for some u in $(0,x_r)$. Then $u\sim x_0$, contradicting that x_r is the least element of its orbit.

G. A. HEUER
Concordia College

Also solved by John T. Annulis, John Atkins & David Trick, California State College at San Bernardino Problem Solving Group, Vasily C. Cateforis, Jesse Deutsch, Michael H. Ecker, Richard A. Groeneveld, Eli L. Isaacson, Thomas E. Iverson, Paul T. Karch & George A. Novacky, Jr., Richard S. Kleber & Robert L. Raymond, Virgil C. Kowalik, Jordan Levy, Frederick M. Liss, James McKim, H. T. Sedinger, Henry Schultz, St. Olaf Problem Solving Group, J. M. Stark, Barbara Turner, Paul Y. H. Yiu (Hong Kong), and the proposer.

A Greatest Integer Integral **September 1976**

994. For n and m positive integers, evaluate $\int_0^1(-1)^{[nt]}(-1)^{[mt]}dt$, where [ ] denotes the greatest integer function. [*Peter Ørno, The Ohio State University.*]

Solution: Let $I(n,m)=\int_0^1(-1)^{[nt]}(-1)^{[mt]}dt$. We observe that for $0\leqslant t\leqslant \frac{1}{2}$, if nt is not an integer, we have $(-1)^{[nt]}=(-1)^{n+1}(-1)^{[n(1-t)]}$ from which it follows that $I(n,m)=0$ if n and m have opposite parity. Now consider n and m with the same parity, and let a be the greatest common divisor of n and m; say $n=an_0$ and $m=am_0$. Putting $u=at$, we see $I(n,m)=a^{-1}\sum_{j=0}^{a-1}\int_j^{j+1}(-1)^{[n_0u]}(-1)^{[m_0u]}du$. Letting $v=u-j$ gives

$$I(n,m)=\frac{1}{a}\sum_{j=0}^{a-1}\int_0^1(-1)^{[n_0v]}(-1)^{[m_0v]}(-1)^{j(m_0+n_0)}dv$$

$$=I(n_0,m_0)\cdot\frac{1}{a}\sum_{j=0}^{a-1}(-1)^{(m_0+n_0)j}.$$

Now, if one of n_0 or m_0 is even, the other must be odd so that $I(n_0,m_0)=0$ and consequently $I(n,m)=0$. On the other hand, if n_0 and m_0 are both odd, then $\sum_{j=0}^{a-1}(-1)^{(m_0+n_0)j}=a$, and hence

$I(n,m)=I(n_0,m_0)$. Thus, to complete the problem we need only compute $I(n,m)$ when n and m are odd and relatively prime.

$$I(n,m)=\int_0^1(-1)^{[nt]}(-1)^{[mt]}dt=\sum_{j=0}^{nm-1}\int_{j/nm}^{(j+1)/nm}(-1)^{[nt]}(-1)^{[mt]}dt$$

$$=\frac{1}{nm}\sum_{j=0}^{nm-1}(-1)^{[j/n]}(-1)^{[j/m]}=\frac{1}{nm}\sum_{j=0}^{nm-1}(-1)^{n[j/n]}(-1)^{m[j/m]}.$$

The last equality holds since n and m are odd. If we let $r_{j,k}$ denote the remainder when j is divided by k, then the last sum becomes

$$\frac{1}{nm}\sum_{j=0}^{nm-1}(-1)^{j-r_{j,n}}(-1)^{j-r_{j,m}}=\frac{1}{nm}\sum_{j=0}^{nm-1}(-1)^{r_{j,n}}(-1)^{r_{j,m}}$$

$$=\frac{1}{nm}\sum_{t=0}^{n-1}\sum_{s=0}^{m-1}(-1)^{r_{sn+t,n}}(-1)^{r_{sn+t,m}}$$

$$=\frac{1}{nm}\sum_{t=0}^{n-1}(-1)^t\left(\sum_{s=0}^{m-1}(-1)^{r_{sn+t,m}}\right).$$

Now, as n and m are relatively prime, the quantities $sn+t$ $(s=0,1,\ldots,m-1)$ form a complete residue system modulo m, and consequently the quantities $r_{sn+t,m}$ $(s=0,1,\ldots,m-1)$ are the numbers $0,1,\ldots,m-1$ in some order. Thus

$$\sum_{s=0}^{m-1}(-1)^{r_{sn+t,m}}=\sum_{s=0}^{m-1}(-1)^s=1.$$

This shows

$$I(n,m)=\frac{1}{nm}\sum_{t=0}^{n-1}(-1)^t\left(\sum_{s=0}^{m-1}(-1)^s\right)=\frac{1}{nm}.$$

Using (n,m) to denote the greatest common divisor of n and m, we can express the value of $I(n,m)$ compactly as

$$I(n,m)=\begin{cases}\dfrac{(n,m)^2}{nm} & \text{if } \dfrac{n}{(n,m)} \text{ and } \dfrac{m}{(n,m)} \text{ are both odd}\\ 0 & \text{otherwise.}\end{cases}$$

M. B. Gregory
J. M. Metzger
University of North Dakota

Also solved by Michael Goldberg, Richard A. Groeneveld, G. A. Heuer, Eli L. Isaacson, Vaclav Konecny, Jordan I. Levy, and the proposer.

R-Symmetric Matrices **September 1976**

995. Call an $n\times n$ matrix $(n\geqslant 2)$ *R-symmetric* if the interchange of any two distinct rows yields a symmetric matrix. Find a characterization of all R-symmetric matrices. [*Edward T. H. Wang, Wilfred Laurier University, Canada.*]

Solution: Let $A=[a_{ij}]$ be an $n\times n$ R-symmetric matrix. It is easily shown that

$$A=\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{11}\end{bmatrix}\quad\text{and}\quad A=\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{13} & a_{11} & a_{12}\\ a_{12} & a_{13} & a_{11}\end{bmatrix}$$

if $n=2$ and 3 respectively. Let $n \geqslant 4$ and let i, j, k, and m be distinct positive integers, all $\leqslant n$. Since interchanging the ith and jth rows gives a symmetric matrix, it follows that (1) $a_{ii}=a_{jj}$, (2) $a_{km}=a_{mk}$, (3) $a_{ik}=a_{kj}$, and (4) $a_{ki}=a_{jk}$. Noting that i, j, k, and m are arbitrary, (1) implies that the diagonal entries are all the same and (2) implies that A is symmetric. Thus (3) and (4) may be rewritten as (5) $a_{ik}=a_{jk}$ and (6) $a_{ki}=a_{kj}$. Now (5) implies that the nondiagonal entries in a column are all the same and (6) implies that the nondiagonal entries in a row are all the same. It follows that the nondiagonal entries of A are all the same. Hence, for $n \geqslant 4$ an R-symmetric matrix is given by

$$a_{ij}=\begin{cases} c & \text{if } i \neq j \\ d & \text{if } i=j, \end{cases}$$

where c,d are fixed numbers.

HARRY SEDINGER
University of Wisconsin, La Crosse

Also solved by Michael W. Ecker, Thomas E. Elsner, Richard A. Groeneveld, Eli L. Isaacson, Richard Troxel, Paul Y. H. Yiu (Hong Kong), and the proposer.

Answers

Solution to the Quickies which appear near the beginning of the Problems section.

Q651. Applying Stewart's theorem (which follows from the law of cosines) to triangle ABC with respect to AA_i, we find

$$a|AA_i|^2 + a \cdot \frac{ai}{n} \cdot \frac{a(n-i)}{n} = \frac{ai}{n} b^2 + \frac{a(n-i)}{n} c^2.$$

If we now sum over i, use the formulas for Σi and Σi^2, we obtain after simplification that

$$\sum_{i=1}^{n} |AA_i|^2 = \frac{n-1}{2}\left(b^2+c^2-\frac{(n+1)a^2}{3n}\right).$$

Adding this to the corresponding results for $\Sigma |BB_i|^2$ and $\Sigma |CC_i|^2$, we obtain the desired equality.

Q652. We see that

$$\sum_{i=1}^{n}(1+\tan\alpha_i) = \sum_{i=1}^{n} \frac{\sin\alpha_i + \cos\alpha_i}{\cos\alpha_i} = \sqrt{2} \sum_{i=1}^{n} \frac{\sin(\alpha_i+\pi/4)}{\cos\alpha_i} \leqslant \sqrt{2} \sum_{i=1}^{n} \sec\alpha_i.$$

Equality holds if and only if $\sin(\alpha_i+\pi/4)=1$ for all i. In particular, if α_1, α_2, and α_3 are angles of a triangle, then $3+\Sigma\tan 3\alpha_i/4 \leqslant \sqrt{2}\,\Sigma\sec 3\alpha_i/4$ with equality if and only if the triangle is equilateral.

In the original answer to Q472 it was also shown that $\sec\alpha+\sec\beta \geqslant 2$. Again, this isn't sharp. Since $\sec x$ is convex in $(-\pi/2, \pi/2)$,

$$\frac{1}{n}\sum_{i=1}^{n}\sec\alpha_i \geqslant \sec\left(\sum_{i=1}^{n}\alpha_i/n\right)$$

with equality if and only if $\alpha_i = \text{constant}$.

REVIEWS

Paul J. Campbell, Editor
Beloit College

Pierre Malraison, Editor
Control Data Corp.

Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of the mathematics literature. Some reviews of books are adapted from the Telegraphic Reviews in the American Mathematical Monthly.

Lewis, Harry R. and Papadimitriou, Christos H., *The efficiency of algorithms*, Scientific American 238 (January 1978) 96-109, 138, 140.

Mathematical problems can be classified in a hierarchy according to the efficiency of algorithms known for their solution. This article sketches the present state of knowledge about polynomial-time and NP-complete problems. Specific problems discussed include the travelling salesman problem, the Euler and Hamilton path problems, the three-color map problem, and the assignment problem.

Reeves, M. Sandra, *The fourth dimension*, Western's World 9 (1978) 38-44.

A popular article dealing primarily with the work of Thomas Banchoff on computer generated images of four-dimensional objects.

Baker, Russell, *A pi in the face*, New York Times Magazine (18 September 1977) 11.

A burlesque of math anxiety, discounting any correlation with sex and attributing it to the use of Greek letters. "It is this fear of sudden assault by incomprehensible tongues and symbols that lies at the heart of math anxiety." The claim is that males suffer equally from male anxiety but their machismo makes them "ashamed to admit that when it comes to pi they are chicken."

Kolata, Gina Bari, *Information theory, a surprising proof*, Science 199 (6 January 1978) 42.

Lászlo Lovász (József Attila University, Szeged, Hungary) has surprised mathematicians with a simple and ingenious solution of a special case of a problem posed by Claude Shannon over 20 years ago. The problem is to determine the rate at which information can be sent error-free over a noisy channel; Lovász found the capacity for a system of signals whose graph is a pentagon. Unfortunately, Kolata gives us no hint of the simple mathematics involved in the proof.

Medical research: statistics and ethics, Science 198 (18 November 1977) 677-705.

Proceedings of a one-day symposium held in May 1977 at the Sloan-Kettering Cancer Center in New York. Among the six articles is one by John W. Tukey on ethical dimensions of design of clinical trials, and another by John P. Gilbert, *et al*, on comparisons of innovations with standard medical treatments. The latter uses the procedure of correcting for sampling error by "shrinking" results toward the mean; this is the procedure that was the subject of "Stein's paradox in statistics," by Bradley Efron and Carl Morris (*Scientific American* 236 (May 1977) 119-127, 148; reviewed here May, 1977).

Mandelbrot, Benoit B., *Physical objects with fractional dimension: seacoasts, galaxy clusters, turbulence and soap*, Bull. of the Institute of Math. and Its Appl. 13 (July/August 1977) 189-196.

Although written in 1973, this paper still has value as a succinct introduction to the author's study of the "fractal geometry of nature," continued in *Fractals: Form, Chance, and Dimension* (Freeman, 1977).

Mullet, Gary M., *Simeon Poisson and the National Hockey League*, American Statistician 31 (February 1977) 8-12.

Fitting of scoring records for 1973-74 season to Poisson distributions. An unexpected result: offense (goals for) and defense (goals against) for any given team appear to be statistically independent.

Zadeh, Norman, *Computation of optimal poker strategies*, Operations Research 25 (July/August 1977) 541-562.

Explains how the system of the author's book *Winning Poker Systems* (Prentice-Hall, 1974) was calculated. Simple but best strategies are given for opening, calling, and raising, depending on size of pot, table position of opponents, and opponents' playing habits. Enlightening and potentially profitable!

Gardner, Martin, Mathematical Magic Show, Knopf, 1977; 284 pp, $8.95.

This eighth collection of Gardner's columns from *Scientific American* spans the years 1967-68 (plus a few outliers). As of now just about half of his 250 or so columns have made it into book form. The introduction to this volume contains a short glossary of mathematical terms, including an unintended singularity ("the point at which something peculiar happens...") which escaped the proofreading: "Irrational numbers: real numbers that are not integers."

Ernst, Bruno, The Magic Mirror of M.C. Escher, Random House, 1976; 112 pp, $15.

An absolutely splendid book, from the copious reproductions to the thoughtful commentary. The mathematician-author discerns much of the mathematical flavor in Escher's work and offers us, through Escher's draft sketches, insight into how he sought and achieved his effects.

Ranucci, E.R. and Teeters, J.L., Creating Escher-Type Drawings, Creative Pub, 1977; 198 pp, $12.80.

An elementary explanation of translation rotation and reflection devices for creating tessellations of the plane based on curvilinear figures. Lavishly illustrated, full of potential for fun and learning.

Gallese, L.R., *A little calculating and a lot of terror equal math anxiety*, Wall Street J, March 14, 1978.

Math anxiety, and how a variety of schools are dealing with it. Some interesting statistics (how many people take 4 years of math in high school: 8% of the women, 52% of the men), and some indications of the wide variety of professions requiring mathematics. Also how often does Al Willcox get quoted on the front page of the *Wall Street Journal?*

Glass, Leon, *Petterns of supernumerary limb regeneration*, Science 198 (21 October 1977) 321-322.

Shows how the index theorem can be used to predict the results of grafting and transplantation of limbs in amphibians and insects, and a footnote points to other applications of the result in biological and chemical contexts.

Swindell, William and Barrett, Harrison H., *Computerized tomography: taking sectional x-rays*, Physics Today 30 (December 1977) 32-41.

Elementary description of both the physics and the mathematics of transverse x-ray scanners, which can isolate a single plane of a patient's body. The mathematical problem is the (computer) reconstruction of the desired image from a multitude of two-dimensional projections from x-rays taken from different angles. Several reconstruction algorithms are discussed, including the filtered back projection, which employs image processing, and iterative techniques. Treated briefly is the mathematics of the Fourier-Radon transform and its "central-slice theorem," which provides the method's basis. More details are given by Kennan T. Smith, *et al.*, in "Practical and mathematical aspects of the problem of reconstructing objects from radiographs," *Bulletin Amer. Math. Soc.* 83 (November 1977) 1227-1271.

Garfinkel, Robert S., *Minimizing wallpaper waste, Part 1: A class of travelling salesman problems*, Operations Research 25 (September/October 1977) 741-751.

A splendid instance of everyday life providing an occasion for mathematics. The author did some wallpapering, was appalled at the amount of wasted paper, and set out to see how to minimize the waste. He shows that the problem is equivalent to a travelling salesman problem for which a "nearest neighbor" algorithm yields an optimal solution. The mathematics is elementary and satisfying.

Das, K.S., *A machine representation of finite topologies*, J. Assoc. Comp. Mach. 24 (1977) 676-692.

Uses matrices to represent similarity classes of topologies and presents an algorithm for counting the number of different matrix representations. A list is given for $n \leq 11$.

Egbert, William E., *Personal calculator algorithms I: Square roots*, Hewlett-Packard J. 28:9 (May 1977) 22-24.

Surprisingly, the core of the square-root algorithm on HP calculators is the (usually forgotten) scheme taught in grade school. A few clever refinements illuminate the important transition from a mathematically sound algorithm to an efficient computational procedure. Accuracy: ± 1 in the tenth digit of the twelve used by the calculator.

Egbert, William E., *Personal calculator algorithms II: Trigonometric functions*, Hewlett-Packard J. 28:10 (June 1977) 17-20.

Input is first scaled to a number between 0 and 2π by repeated subtraction. (If the original argument was of order of magnitude n, then n digits of accuracy are lost at this stage--so don't believe the number your calculator gives for $\sin(10^{22})$!). Then the new angle is divided into smaller angles whose tangents are powers of 10; and trigonometric identities are implemented, using shifting and adding, to arrive at direction numbers. The desired trigonometric function is then computed using elementary operations.

Egbert, William E., *Personal calculator algorithms III: Inverse trigonometric functions*, Hewlett-Packard J. 29:3 (November 1977) 22-23.

The process used is virtually the reverse of that for the trigonometric functions (and an understanding of the algorithm for the latter is assumed): vector rotation is followed by pseudo-multiplication (the summing of inverse tangents of powers of 10) to get the inverse tangent, and the other inverse functions are calculated from this.

Egbert, William E., *Personal calculator algorithms IV: Logarithmic functions,* Hewlett-Packard J. 29:8 (April 1978) 29-32.

The approximation process used--pseudo-multiplication, or "shift and add"--is much the same as the one used to compute trigonometric functions.

Browne, Malcolm W., *Experts debate the prediction of disasters,* New York Times (19 November 1977) 1, 14.

Popularized (and slightly distorted) report on catastrophe theory and the controversy surrounding it.

Lacher, R.C., McArthur, Robert and Buzyna, George, *Catastrophic changes in circulation flow patterns,* American Scientist 65 (September-October 1977) 614-621.

Careful application of catastrophe theory to abrupt changes in large-scale atmospheric circulation patterns, as simulated in the laboratory.

Rivier, Nicolas, *Alice through the renormalisation group,* New Scientist 77 (9 February 1978) 353-355.

Renormalization, or scaling, has proved to be an enormously successful mathematical technique for looking at a given principal system at different levels of magnification. The renormalization group (really a semigroup: no inverses) was introduced in the 1950's to deal with spurious infinities in particle physics; this article relates its success in describing the behavior of condensed matter near a critical point. Other applications hinted at include cosmology, percolation, the Kondo effect, curling of a polymer chain, transformational grammars, and societal structural analyses. Too bad no references are provided for further reading.

Bookstein, Fred L., *The study of shape transformation after D'Arcy Thompson,* Mathematical Biosciences 34 (1977) 177-219.

D'Arcy Thompson's *On Growth and Form* (1917, 1942) tried to interpret both short-term and evolutionary change in biological form by means of unquantified mathematical transformations. The present article surveys other attempts to quantify Thompson's work and then describes the author's own theory of biorthogonal grids.

Metz, William D., *Midwest computer architect struggles with speed of light,* Science 199 (27 January 1978) 404-409.

A profile of Seymour R. Cray and his unusual-looking Cray-1 computers: at $8 million each, they are four times faster than anything previously built. A circular cross-section was chosen so as to minimize time delay of electrical signals; a Cray takes up only eight square feet of floor space.

The computer society, Time (20 February 1978) 44-59.

A series of illustrated stories glorifying the "miracle chip" and the transformations computer technology will work upon society. Optimists' hopes will be enhanced, and pessimists' worst fears confirmed.

Special Morley Issue, Eureka (Ottawa) 3 (December 1977) 271-300.

Beautiful because of its simplicity, and extraordinary because it eluded the Greeks, the subject of this special issue is Morley's Theorem: the intersections of adjacent trisectors of the angles of a triangle are the vertices of an equilateral triangle. Morley's original proof is reproduced and analyzed, several more elementary proofs are presented, and a comprehensive bibliography is provided.

NEWS & LETTERS

SCAFFOLDING 2

We are pleased to feature on our cover this month an etching created in 1971 by Jonathan Talbot (R.D. 2, Pine Island Turnpike, Warwick, NY 10990). Talbot's work has been exhibited in New York's Museum of Modern Art, and he has won many first place and best-of-show awards.

Each pair of workmen in Scaffolding commenced work independently and, according to Talbot, only as their work neared completion did they discover the existence of their counterparts. "They then joined their efforts to produce the unified structure represented in this etching. Now two of the workmen are enjoying a well deserved rest while their partners ascend to the top of the structure to better observe the fruits of their labors. All four seem to be totally oblivious to the fact that the scaffolding they have erected will serve no useful function in a three dimensional world, except perhaps to indicate the existence of some further dimension of which we are, as yet, unaware."

1978 CHAUVENET PRIZE

Professor Sheeram Shankar Abhyankar of Purdue University has been awarded the 1978 Chauvenet Prize for noteworthy exposition for his paper "Historical Ramblings in Algebraic Geometry and Related Algebra" (*Amer. Math. Monthly*, 83 (1976) 409-448). This prize, represented by a certificate and an award of $500, is the twenty-sixth award of the Chauvenet Prize since its institution by the Mathematical Association of America in 1925.

Professor Abhyankar was born on July 22, 1930, in Ujjain, India. He received the B.Sc. from Bombay University in 1951 and A.M. and Ph.D. degrees from Harvard University in 1952 and 1955, respectively. He has held regular faculty positions at Columbia, Cornell, Johns Hopkins, and Purdue and visiting faculty positions at Columbia, Princeton, Harvard, Münster, Erlangen, Matscience Madras, India, Yale, Tata Institute (Bombay), and Kyoto (Japan). Currently he is a Distinguished Professor at Purdue University.

Professor Abhyankar has authored seven texts and numerous papers in the area of algebraic geometry, commutative algebra, local algebra, theory of functions of several complex variables, and circuit theory. He has recently helped found Bhaskaracharya Pratishthana, a mathematical research institute at 39/5 Erandavane, Pune 4, India. Professor Abhyankar serves as Prime Member and Director of the Governing Body of this institute, which is named after the twelfth century Indian poet-mathematician whose book on geometry called *Lilavati*, composed in Sanskrit poetry around 1114 A.D., inspired Abhyankar's early interest in writing readable mathematics.

MATH FOR POETS

CUPM has established a panel to prepare recommendations concerning the content of those courses that treat mathematics appreciation for liberal arts students. The panel is interested in receiving information from instructors who have had experience with such courses, or who have suggestions concerning objectives, content, or resources. *Magazine* readers interested in this endeavor are urged to write with suggestions to the panel chairman, Professor Jerome A. Goldstein, Department of Mathematics, Tulane University, New Orleans, LA 70118.

STATISTICS AND MATHEMATICS

The sixth annual Mathematics and Statistics Conference at Miami University, Oxford, Ohio will be held on September 29-30, 1978 on the topic "Applications of Statistics and Mathematics." George Carrier (Harvard University), Victor Klee (University of Washington), and Frederick Mosteller (Harvard University) will be the featured speakers. There will be sessions of contributed papers. Abstracts should be sent to Charles Holmes or Elwood Bohn, Department of Mathematics and Statistics, Miami University, Oxford, Ohio 45056. The deadline for abstracts is June 15, 1978. Late abstracts may be considered. Information concerning preregistration, housing, etc., can be obtained from the above address.

MORE FALSE COIN CLUES

My copy of Dover's *Mathematical Recreations* by Kraitchik is clearly marked *Second Revised Edition*. The first edition of the work does not contain the False Coin problem. In the Dover edition the problem number 57 and the problem title are given on page 43 along with instructions to see page 324. This required the renumbering of only a few pages. The obvious difference in type is clue enough that the material on the false coin problem is an addition.

Two other additions in the Dover edition are an update of information on Mersenne primes on pages 70 and 71, and a footnote and a shift of one line of text on pages 122 and 123.

It would be interesting to see a complete copy of Dover's 1953 ad for the *Revised Edition*. The inside front cover of my paperback copy contains the following:

> "From Clavius' 17th century problem of the beggars (page 23) to the modern "false-coin" puzzle (page 324),..."

It seems that the inclusion of the false-coin problem was a featured selling point for the revised edition. It amuses me that Dover used the word "modern" in the ad after having placed the problem at the end of the chapter on ancient and curious problems.

Joseph D.E. Konhauser
Macalester College
St. Paul
Minnesota 55105

SOLUTIONS TO THE 1977 PUTNAM EXAM

In January we printed in this column questions from the 1977 Putnam Examination. To assist those who have been puzzling over these problems, we provide here hints and answers. The official report on the results of the competition, including names of winners and complete sample solutions, will be published later this year in the American Mathematical Monthly.

A-1. Consider all lines which meet the graph of

$$y = 2x^4 + 7x^3 + 3x - 5$$

in four distinct points, say

$$(x_i, y_i),\ i = 1, 2, 3, 4 .$$

Show that $(x_1+x_2+x_3+x_4)/4$ is independent of the line and find its value.

Sol. Let $y = mx + b$ be such a line. Then the x_i are the roots of

$$2x^4 + 7x^3 + (3-m)x - (5+b) = 0,$$

and so their sum is -7/2, and their arithmetic mean is -7/8.

A-2. Determine all solutions in real numbers x, y, z, w of the system

$$x + y + z = w,$$
$$\frac{1}{x} + \frac{1}{y} + \frac{1}{z} = \frac{1}{w} .$$

Sol. Substituting w from the first equation into the second and simplifying gives $0 = (x+y)(x+z)(y+z)$. From this it is easy to deduce that w must equal one of x, y, z and that the remaining two unknowns must be negatives of each other.

A-3. Let u, f, and g be functions, defined for all real numbers x such that

$$\frac{u(x+1) + u(x-1)}{2} = f(x)$$

and

$$\frac{u(x+4) + u(x-4)}{2} = g(x) .$$

Determine $u(x)$ in terms of f and g.

Sol. One possible answer, among others, is $u(x) = g(x) - f(x+3) + f(x+1) + f(x-1) - f(x-3)$.

A-4. For $0 < x < 1$, express

$$\sum_{n=0}^{\infty} \frac{x^{2^n}}{1-x^{2^{n+1}}}$$

as a rational function of x.

Sol. Let

$$S_n = \sum_{k=0}^{n} \frac{x^{2^k}}{1 - x^{2^{k+1}}}$$

$$= \sum_{k=0}^{n} \left\{ \frac{1}{1-x^{2^k}} - \frac{1}{1-x^{2^{k+1}}} \right\}$$

$$= \frac{1}{1-x} - \frac{1}{1-x^{2^{n+1}}} .$$

For $0 < x < 1$, the desired sum is

$$\lim_{n\to\infty} S_n = \frac{1}{1-x} - 1 = \frac{x}{1-x} .$$

A-5. Prove that

$$\binom{pa}{pb} \equiv \binom{a}{b} \pmod p$$

for all integers p, a, and b with p a prime, $p > 0$, and $a \geq b \geq 0$.

Sol. We know that $(x+1)^p = x^p + 1$ in $Z_p[x]$, where Z_p is the field of integers modulo p. Thus in $Z_p[x]$,

$$\sum_{k=0}^{pa} \binom{pa}{k} x^k = (x+1)^{pa} = [(x+1)^p]^a$$

$$= (x^p+1)^a = \sum_{j=0}^{a} \binom{a}{j} x^{jp} .$$

Since coefficients of x^{bp} must be congruent modulo p, for $b = 0, 1, \ldots, a$, we see that

$$\binom{pa}{pb} \equiv \binom{a}{b} \pmod p .$$

A-6. Let $f(x,y)$ be a continuous function on the square

$$S = \{(x,y): 0 \leq x \leq 1, 0 \leq y \leq 1\} .$$

For each point (a,b) in the interior of S, let $S_{(a,b)}$ be the largest square that is contained in S, is centered at (a,b), and has sides parallel to those of S. If the double integral $\iint f(x,y)dx\,dy$ is zero when taken over each square $S_{(a,b)}$, must $f(x,y)$ be identically zero on S?

Sol. For (a,b) in S, let $I(a,b)$ be $\iint f(x,y)dx\,dy$ over the rectangle $0 \leq x \leq a$, $0 \leq y \leq b$. Also let (a,b) define inductively a sequence (a_n,b_n) using $a_1 = a$, $b_1 = b$, $a_{n+1} = a_n - b_n$ and $b_{n+1} = b_n$ when $0 \leq b_n \leq a_n$, and $a_{n+1} = a_n$ and $b_{n+1} = b_n - a_n$ when $0 \leq a_n < b_n$. Then the hypothesis implies that $I(a,b) = I(a_n,b_n)$ for all n. Since f is bounded on S and $\lim_{n\to\infty} a_n = 0 = \lim_{n\to\infty} b_n$, it follows that $I(a,b) = 0$ for all (a,b) in S.

If $f(x,y)$ is not zero for all (x,y) in S, then f must be positive (or negative) in some rectangle $R = \{(x,y): c \leq x \leq d, h \leq y \leq k\}$ and hence $I = \iint_R f(x,y)dx\,dy$ must be positive (or negative). But this contradicts

$$I = I(d,k) - I(d,h) - I(c,k) + I(c,h) = 0.$$

Thus f is identically zero on S.

B-1. Evaluate the infinite product

$$\prod_{n=2}^{\infty} \frac{n^3-1}{n^3+1} .$$

Sol. The product equals

$$\lim_{k\to\infty} \left[\frac{1\cdot 7}{3\cdot 3} \cdot \frac{2\cdot 13}{4\cdot 7} \cdot \frac{3\cdot 21}{5\cdot 13} \cdots \frac{(k-1)(k^2+k+1)}{(k+1)(k^2-k+1)}\right] =$$

$$\lim_{k\to\infty} \left[\frac{2}{3} \cdot \frac{k^2+k+1}{k(k+1)}\right] = \frac{2}{3} .$$

B-2. Given a convex quadrilateral $ABCD$ and a point O not in the plane of $ABCD$, locate point A' on line OA, point B' on line OB, point C' on line OC, and point D' on line OD so that $A'B'C'D'$ is a parallelogram.

Sol. Let O' be any point different from O on the line of intersection of planes AOC and BOD, e.g., O' may be the intersection of lines AC and BD. Let A' be the intersection of line OA with the line through O' and parallel to OC. Let C' be the intersection of line OC with the line through O' which is parallel to OA. Then $OA'O'C'$ is a parallelogram and its diagonals OO' and $A'C'$ bisect each other at a point M.

Choosing B' and D' in the same way, one obtains a parallelogram $OB'O'D'$ whose diagonals OO' and $B'D'$ also bisect each other at the midpoint M of segment OO'. Hence segments $A'C'$ and $B'D'$ bisect each other (at M) and $A'B'C'D'$ is a parallelogram. (The parallelogram is not unique.)

B-3. An (ordered) triple (x_1, x_2, x_3) of positive *irrational* numbers with $x_1 + x_2 + x_3 = 1$ is called "balanced" if each $x_i < 1/2$. If a triple is not balanced, say if $x_j > 1/2$, one performs the following "balancing act"

$$B(x_1, x_2, x_3) = (x_1', x_2', x_3'),$$

where $x_i' = 2x_i$ if $i \neq j$ and $x_j' = 2x_j - 1$. If the new triple is not balanced, one performs the balancing act on it. Does continuation of this process always lead to a balanced triple after a finite number of performances of the balancing act?

Sol. Let $x_i = \sum_{j=1}^{\infty} a_{ij} 2^{-j}$, with $a_{ij} \in \{0,1\}$, be the binary expansion of x_i. An unbalanced triple that remains unbalanced after any finite number of balancing acts is constructed by choosing the a_{ij} so that exactly one of a_{1j}, a_{2j}, a_{3j} equals 1 for each j while taking care that no one of sequences $a_{i1}, a_{i2}, \ldots$ repeats in blocks, i.e., that each x_i is irrational. One such solution has $a_{1j} = 1$ if and only if $j \in \{1, 9, 25, 49, \ldots\}$, $a_{2j} = 1$ if and only if $j \in \{4, 16, 36, 64, \ldots\}$, $a_{3j} = 1$ if and only if $j \in \{2, 3, 5, 6, \ldots\}$.

B-4. Let C be a continuous closed curve in the plane which does not cross itself and let Q be a point inside C. Show that there exists points P_1 and P_2 on C such that Q is the midpoint of the line segment P_1P_2.

Sol. We can assume that $Q = 0$, the origin. Let $-C$ be the image of C under the reflection $P \to -P$. $-C$ is again a continuous closed curve surrounding 0 and $C \cap -C \neq \emptyset$ since they have the same diameter and both surround 0 (hence neither can be exterior to the other). Let $P_1 \in C \cap -C$. Then there exists $P_2 \in C$ such that $P_1 = -P_2$. These two points do the trick.

B-5. Suppose that $a_1, a_2, \ldots, a_n$ are real $(n>1)$ and

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

Prove that $A < 2a_i a_j$ for $1 \leq i < j \leq n$.

Sol. From the Cauchy-Schwarz Inequality, one has

$$[(a_1+a_2)+a_3+a_4+\ldots+a_n]^2$$
$$\leq [1^2+1^2+\ldots+1^2][(a_1+a_2)^2+a_3^2+\ldots+a_n^2]$$

or

$$(\Sigma a_i)^2 \leq (n-1)[(\Sigma a_i^2)+2a_1a_2]$$

or

$$[1/(n-1)](\Sigma a_i)^2 \leq (\Sigma a_i^2)+2a_1a_2.$$

Using the hypothesis, one then has

$$A < -(\Sigma a_i^2)+\frac{1}{n-1}(\Sigma a_i)^2$$
$$\leq -(\Sigma a_i^2)+(\Sigma a_i^2)+2a_1a_2=2a_1a_2 .$$

Similarly, $A < 2a_i a_j$ for $1 \leq i < j \leq n$.

B-6. Let H be a subgroup with h elements in a group G. Suppose that G has an element a such that for all x in H, $(xa)^3 = 1$, the identity. In G, let P be the subset of all products $x_1ax_2a\ldots x_na$, with n a positive integer and the x_i in H.

(a) Show that P is a finite set.

(b) Show that, in fact, P has no more than $3h^2$ elements.

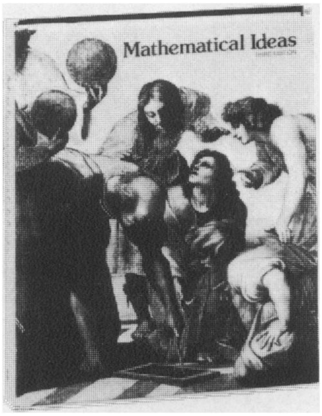
Sol. Since $1 \in H$, the hypothesis implies that $a^{-1} = a^2$. Also, it is easy to show that $axa = x^{-1}a^2x^{-1}$ and $a^2xa^2 = x^{-1}ax^{-1}$. Using these identities and the fact that $x \in H$ implies $x^{-1} \in H$, an easy induction shows that any element of P can be reduced to one of the following forms:

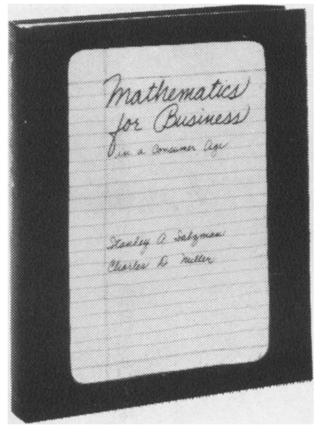
(i) xay, $x,y \in H$,

(ii) xa^2y, $x,y \in H$, or

(iii) xa^2ya, $x,y \in H$.

The result follows by noting that there are at most h^2 elements of each of these types.





Mathematical Ideas

Third Edition

Charles D. Miller/Vern E. Heeren
March 1978, 528 pages, illustrated
hardbound $13.95

Now in its Third Edition, *Mathematical Ideas* is a proven success for general, liberal arts, and teacher training courses. New chapters on the real number system, algebra, and geometry make the survey more complete. New topic sequence, chapter reviews and tests, and illustration program make this edition even more teachable. Still highly readable and informal with worked-out examples, lots of exercises and applications, and historical perspective on each topic. Instructor's Guide with alternate tests and additional topic material.

Mathematics: An Everyday Experience

Charles D. Miller/Vern E. Heeren
1976, 512 pages, illustrated, hardbound
$12.95

Eleven fundamental, lively topics are presented informally, with emphasis on their applications for an appreciation of the usefulness of mathematics in everyday life. Extensive Instructor's Guide.

Mathematics for Business

In a Consumer Age

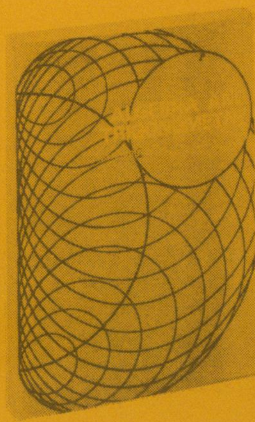
Stanley A. Salzman/Charles D. Miller
January 1978, 448 pages, illustrated
hardbound $12.95

All the basic mathematical ideas needed for success in today's business world are covered in a down-to-earth, non-technical presentation. All student materials are included in the text itself: pretests, section quick-checks, chapter tests, learning objectives for each section, a glossary, and a large number of practical problems with selected solutions and answers. Instructor's Guide with Tests.

Technical Mathematics

Philip M. Jaffe/Rodolfo Maglio
Oakton Community College
March 1978, 576 pages, hardbound $13.95

Class-tested material and a large number of topics meet the very practical needs of vocational-technical students and instructors. Short explanations followed by worked-out examples and exercises with word problems for general, technical, and shop students promote skills plus understanding. Instructor's Guide with outlines, tests, and answers.



Algebra and Trigonometry

Margaret L. Lial/Charles D. Miller
January 1978, 576 pages, hardbound $13.95

Accessible, thorough coverage of algebra and trigonometry ensures complete preparation for calculus. Numerous applications, over 400 worked-out examples, and over 4,000 graded exercises with many word problems are provided. Instructor's Guide with quizzes and tests. Solutions Guide and Study Guide.

Also available is Lial and Miller's completely revised precalculus sequence—each text complete with Instructor's Guides, Solutions Guides, Study Guides, and MathLabs (complete testing program for unit mastery):

Beginning Algebra
Second Edition
1976, 334 pages, hardbound $11.50

Intermediate Algebra
Second Edition
1976, 432 pages, hardbound $12.50

College Algebra
Second Edition
1977, 384 pages, hardbound $12.95

Trigonometry
1977, 320 pages, hardbound $11.95

Finite Mathematics With Applications in Business, Biology, and Behavioral Sciences

Margaret L. Lial/Charles D. Miller
1977, 448 pages, hardbound $13.95

Seventeen actual case studies and questions from past CPA examinations help show how mathematics is used in real-world situations. Examples, word problems, and applications throughout. Instructor's Guide with quizzes and tests list models.

Essential Calculus With Applications in Business, Biology, and Behavioral Sciences

Margaret L. Lial/Charles D. Miller
1975, 346 pages, hardbound, $14.50

Includes 13 case studies, diagnostic pretest, algebra review, chapter pretests, word problems throughout, optional sections for longer courses, and Instructor's Guide with test program.

For further information write
Jennifer Toms, Department SA
1900 East Lake Avenue
Glenview, Illinois 60025

Scott, Foresman College Division

ORGANIZING FOR MATHEMATICS INSTRUCTION

1977 YEARBOOK

edited by F. Joe Crosswhite

Here's a book that will be your key to new vistas! Read it and gather a multitude of ideas for teaching approaches —see how you can individualize inexpensively in your classroom, how you can centralize learning activities in a mathematics resource area, how you can form a learning cooperative. . . how you can ORGANIZE for teaching mathematics effectively!

This 256-page hardcover features 12 thematic essays that analyze and interpret organizational alternatives for teaching mathematics in the elementary and the secondary schools. Authors have written from their own personal experiences. Each advocates a specific teaching model that he or she believes is viable and effective—it could be just what you need in your classroom or school. Some of the essay titles are

- **Organizing for Individualization: A Classroom Model**
- **Organizing Independent Learning Units**
- **Organizing a Learning Cooperative: Survival Groups**
- **Organizing for Mastery Learning: A Group-based Approach**
- **Organizing for Alternative Schools: An Integrated Approach**
- **The Teacher-centered Classroom**

Also read the bonus nonthematic essay, "Hand-held Calculators: Past, Present, and Future." There couldn't be a timelier topic than this, the potential impact of the minicalculator on mathematics education.

National Council of Teachers of Mathematics
1906 Association Drive
Reston, VA 22091 1977 B

Please send me the 1977 Yearbook,
Organizing for Mathematics Instruction.

Quantity	Each	Total
______	$10.00	______
______	$ 9.00*	______

* Special price for NCTM members
Virginia residents add 4% sales tax

Name (please print) ____________________

Address ____________________

City ____________ State or Province ____________ ZIP Code ______

All orders totaling $20 or less must be accompanied by full payment in U S currency or equivalent Make checks payable to the NCTM Shipping and handling charges will be added to all billed orders

An annotated listing of NCTM publications is free on request

THE MATHEMATICAL ASSOCIATION OF AMERICA
1225 Connecticut Avenue, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 51, NO. 2, MARCH 1978