MATHEMATICS MAGAZINE

CONTENTS

VOLUME 45 • JANUARY 1972 • NUMBER 1

M M

MATHEMATICS MAGAZINE

Published by

THE MATHEMATICAL ASSOCIATION OF AMERICA

(INCORPORATED)

A JOURNAL OF COLLEGIATE MATHEMATICS

VOLUME 45

1972

Published at

MENASHA, WIS. AND WASHINGTON, D. C.

MATHEMATICS MAGAZINE

GERHARD N. WOLLAN, *EDITOR*

EDITORIAL CORRESPONDENCE should be sent to the EDITOR, G. N. WOLLAN, Department of Mathematics, Purdue University, Lafayette, Indiana 47907. Articles should be typewritten and triple-spaced on $8\frac{1}{2}$ by 11 paper. The greatest possible care should be taken in preparing the manuscript, and authors should keep a complete copy. Figures should be drawn on separate sheets in India ink and of a suitable size for photographing.

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

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

The MATHEMATICS MAGAZINE is published by the Mathematical Association of America at Washington, D. C., bi-monthly except July–August. Ordinary subscriptions are: 1 year \$4.00. Members of the Mathematical Association of America and of Mu Alpha Theta may subscribe at the special rate of 2 years for \$6.00. Single copies are 80¢.

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

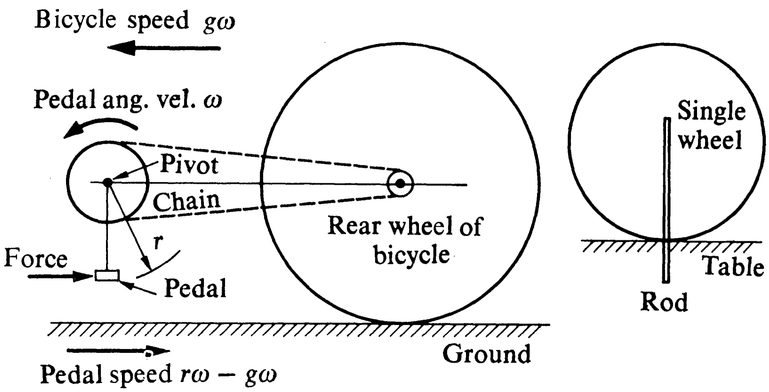
THE BICYCLE PROBLEM

D. E. DAYKIN, University of Reading, England

The following old problem is very instructive:

A bicycle is standing at rest with its pedal cranks vertical. If the bottom pedal is gently pushed backwards, which way will the bicycle move?

The force is trying to rotate the pedal in the same way as it would go if the bicycle were being ridden, so most people answer, "Forward, of course!" However, when they take any bicycle in the street and try it, they find that the bicycle moves backward. Now the fun starts. To try to explain what happens they introduce friction forces, couples, Newton's laws, and even count the teeth on the gear wheels. The following, however, is an easy solution.



As shown in the diagram, let the force cause the angular velocity of the pedal about its pivot to be ω. If g is the gear ratio of the bicycle, then the forward speed of the bicycle relative to the ground is $g\omega$. Moreover if r is the radius of the pedal, because the pedal is at its lowest position, the backward speed of the pedal relative to the ground is $r\omega - g\omega$. Now the pedal is the point of application of the force and so it cannot move in a direction opposite to that of the force. In other words the pedal moves backward if it moves at all and $(r-g)\omega \geqq 0$. Hence either $r > g$ and $\omega \geqq 0$ or $r < g$ and $\omega \leqq 0$ or $r = g$. But we already pointed out that the speed of the bicycle is $g\omega$, and $g > 0$, so the bicycle moves forward if $\omega > 0$ and backward if $\omega < 0$. For conventional bicycles $r < g$ and the bicycle moves backward. Notice that if $r = g$ the bicycle will remain at rest in unstable equilibrium, for the speed of the pedal is zero and so the force can do no work.

To further clarify the situation one can consider a single wheel standing in the vertical plane at the edge of a table. Suppose the wheel has a long rod fixed radially to it so that the rod hangs vertically from the center of the wheel to below the level of the table. If we turn the wheel by pushing on the rod, the point of contact between the wheel and the table will be instantaneously at rest, so the direction of turning depends upon whether we push above or below the level of the table.

SOME CANTOR SETS AND CANTOR FUNCTIONS

R. B. DARST, Purdue University

1. Introduction. Not everybody is aware that a function can be continuous and increase from 0 at 0 to 1 at 1 and have its graph of length 2. We shall construct an example using the Cantor set, some of whose interesting properties we recall. Then, we notice that an apparently trivial change in the definition produces sets and functions with quite different properties.

2. Some properties of continuous nondecreasing functions. Let us begin by recalling that the length, $l(G_f)$, of the graph, G_f, of a continuous nondecreasing map, f, of an interval $[a, b]$ onto an interval $[f(a), f(b)]$ is defined by

$$l(G_f) = \sup \left\{ \sum_{i-1}^{n} [(x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2]^{\frac{1}{2}} : \\ 0 = x_0 < x_1 < \cdots < x_n = 1 \right\}.$$

We also remember that when both α and β are real numbers $[\alpha^2 + \beta^2]^{\frac{1}{2}} \leqq |\alpha| + |\beta|$, and strict inequality obtains when both α and β are nonzero. Consequently,

$$\begin{aligned}
&\sum_{i=1}^{n} [(x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2]^{\frac{1}{2}} \\
&\quad < \sum_{i=1}^{n} [(x_i - x_{i-1}) + (f(x_i) - f(x_{i-1}))] \\
&\quad = (b - a) + (f(b) - f(a)),
\end{aligned}$$

so $l(G_f) \leqq (b-a) + (f(b) - f(a))$. Particular choices of a, b, $f(a)$ and $f(b)$ do not affect the qualitative properties that we wish to consider, so we set $f(a) = a = 0$ and $f(b) = b = 1$. Thus, f is a continuous nondecreasing map of $I = [0, 1]$ onto I.

When f is continuously differentiable, we can use the mean value theorem to show that $l(G_f) < 2$ as follows: either $f(x) = x$ for all $x \in I$ and $l(G_f) = \sqrt{2}$ or there are positive numbers γ and δ and an interval $[\alpha, \beta] \subset (0, 1)$ such that $0 < \gamma \leqq f'(t) \leqq \delta < 1$ for all $t \in [\alpha, \beta]$. Denote by l_1, l_2 and l_3 the length of the part of G_f lying above $[0, \alpha]$, $[\alpha, \beta]$, $[\beta, 1]$. A basic property of length is $l(G_f) = l_1 + l_2 + l_3$; the obvious estimates $l_1 \leqq \alpha + f(\alpha)$, $l_2 \leqq (\beta - \alpha) + (f(\beta) - f(\alpha))$, and $l_3 \leqq (1 - \beta) + (1 - f(\beta))$ only give $l(G_f) \leqq 2$. But, if we can establish $l_2 \leqq A[(\beta - \alpha) + (f(\beta) - f(\alpha))]$, where $0 < A < 1$, then we have $l(G_f) < 2$. Thus we notice that $\xi \in [\gamma, \delta]$, implies

$$(1 + \xi) = [1 + \xi]^{\frac{1}{2}}[1 + \xi]^{\frac{1}{2}} \geqq [1 + \gamma]^{\frac{1}{2}}[1 + \xi^2]^{\frac{1}{2}}.$$

$$[1 + \xi^2]^{\frac{1}{2}} \leqq A(1 + \xi), \text{ where } 0 < A = [1 + \gamma]^{-\frac{1}{2}} < 1.$$

Hence when $\alpha \leqq u < v \leqq \beta$, there is a point r in (u,v) with $f(v)-v(u) = f'(r)(v-u)$, and $f'(r) \in \gamma[,\delta]$. Hence $[(v-u)^2 + (f(v)-f(u))^2]^{\frac{1}{2}} = [1 + (f'(r))^2]^{\frac{1}{2}}(v-u)$

$$\leqq A[1+f'(r)](v-u)$$

$$= A[(v-u) + (f(v) - f(u))].$$

Applying this inequality to the pairs of consecutive elements of any partition $\alpha = t_0 < t_1 < \cdots < t_n = \beta$ of $[\alpha,\beta]$ and summing yields $l_2 \leqq A[(\beta-\alpha) + (f(\beta) - f(\alpha))]$, which permits us to claim $l(G_f) < 2$.

Nevertheless, if $f_n(x) = x^n$ for $0 \leqq x \leqq 1$, then for each p with $0 < p < 1$,

(i) $l(G_{f_n}) \geqq (1-p) + 1 - f_n(p)$ and

(ii) $\lim_n f_n\ (p) = 0$;

so

(iii) $\lim_n l(G_{f_n}) = 2$.

3. The standard Cantor set. Let us recall that the standard Cantor set, C, can be obtained by deleting a sequence $\{(a_i,b_i)\}_{i=1}^{\infty}$ of pairwise disjoint segments from the interior of the interval $I = [0,1]$ as follows. Firstly, we delete a segment which comprises the open middle third of I, i.e., we delete the segment (1/3, 2/3) and leave two intervals $I_1 = [0,1/3]$ and $I_2 = [2/3,1]$, each of length 1/3. The second step consists in deleting the open middle thirds, (1/9, 2/9) and (7/9, 8/9), from each of I_1 and I_2 to leave 2^2 intervals, denoted by $I_{11}, I_{12}, I_{21}, I_{22}$, each of length $(1/3)^2$. Mathematical induction permits us to continue this process as follows.

Suppose that we have completed k steps in the process, where k denotes a positive integer, and 2^k intervals, each of length $(1/3)^k$, remain. The $(k+1)$-st step consists in deleting the open middle third of each remaining interval to leave $2^{(k+1)}$ intervals, each of length $(1/3)^{(k+1)}$. We continue the process and define C to be the set of points in I which are removed at *no* step in the process.

The points 0, 1, 1/3, 2/3, 1/9, 2/9, 7/9, 8/9, 1/27, 2/27, $\cdots$ all belong to C, and there are many more points in C. Indeed, specifying a point in C is equivalent to stating whether it is in a first third or a last third at each step in the construction. For instance, if p is the point which is in a first third at each odd step and in a last third at each even step, then $p = 2/9 + 2/(9)^2 + \cdots = 2/9[1 + (1/9) + (1/9)^2 + \cdots] = 1/4$. Suppose $p_1, \cdots, p_k, \cdots$ is an enumeration of points in C. Then we can specify a point x in C which is different from each p_k as follows: x is in a first third at the kth step if p_k is in a last third at the kth step and x is in a last third at the kth step otherwise.

4. The length of the graph of the standard Cantor function is 2. We shall define the standard Cantor function, ϕ, and show $l(G_\phi) = 2$. To begin a definition of ϕ, define $\phi(0) = 0$ and $\phi(1) = 1$; ϕ is thus specified at the endpoints of I. Now recall that the first step in our construction of C amounts to splitting the interval into three pieces: the middle segment is removed and two intervals, I_1 and I_2, of equal length remain. Define $\phi(x) = 1/2$ if $1/3 \leqq x \leqq 2/3$. Thus, $\phi(x)$ is the average of the values of ϕ at the endpoints of I when x is in the middle 1/3 of I and ϕ is specified at the endpoints of I_1 and I_2. We iterate this procedure and thereby define ϕ on each deleted segment and at every endpoint of every interval which is retained at some step in our construction of C. Suppose ϕ is not yet defined at x. Then for each positive integer k, x is in the interior of exactly one, say $[\alpha_k, \beta_k]$, of the 2^k intervals of length $(1/3)^k$ which remain after the kth step in our construction of C. Moreover, $\beta_k = \alpha_k + (1/3)^k$, $\phi(\beta_k) = \phi(\alpha_k) + (1/2)^k$, $\alpha_k \leqq \alpha_{k+1} < \beta_{k+1} \leqq \beta_k$, and $\phi(\alpha_k) \leqq \phi(\alpha_{k+1}) < \phi(\beta_{k+1}) \leqq \phi(\beta_k)$, so $\phi(x)$ is defined by

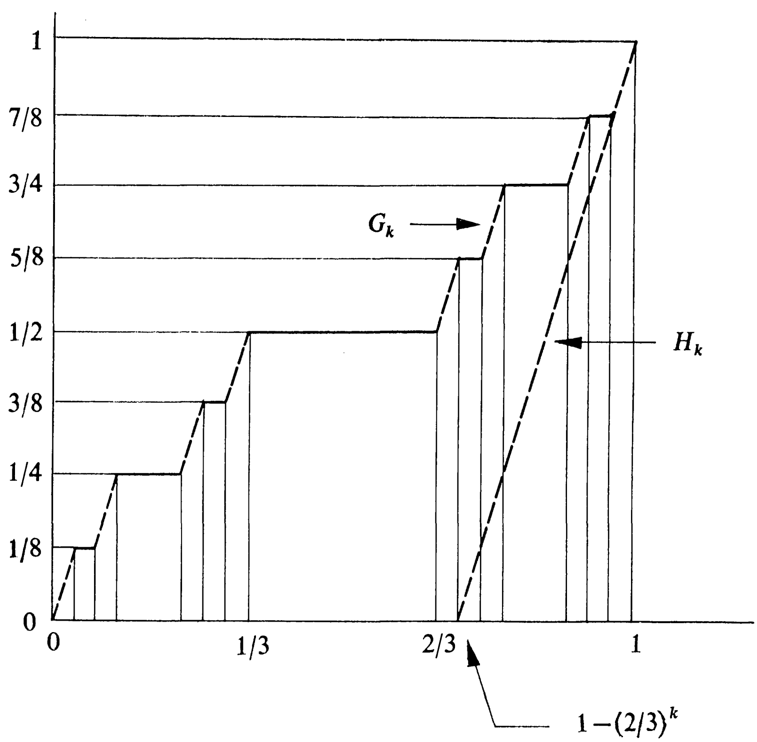
$$\lim_{k\to\infty} \phi(\alpha_k) = \phi(x) = \lim_{k\to\infty} \phi(\beta_k).$$

Furthermore, ϕ is a continuous nondecreasing map of I onto I. Now that ϕ has been defined, we can show $l(G_\phi) = 2$ by showing that it is not less than 2. To this end, we pick a positive integer k and compute the approximation

$$L_k = \sum_{i=2}^{2(k+1)} [(x_i - x_{i-1})^2 + (\phi(x_i) - \phi(x_{i-1}))^2]^{\frac{1}{2}}$$

to the length of G_ϕ obtained from the sequence $0 = x_1 < x_2 < \cdots < x_{2(k+1)} = 1$, where the points x_i are the endpoints of the 2^k intervals which remain after the kth step in our construction of C. Recall that L_k is the length of a polygonal approximation, P_k, to ϕ. Since $x_{2j} - x_{2j-1} = (1/3)^k$, $\phi(x_{2j}) - \phi(x_{2j-1}) = (1/2)^k$, and

$(k = 3)$



$\phi(x_{2j+1}) - \phi(x_{2j}) = 0$, the graph, G_k, of P_k is comprised of 2^k horizontal intervals, with total length $1 - (2/3)^k$, plus 2^k congruent intervals with positive slope and total length $[(2/3)^{2k} + 1]^{\frac{1}{2}}$. By direct computation or merely noticing that the length, L_k, of G_k is the length of the graph, H_k, of the function which takes the value zero on $[0, 1 - (2/3)^k]$, the value one at one, and is linear on $[1 - (2/3)^k, 1]$, we obtain

$$L_k = 1 - (2/3)^k + [(2/3)^{2k} + 1]^{\frac{1}{2}} \leqq l(G\phi).$$

Since $\lim_{k \to \infty} L_k = 2$, $l(G_\phi) = 2$.

Next, we make an observation about the derivative of ϕ. Firstly, notice that if x is in a segment which is deleted during our construction of C, then ϕ is constant near x, so $\phi'(x) = 0$. Otherwise, x is in C and for each positive integer k, x is in an interval $[x_{2j-1}, x_{2j}]$ of the form considered in the proof that $l(G_\phi) = 2$ and

$$\phi(x_{2j}) - \phi(x_{2j-1}) = (3/2)^k(x_{2j} - x_{2j-1}).$$

Now recall the elementary inequality:

$$(*) \qquad \max\left\{\frac{\alpha}{\beta}, \frac{\gamma}{\delta}\right\} \geqq \frac{\alpha + \gamma}{\beta + \delta}, \quad \text{when } \alpha \geqq 0, \beta > 0, \gamma \geqq 0, \delta > 0,$$

which permits us to assert that

$$\max\left\{\frac{\phi(x_{2j}) - \phi(x)}{x_{2j} - x}, \frac{\phi(x) - \phi(x_{2j-1})}{x - x_{2j-1}}\right\} \geqq (3/2)^k, \text{ when } x_{2j-1} < x < x_{2j}.$$

Therefore ϕ is differentiable at *no* point of C.

The standard Cantor set is an instance of a type of Cantor set which we consider in the next section.

5. Some basic properties of a family of Cantor sets. To describe the type of Cantor set which we consider, suppose we choose a number λ satisfying $0 < \lambda \leqq 1$. Then we repeat our construction of C with the following modification. At the kth step, instead of taking out a segment of length $1/3^k$ from the center of each of $2^{(k-1)}$ intervals of equal length, we take out a segment of length $\lambda/3^k$. As in the case $\lambda = 1$, when we obtain C and ϕ, for each λ satisfying $0 < \lambda \leqq 1$ we obtain a Cantor set C_λ and a corresponding Cantor function ϕ_λ. Notice that the sum of the lengths of the segments taken out in the first k steps in the construction of C_λ is

$$\lambda/3 + 2(\lambda/3^2) + \cdots + 2^{k-1}(\lambda/3^k) = \lambda[1 - (2/3)^k].$$

It will be convenient to denote $\lambda[1 - (2/3)^k]$ by λ_k.

Let us fix $\lambda < 1$ and compute $l(G_{\phi_\lambda})$. To this end, choose a positive integer k and, as before, consider the points $0 = x_1 < x_2 < \cdots < x_{2(k+1)} = 1$, where x_i is an endpoint of one of the 2^k intervals which remain after k steps in the construction of C_λ. Again $\phi_\lambda(x_{2j}) - \phi_\lambda(x_{2j-1}) = 2^{-k}$ and $\phi_\lambda(x_{2j+1}) = \phi_\lambda(x_{2j})$, where $j = 1, 2, \cdots, 2^k$ so there are 2^k horizontal intervals, with total length λ_k, and 2^k congruent intervals with positive slope and total length $[(1 - \lambda_k)^2 + 1]^{\frac{1}{2}}$ in the consequent polygonal approximation, $P_k(\lambda)$, to ϕ_λ. The corresponding length

$$L_k(\lambda) = \sum_{i=2}^{2(k+1)} [(x_i - x_{i-1})^2 + (\phi_\lambda(x_i) - \phi_\lambda(x_{i-1})^2)]^{\frac{1}{2}}$$
$$= \lambda_k + [(1-\lambda_k)^2 + 1]^{\frac{1}{2}}.$$

We record

$$\lim_{k\to\infty} L_k(\lambda) = \lambda + [(1-\lambda)^2 + 1]^{\frac{1}{2}}$$

and remember that this number is certainly not greater than $l(G_{\phi_\lambda})$. To establish the opposite inequality, notice that

$$\frac{P_k(\lambda)(y) - P_k(\lambda)(x)}{y-x} \leqq \frac{1}{1-\lambda_k} < \frac{1}{1-\lambda}$$

which implies

$$(**) \qquad \frac{\phi_\lambda(y) - \phi_\lambda(x)}{y-x} \leqq \frac{1}{1-\lambda}, \text{ where } y \neq x,$$

because $P_k(\lambda)$ converges to ϕ_λ (pointwise convergence is sufficient although the convergence is uniform). Thus, whenever we add points to the sequence $\{x_i\}_{i=1}^{2k+1}$, the consequent approximation to $l(G_{\phi_\lambda})$ is at most

$$\lambda_k + (1-\lambda_k)\left[1 + \frac{1}{(1-\lambda)^2}\right]^2;$$

using these approximations, an elementary argument with limits establishes

$$l(G_{\phi_\lambda}) = \lambda + [(1-\lambda)^2 + 1]^{\frac{1}{2}}.$$

(Should the reader experience difficulty in verifying any of the preceding assertions he might find chapter 6 of [3] to be a helpful reference.)

Now we make a preliminary observation about the derivative of ϕ_λ. From (**) it follows if ϕ_λ is differentiable at a point x of C_λ, then $\phi'_\lambda(x) \leqq 1/(1-\lambda)$. Of course, $\phi'(x) = 0$ if x is a point of $I - C_\lambda$.

To continue our consideration of the derivative of ϕ_λ, we assume hereafter some familiarity with measure theory: see for example, Section 5.1–4 of [2].

Any continuous nondecreasing function on I (in particular ϕ_λ) is differentiable at almost all points of I. Moreover, a measure μ_λ is defined by specifying that its values on sets $[0,x)$ be given by

$$\mu_\lambda([0,x)) = \phi_\lambda(x).$$

Since $\lambda < 1$, (**) implies that μ_λ is absolutely continuous. Hence for every measurable set, E,

$$\mu_\lambda(E) = \int_E \phi'_\lambda$$

In particular, since $\phi_\lambda'(x) = 0$ when $x \in (I - C_\lambda)$,

$$\mu_\lambda(I - C_\lambda) = 0,$$

and when E is a measurable subset of I,

$$\mu_\lambda(E) = \int_E \phi_\lambda' \leqq m(E)/(1-\lambda),$$

where $m(E)$ denotes the Lebesgue measure of E. Now suppose that E is a measurable subset of C_λ. Then $m(C_\lambda) = 1 - \lambda$ and

$$\begin{aligned} 1 = \mu_\lambda(I) &= \mu_\lambda(C_\lambda) + \mu_\lambda(I - C_\lambda) \\ &= \mu_\lambda(C_\lambda) \\ &= \mu_\lambda(E) + \mu_\lambda(C_\lambda - E) \\ &\leqq \int_E \phi_\lambda' + m(C_\lambda - E)/(1-\lambda) \\ &= \int_E \phi_\lambda' + [m(C_\lambda) - m(E)]/(1-\lambda) \\ &= \int_E \phi_\lambda' + 1 - m(E)/(1-\lambda). \end{aligned}$$

Thus

$$m(E)/(1-\lambda) \leqq \int_E \phi_\lambda'$$

and we conclude that

$$\phi_\lambda'(x) = 1/(1-\lambda)$$

at almost every point of C.

It is easy to check that ϕ_λ is differentiable at no end point of any of the segments which are removed when C_λ is constructed. We leave the reader with the problem of characterizing the set of points in I at which ϕ_λ is not differentiable and close by suggesting [**1**] and [**4**] as leads to the extensive literature of Cantor sets.

References

1. J. P. Kahane and R. Salem, Ensembles parfaits et séries trigonométriques, Hermann, Paris, 1963.

2. H. L. Royden, Real Analysis, 2nd ed., Macmillan, New York, 1968.

3. W. Rudin, Principles of Mathematical Analysis, 2nd ed., McGraw-Hill, New York, 1964.

4. A. Zygmund, Trigonometric Series, vol. 1, 3rd ed., Cambridge University Press, London, 1968.

EARLY SUNDIALS AND THE DISCOVERY OF THE CONIC SECTIONS

W. W. DOLAN, Linfield College, McMinnville, Oregon

The earliest knowledge of the plane curves we call conic sections is obscure, but is usually attributed to the Greek mathematician Menaechmus in the fourth century B. C. [1]. One may speculate about the circumstances that might have led Menaechmus to discover the curves, which seem to have been associated with the cone and with each other from the beginning. Conceivably he could have developed the idea from observing a volcano, or an anthill scuffed off by his sandal, or some artifact like a sharpened stake. However, there is much to recommend the conjecture of O. Neugebauer [2] who felt strongly that the sundial was the most probable basis for discovery of the conics. The purpose of the present comment is to examine this possibility from several points of view.

Neugebauer, to be sure, was not the first to consider this relationship. Writing a sundial treatise in 1682, Philippe de la Hire [3] said (as I translate him), "I could easily show that we owe to sundials the discovery of these wonderful curves which we find so useful in all parts of mathematics, ··· but this would take me too far from my subject." And in one of the most recent sundial books, F. Cousins [4] writes briefly of the connection with conics, but without reference to the historical question.

Sundials began to appear in the Greek world several centuries before the Christian era. The earlier ones consisted of a gnomon in the form of a vertical post or peg set in a flat surface, upon which the shadow of the gnomon served to indicate the time. Dials on a vertical wall, with horizontal gnomon, came along later and had similar characteristics. There were other types, like the hollow hemisphere of Berossos, not related to the present inquiry. None of these had the pervasive feature of the modern dial, first popular in about the fifteenth century: the slanted gnomon with its edge

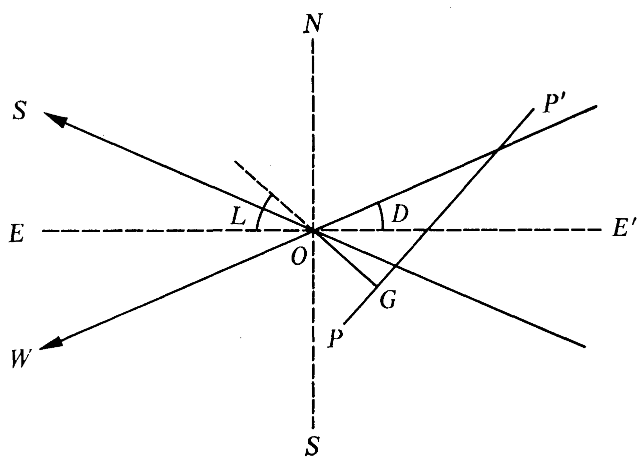


FIG. 1. Axial section of diurnal cone formed by sun's rays with vertex at tip of gnomon, intersected by plane of horizontal sundial. *NS*, earth's axis; *EE'*, equatorial plane; *S*, *W*, directions of summer and winter sun; *PP'*, dial plane; *OG*, gnomon; *L*, latitude; *D*, declination of sun.

parallel to the earth's axis. The principal advantage of the modern form is that the hour lines are straight and radiate from a central point. The shadow of the gnomon's edge lies along these lines without essential variation as the sun's declination changes through the year.

In the ancient dials with vertical gnomon, the direction of the shadow at any given time of day varied with the seasons. Thus it was the position of the tip of the shadow that was essential to the determination of the hour. The shadow's tip traced a curve on the dial plane as the sun moved, a curve which changed from summer to winter. It is easy for us to see the implication for the question of conics (Figure 1). The sun traces a circular path in the sky in its daily motion. The tip of the gnomon is the vertex of a cone with the sun's rays as elements, and since the dial plane cuts the cone, the shadow path is a conic section. If Menaechmus or someone else marked this path with a series of dots on a given day, he would "discover" a hyperbola.

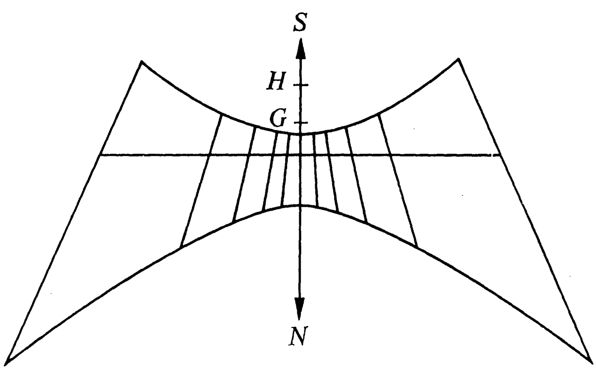


FIG. 2. Reconstruction of *pelekinon*, ancient horizontal sundial face, calculated for latitude of Rome. *G*, base of gnomon; *GH*, height of gnomon. Upper and lower curves, shadow paths at summer and winter solstices; horizontal line, shadow path at equinoxes. Oblique lines denote seasonal hours, noon at center.

A kind of standard pattern developed for marking the dial plane, illustrated in Figure 2 as it might appear on a horizontal sundial. The upper curved boundary, toward the south, is the diurnal shadow path on the longest day of summer when the sun is high in the sky. Similarly the lower (northern) boundary is the path on the date of the winter solstice, when the day is actually shorter but the low sun casts long shadows. Between the two is the horizontal straight line that marks the shadow path at the spring and fall equinoxes. The vertical and oblique lines denote the hours, with noon at the center. These were the temporary or seasonal hours, which varied in duration through the year because there were always exactly twelve of them from sunrise to sunset. The hour lines should not be quite straight, but were often drawn so. Only five morning and afternoon hours were shown, because the sixth would be infinitely distant.

This configuration has had some interesting names. The Greeks called it *pelekinon*, meaning double-bitted axe. It was also called the spider-web, an especially appropriate description when the boundary curves were drawn in broken line segments—a form, incidentally, suggesting that the maker did not recognize them as curves having any inherent interest of their own. The German historian Diels [5] liked the term Schwalbeschwanz or swallow-tail, and it would surprise me if someone had not called it a butterfly.

Diels showed a photograph of such a dial preserved from the vicinity of Rome, without giving it a date. He considered that the layout might have been either empirical or calculated. The former method would have been straightforward enough, requiring only time and patience to trace the shadow paths at the principal dates, and once established it could be copied for new dials at the same latitude. On the other hand, methods for trigonometric calculation were first formalized at the time of Ptolemy in the second century A. D. Delambre [6] recalculated several surviving ancient dials using Ptolemy's methods, and found his results to compare quite satisfactorily with the actual layouts. He noted that the methods might well have been known earlier, perhaps from the time of Hipparchus some 200 years before Ptolemy.

Figure 2 was calculated anew for the present purpose, using the appropriate latitude for comparison with Diels' example. It corresponds rather closely, except that the lower corners are slightly more elongated than in Diels' photograph. The question arises as to whether the ancient designers could have placed their hour lines so accurately if they had been restricted to empirical methods. Some independent determination would be needed of the moments when the successive hours had arrived, allowing for hours which in summer would equal 75 modern minutes and in winter only 46. Perhaps a clepsydra or water-clock could have been calibrated for the varying lengths of the day with fairly good results. It seems more likely however, that the *pelekinon* dials appeared after the trigonometric method became known. Drecker [7] has also treated this question in detail.

In his study of the sundial and conics, Neugebauer was much concerned with another circumstance that needs explanation. In the earliest writings on the conics, it was customary to consider the cutting plane at right angles to an element of the cone, and the occurrence of the three types of curve then depended on the vertex angle of the cone itself, whether acute, right or obtuse. Thus the ellipse, for example, would be referred to as the right section of an acute-angled cone. Our present conceit of using a single arbitrary cone and varying the angle of intersection of the plane did not appear until later. Neugebauer therefore felt that any theory of conics as having been discovered through sundials would have to be rooted in an arrangement allowing for only a right-angled section. He proposed a movable dial, hinged at one edge of the base, to be adjusted every few days so that the gnomon would point directly at the noon sun. Thus the gnomon would coincide with an element of the cone, and the dial plane would be a right section.

Neugebauer recognized two difficulties with this hypothesis. The first was that no description or archaeological find has come down to us suggesting the existence of this type of movable oblique dial plane. The second was the fact that, since the sun is always at least $66\frac{1}{2}°$ from the polar axis, the cone is always very obtuse and can yield only the hyperbola as a right section. As to the latter objection, he conjectured that once the idea of the conic section had occurred, the imagination of the discoverer would move readily to the implications for the right and acute-angled cones, thus developing the parabola and ellipse.

It seems to me that there is another strong argument against the movable dial besides the lack of objective evidence. Dials that do survive from the early period are graven on stone blocks. They would be massive, awkward to move, and what motive would there be for making them in this form? No advantage in timekeeping is evident as compared with a stationary dial. Moreover, the observer's imagination could just as well work toward changing the angle of intersection of the plane as toward holding the intersection constant and varying the vertex angle. The idea of the right section could then have become standardized in the geometry for some quite different reason, still obscure.

Could the ellipse and parabola ever appear directly as shadow paths on a sundial plane? A moment's thought indicates that the ellipse, the only bounded member of the family, requires the sun to be above the dial plane for its entire daily circle. For a horizontal dial this can occur only north of the arctic circle for a part of the summer, or correspondingly near the south pole. The parabola appears as a transition case when the sun barely grazes the horizon. Clearly no fourth-century Greeks were setting up sundials in such a location. They did understand the effects of changing latitude, however, and may well have made the journey in thought.

On the other hand, a south-facing vertical dial will display an ellipse, or a part of one, at certain seasons between the Tropic of Cancer and the equator. Here the sun dips below the horizon, to be sure, but does not pass behind the dial plane in the winter months. This region, too, was outside the normal range of activity of the Mediterranean Greeks. Eratosthenes, however, in his famous earth-measuring experiment, went as far as Syene on the upper Nile, near the present site of the Aswan Dam and just at the Tropic. Here he noted that a vertical post cast no shadow at the date of the summer solstice. Some have supposed that his post was the gnomon of a sundial. Perhaps he went still farther south, in imagination if not in person. There is no implication that Eratosthenes discovered the conics in this manner, for they were already known before his time.

The relationship between the type of conic and the latitude of the dial may be expressed simply from the modern viewpoint in terms of eccentricity. (I have not found this question treated in any work on sundials, but there are many old ones that I have not been able to examine.) Considering the case of the horizontal dial, one sees in Figure 1 that the angle between the cone element and the equatorial plane is the sun's declination, while the angle between the dial plane and the equatorial is the colatitude. It is apparent [4] that the section cut from the cone by the dial plane will be ellipse, parabola or hyperbola according as the colatitude is less than, equal to or greater than the declination.

More precisely, one may calculate the eccentricity of the corresponding curve for any combination of latitude L and declination D, with the rather elegant result that $e = \cos L/ \sin D$. For a given declination the variation is then continuous from a circle at the pole, through ellipse, parabola and hyperbola to a maximum eccentricity of csc D at the equator. For a south vertical dial the eccentricity of the corresponding shadow curve is $e = \sin L / \sin D$. Menaechmus and his contemporaries had no knowledge of the eccentricity concept nor of the analytic processes involved in developing this formula, but they might indeed have seen the descriptive implications of the latitude variation.

A dial traced on an appropriate oblique plane could show all three conics at any latitude, but apparently such dials are not known to have existed until much later [**6**]. Thus the whole notion of the discovery of the conics through sundials is full of uncertainties, though it surely remains at least as plausible as any other hypothesis. If the development did occur in this manner we are led, as in other cases, to a profound respect for the ingenuity and intuition of the Greeks. For while the sundial itself is a solid, palpable object, the daily circle of the sun is not, and even less so is the cone of which its rays are the elements. Without other background the leap of abstraction here would be most remarkable. Knowing some of the other astonishing accomplishments of the Greek geometers, however, we should not place this one beyond their capacity.

References

1. T. L. Heath, History of Greek Mathematics, vol. 1, Oxford, 1921, p. 255.
2. O. Neugebauer, The astronomical origin of the theory of conic sections, Proc. Amer. Philos. Soc., 92 (1948) 136.
3. Philippe de la Hire, La Gnomonique, Paris, 1682.
4. F. W. Cousins, Sundials, John Baker, London, 1969 (also Pica Press, New York, 1970).
5. H. Diels, Antike Technik, Leipzig, 1919, p. 179.
6. J. B. J. Delambre, Histoire de l'Astronomie Ancienne, Paris, 1817; repr. by Johnson Reprint, New York, 1965, vol. 2, p. 462.
7. J. Drecker, Die Theorie der Sonnenuhren, Berlin and Leipzig, 1925, p. 2.

FERMAT'S EQUATION $A^p+B^p=C^p$ FOR MATRICES OF INTEGERS

J. L. BRENNER, University of Arizona and J. DE PILLIS, University of California, Riverside

1. Introduction. For more than two hundred years the diophantine equation (1) $X^r+Y^r=Z^r$ has interested number theorists. If X, Y, Z are taken to be rational integers, we call (1) the classical problem; if $XYZ\neq 0$ we call the solution (X, Y, Z) nontrivial. Nontrivial solutions are known for $r = 2$, $[a^2-b^2, 2ab, a^2+b^2]$, but not for any other value of r. Nonexistence proofs have been given as follows. For $0<XYZ<10^{45}$ there are no solutions if $(r, XYZ)=1$ (common factor); also for $(r, XYZ)=1$ whenever $r<2^{31}$ and $0<XYZ$ [**1**]. For $2<r<25{,}001$ there are no nontrivial solutions to the classical problem at all [**3**].

that $e = \cos L / \sin D$. For a given declination the variation is then continuous from a circle at the pole, through ellipse, parabola and hyperbola to a maximum eccentricity of $\csc D$ at the equator. For a south vertical dial the eccentricity of the corresponding shadow curve is $e = \sin L / \sin D$. Menaechmus and his contemporaries had no knowledge of the eccentricity concept nor of the analytic processes involved in developing this formula, but they might indeed have seen the descriptive implications of the latitude variation.

A dial traced on an appropriate oblique plane could show all three conics at any latitude, but apparently such dials are not known to have existed until much later [**6**]. Thus the whole notion of the discovery of the conics through sundials is full of uncertainties, though it surely remains at least as plausible as any other hypothesis. If the development did occur in this manner we are led, as in other cases, to a profound respect for the ingenuity and intuition of the Greeks. For while the sundial itself is a solid, palpable object, the daily circle of the sun is not, and even less so is the cone of which its rays are the elements. Without other background the leap of abstraction here would be most remarkable. Knowing some of the other astonishing accomplishments of the Greek geometers, however, we should not place this one beyond their capacity.

References

1. T. L. Heath, History of Greek Mathematics, vol. 1, Oxford, 1921, p. 255.
2. O. Neugebauer, The astronomical origin of the theory of conic sections, Proc. Amer. Philos. Soc., 92 (1948) 136.
3. Philippe de la Hire, La Gnomonique, Paris, 1682.
4. F. W. Cousins, Sundials, John Baker, London, 1969 (also Pica Press, New York, 1970).
5. H. Diels, Antike Technik, Leipzig, 1919, p. 179.
6. J. B. J. Delambre, Histoire de l'Astronomie Ancienne, Paris, 1817; repr. by Johnson Reprint, New York, 1965, vol. 2, p. 462.
7. J. Drecker, Die Theorie der Sonnenuhren, Berlin and Leipzig, 1925, p. 2.

FERMAT'S EQUATION $A^p + B^p = C^p$ FOR MATRICES OF INTEGERS

J. L. BRENNER, University of Arizona and J. DE PILLIS,
University of California, Riverside

1. Introduction. For more than two hundred years the diophantine equation (1) $X^r + Y^r = Z^r$ has interested number theorists. If X, Y, Z are taken to be rational integers, we call (1) the classical problem; if $XYZ \neq 0$ we call the solution (X, Y, Z) nontrivial. Nontrivial solutions are known for $r = 2$, $[a^2 - b^2, 2ab, a^2 + b^2]$, but not for any other value of r. Nonexistence proofs have been given as follows. For $0 < XYZ < 10^{45}$ there are no solutions if $(r, XYZ) = 1$ (common factor); also for $(r, XYZ) = 1$ whenever $r < 2^{31}$ and $0 < XYZ$ [1]. For $2 < r < 25{,}001$ there are no nontrivial solutions to the classical problem at all [3].

Kummer generalized the last nonexistence theorem to solutions over a cyclotomic field. This is a field obtained by adjoining a root of unity to the rational field. For example Kummer showed that for $2<r<11$, nontrivial gaussian integers

$$(x_1 + x_2 i,\ y_1 + y_2 i,\ z_1 + z_2 i) \quad [x_1,\ x_2,\ y_1,\ y_2,\ z_1,\ z_2 \text{ rational integers}]$$

do not exist satisfying (1).

Interest attaches to (1) in case X, Y, Z are integers from more complicated algebraic fields. We have not found any report in the literature concerning this extension of the "Fermat problem." Sophisticated readers will not be astonished to learn from the present paper that this extension of the problem can be systematized by the use of matrix theory. We show that there is a certain connection between this extension and the equation (1) for nonsingular *matrices of rational integers*. Without real loss of generality, attention can be restricted to the case of prime exponent r. We assume $XY=YX$. Although this is a genuine restriction, it is a natural one, i.e., it holds in all the above cases as well.

In what follows, A, B, and C will be nonsingular elements of $M_n(Z)$, the ring of $n\times n$ matrices with (rational) integer entries. The symbols a, b, and c will denote algebraic integers (roots of monic polynomials having integer coefficients). We shall be concerned with the question: When does the existence of a nonsingular solution triple A, B, and $C\in M_n(Z)$ to the Fermat matrix equation $A^p+B^p=C^p$, $p>2$ ensure the existence of a nontrivial solution triple a, b, and c of algebraic integers to the corresponding Fermat equation $a^p+b^p=c^p$, and conversely? In case $p=n$, the situation is trivial. In fact for any triple of algebraic integers a, b, and $c=a+b$, set

$$A = \begin{bmatrix} 0 & a & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \\ \vdots & & & \ddots & \\ 0 & & & & 1 \\ 1 & \cdots & & & 0 \end{bmatrix}. \tag{2.1}$$

Define B and C similarly. Observe that

$$A^n = \begin{bmatrix} a & & & 0 \\ & a & & \\ & & \ddots & \\ 0 & & & a \end{bmatrix},$$

so that $A^n+B^n=C^n$ follows.

Also, if p divides n (say $n=pq$), then consider the $p\times p$ matrices of the form (2.1); denote them by A_p, B_p and C_p. It follows that the q-fold direct sums

$$A = \underbrace{A_p\oplus \cdots \oplus A_p}_{q} \quad \text{(similarly for } B_p \text{ and } C_p\text{)},$$

have the property that $A^p+B^p=C^p$. (By a direct sum we mean the $pq\times pq$ matrix partitioned into blocks with A_p appearing q times along the generalized diagonal; other blocks zero.) Thus, for suitable p it is easy to find A, B and $C\in M_n(Z)$ satisfying the equation above. In Section 2, we show that if A and $B\in M_n(Z)$ commute, then algebraic integers a, b, and c, of degree no more than n, exist such that $A^p+B^p=C^p \Rightarrow a^p+b^p=c^p$ (Theorem I). In Section 3, we show a partial converse, viz., if a triple a, b, and c, of algebraic integers exists such that $a^p+b^p=c^p$, and if a, b and c are of degree 2, and all belong to the extension field $Q(a)$, Q = rationals, then A, B and $C\in M_2(Z)$ can be found such that $A^p+B^p=C^p$ (Theorem II).

2. $A^p+B^p=C^p$. We consider the case of when a matrix solution triple A, B, C forces an algebraic integer solution triple a, b, c.

THEOREM I. *Suppose A and B commute in $M_n(Z)$. Suppose, moreover, that there exists C in $M_n(Z)$ such that $A^p+B^p=C^p$. Then if A, B, C are nonsingular, there exists a nontrivial triple of algebraic integers a, b and c, such that $a^p+b^p=c^p$. Moreover, the degrees of a, b, and c are $\leqq n$.*

Proof. In assuming the commutativity of A and B, we guarantee the existence of a common eigenvector x for A and B (cf. [2], p. 77, Section 4.21.1). The entries of x, considered as an n-column vector, are algebraic integers with corresponding eigenvalues a and b, which are algebraic integers. That is,

$$A^px = a^px \quad \text{and} \quad B^px = b^px.$$

Now to say $A^p+B^p=C^p$, is to assure the fact that x is also an eigenvector for C^p, with corresponding (algebraic integer) eigenvalue c'. By defining the algebraic integer $c=(c')^{1/p}$ we have $a^p+b^p=c^p$.

Each of the algebraic integers a, b, and c is an eigenvalue of a matrix in $M_n(Z)$, each degree is no more than n. This proves the theorem.

3. $a^p+b^p=c^p$.

THEOREM II. *Given nonzero algebraic integers a, b, c satisfying the equation $a^p+b^p=c^p$, $p\geqq 2$. Suppose the elements b and c belong to the extension field $Q(a)$, where a is of degree two. Then there exists a triple of (nonsingular) matrices A, B and C in $M_2(Z)$ such that $A^p+B^p=C^p$.*

Proof. Since b and c are assumed to be of the form $r+s\cdot a\in Q(a)$, it follows that b and c are also algebraic integers of degree two. Supposing a satisfies the equation

$$(*) \quad x^2+r_1x+r_2=0, \text{ for integers } r_1 \text{ and } r_2,$$

it follows that there exists an $S\in M_2(C)$ such that

$$S\begin{bmatrix} a & 0 \\ 0 & \bar{a} \end{bmatrix} S^{-1} = \begin{bmatrix} 0 & 1 \\ -r_2 & -r_1 \end{bmatrix},$$

the companion matrix to the polynomial (*), where $\bar{a}$ denotes the conjugate (root) of a. Thus, for $b=t_1+t_2a$ and its conjugate $\bar{b}=t_1+t_2\bar{a}$, we have

$$S\begin{bmatrix} b & 0 \\ 0 & \bar{b} \end{bmatrix} S^{-1} = S\left(t_1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t_2\begin{bmatrix} a & 0 \\ 0 & \bar{a} \end{bmatrix}\right) S^{-1}$$

$$= t_1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t_2\begin{bmatrix} 0 & 1 \\ -r_2 & -r_1 \end{bmatrix}.$$

Since $a^p + b^p = c^p$, it follows that $\bar{a}^p + \bar{b}^p = \bar{c}^p$, so that

$$\left(S\begin{bmatrix} a & 0 \\ 0 & \bar{a} \end{bmatrix} S^{-1}\right)^p + \left(S\begin{bmatrix} b & 0 \\ 0 & \bar{b} \end{bmatrix} S^{-1}\right)^p = \left(S\begin{bmatrix} c & 0 \\ 0 & \bar{c} \end{bmatrix} S^{-1}\right)^p.$$

The matrices being raised to pth power all belong to $M_2(Z)$; see the penultimate matrix equation above. The proof is done.

An application of the above theorem is the following equation, which was pointed out to us by Professor and Mrs. D. H. Lehmer, valid for all primes p exceeding 3. If $(p, 6)$ equals 1, then

$$\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}^p + \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}^p = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}^p.$$

Added in proof. B. Divis has kindly called our attention to work of D. K. Faddeev, suggesting that there may be at most finitely many solutions to $A^p + B^p = C^p$ for $n = 2$, $p = 3,4$ ($AB = BA$, A, B, C invertible). See Dokl. Akad. Nauk USSR 134 (776–777); 136 (296–298).

Partial support for this paper was supplied under NSF GP 9483 and GP 8758. MOS classification: 10B15, 10B30.

References

1. J. Brillhart, Proc. of the Second Atlas Symposium, 1970.

2. M. Marcus and H. Mintz, A Survey of Matrix Theory and Matrix Inequalities, Allyn & Bacon, Boston, 1964.

3. R. Selfridge and H. Pollack, Notices, Amer. Math. Soc., Abstract #608-138 (1967) 97.

4. V. V. Kobelev, A proof of Fermat's last theorem for all prime exponents less than 5500, Dokl. Akad. Nauk SSSR, MR 41, 3363, (1970) 767–768.

5. I. A. Barnett and H. M. Weitkamp, The equation $X^n + Y^n + Z^n = 0$ in rational binary matrices, An. Sti. Univ. "Al. I. Cuza" Iasi Sect. I a Mat., 7 (1961) 1–64.

WHEN WILL THE NEXT RECORD RAINFALL OCCUR?

DONALD R. BARR, Naval Postgraduate School, Monterey, California

1. Introduction. We consider the occurrences of extremes in a sequence of past observations and their use in making simple inferences about new extremes. Problems concerning extremes fall generally in the area of "extreme value theory" [**7**] or "dam theory" [**5**], about which a great deal has been written. (An extensive bibliography is given by Gumbel [**3**].) A basic problem considered in dam theory is that of deciding how "large" or "strong" a dam should be in order

$$S\begin{bmatrix} b & 0 \\ 0 & \bar{b} \end{bmatrix} S^{-1} = S\left(t_1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t_2\begin{bmatrix} a & 0 \\ 0 & \bar{a} \end{bmatrix}\right)S^{-1}$$

$$= t_1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t_2\begin{bmatrix} 0 & 1 \\ -r_2 & -r_1 \end{bmatrix}.$$

Since $a^p+b^p=c^p$, it follows that $\bar{a}^p+\bar{b}^p=\bar{c}^p$, so that

$$\left(S\begin{bmatrix} a & 0 \\ 0 & \bar{a} \end{bmatrix} S^{-1}\right)^p + \left(S\begin{bmatrix} b & 0 \\ 0 & \bar{b} \end{bmatrix} S^{-1}\right)^p = \left(S\begin{bmatrix} c & 0 \\ 0 & \bar{c} \end{bmatrix} S^{-1}\right)^p.$$

The matrices being raised to pth power all belong to $M_2(Z)$; see the penultimate matrix equation above. The proof is done.

An application of the above theorem is the following equation, which was pointed out to us by Professor and Mrs. D. H. Lehmer, valid for all primes p exceeding 3. If $(p, 6)$ equals 1, then

$$\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}^p + \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}^p = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}^p.$$

Added in proof. B. Divis has kindly called our attention to work of D. K. Faddeev, suggesting that there may be at most finitely many solutions to $A^p+B^p=C^p$ for $n=2$, $p=3,4$ ($AB=BA$, A, B, C invertible). See Dokl. Akad. Nauk USSR 134 (776–777); 136 (296–298).

Partial support for this paper was supplied under NSF GP 9483 and GP 8758. MOS classification: 10B15, 10B30.

References

1. J. Brillhart, Proc. of the Second Atlas Symposium, 1970.

2. M. Marcus and H. Mintz, A Survey of Matrix Theory and Matrix Inequalities, Allyn & Bacon, Boston, 1964.

3. R. Selfridge and H. Pollack, NOTICES, Amer. Math. Soc., Abstract #608-138 (1967) 97.

4. V. V. Kobelev, A proof of Fermat's last theorem for all prime exponents less than 5500, Dokl. Akad. Nauk SSSR, MR 41, 3363, (1970) 767–768.

5. I. A. Barnett and H. M. Weitkamp, The equation $X^n + Y^n + Z^n = 0$ in rational binary matrices, An. Sti. Univ. "Al. I. Cuza" Iasi Sect. I a Mat., 7 (1961) 1–64.

WHEN WILL THE NEXT RECORD RAINFALL OCCUR?

DONALD R. BARR, Naval Postgraduate School, Monterey, California

1. Introduction. We consider the occurrences of extremes in a sequence of past observations and their use in making simple inferences about new extremes. Problems concerning extremes fall generally in the area of "extreme value theory" [**7**] or "dam theory" [**5**], about which a great deal has been written. (An extensive bibliography is given by Gumbel [**3**].) A basic problem considered in dam theory is that of deciding how "large" or "strong" a dam should be in order to withstand, with high probability, the most severe flood to occur in the next n years. As Moran [5] points out, a solution for this flood *severity* problem requires that assumptions be made concerning the distribution of yearly flows. We shall be interested in flood *frequency*, where "flood" is defined in terms of past "flows."

Among many phenomena for which records are kept, new all time highs (lows, etc.) are observed from time to time. For example, it is occasionally announced that this year's annual rainfall in a given location has exceeded the record established in some specified previous year. Such an announcement is sometimes followed by a comment to the effect that the annual rainfall therefore appears to be generally increasing. Of course, this need not be the case, for suppose the annual rainfall in the kth year of record keeping is modeled as an outcome on the random variable X_k. Assume moreover that $X_1, X_2, \cdots$ are independent with identical continuous distributions. Then it is easy to show that the largest annual rainfall in j years of record keeping, $\max_{k \leqq j} \{X_k\}$, will almost surely be exceeded by that in some subsequent year, $X_{j+1}, X_{j+2}, \cdots$. Hence, one would expect to observe new record rainfalls from time to time as new annual rainfalls are observed.

Let us define two sequences of random variables associated with a sequence $\{X_k\}$ of the type described above (independent, identically and continuously distributed): the sequence $\{U_j\}$ of serial numbers of the successive record high values in the sequence $\{X_k\}$, and the sequence $\{\Delta_j\}$ of lengths between record values. For example, if $x_1=3$, $x_2=5$, $x_3=1$, $x_4=2$, $x_5=7, \cdots$ were observed, then $u_1=1$, $u_2=2$, $u_3=5, \cdots$ and $\delta_2=u_2-u_1=1$, $\delta_3=3, \cdots$ would be observed. The distributions of U_j and Δ_j have been obtained by Chandler [1], Neuts [6] and others, using elementary methods. Surprisingly, these distributions do not depend upon the distribution of the X_k. The sequence $\{\Delta_j\}$ is interesting in that each term tends to dominate the sum of the preceding terms. For example, approximate values of n such that $P[U_j \leqq n] \approx .5$ and values of n such that $P[\Delta_j \leqq n] \approx .5$ are shown in Table 1 for several values of j.

j	2	3	4	5	6	7	8	9
$P[U_j \leqq n] \approx .5$	$n = 2$	7	21	60	200	500	2000	5000
$P[\Delta_j \leqq n] \approx .5$		4	10	26	70	200	500	1000

TABLE 1. Approximate medians of U_j and Δ_j for $j<10$.

These entries verify that, as one would expect, new record values tend to get scarce, so that the duration between them tends to increase, as the total length of time records have been kept increases. Incidentally, Holmes and Strawderman [4] prove the rather remarkable fact that $\sqrt[j]{\Delta_j}$ almost surely converges to e.

The preceding observations suggest the tractability of the following closely related questions: If the previous record high observed in year y was first broken this year,

(i) how long have records been kept, and

(ii) how long will it be until the present record is broken?

We wish to emphasize that answers to these questions are to be based only upon a knowledge of the length of time from the immediately previous record high to the present one. As we shall see, for these questions, inferences may be made without distribution assumptions beyond those already stated above.

2. Estimating how long records have been kept. We wish to estimate j, based upon the number of additional observations required to first exceed $\max_{k\leqq j}\{X_k\}$, where $\{X_1, X_2, \cdots\}$ is a sequence of independent identically distributed continuous random variables. Assume that $X_j=\max_{k\leqq j}\{X_k\}$, so that x_j is the observed record value in j years of record keeping. Let N denote the (random) number of additional observations required to first exceed x_j. The distribution of N, for given j, may be obtained as follows:

$$
\begin{aligned}
p_j(n) &= P[N = n \mid X_j = \max_{k\leqq j}\{X_k\}] \\
&= P[X_{j+1}, \cdots, X_{j+n-1} < X_j < X_{j+n} \mid X_j > X_1, X_2, \cdots, X_{j-1}] \qquad (1)\\
&= \frac{P[X_1, X_2, \cdots, X_{j-1}, X_{j+1}, \cdots, X_{j+n-1} < X_j < X_{j+n}]}{P[X_1, \cdots, X_{j-1} < X_j]}.
\end{aligned}
$$

Under the assumptions that the X_k are independent and identically distributed, it follows that each possible ordering of a collection of these random variables is equally likely. With probability 1, there will be no ties among the X_k's since their distribution is assumed to be continuous. Since there are $j!$ possible orderings of $\{x_1, x_2, \cdots, x_j\}$, the denominator in (1) is simply $(j-1)!/j!$. A similar argument may be used with the numerator of (1), so that

$$
p_j(n) = j/(j+n)(j+n-1); \qquad n = 1, 2, \cdots, \tag{2}
$$

where the parameter space is $j=1, 2, \cdots$. If x_j was first exceeded by x_{j+n}, we may regard n as an outcome on N and seek estimates of j based on this sample of size 1 on N. We consider two principles of estimation: maximum likelihood and the method of moments, in that order.

For observed n, the likelihood function is given by

$$
f(j)=j/(j+n)(j+n-1); \qquad j = 1, 2, \cdots.
$$

It is easily seen that

$$
f(j+1)/f(j) \begin{cases} >1 & \text{for } j < n-1 \\ =1 & \text{for } j = n-1 \\ <1 & \text{for } j > n-1, \end{cases}
$$

so it follows that $\max_j f(j)$ occurs at $n-1$. Thus the maximum likelihood estimate $\hat{j}$ of j is given by $n-1$. The method of moments cannot be used in this case since N has no mean ($\sum_{n=1}^{\infty} np_j(n)$ is infinite for each j by comparison with the harmonic series).

Remark. If it is known only that the record value x_j was exceeded once in n subsequent years of record keeping, the distribution given in (2) is no longer appropriate. An alternative can be developed using a combinatorial argument as above; for variety we use an approach involving the probability integral transformation. Let $m_{j,n}(z)$ denote the probability that z observations in a random sample of size n, $X_{j+1}, \cdots, X_{j+n}$, exceed the largest value in a different random sample of size j, $X_1, \cdots, X_j$. Under the assumptions stated above, the number Z of observations exceeding $\max_{k\leq j}\{X_k\}$ has a binomial distribution,

$$m_{j,n}(z \mid Q = q) = \binom{n}{z} q^z(1-q)^{n-z}; \qquad z = 0, 1, 2, \cdots, n, \tag{3}$$

where Q is the probability that an individual random observation exceeds $\max_{k\leq j}\{X_k\}$. Now

$$Q = 1 - F(\max_{k\leq j}\{X_k\}), \tag{4}$$

where F is the distribution function of the X_k's. Since $F(X_k)$ has a uniform distribution over (0, 1) and $F(\max_{k\leq j}\{X_k\}) = \max_{k\leq j}\{F(X_k)\}$, $1-F(\max_{k\leq j}\{X_k\})$ has a beta distribution [7, page 75], so the unconditional distribution of Z is obtained as

$$\begin{aligned} m_{j,n}(z) &= \int_0^1 m_{j,n}(z \mid Q = q) f_Q(q)dq \\ &= \binom{n}{z}\int_0^1 q^z(1-q)^{n-z} j(1-q)^{j-1}dq \\ &= \binom{n}{z} j\beta(z+1, n+j-z) \\ &= \frac{n!j(n+j-z-1)!}{(n-z)!(n+j)!}; \qquad z = 0, 1, \cdots, n. \end{aligned} \tag{5}$$

(This distribution has been tabulated by Epstein [2] for the case $n=j$, and is discussed in a more general setting by Gumbel [3, page 59].) Since the probabilities (3) and (4) do not depend upon *which* of $X_1, \cdots, X_j$ is the largest, (5) is appropriate when it is known that $X_j = \max_{k\leq j}\{X_k\}$. Now suppose that in n observations one value larger than $\max_{k\leq j}\{X_k\} = X_j$ is observed. Using (5) with $z=1$, the maximum likelihood estimate of j may be found (using the method of the preceding paragraph) to be $n-1$. Since the mean of the distribution given in (5) is

$$E(Z) = n/(j+1),$$

the method of moments may be used in this case to obtain the same estimate.

While the approaches discussed above differ by a subtle change in assumptions, the estimates found by applying the various principles in them are the same: estimate j by $n-1$. Thus, a reasonable answer for question (i) seems to be: If this year's rainfall is the first to exceed the record of n years ago, estimate the total duration that records have been kept to be $\hat{j}+n$, or about $2n-1$ years.

3. Predicting when the present record will be broken. As suggested in the introduction, question (ii) is closely related to (i). Obviously, under the assumptions leading to (2), we cannot consider the mean time until the present record (say x_{j+n}) is broken. The most likely length of time is 1 for all j, and, depending upon one's objectives, this could be the best prediction. The median time may be taken to be the midpoint of $F_{j+n}^{-1}(\{1/2\})$, where

$$F_j(x) = P[N \leqq x \mid X_j = \max_{k \leqq j} \{X_k\}] = \sum_{i=1}^{[x]} p_j(i) = n/(j+n)$$

for $n \leqq x < n+1$. This median is given by $n+j+1/2$, which, again depending on circumstances, could serve as a reasonable prediction if j were known. At the expense of notational simplicity, our development of equation (2) (and consequently the function F_{j+n}) may be modified and used to show that this prediction is the same regardless of *which* of the X_k is $\max_{k \leqq j+n} \{X_k\}$. That is to say, at *any* time, the median time to exceed the record value to date is approximately the total duration of record keeping. For example, the largest recorded annual rainfall in Monterey (over 33 inches) occurred in the "year" 1906–07. Records have been kept here since 1896–1897 (take $n+j=74$), so we may predict that the record of 1907 will stand for roughly another 74 years. If j is unknown, $n+\hat{j}+\frac{1}{2}$, or about $2n$ years, appears to be a reasonable answer to (ii).

References

1. K. N. Chandler, The distribution and frequency of record values, J. Roy. Statist. Soc. Ser. B, 14 (1952) 220–228.

2. Benjamin Epstein, Tables for the distribution of the number of exceedances, Ann. Math. Statist., 25 (1954) 762–768.

3. E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York, 1958.

4. P. T. Holmes and W. E. Strawderman, A note on the waiting times between record observations, J. Appl. Probability, 6 (1969) 711–714.

5. P. A. P. Moran, The Theory of Storage, Wiley, New York, 1959.

6. M. F. Neuts, Waitingtimes between record observations, J. Appl. Probability, 4 (1967) 206–208.

7. A. E. Sarhan and B. G. Greenberg, Contributions to Order Statistics, Wiley, New York, 1962.

ONE-SIDED LIMITS AND INTEGRABILITY

W. R. JONES and M. D. LANDAU, Lafayette College

In an approach to the Riemann integral which uses upper and lower sums, it is a fundamental theorem that a bounded function is integrable if and only if the upper and lower sums are arbitrarily close for some partition. Utilizing this theorem, we obtain sufficient conditions for integrability which immediately provide a broad class of integrable functions including, among others, the continuous, piecewise continuous, and monotonic functions on closed bounded intervals. The proof of Theorem 1 provides a simpler derivation of a result which

3. Predicting when the present record will be broken. As suggested in the introduction, question (ii) is closely related to (i). Obviously, under the assumptions leading to (2), we cannot consider the mean time until the present record (say x_{j+n}) is broken. The most likely length of time is 1 for all j, and, depending upon one's objectives, this could be the best prediction. The median time may be taken to be the midpoint of $F_{j+n}^{-1}(\{1/2\})$, where

$$F_j(x) = P[N \leqq x \mid X_j = \max_{k \leqq j} \{X_k\}] = \sum_{i=1}^{[x]} p_j(i) = n/(j+n)$$

for $n \leqq x < n+1$. This median is given by $n+j+1/2$, which, again depending on circumstances, could serve as a reasonable prediction if j were known. At the expense of notational simplicity, our development of equation (2) (and consequently the function F_{j+n}) may be modified and used to show that this prediction is the same regardless of *which* of the X_k is $\max_{k \leqq j+n} \{X_k\}$. That is to say, at *any* time, the median time to exceed the record value to date is approximately the total duration of record keeping. For example, the largest recorded annual rainfall in Monterey (over 33 inches) occurred in the "year" 1906–07. Records have been kept here since 1896–1897 (take $n+j=74$), so we may predict that the record of 1907 will stand for roughly another 74 years. If j is unknown, $n+j+\frac{1}{2}$, or about $2n$ years, appears to be a reasonable answer to (ii).

References

1. K. N. Chandler, The distribution and frequency of record values, J. Roy. Statist. Soc. Ser. B, 14 (1952) 220–228.

2. Benjamin Epstein, Tables for the distribution of the number of exceedances, Ann. Math. Statist., 25 (1954) 762–768.

3. E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York, 1958.

4. P. T. Holmes and W. E. Strawderman, A note on the waiting times between record observations, J. Appl. Probability, 6 (1969) 711–714.

5. P. A. P. Moran, The Theory of Storage, Wiley, New York, 1959.

6. M. F. Neuts, Waitingtimes between record observations, J. Appl. Probability, 4 (1967) 206–208.

7. A. E. Sarhan and B. G. Greenberg, Contributions to Order Statistics, Wiley, New York, 1962.

ONE-SIDED LIMITS AND INTEGRABILITY

W. R. JONES and M. D. LANDAU, Lafayette College

In an approach to the Riemann integral which uses upper and lower sums, it is a fundamental theorem that a bounded function is integrable if and only if the upper and lower sums are arbitrarily close for some partition. Utilizing this theorem, we obtain sufficient conditions for integrability which immediately provide a broad class of integrable functions including, among others, the continuous, piecewise continuous, and monotonic functions on closed bounded intervals. The proof of Theorem 1 provides a simpler derivation of a result which dates back to Dini [1]. The scope of these results suggests their suitability for possible inclusion in an advanced calculus course.

THEOREM 1. *If f is bounded on $[a, b]$ and the limit from the right $f(x+)$ exists at each point of (a, b), then f is Riemann integrable on $[a, b]$.*

Proof. Given $\epsilon>0$, we shall show the existence of a partition of $[a, b]$ for which the upper and lower sums differ by less than ϵ. Let M be such that $|f(x)|<M$ for each x in $[a, b]$ and $\epsilon/16M<b-a$. Then let $k>0$ be such that $16M(b-a)k<\epsilon$. For each x in (a, b) let $\delta(x)>0$ satisfy $x-\delta(x)k>a$, $x+\delta(x)<b$ and be such that $x<t<x+\delta(x)$ implies $|f(t)-f(x+)|<\epsilon/4(b-a)\equiv\epsilon_1$. Let $N(x)=(x-\delta(x)k, x+\delta(x))$ for each x in (a, b) and

$$N(a) = [a, a + \epsilon/16M), \qquad N(b) = (b - \epsilon/16M, b].$$

Since $\{N(x) \mid a\leqq x\leqq b\}$ is an open covering of $[a, b]$, the Heine-Borel theorem assures that there is a finite subcovering, which may be reduced (by discarding any interval contained in the union of the remaining intervals) so that we may denote by $\{N(x_0), N(x_1), \cdots, N(x_n)\}$ a covering with $a=x_0<x_1<\cdots<x_n=b$. Moreover the construction of the intervals $N(x)$ assures us that this subcovering is a simple chain, i.e., $N(x_i)\cap N(x_j)\neq\varnothing$ iff $|i-j|\leqq 1$.

Let $c_i\in(x_{i-1}, x_i)\cap N(x_{i-1})\cap N(x_i)$ for $i=1, \cdots, n$. Then with $\delta_i=\delta(x_i)$ we have $x_i-c_i<\delta_i k$ so that for $i=1, \cdots, n-1$ we may let d_i satisfy both $d_i-c_i<\delta_i k$ and $x_i<d_i<c_{i+1}$. The collection $\{a, c_1, d_1, c_2, d_2, \cdots, c_n, b\}$ will serve as the desired partition of $[a, b]$. Both c_1-a and $b-c_n$ are less than $\epsilon/16M$, and, hence, because of the bound on f, the contribution to the difference of the upper and lower sums from each of the subintervals $[a, c_1]$, $[c_n, b]$ is less than $\epsilon/8$. Next since $x_i<d_i<c_{i+1}<x_i+\delta_i$ for $i=1, \cdots, n-1$, then $|f(s)-f(t)|<2\epsilon_1$ for any s, t in $[d_i, c_{i+1}]$, and, hence, the total contribution of these $n-1$ subintervals to the difference of the upper and lower sums of f is at most $2\epsilon_1\sum_{i=1}^{n-1}(c_{i+1}-d_i)<\epsilon/2$. The total contribution of the remaining $n-1$ subintervals $[c_i, d_i]$ is, by the bound on f, no more than $2M\sum_{i=1}^{n-1}(d_i-c_i)$. However, $d_i-c_i<\delta_i k<(x_{i+2}-x_i)k$ for $i=1, \cdots, n-2$ since the covering is a simple chain. By construction $d_{n-1}-c_{n-1}<\delta_{n-1}k<(x_n-x_{n-1})k$. Hence,

$$2M\sum_{i=1}^{n-1}(d_i - c_i) < 2Mk\left[\sum_{i=1}^{n-2}(x_{i+2} - x_i) + (x_n - x_{n-1})\right] < 4Mk(b - a) < \epsilon/4$$

by the choice of k. Thus the difference of the upper and lower sums for f and this partition is less than $2\epsilon/8+\epsilon/2+\epsilon/4=\epsilon$, and the proof of integrability is complete.

The following result is a simple extension of Theorem 1:

THEOREM 2. *If $[a, b]$ is the union of finitely many abutting closed subintervals such that on the interior of each of these subintervals either all the left or all the right limits of f exist, and if f is bounded on $[a, b]$, then f is Riemann integrable on $[a, b]$.*

THEOREM 3. *If the one-sided limits of f exist at each point of $[a, b]$, then f is Riemann integrable on $[a, b]$.*

Proof. By Theorem 1 we need show only that f is bounded.

Let $f(x+)=\lim_{t\to x^+} f(t)$ for each x in $[a, b)$ and $f(x-)=\lim_{t\to x^-} f(t)$ for each x in $(a, b]$. For each x in $[a, b)$ let $\delta_1(x)>0$ be such that $x<t<x+\delta_1(x)$ implies $|f(t)-f(x+)|<1$; for each x in $(a, b]$ let $\delta_2(x)>0$ be such that $x-\delta_2(x)<t<x$ implies $|f(t)-f(x-)|<1$. Define $N(a)=[a, a+\delta_1(a))$, $N(b)=(b-\delta_2(b), b]$ and $N(x)=(x-\delta_2(x), x+\delta_1(x))$ for each x in (a, b) so that $\{N(x)|a\leqq x\leqq b\}$ is an open covering of $[a, b]$. By the Heine-Borel theorem there is a finite subcovering, say, $\{N(x_1), \cdots, N(x_n)\}$. For $i=1, \cdots, n$ let $M_i=\max\{f(x_i-)+1, f(x_i), f(x_i+)+1\}$ with $f(x_i-)+1$ deleted if $x_i=a$ and $f(x_i+)+1$ deleted if $x_i=b$. It follows that $M=\max\{M_i|i=1, \cdots, n\}$ is a bound on f on $[a, b]$.

For classroom use it may be noted that Theorem 3 can be proved independently by an argument similar to that used in proving Theorem 1 and with some gain in simplicity.

The oft-mentioned function defined on $[0, 1]$ by $f(0)=1$, $f(x)=0$ if x is irrational, $f(x)=1/q$ if $x=p/q$ and the positive integers p, q are relatively prime, falls in the scope of these theorems and is integrable.

Reference

1. Ulisse Dini, Grundlagen für eine Theorie der Functionen einer Veränderlichen Reelen Grösse, B. G. Teubner, Leipzig, 1892, pp. 335–338.

RELATIVELY PRIME AMICABLE NUMBERS WITH TWENTY-ONE PRIME DIVISORS

PETER HAGIS, JR., Temple University

1. Introduction. A pair of positive integers, m and n, is said to be amicable if

$$m + n = \sigma(m) = \sigma(n) \tag{1}$$

where $\sigma(k)$ is the sum of the positive divisors of k. At present more than 1000 pairs of amicable numbers have been discovered, none of which is relatively prime. Kanold [3] has shown that if a pair of relatively prime amicable numbers exists then their product must be divisible by *at least* twenty-one different primes (see also [1] and [2]). In the present paper we investigate some of the properties which must be possessed by a pair of relatively prime amicable numbers whose product contains *exactly* twenty-one different primes. To be precise, we shall prove the following

THEOREM. *Let m and n be a pair of relatively prime amicable numbers such that mn is divisible by precisely* 21 *primes. Then*:

(a) *m and n are of opposite parity*;

(b) *mn is divisible by every prime p such that $5\leqq p\leqq 61$, and at least three of the primes* 67, 71, 73, 79, 83 *divide mn including at least two of the primes* 67, 71, 73;

(c) *if q is the largest prime dividing mn, then $79\leqq q\leqq 113$*;

(d) *if m is even then m has at least two, but not more than fourteen, odd prime divisors*;

(e) $0.875<\min(m/n, n/m)<1<\max(m/n, n/m)<1.142$;

(f) $mn>2\cdot 10^{177}$, $m>4\cdot 10^{88}$, $n>4\cdot 10^{88}$;

Let $f(x+) = \lim_{t \to x^+} f(t)$ for each x in $[a, b)$ and $f(x-) = \lim_{t \to x^-} f(t)$ for each x in $(a, b]$. For each x in $[a, b)$ let $\delta_1(x) > 0$ be such that $x < t < x + \delta_1(x)$ implies $|f(t) - f(x+)| < 1$; for each x in $(a, b]$ let $\delta_2(x) > 0$ be such that $x - \delta_2(x) < t < x$ implies $|f(t) - f(x-)| < 1$. Define $N(a) = [a, a + \delta_1(a))$, $N(b) = (b - \delta_2(b), b]$ and $N(x) = (x - \delta_2(x), x + \delta_1(x))$ for each x in (a, b) so that $\{N(x) | a \leq x \leq b\}$ is an open covering of $[a, b]$. By the Heine-Borel theorem there is a finite subcovering, say, $\{N(x_1), \cdots, N(x_n)\}$. For $i = 1, \cdots, n$ let $M_i = \max\{f(x_i-) + 1, f(x_i), f(x_i+) + 1\}$ with $f(x_i-) + 1$ deleted if $x_i = a$ and $f(x_i+) + 1$ deleted if $x_i = b$. It follows that $M = \max\{M_i | i = 1, \cdots, n\}$ is a bound on f on $[a, b]$.

For classroom use it may be noted that Theorem 3 can be proved independently by an argument similar to that used in proving Theorem 1 and with some gain in simplicity.

The oft-mentioned function defined on $[0, 1]$ by $f(0) = 1$, $f(x) = 0$ if x is irrational, $f(x) = 1/q$ if $x = p/q$ and the positive integers p, q are relatively prime, falls in the scope of these theorems and is integrable.

Reference

1. Ulisse Dini, Grundlagen für eine Theorie der Functionen einer Veränderlichen Reelen Grösse, B. G. Teubner, Leipzig, 1892, pp. 335–338.

RELATIVELY PRIME AMICABLE NUMBERS WITH TWENTY-ONE PRIME DIVISORS

PETER HAGIS, JR., Temple University

1. Introduction. A pair of positive integers, m and n, is said to be amicable if

$$m + n = \sigma(m) = \sigma(n) \tag{1}$$

where $\sigma(k)$ is the sum of the positive divisors of k. At present more than 1000 pairs of amicable numbers have been discovered, none of which is relatively prime. Kanold [3] has shown that if a pair of relatively prime amicable numbers exists then their product must be divisible by *at least* twenty-one different primes (see also [1] and [2]). In the present paper we investigate some of the properties which must be possessed by a pair of relatively prime amicable numbers whose product contains *exactly* twenty-one different primes. To be precise, we shall prove the following

THEOREM. *Let m and n be a pair of relatively prime amicable numbers such that mn is divisible by precisely* 21 *primes. Then:*

(a) *m and n are of opposite parity;*

(b) *mn is divisible by every prime p such that $5 \leq p \leq 61$, and at least three of the primes* 67, 71, 73, 79, 83 *divide mn including at least two of the primes* 67, 71, 73;

(c) *if q is the largest prime dividing mn, then $79 \leq q \leq 113$;*

(d) *if m is even then m has at least two, but not more than fourteen, odd prime divisors;*

(e) $0.875 < \min(m/n, n/m) < 1 < \max(m/n, n/m) < 1.142$;

(f) $mn > 2 \cdot 10^{177}$, $m > 4 \cdot 10^{88}$, $n > 4 \cdot 10^{88}$;

(g) $.7497 < \phi(mn)\sigma(mn)/(mn)^2 < .7500$.

In (g) ϕ is the Euler phi-function (see Section 11.4 in [4]).

In conjunction with Kanold's lower bound of 21 primes mentioned above (a) yields the following corollary which improves Kanold's result (if $2 \nmid mn$).

COROLLARY. *If m and n are relatively prime odd amicable numbers then mn is divisible by at least 22 different primes.*

2. Notation and groundwork. In the sequel p and q will always denote primes while the symbol P_j will be used to represent the jth odd prime. Thus, $P_1 = 3$ and $P_{17} = 61$. If $p^a | k$ but $p^{a+1} \nmid k$ we shall write $p^a \| k$ *or* $a = \text{EXP}(p)$. When the latter notation is used k will be clear from the context.

m and n will be understood to be a pair of relatively prime amicable numbers. It then follows from (1) that

$$4 < 2 + m/n + n/m = \sigma(mn)/mn = \prod \sigma(p^a)/p^a \tag{2}$$

where the product is taken over all the prime divisors of mn and $p^a \| mn$.

We next state four propositions concerning m and n which will be needed later. The proofs of the first two appear in [1]; the proofs of the other two may be found in [2].

PROPOSITION 1. *If $q | mn$ and $p^a \| mn$ then $q \nmid \sigma(p^a)$, and if $p = qk - 1$ then $2 | a$.*

PROPOSITION 2. *If m and n are odd and $3 \nmid mn$, then mn has at least 140 prime factors.*

PROPOSITION 3. *If m and n are of opposite parity then $mn = 2K^2$ where $(6, K) = 1$.*

PROPOSITION 4. *If $2 | mn$ and $p^a \| mn$ then* (i) *if $p = 8k+1$ then $a \equiv 0, 2 \pmod 8$*; (ii) *if $p = 8k+3$ then $4 | a$*; (iii) *if $p = 8k+5$ then $a \equiv 0, 6 \pmod 8$*; (iv) *if $p = 8k+7$ then $2 | a$.*

Finally we recall two facts from elementary number theory.

LEMMA 1. *If k is odd then $k^2 \equiv 1 \pmod 8$.*

LEMMA 2. *$\sigma(k)$ is odd if and only if $k = N^2$ or $k = 2N^2$.*

3. The prime divisors of *mn*. From Propositions 2 and 3, (2), and the fact that $\sigma(p^a)/p^a < p/(p-1)$ we see that if mn has exactly 21 prime divisors then

$$4 < 1.5 \prod_{i=2}^{21} p_i/(p_i - 1) \tag{3}$$

where $p_2 < p_3 < \cdots < p_{21}$ are the prime divisors of mn which exceed 3. If $p_r \neq P_r$ then it follows from (3) and the monotonic decreasing nature of $x/(x-1)$ that for $2 \leq r \leq 22$

$$4 < F(r) = 1.5 \prod_{i=2}^{22} P_i/(P_i - 1) \qquad (i \neq r). \tag{4}$$

Making use of the CDC 6400 at the Temple University Computing Center it was found that $F(r)<4$ for $2\leqq r\leqq 17$ from which we conclude that $p_r=P_r$ for $r=2, \cdots, 17$ or that $p|mn$ for $5\leqq p\leqq 61$.

By a similar argument if one or none of the primes 67, 71, 73 divides mn, then

$$4 < \prod_{i=1}^{18} P_i/(P_i - 1)\cdot \prod_{i=21}^{23} P_i/(P_i - 1) = 3.9994$$

which is a contradiction. Therefore, mn is divisible by at least two of the primes 67, 71, 73.

If only two of the primes 67, 71, 73, 79, 83 divide mn, then

$$4 < \prod_{i=1}^{19} P_i/(P_i - 1)\cdot(89\cdot 97)/(88\cdot 96) = 3.9981$$

which is again a contradiction and which completes the proof of part (b) of our theorem.

If q is the largest prime dividing mn then

$$4 < q/(q-1)\cdot \prod_{i=1}^{20} P_i/(P_i - 1) < 3.9668q/(q-1).$$

It follows that $q<121$ and, since q is a prime, that $q\leqq 113$.

4. The parity of *m* and *n*. Let us assume that m and n are both odd. Since $m+n$ is then even we see from (1) and Lemma 2 that neither m nor n is a perfect square. Since every odd prime which does not exceed 113, except 3, 7, and 31, is congruent to -1 modulo a prime q, where $3\leqq q\leqq 37$, it follows from the discussion in Section 3 and Proposition 1 that

$$mn = 3^a 7^b 31^c J^2 \tag{5}$$

where J is not divisible by 2, 3, 7, 31 and a, b, c are positive integers. From Lemma 1 we have

$$mn \equiv \pm 3^a \pmod 8. \tag{6}$$

Now either b or c is odd since otherwise either m or n would be a perfect square. And since if $p\equiv -1 \pmod 8$ and k is an odd natural number $\sigma(p^k) \equiv 0 \pmod 8$ we see from (1) that

$$m + n \equiv 0 \pmod 8. \tag{7}$$

Also, a is even since otherwise $mn\equiv \pm 3 \pmod 8$ from (6), which is incompatible with (7). We conclude then that $m=7^bM^2$ and $n=31^cN^2$ where b, c, M, N are odd positive integers. Again using Lemma 1 we see that $m\equiv n\equiv -1 \pmod 8$ which contradicts (7). Therefore, the assumption of the first sentence of this section is false, and m and n (if they exist) are of opposite parity.

5. Bounds on the quotient of *m* and *n*. Since we now know that m and n are of opposite parity we may assume that m is even and n is odd. From (2) we see that

$$2 + m/n + n/m = \sigma(mn)/mn < \prod_{i=1}^{21} P_i/(P_i - 1) < 4.017634.$$

Therefore, if $x = \max(m/n, n/m)$ we have $x^2 - 2.017634x + 1 < 0$ from which an application of the quadratic formula yields $1 < x < 1.142$. Also, $1 > \min(m/n, n/m) = 1/x > 0.875$ which completes the proof of (e).

If $m = 2p^a$ then $p \geqq 5$ and it follows from (1) and (e) that $1.875 < (m+n)/m = \sigma(m)/m < (3/2)(5/4) = 1.875$ which is a contradiction. Therefore, m has at least two odd prime divisors.

If n has only five prime divisors then

$$1.875 < (m + n)/n = \sigma(n)/n < \prod_{i=2}^{6} P_i/(P_i - 1) < 1.85$$

which is a contradiction and completes the proof of (d).

Remark. If $n < m$ then n has at least seven prime factors. For otherwise

$$2 < (m + n)/n = \sigma(n)/n < \prod_{i=2}^{7} P_i/(P_i - 1) < 1.95$$

which is impossible.

6. Lower bounds for m, n, mn. The restrictions imposed on mn by parts (a), (b), (c) of our theorem as well as by Propositions 1, 3, and 4 will now be utilized to obtain a lower bound for mn, and this bound in conjunction with (e) will be used to bound m and n from below.

We begin by investigating the divisibility of $\sigma(p^a)$ by q where both p and q are between 5 and 113 (both bounds being inclusive) and where a is restricted in accordance with the conclusions of Proposition 4. Due to the magnitude of the numbers involved and the multiplicity of cases which had to be studied the investigation was carried out on the CDC 6400 using modular arithmetic. It will be convenient to discuss separately the cases p congruent to 1, 3, 5, 7 modulo 8.

If $p = 17, 41, 73, 89, 97$ we find that $\sigma(p^2)$ is not divisible by q, $5 \leqq q \leqq 113$, so that $\mathrm{EXP}(p) \geqq 2$ by Propositions 1 and 4. Since $13 | \sigma(113^2)$ and $13 | \sigma(113^8)$ while $q \nmid \sigma(113^{10})$ we see that $\mathrm{EXP}(113) \geqq 10$.

If $p = 19, 43, 67, 83, 107$ then $\sigma(p^4)$ is not divisible by q, $5 \leqq q \leqq 113$. Therefore, $\mathrm{EXP}(p) \geqq 4$ by Propositions 1 and 4. $5 | \sigma(11^4)$; $7, 19 | \sigma(11^8)$; $q \nmid \sigma(11^{12})$ so that $\mathrm{EXP}(11) \geqq 12$. $11, 41 | \sigma(59^4)$ while $q \nmid \sigma(59^8)$. Therefore, $\mathrm{EXP}(59) \geqq 8$.

The prime divisors (between 5 and 113) of $\sigma(p^a)$ for $p \equiv 5, 7 \pmod 8$ are given in Tables I, II respectively. If $q \nmid \sigma(p^a)$, $5 \leqq q \leqq 113$, and $b > a$ then the divisors of $\sigma(p^b)$ are not tabulated.

From these tables we see, for example, that $\mathrm{EXP}(47) \geqq 10$, while $\mathrm{EXP}(37) \geqq 6$ or 16 according as 71 does not or does divide mn.

From (a), (b), (c), Proposition 3, and the discussion just concluded we have

$$mn \geqq 2 \cdot 5^6 7^4 11^{12} 13^6 17^2 19^4 23^4 29^{16} 31^2 37^6 41^2 43^4 47^{10} 53^{16} 59^8 61^6 67^4 73^2 79^4 89^2$$

from which it follows that $mn > 2 \cdot 10^{177}$.

Since $m < 1.2n$ we have $1.2n^2 > mn > 2 \cdot 10^{177}$ so that $n > 4 \cdot 10^{88}$. Similarly, $m > 4 \cdot 10^{88}$ and the proof of (f) is complete.

TABLE I

The Relevant Prime Divisors of $\sigma(p^a)$

p \ a	6	8	14	16
5	None			
13	None			
29	7	13, 67	13, 67	None
37	71	7, 67, 73	7, 11, 41, 67	None
53	29	7, 37	7, 11	None
61	None			
101	71	19	5, 31	None
109	113	7	7, 31	None

TABLE II

The Relevant Prime Divisors of $\sigma(p^a)$

p \ a	2	4	6	8	10
7	19	None			
23	7, 79	None			
31	None				
47	37, 61	11, 31	43	19, 37, 61	None
71	None				
79	7, 43	None			
103	None				

7. Bounds on $\phi(mn)\sigma(mn)/(mn)^2$. If, for convenience, we let $K=mn$, and if $\phi(K)$ denotes the Euler phi-function, then using the familiar formulas for $\sigma(K)$ and $\phi(K)$ (see Sections 10.2 and 11.4 in [4]) we have

$$\phi(K)\sigma(K)/K^2 = .75 \prod (1 - 1/p^{a+1})$$

where the product is taken over all the odd primes dividing mn and $a=\mathrm{EXP}(p)$. We see immediately that $\phi(K)\sigma(K)/K^2<.75$, while from the discussion in Section 6 we have

$$\phi(K)\sigma(K)/K^2 \geqq .75AB \tag{8}$$

where, with $p^{(b)}=1-1/p^b$,

$$A = 5^{(7)}7^{(5)}17^{(3)}31^{(3)}41^{(3)} > .99967, \tag{9}$$

$$\begin{aligned} B &= 11^{(13)}13^{(7)}19^{(5)}23^{(5)}29^{(17)}37^{(17)}43^{(5)}47^{(11)}53^{(17)}59^{(9)}61^{(7)} \\ &\quad \cdot 67^{(5)}71^{(3)}73^{(3)}89^{(3)} > (1-1/71^3)^{15} > 1-15/71^3 > .99995. \end{aligned} \tag{10}$$

From (8), (9), (10) we have $\phi(K)\sigma(K)/K^2>.7497$ which completes the proof of (g) and our theorem.

References

1. P. Hagis, Jr., On relatively prime odd amicable numbers, Math. Comp., 23 (1969) 539–543.
2. ———, Relatively prime amicable numbers of opposite parity, this MAGAZINE, 43 (1970) 14–20.
3. H-J. Kanold, Untere Schranken für teilerfremde befreudete Zahlen, Arch. Math., 4 (1953) 399–401.
4. B. M. Stewart, Theory of Numbers, second ed., Macmillan, New York, 1964.

MONOPOLY AS A MARKOV PROCESS

ROBERT B. ASH and RICHARD L. BISHOP, University of Illinois

1. Introduction. Practically everyone is familiar with the rudiments, anyway, of the game of Monopoly. Pawns are marched around a board with 40 positions according to the dictates of a pair of dice. The person in control of the pawn pays out or receives money to or from other players and the bank, depending on the position of the pawn and the state of development of the particular position. Thus, in order to assess the relative value of controlling the various properties, one must have an idea of the frequencies with which it can be expected that pawns will occupy each position. The basic problem is that these frequencies are not all the same.

By making minor modifications of the rules we are able to render the motion of the pawns a Markov process. We have used a standard method for calculating the frequencies displayed in Table I. Combining this data with rules and rates for money exchanges with the bank and other players, we can predict the average net income from the bank and from rentals on each property group (fully developed), as given in Tables II and III.

We feel that our results give the only feasible quantitative basis for strategy in playing Monopoly. One should expect that these results will be useful only over a sequence of many games and that they will apply better to long games. This is because convergence to the limit frequencies is good at about 24 turns, while decisive events can take place in a much shorter time. Further development of strategy on the foundation we give would have to discuss topics which we have not considered, such as bargaining and development financing.

where the product is taken over all the odd primes dividing mn and $a = \mathrm{EXP}(p)$. We see immediately that $\phi(K)\sigma(K)/K^2 < .75$, while from the discussion in Section 6 we have

$$\phi(K)\sigma(K)/K^2 \geqq .75AB \tag{8}$$

where, with $p^{(b)} = 1 - 1/p^b$,

$$A = 5^{(7)}7^{(5)}17^{(3)}31^{(3)}41^{(3)} > .99967, \tag{9}$$

$$\begin{aligned} B = {} & 11^{(13)}13^{(7)}19^{(5)}23^{(5)}29^{(17)}37^{(17)}43^{(5)}47^{(11)}53^{(17)}59^{(9)}61^{(7)} \\ & \cdot 67^{(5)}71^{(3)}73^{(3)}89^{(3)} > (1 - 1/71^3)^{15} > 1 - 15/71^3 > .99995. \end{aligned} \tag{10}$$

From (8), (9), (10) we have $\phi(K)\sigma(K)/K^2 > .7497$ which completes the proof of (g) and our theorem.

References

1. P. Hagis, Jr., On relatively prime odd amicable numbers, Math. Comp., 23 (1969) 539–543.
2. ———, Relatively prime amicable numbers of opposite parity, this MAGAZINE, 43 (1970) 14–20.
3. H-J. Kanold, Untere Schranken für teilerfremde befreudete Zahlen, Arch. Math., 4 (1953) 399–401.
4. B. M. Stewart, Theory of Numbers, second ed., Macmillan, New York, 1964.

MONOPOLY AS A MARKOV PROCESS

ROBERT B. ASH and RICHARD L. BISHOP, University of Illinois

1. Introduction. Practically everyone is familiar with the rudiments, anyway, of the game of Monopoly. Pawns are marched around a board with 40 positions according to the dictates of a pair of dice. The person in control of the pawn pays out or receives money to or from other players and the bank, depending on the position of the pawn and the state of development of the particular position. Thus, in order to assess the relative value of controlling the various properties, one must have an idea of the frequencies with which it can be expected that pawns will occupy each position. The basic problem is that these frequencies are not all the same.

By making minor modifications of the rules we are able to render the motion of the pawns a Markov process. We have used a standard method for calculating the frequencies displayed in Table I. Combining this data with rules and rates for money exchanges with the bank and other players, we can predict the average net income from the bank and from rentals on each property group (fully developed), as given in Tables II and III.

We feel that our results give the only feasible quantitative basis for strategy in playing Monopoly. One should expect that these results will be useful only over a sequence of many games and that they will apply better to long games. This is because convergence to the limit frequencies is good at about 24 turns, while decisive events can take place in a much shorter time. Further development of strategy on the foundation we give would have to discuss topics which we have not considered, such as bargaining and development financing.

TABLE I

n Position	RIJ	LJ
0 Go	.0346	.0368
1 Mediterranean Ave.	.0237	.0252
2 Community Chest 1	.0218	.0223
3 Baltic Ave.	.0241	.0256
4 Income tax	.0261	.0277
5 Reading RR	.0332	.0352
6 Oriental Ave.	.0253	.0268
7 Chance 1	.0096	.0102
8 Vermont Ave.	.0258	.0274
9 Connecticut Ave.	.0237	.0252
10 Visiting jail	.0254	.0269
11 St. Charles Place	.0304	.0321
12 Electric Co.	.0311	.0310
13 State Ave.	.0258	.0281
14 Virginia Ave.	.0288	.0293
15 Pennsylvania RR	.0313	.0346
16 St. James Place	.0318	.0331
17 Community Chest 2	.0272	.0307
18 Tennessee Ave.	.0335	.0349
19 New York Ave.	.0334	.0366
20 Free parking	.0336	.0343
21 Kentucky Ave.	.0310	.0336
22 Chance 2	.0125	.0125
23 Indiana Ave.	.0305	.0325
24 Illinois Ave.	.0355	.0377
25 B and O RR	.0344	.0364
26 Atlantic Ave.	.0301	.0321
27 Ventnor Ave.	.0299	.0318
28 Water works	.0315	.0333
29 Marvin Gardens	.0289	.0307
30 Jail	.1123	.0469
31 Pacific Ave.	.0300	.0318
32 North Carolina Ave.	.0294	.0312
33 Community Chest 3	.0263	.0282
34 Pennsylvania Ave.	.0279	.0297
35 Short Line RR	.0272	.0289
36 Chance 3	.0096	.0102
37 Park Place	.0245	.0260
38 Luxury tax	.0245	.0260
39 Boardwalk	.0295	.0312

Mathematically, the problem presents a nice illustration of a finite Markov chain, particularly showing how well iterative techniques converge even when the size is rather large (120 states). We do not have space here to give the mathematical details and describe the complications which arose in, for example, estimating the rapidity of convergence. These details can be obtained in mimeographed form by writing an author.

2. Markov chains. For the technical details of what constitutes a Markov

TABLE II

Expected income from bank, dollars per turn

	RIJ	LJ
Salary (pass Go)	38.48	41.16
Income and Luxury taxes	−7.06	−7.49
Community Chest	1.94	2.03
Chance	1.25	1.29
Get out of jail charge	−1.29	−2.35
Total	33.37	34.64
Assessment and General Repairs:		
each house	− .35	− .37
each hotel	−1.15	−1.22

TABLE III

Expected property-group income, with hotels, dollars per opponents' turn

Position numbers	RIJ	LJ
1, 3	16.8	17.8
6, 8, 9	42.3	44.9
11, 13, 14	68.1	71.5
16, 18, 19	95.4	101.2
21, 23, 24	103.6	110.9
26, 27, 29	103.7	110.3
31, 32, 34	114.8	121.9
37, 39	95.8	101.4
Railroads (one owner)	27.3	29.2
Utilities (one owner)	4.4	4.5

chain we refer the reader to [1] or [2]. What concerns us here is the extent to which Monopoly movements fail to be Markovian and how we modify the rules to make them so.

In a Markov process free-will choices are forbidden, that is, the probability of passing from state i to state j is some definite number p_{ij}, unalterable by a fickle human. The only true violation of this principle in Monopoly is the choice given a person in jail to leave or not provided he does not throw doubles. We have resolved this by assuming that one of two possible strategies is invariably followed:

1. The remain-in-jail (RIJ) rule. In this case the player stays in jail as long as possible, that is, until the third turn or doubles occurs.
2. The leave-jail (LJ) rule. Here the player always exits on the first turn.

Thus we may obtain frequencies and expected income under each strategy, as given in the tables.

Because of the repeated throws allowed in a turn when doubles are thrown, there is the possiblity of passing through three positions on the board in a single turn. It is therefore convenient to take a *throw* (of the dice) as the basic unit of time for our Markov chain, and to determine a state not only by the position on the board, but also by whether the first, second, or third throw of the turn is coming up. The jail position does not have these three within-a-turn states, but it does have three states in the RIJ mode, depending on how many turns the jail stay has lasted. Thus there are 120 states, two of which cannot be attained in the LJ mode.

The actual rule concerning Chance and Community Chest cards would force us to consider impossibly many more states than we have described, since the outcome of a throw will be influenced by the particular cards at the top of the Chance and Community Chest stacks. A further complication is that the "get out of jail free" cards may be removed from the stacks. A simple modification of the rules eliminates this difficulty. Instead of replacing the card at the bottom of the deck after it is drawn, the whole deck is shuffled after each draw. The "get out of jail free" cards are replaced immediately, with a system of credit replacing the physical use of these cards.

3. Frequencies. A basic initial goal in analyzing a Markov chain is the computation of the limit frequencies of the states. If $R_i(n)$ is the number of times state i is visited in the first n steps, then (under assumptions of irreducibility and aperiodicity that are satisfied by our chain), regardless of the initial state, $n^{-1}R_i(n)$ approaches a limit f_i with probability 1. The number f_i is called the *limit frequency of state i*. However, for us the relevant data are not the limit frequencies of our *per throw* states, but the limit frequencies of the positions *per turn*. The sum of these limit frequencies will be greater than 1 since more than one position is possible in each turn. Thus we multiply the per throw frequencies by a scale factor so that the sum of the frequencies of states corresponding to the beginning of a turn is 1. Then the sum of the scaled frequencies of the three states of a given position will be the per turn frequency of that position. It is these numbers that are presented in Table I.

An examination of these figures leads to the following conclusions. The main reason for the unevenness of frequencies is the transfer-to-jail spot. If we think of the movement around the board as a fluid flow, then position 30 seems to be a sink which feeds a source at position 10. The positions deprived by this arrangement are those likely to follow 30 in one or two steps, say positions 1 or 3, which show frequencies 20% below average. On the other hand, positions 16 through 31 have above average frequencies, again 20% at the maximum (c. position 20). More surprising is that the position-transfer cards (mostly from the Chance pile) have almost as great an effect. This explains why the frequency of position 24 is 16% greater than its neighbor 23. The Shortline RR is much less desirable than the other three railroads because it is not accessible from a Chance card. When computing expected income one must separate the portion of income derived from these railroad transfer cards, since they direct the renter to pay double.

References

1. R. B. Ash, Basic Probability Theory, Wiley, New York, 1970, Chapter 7.

2. W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, Wiley, New York, 1950, Chapters 15 and 16.

THE MELANCHOLY OCTAHEDRON

P. J. FEDERICO, Washington, D. C.

Albrecht Dürer's classic engraving of 1514, "Melencolia I," has been referred to many times, and often reproduced, in works dealing with magic squares, on account of the magic square of order 4, with special properties, which appears in the upper right. References [1], [2] and [3] are merely a few examples. The original publication of the engraving may have constituted the first appearance of a magic square in print. It would be out of place here to present any comments on the status of the engraving in the world of art, which appears to be considerable, and none are attempted (see References [2] and [4–9]). There is another mathematical feature in the work which has sometimes been mentioned but apparently seldom discussed. The left center shows a large block, taken as being of stone, with curiously arranged plane faces, a polyhedron of some sort. Figure 1 is a tracing of this feature of the engraving, with added reference numerals and a ground line. This note, which is not intended to be taken too seriously, attempts to identify Dürer's polyhedron. Reproductions of the engraving are found in each of the references except [3].

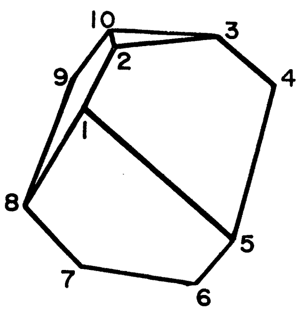


FIG. 1.

Of course you can place anything your fancy pleases on the back side of the stone block, a *bas relief* of St. George slaying the dragon if you wish. The assumption must be made that the part of a figure or object which does not show is consistent in form and nature with the part which is visible; that the hourglass, bell, millstone and sphere are still such on the sides which do not appear; that the emaciated sleeping dog in the foreground is still dog all around; the putto sitting on the millstone and writing on a tablet is putto on the other side, etc. Hence the stone block with plane faces and no holes or re-entrant concavities as placed by the artist is taken as having these same characteristics no matter from which side viewed, in other words a convex polyhedron.

The visible part of the figure of the polyhedron as drawn by Dürer (see Figure 1) shows 4 faces, 10 vertices (corners), which have been numbered in the figure, and 13 edges. Some higher lower limits can be deduced. The bottom face on which the block rests shows only 2 vertices; there must be at least one other vertex on this bottom face which does not coincide with any of those showing, otherwise the block would topple over. Hence the polyhedron must have at least 11 vertices. Eight of the edges which show are around the rim of the drawing and each of these edges borders 2 faces, one which shows and one which does not. But there are not necessarily 8 additional faces since some of them could be the same; for example, the face on the other side of edge 3–10 could be the same as the face on the other side of edge 3–4, but the face on the other side of edge 4–5 could not be the same as the face on the other side of edge 3–4. A study of the connections of the edges around the rim of the drawing in the above manner shows that there must be at least 4 faces in addition to the 4 showing, and hence the polyhedron must have at least 8 faces. Of the 4 faces which show, one is a triangle and three are pentagons; the face on the left is considerably foreshortened and is in shadow in the engraving, but it definitely has 5 sides.

Hence we must seek a polyhedron with at least 11 vertices, and with at least 8 faces at least one of which is a triangle and at least three of which are pentagons, these 4 faces being so arranged that the body could be fashioned to give a view corresponding to the engraving. There would be an infinite number of these and some further assumption must be made in order to arrive at any reasonable result. It is assumed that the simplest and most uniform figure possible, conforming to the requirements, would be the one which was used. Such an assumption is reasonable from artistic considerations and it should not be overlooked that Dürer was the author of a work on applied geometry [9]. Since the smallest number of faces possible for the block is eight, the octahedra will be examined.

The number of vertices V, faces F, and edges E, of any polyhedron must satisfy Euler's relationship:

$$V + F = E + 2.$$

It is easy to show from this equation and the fact that each face must have at least 3 edges surrounding it and each vertex must have at least 3 edges leading from it, that an 8-faced polyhedron cannot have more than 12 vertices (see Note 1 at end). As has been indicated, the number of vertices on the polyhedron of the engraving must be 11 or more, hence there are two cases to consider, octahedra with 11 vertices and octahedra with 12 vertices.

With 12 vertices, Euler's formula shows that an 8-faced polyhedron must have 18 edges. Ten vertices and 13 edges are visible in the engraving and hence there are 2 vertices and 5 edges on the rear side. One of the 2 invisible vertices belongs to the base on which the block rests and the other could also be assigned to the base. There are, then, two cases, one with a triangular base and one with a quadrilateral base. In each of these two cases only one configuration is possible. These are shown in Figure 2A, the triangular base, and Figure 2B, the quadrilateral base; these figures are duplicates of Figure 1, with the back sides shown in dotted lines. In placing the 5 additional edges in the rear of the figures it is helpful to utilize the fact that when an octahedron has 12 vertices, each vertex can have only 3 edges leading from it, see Note 2.

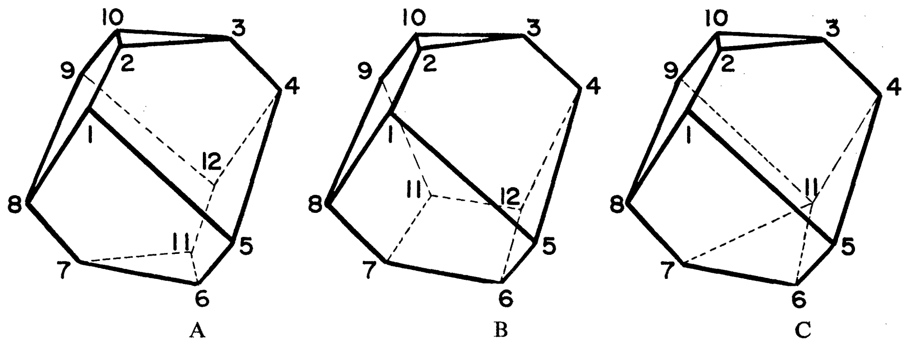


FIG. 2.

Consider now the case of octahedra with 11 vertices. Euler's formula shows that there are 17 edges and hence there are 4 edges, and the one additional vertex, on the rear side. The eleventh vertex must be on the base, and the base is hence a triangle. Two of the four invisible edges form part of the base and the remaining two remain to be located. For an octahedron with 11 vertices, one vertex must have 4 edges leading from it and the other 10 only 3, see Note 2. There are four different ways of placing the remaining 2 edges in the rear of the figure, resulting in four different types. One of these is shown in Figure 2C. The other three are not illustrated but can be visualized from Figure 2C as follows. In Figure 2C, remove dotted edge 9–11, retaining the three other dotted edges: for octahedron D, connect vertices 3 and 9 by a dotted line; for octahedron E, connect vertices 4 and 9 instead; and for octahedron F, connect vertices 7 and 9 instead. There are three additional ways in which the two new edges which do not form part of the base can be placed but these only result in reflections of D, E and F.

There are, then, six distinct types of polyhedra which are candidates for Dürer's polyhedron on the assumption that it is an octahedron. The word "type" is used to indicate that some squashing or stretching is permissible without destroying or changing the nature of the figure. As long as the number of edges, vertices and faces remain the same, with the types of the faces remaining the same (that is, a triangle is still a triangle, etc.) and the connections and relationships of the faces, vertices and edges remain the same, the figure is the same type. A selection is to be made from the six candidates and they will be described further.

Octahedron A, Figure 2A, has 2 triangles and 6 pentagons for faces. It can be constructed so that the 2 triangles are congruent equilateral triangles and the 6 pentagons are congruent to each other (but not equilateral). When so constructed there would be 3 planes of symmetry about each of which the figure could be reflected and remain identical with its initial position; the 3 planes pass through one axis (which would be vertical if the body sits on one of the triangular faces) about which the figure could be rotated into two other positions in which it would be identical with its initial position. There would also be three other axes, lying in a plane midway between the top and bottom triangular faces, about each of which the body could be rotated 180° and coincide with its initial position.

Octahedron *B*, Figure 2B, has one triangle, 3 quadrilaterals, 3 pentagons, and one hexagon for faces. The body can be constructed so as to have a plane of symmetry; the plane would pass through vertices 1, 2 and bisect edges 3–10, 6–7, 11–12. Octahedron *C*, Figure 2C, has 2 triangles, 2 quadrilaterals, and 4 pentagons for faces. It, likewise, can be constructed so as to have a plane of symmetry; the plane would pass through vertices 1, 2, 11 and bisect edges 3–10 and 6–7. *B* and *C* do not have any symmetry other than the single one mentioned. *D* and *F*, not illustrated, each have 3 triangles, one quadrilateral, 3 pentagons and one hexagon for faces; *E* has 2 triangles, 2 quadrilaterals, and 4 pentagons. These three are not symmetrical.

The high degree of symmetry of octahedron *A*, in comparison with the others (the order of the automorphism group, which is a measure of the degree of symmetry, of *A* is 18, whereas for *B* and *C* it is but 2, and the remaining ones can have no symmetry), points to it as the most appropriate one for the artist to have had in mind.

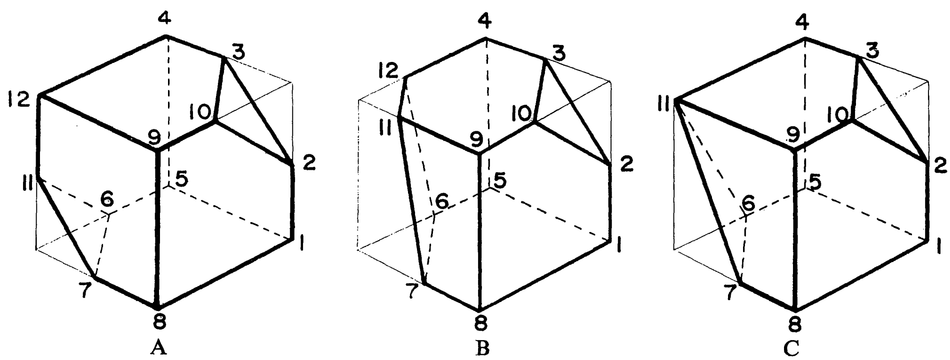


FIG. 3.

There is another feature of the three octahedra *A*, *B* and *C*, not possessed by *D*, *E* and *F*, which deserves mention. This is that each of the three can be made in a simple manner by sawing two pieces from a cube type hexahedron. This is shown in Figures 3A, 3B and 3C, corresponding to 2A, 2B and 2C. The vertices have been numbered so that corresponding figures can be compared readily, and the pieces to be discarded are shown in light lines. As seen in Figure 3A, *A* can be made by simply cutting two opposite corners from a cubical block and the two pieces cut off can be symmetrical and identical. If 3A is stood on face 6–7–11 as base, with face 1–5–6–7–8 in front, very little adjustment of length of lines would be needed to make it coincide with 2A. For 3C one of the cuts must pass through a vertex of the cube, as shown. If stood on face 6–7–11 as base with face 1–5–6–7–8 in front, some distortion would be needed to make it coincide with 2C. And for 3B, one of the cuts must remove a ridge rather than a pyramidal piece. If stood on face 6–7–11–12 as base with face 1–5–6–7–8 in front, considerable adjustment of lengths of lines would be needed to make it coincide

MELENCOLIA I, by Albrecht Dürer

(Courtesy of the National Gallery of Art, Washington, D. C. Gift of R. Horace Gallatin)

with 2B. The simplicity and regularity of the formation of A also point to this one as the most probable one to have been used.

The octahedron A does not appear to have any special name, but "Dürer's Octahedron" is appropriate, even though some subjective considerations are involved in selecting it as the prototype of the stone block of Melencolia I. The works on Dürer which have been consulted merely refer to the block as a "polyhedron", "a granite polyhedron", or a geometric solid, except for Panofsky [**9**]. He refers to it as a "truncated rhombohedron of stone" (page 156) and as a "truncated rhomboid" (page 161). While these terms do not completely identify any specific figure but would include a large number of different types, it is quite apparent that he had the same figure in mind as is suggested in this note, although he does not indicate how he arrived at his conclusion.

There has been a good deal of speculation on the symbolisms in Dürer's Melencolia I, but as for the octahedron the works consulted merely refer to its geometrical significance without any specific intrinsic symbolism relative to the mood of despondency of the engraving. Perhaps, as suggested by my wife, it might be intended for "the stone which the builders refused" of Psalm 118:22, or "the stone of stumbling" of Isaiah 8:14,15.

NOTE 1. If the number of edges on each face of a polyhedron are counted and added together, the result will be twice the number of edges in the body since each edge will have been counted twice. Each face must have at least 3 edges and hence $2E$ must be equal to or greater than $3F$:

$$2E \geqq 3F. \tag{1}$$

Similarly, $2E$ must be equal to or greater than $3V$ since each vertex must have at least 3 edges leading from it:

$$2E \geqq 3V. \tag{2}$$

Euler's equation is multiplied throughout by 2 and written as:

$$2E = 2V + 2F - 4. \tag{3}$$

If this value for $2E$ is substituted in equation (1) the result is:

$$2V + 2F - 4 \geqq 3F.$$

On adding 4 to each side, and subtracting $2F$ from each side, this reduces to:

$$2V \geqq F + 4. \tag{4}$$

Similarly, the substitution of (3) in equation (2) results in:

$$2F \geqq V + 4. \tag{5}$$

With F equal to 8, equation (4) shows that V must be equal to or greater than 6 and equation (5) shows that V must be equal to or less than 12.

NOTE 2. Let v_3 be the number of vertices which are 3-valent, that is, have 3 edges leading from them, v_4 the number which are 4-valent, etc. Two equations can be written:

$$V = v_3 + v_4 + v_5 + \cdots,$$

$$2E = 3v_3 + 4v_4 + 5v_5 + \cdots.$$

For octahedra with 12 vertices, which would have 18 edges, these two equations become:

$$12 = v_3 + v_4 + v_5 + \cdots,$$

$$36 = 3v_3 + 4v_4 + 5v_5 + \cdots.$$

These have only one solution for the v's (which must be positive integers or zero), namely, $v_3 = 12$, with the remaining ones all zero.

For octahedra with 11 vertices, which would have 17 edges, the two equations become:

$$11 = v_3 + v_4 + v_5 + \cdots,$$

$$34 = 3v_3 + 4v_4 + 5v_5 + \cdots.$$

These two equations have only one solution in positive integers and zeros for the v's, namely, $v_3 = 10$, $v_4 = 1$ and the remaining ones all zero. (This is readily seen if the first equation is multiplied throughout by 3 and subtracted from the second.)

References

1. D. E. Smith, History of Mathematics, vol. II, Ginn, Boston, 1925 p. 598, (also paperback reprint).
2. Martin Gardner, The Second Scientific American Book of Mathematical Puzzles and Diversions, Simon and Schuster, New York, 1961, pp. 132–4 (also paperback reprint).
3. Elizabeth H. Agnew, Two Problems on Magic Squares, this MAGAZINE, 44 (1971) 13–15.
4. Kenneth Clark, Civilisation, A Personal View, Harper & Row, New York, 1970, pp. 152, 155.
5. Karl-Adolf Knappe, Dürer, The Complete Engravings, Etchings, and Woodcuts, H. N. Adams, New York, 1965, pp. 42–44.
6. Michael Levey, Dürer, Norton, New York, 1964, pp. 106–7.
7. Marcel Brion, Dürer, His Life and Work, Tudor, New York, 1960, pp. 206–212.
8. Francis Russell, The World of Dürer, Time, Inc., New York, 1967, pp. 11–13.
9. Erwin Panofsky, The Life and Art of Albrecht Dürer, 4th ed., Princeton University Press, 1955 (also paperback reprint, 1971). Pages 156–171 discuss the Melencolia; part of chapter 8, which contains a review of Dürer's book on geometry, is reprinted under the title "Dürer as a Mathematician" in The World of Mathematics, James R. Newman, ed., Simon & Schuster, New York, 1956, pages 603–621.

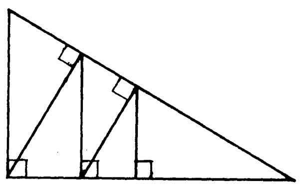
PARTITIONING A TRIANGLE INTO 5 TRIANGLES SIMILAR TO IT

ZALMAN USISKIN, University of Chicago and
STANLEY G. WAYMENT, Utah State University

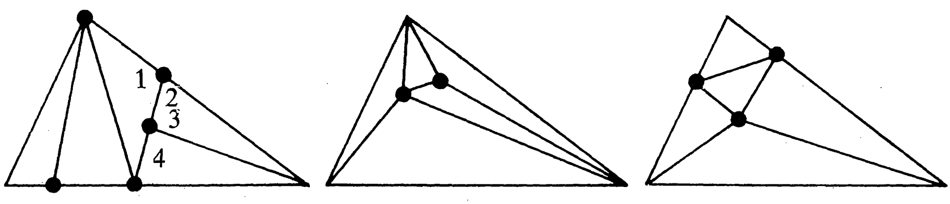
When can a triangle be partitioned into a given number n of triangles similar to it? This question was posed and answered by Freese, Miller, and Usiskin [1] for $n \neq 5$.

Specifically, it was found for $n = 4$ and $n \geqq 6$, any triangle can be partitioned into n triangles similar to it. For $n = 2$ and $n = 3$, such a partitioning can be accomplished only if the given triangle is a right triangle.

For $n = 5$, a *right* triangle can be partitioned in the desired manner, as illustrated here. We now consider whether any other triangle can be partitioned into 5 triangles similar to it, and obtain the remarkable result that there is precisely one other triangle with this property.



There are many essentially different ways of partitioning a triangle into 5 triangles. Several of these are drawn here.



(a) (b) (c)

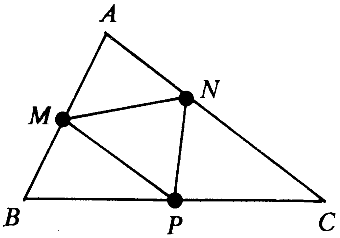
Call the original triangle ABC. Let each point other than A, B, or C which is a vertex of a triangle of the partition be called a p-point.

LEMMA 1. *Each p-point is a vertex of at least* 3 *triangles of the partition.*

Proof. If the p-point is on a side of $\triangle ABC$ and is only a vertex of two triangles, then a situation occurs as with angles 1 and 2 in drawing (a) above. These angles are congruent to two angles of $\triangle ABC$. If congruent to the same angle, then $\triangle ABC$ is a right triangle. If congruent to different angles, then two angles of $\triangle ABC$ are supplementary—this is impossible.

Angles 3 and 4 in drawing (a) (p. 37) show a potential p-point which is a vertex of only 2 triangles of the partition and is interior to $\triangle ABC$. This yields the same situation as with angles 1 and 2. The angles about the p-point must otherwise add to 360, and hence, if these are to be angles of a triangular partition, there must be at least 3 angles.

As there are 5 small triangles, they together have 15 vertices, of which at least three are the points A, B, and C. Hence p-points account for at most 12 vertices. By Lemma 1, each p-point accounts for at least 3 vertices. Therefore there are at most 4 p-points.



When there are exactly 4 p-points, A, B and C account for only 3 vertices of the triangles of the partition. Hence they cannot be divided in the partition. It is quickly seen that a p-point must lie on each side of $\triangle ABC$, as above. But there is no place for a 4th p-point Q. If Q is exterior to $\triangle MNP$, then since Q cannot be connected to A, B, or C (this would make A, B, or C between them account for more than 3 vertices), no partition occurs. If Q is on a side of $\triangle MNP$, then a situation violating Lemma 1 occurs. If Q is interior to $\triangle MNP$, then it must be connected to all three vertices in order to form a partition, and we then have six triangles. This proves:

LEMMA 2. *There are at most* 3 *p-points.*

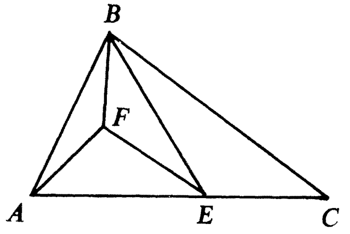
In figure (b) earlier in the article, a figure is given in which all three vertices of the original triangle are divided in the partition. In this case it is impossible for all of the triangles to be similar to the original triangle, as proved in the next lemma.

LEMMA 3. *The vertices A, B, and C cannot all be vertices of more than one triangle of the desired partition.*

Proof. If all angles are divided by the partition, then a smallest angle is so divided. Hence a triangle of the partition contains an angle smaller than any angle of the original triangle. This triangle cannot be similar to the original.

LEMMA 4. *There are at least* 3 *p-points.*

Proof. Suppose there are exactly 2 p-points E and F in the interior of $\triangle ABC$. They cannot be connected to all three vertices of triangle ABC, so suppose they are connected to A and B only. The only triangles possible are ABE, ABF, BEF, and AEF. Hence there are not five triangles for a partition. It is easy to see that the two points cannot both lie on sides of the triangle. Finally, suppose one of them, say E, lies on a side, say AC. Then it must be connected both to B and to F or Lemma 1 is violated. F can be connected to two of the vertices but not the third, so five triangles again cannot be formed.



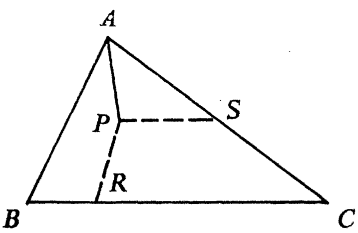
Combining Lemmas 2 and 4:

LEMMA 5. *If there is to be a partition of a nonright triangle into five triangles similar to it, then there are six points which are vertices of the partition.*

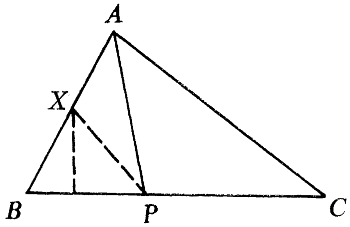
Three of the six points are A, B, and C; the other three are p-points. The 3 p-points account for at least 9 vertices of the 15 needed in the partition. Hence A, B, and C account for at most 6 vertices. If we adopt the convention that angle C will be the smallest angle of the triangle, then angle C cannot be divided (Lemma 3) and so the two points A and B account for a maximum of 5 vertices.

Let a denote the number of vertices accounted for by point A; let b denote the number of vertices accounted for by B. Then there exist the following possibilities for (a, b): (1, 1), (2, 1), (3, 1), (4, 1), (2, 2), and (3, 2). (We ignore (1, 2) as being identical to (2, 1), etc.)

Case 1. (a, b) = (1, 1). Here neither angle A nor angle B is divided. The situation is pictured in the drawing preceding Lemma 2, where such a situation was shown to be impossible.

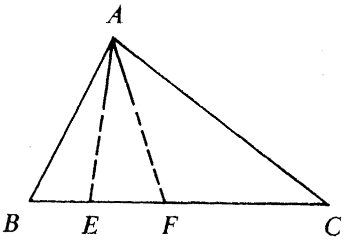


(2, 1). Here angle A is divided by one segment, angle B by egment AP divides A, and P is interior to $\triangle ABC$. Since P must be the vertex of at least three triangles, there are at least two additional segments PR and PS emanating from P. Both R and S must lie on sides of $\triangle ABC$, for there are at most 3 p-points. If R and S lie on the same side of $\triangle ABC$ then another p-point is required to form a partition, contrary to Lemma 5. In fact, if either APR or APS is not a straight line, then another p-point is required to form a partition, again contradicting Lemma 5. However, if APR is a straight line, then another segment PT must emanate from P, and P, R, S, and T are p-points contradicting Lemma 2. Hence we conclude P lies on BC.

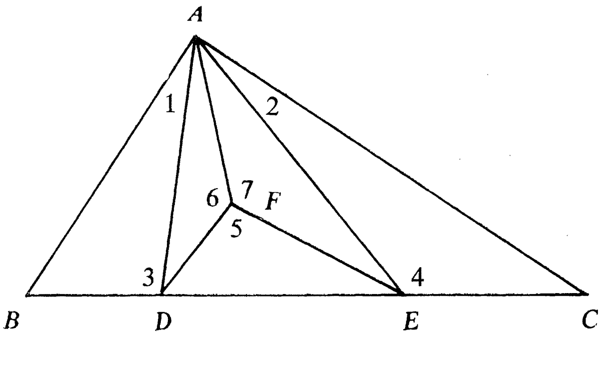


By Lemma 1, another segment must be drawn from P (whether towards AB or AC is immaterial). By an argument similar to the previous one, this segment must intersect a side of the triangle in a p-point F. Then still a third segment must be drawn from F intersecting BP (in the drawing above). And so the process continues, and it is impossible to obtain exactly 3 p-points which do not violate Lemma 1.

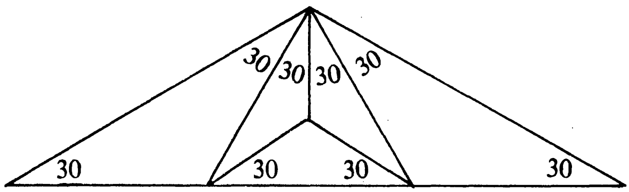
Case 3. (a, b) = (3, 1). Here two segments divide angle A, no segment divides angle B. The two segments must each extend to BC, and an impossible situation results again as in Case 2.



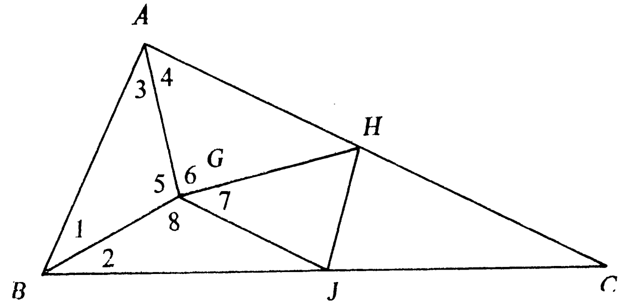
Case 4. (a, b) = (4, 1). Here three segments divide angle A while none divide angle B. The two outer dividing segments must intersect BC but the third cannot. (If it did, the 3 p-points are collinear and 5 triangles cannot be formed.) This determines the situation drawn below. Suppose the triangles are similar. Then, in order, the following conclusions can be reached; $\angle 1 = \angle C$ and $\angle 2 = \angle B$; $\angle A$ is



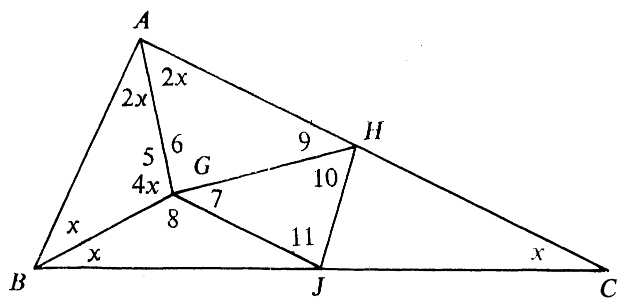
obtuse; $\angle 3 = \angle A$ and $\angle 4 = \angle A$; $\angle 5 = \angle 6 = \angle 7 = 120°$ (otherwise two angles about F add to less than 180° and there is no partition); $\angle A = 120°$. Then $\angle ADE = \angle AED = 60°$. So $\triangle ADE$ is equilateral. It is known that there is only one point X in a triangle PQR (the Fermat point) such that $\angle PXQ = \angle QXR = \angle RXP$. Hence F is uniquely determined and the 5 triangles must have angles 30–30–120. An accurate drawing indicates the similarity of the triangles in the partition.



Case 5. (a, b) = (2, 2). The two angles A and B are each split by one segment. These must intersect at a p-point G. Segments must also extend from G to AC and BC, yielding p-points H and J. Connecting H to J yields a second configuration which does not violate previous constrictions. This configuration is symmetric in

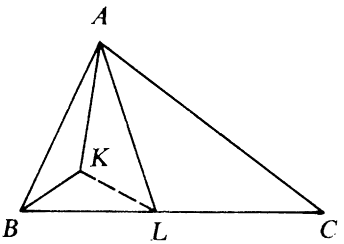


A and B. Suppose $\angle A \geqq \angle B \geqq \angle C$. Since $\angle B$ is divided, we have $\angle B > \angle C$. If $\angle 3 + \angle 4 = \angle B + \angle C$, then $\angle A = 90$ and right triangles are known to be partitionable. Hence we suppose $\angle A$ is not a right angle, and so $\angle 3 = \angle 4$. Since $\angle 1 = \angle 2 = \angle C$, we conclude $\angle 3 = \angle 4 = \angle B$; hence $\angle A > \angle B$. More precisely, the triangle is seen to have angles x, $2x$, and $4x$. Hence $7x = 180$ and the angles about G must add to $14x$. Of the angles about G, immediately $\angle 5 = 2x$, and two of angles 6, 7, and 8 must have measure $4x$, the other having measure $2x$. Immediately $\angle 6 = 4x$ and hence $\angle 9 = x$.



Note that $\triangle ABG$ is congruent to $\triangle AHG$ (by Angle-Side-Angle); hence $BG = GH$. Now recall that the shortest side of a triangle is opposite the smallest angle. Each triangle in the partition must have angles with measure x, $2x$, $4x$. It is immediately seen that in $\triangle BGJ$, $GJ < BG$. Now $\angle 10 \neq x$, so $\angle 11 = x$ and GH is the shortest side of $\triangle GHJ$. Hence $GH < GJ < BG$. This contradicts $GH = BG$ and shows that this case is impossible.

Case 6. $(a, b) = (3, 2)$. Angle A is divided by 2 segments. One of these segments intersects the segment dividing B; the other must extend to BC. The two p-points K and L formed in this way must be connected. There are now 4 triangles; one of these must be divided to form the fifth triangle, and it is impossible to do so without violating Lemma 1.



Summary. The question has been posed: When can a triangle be partitioned into n triangles similar to it? Previous results had shown that for $n \neq 5$, such a partition can be done only with right triangles or with all triangles, depending upon the particular value of n. In this note, the case $n = 5$ has been considered and shown to be oddball. We have proved [2]:

THEOREM. *A triangle can be partitioned into* 5 *triangles similar to it if and only if the triangle is a right triangle or the angles of the triangle are* 30°, 30° *and* 120°.

Generalizations. The oddity of the case $n = 5$ makes one curious to consider generalizations. For example, which triangles can be partitioned into n similar triangles, with no restriction on their similarity to the original triangle? For which n can a quadrilateral be partitioned into n quadrilaterals similar to each other and/or to it? For which n can a (convex?) k-gon be partitioned into n similar k-gons? There seem not to be as many solutions to this last question as one might at first expect.

References

1. R. W. Freese, Ann K. Miller, and Zalman Usiskin, Can every triangle be divided into n triangles similar to it?, Amer. Math. Monthly, 77 (1970) 867–869.

2. While this article was in preparation, an outline of a slightly different proof of the same result was conveyed to the authors by James Kiefer. Other solutions or partial solutions by Peter Kornya, Jim Morris, W. D. Bouwsma, and Ray Killgrove have also been called to our attention.

SOME INEQUALITIES FOR TWO TRIANGLES

L. CARLITZ, Duke University

1. Let a, b, c denote the sides of the triangle ABC and let a', b', c' denote the sides of the triangle $A'B'C'$. Let F, F' denote the respective areas. Pedoe has proved that

$$(1) \qquad a^2(-a'^2 + b'^2 + c'^2) + b^2(a'^2 - b'^2 + c'^2) + c^2(a'^2 + b'^2 - c'^2) \geqq 16FF'$$

with equality if and only if the triangles ABC, $A'B'C'$ are similar. For related results and references see [**4**]. The present writer [**1**] has given a simple algebraic proof of (1).

It may be of interest to point out that (1) can be proved very rapidly by making use of an inequality of Aczél [**3**, p. 52]. Let $a = (a_1, \cdots, a_n)$ and $b = (b_1, \cdots, b_n)$ be two sequences of real numbers, such that

$$(2) \qquad a_1^2 - a_2^2 - \cdots - a_n^2 > 0 \quad \text{or} \quad b_1^2 - b_2^2 - \cdots - b_n^2 > 0.$$

Then

$$(3) \qquad (a_1^2 - a_2^2 - \cdots - a_n^2)(b_1^2 - b_2^2 - \cdots - b_n^2) \leqq (a_1b_1 - a_2b_2 - \cdots - a_nb_n)^2$$

with equality if and only if the sequences a and b are proportional.

We remark that if we replace (2) by

$$(2)' \qquad a_1^2 - a_2^2 - \cdots - a_n^2 > 0 \text{ and } b_1^2 - b_2^2 - \cdots - b_n^2 > 0$$

then

$$(4) \qquad a_1b_1 > a_2b_2 + \cdots + a_nb_n.$$

For, if we assume that

$$a_1b_1 \leqq a_2b_2 + \cdots + a_nb_n,$$

then

$$a_1^2b_1^2 \leqq (a_2^2 + \cdots + a_n^2)(b_2^2 + \cdots + b_n^2) < a_1^2b_1^2.$$

To prove (1) we take

$$\begin{aligned} 16F^2 &= 2b^2c^2 + 2c^2a^2 + 2a^2b^2 - a^4 - b^4 - c^4 \\ &= (a^2 + b^2 + c^2)^2 - 2(a^4 + b^4 + c^4), \\ 16F'^2 &= (a'^2 + b'^2 + c'^2)^2 - 2(a'^4 + b'^4 + c'^4). \end{aligned}$$

Now put

$$\begin{aligned} &a_1 = a^2 + b^2 + c^2, \quad a_2 = 2^{\frac{1}{2}}a^2, \quad a_3 = 2^{\frac{1}{2}}b^2, \quad a_4 = 2^{\frac{1}{2}}c^2, \\ &b_1 = a'^2 + b'^2 + c'^2, \quad b_2 = 2^{\frac{1}{2}}a'^2, \quad b_3 = 2^{\frac{1}{2}}b'^2, \quad b_4 = 2^{\frac{1}{2}}c'^2. \end{aligned}$$

Then (2)$'$ holds and (4) becomes

$$(a^2 + b^2 + c^2)(a'^2 + b'^2 + c'^2) > 2a^2a'^2 + 2b^2b'^2 + 2c^2c'^2,$$

that is,

(5) $$a^2(-a'^2+b'^2+c'^2)+b^2(a'^2-b'^2+c'^2)+c^2(a'^2+b'^2-c'^2)>0.$$

Also (3) becomes

(6) $$256F^2F'^2 \leqq \{a^2(-a'^2+b'^2+c'^2)+b^2(a'^2-b'^2+c'^2)+c^2(a'^2+b'^2-c'^2)\}^2.$$

In view of (5), it is clear that (6) implies (1).

2. By means of Aczél's inequality we can also prove other inequalities for two triangles. For example if O is the circumcenter, G the centroid and R the circumradius of ABC and similarly for O', G', R', it is known [**2**, p. 71] that

$$R^2-\frac{1}{9}(a^2+b^2+c^2)=OG^2,\ R'^2-\frac{1}{9}(a'^2+b'^2+c'^2)=O'G'^2.$$

Therefore, by (3),

$$\left\{RR'-\frac{1}{9}(aa'+bb'+cc')\right\}^2 \geqq \left\{R^2-\frac{1}{9}(a^2+b^2+c^2)\right\}\left\{R'^2-\frac{1}{9}(a'^2+b'^2+c'^2)\right\}.$$

Now using (4), we get

(7) $$RR'-\frac{1}{9}(aa'+bb'+cc')\geqq OG\cdot O'G',$$

with equality if and only if the triangles are similar.

Next let I denote the incenter, I_a, I_b, I_c the excenters of ABC and similarly for I', I'_a, I'_b, I'_c. It is known [**2**, p. 87] that

$$12R^2-OI_a^2-OI_b^2-OI_c^2=OI^2,\ 12R'^2-O'I_a'^2-O'I_b'^2-O'I_c'^2=O'I'^2.$$

Therefore, by (3) and (4), we get

(8) $$12RR'-OI_a\cdot O'I'_a-OI_b\cdot O'I'_b-OI_c\cdot O'I'_c\geqq OI\cdot O'I',$$

with equality if and only if the triangles are similar.

Supported in part by NSF grant GP–17071.

References

1. L. Carlitz, An equality involving the area of two triangles, Amer. Math. Monthly, 78 (1971) 772.

2. N. A. Court, College Geometry, 2nd ed., Barnes and Noble, New York, 1952.

3. D. S. Mitrinović, Analytic Inequalities, Springer, Berlin-Heidelberg-New York, 1970.

4. D. Pedoe, Thinking geometrically, Amer. Math. Monthly, 77 (1970) 711–721.

A NOTE ON THE GRAPHS OF GROUPS, I

BILLY E. MILNER, Washburn University of Topeka

Herein, the graph of a group is in the sense of Steven Bryant, that is, a directed graph mapping each element of the group to its square. Bryant [1] proved that Abelian groups with fewer than $(1093)^2$ elements are isomorphic if they have isomorphic graphs. In fact, he established that the preceding result held for Abelian groups whose order is not divisible by the square of any bad prime. (A bad prime is a prime p satisfying $2^{p-1} \equiv 1 \pmod{p^2}$ and 1093 is the first such prime.) It may be of interest to note there exist groups of much smaller order than $(1093)^2$ which illustrate that Bryant's result does not hold for arbitrary groups.

The smallest order for which there exist nonisomorphic groups that have isomorphic graphs is sixteen. Of the fourteen groups of order sixteen, five merit comment on the basis of their graphs. Let

$$G_1 = Z_2 \times Z_8, \qquad G_2 = Z_2 \times Z_2 \times Z_4,$$

$$G_3 = \langle a,b,c,d \mid a^2 = b^2 = e,\ c^2 = d,\ d^2 = a,\ b^{-1}c^{-1}bc = a \rangle,$$

$$G_4 = \langle a,b,c,d \mid a^2 = b^2 = c^2 = e,\ d^2 = a,\ b^{-1}c^{-1}bc = a \rangle,$$

$$G_5 = \langle a,b,c,d \mid a^2 = c,\ b^2 = d,\ a^{-1}b^{-1}ab = c,\ c^2 = d^2 = e \rangle.$$

Then the graph of G_1 is isomorphic to the graph of G_3, but G_1 is Abelian and G_3 is non-Abelian and, hence, are nonisomorphic. Similarly, the graph of G_2 is isomorphic to the graph of G_4 while G_2 and G_4 are not isomorphic. Thus, we have two cases where an Abelian group has the same graph as a non-Abelian group.

With the above graph isomorphisms, two examples of two non-Abelian groups of order ninety-six can be obtained.

Letting $H_i = G_i \times \mathrm{Sym}(3)$, where $i = 1,2,3,4$, H_1 and H_3 are nonisomorphic non-Abelian groups which have isomorphic graphs. A similar statement can be made for H_2 and H_4.

The fifth group of order sixteen listed, G_5, is worth noting since it is the first group whose graph suffers a loss of the complete symmetry apparent in the graphs of other groups.

The elements c, d and cd of G_5 are the only elements which map to the identity e (other than e itself.). But c has four preimages, d has eight preimages, and cd has no preimage. Thus, none of the branches are isomorphic. If $G \neq G_5$ is a group of order less than eighteen and the graph of G has more than one branch then at least two of the branches are isomorphic.

Reference

1. Steven Bryant, Groups, graphs, and Fermat's last theorem, Amer. Math. Monthly, 74 (1967) 152–156.

PROBLEMS AND SOLUTIONS

EDITED BY ROBERT E. HORTON, Los Angeles Valley College

ASSOCIATE EDITOR, J. SUTHERLAND FRAME, Michigan State University

Readers of this department are invited to submit for solution problems believed to be new that may arise in study, in research, or in extra-academic situations. Problems may be submitted from any branch of mathematics and ranging in subject content from that accessible to the talented high school student to problems challenging to the professional mathematician. Proposals should be accompanied by solutions, when available, and by any information that will assist the editor. Ordinarily, problems in well-known textbooks should not be submitted.

The asterisk (*) will be placed by the problem number to indicate that the proposer did not supply a solution. Readers' solutions are solicited for all problems proposed. Proposers' solutions may not be "best possible" and solutions by others will be given preference.

Solutions should be submitted on separate, signed sheets. Figures should be drawn in India ink and exactly the size desired for reproduction.

Send all communications for this department to Robert E. Horton, Los Angeles Valley College, 5800 Fulton Avenue, Van Nuys, California 94101.

To be considered for publication, solutions should be mailed before July 15, 1972.

PROPOSALS

817. *Proposed by Charles W. Trigg. San Diego, California.*

In the following cryptarithm, each letter represents a distinct decimal digit, $3(B\ I\ D\ F\ O\ R) = 4(F\ O\ R\ B\ I\ D)$. Find the digits.

818. *Proposed by W. K. Viertel, SUNY Agricultural and Technical College, Canton, New York.*

Find the angle of repose of a rod of length d and of negligible thickness, when placed in a hemispherical bowl of diameter d. Assume no friction.

819. *Proposed by Elkedagmar Heinrich, Frankfurt, Germany.*

Find a formula for $\sum_{n=0}^{m} n^k$, k any positive integer.

820. *Proposed by Ronald Alter, University of Kentucky.*

Given an ordinary poker deck of 52 cards, choose 25 cards from the deck at random.

a. What is the probability of getting five full houses?

b. What is the maximum number of different arrangements of five full houses that could occur?

821. *Proposed by Rina Rubenfeld, New York City Community College.*

A and B are two given points. Looking at them as two vertices of a square, find the other two vertices of the square using a compass only.

822. *Proposed by C. S. Venkataraman, Sree Kerala Varma College, Trichur, South India.*

Given a_k $(k = 1,2,\cdots,n)$ is zero or any positive integer and A_n is their arithmetic mean, prove that $\prod_{k=1}^{n} a_k! \geqq [A_n]^n!$ with equality arising when $n = 1$ or when all the a's are equal, and where $[x]$ is the integral part of x.

823. *Proposed by Ira Gessel, Dayton, Ohio.*

Find a function f such that $\Sigma_{k=1}^{m} f(k)[m/k] = m(m+1)/2$.

QUICKIES

From time to time this department will publish problems which may be solved by laborious methods, but with the proper insight may be disposed of with dispatch. Readers are urged to submit their favorite problems of this type, together with the elegant solution and the source, if known.

Q532. Without evaluating either of the following integrals show why

$$2\int_{-1}^{1} \sqrt{1-x^2}\, dx = \int_{-1}^{1} 1/\sqrt{1-x^2}\, dx.$$

[*Submitted by Peter A. Lindstrom*]

Q533. In the finite field Z_{11}, the set $\{1,3,4,5,9\}$ is a subgroup under multiplication. Does this fact violate Lagrange's theorem?

[*Submitted by Eugene McGovern*]

Q534. If the probability of winning a game is p, what is the probability of winning the best $(n+1)$ games of $(2m+1)$?

[*Submitted by John M. Howell*]

Q535. Is e^{2737} approximately equal to $1000!/70!\,270!\,300!\,220!\,140!$? An engineering student had to answer this question in order to receive credit for his answer (the latter number) when the correct answer was e^{2737}

[*Submitted by Eugene Wermer*]

Q536. Show that the square roots of three distinct prime numbers cannot be terms of a common geometric progression.

[*Submitted by Murray S. Klamkin*]

(Answers on page 56)

SOLUTIONS

Distinct Prime Divisors

788. [March, 1971] *Proposed by R. S. Luthar, University of Wisconsin, Waukesha.*

Show that $(p_1 \cdot p_1 \cdot p_3 \cdots p_n + 1)^{2^k} - 1$ has at least $n + k$ distinct prime divisors, where $p_1, p_2 \ldots p_n$ are the first n primes.

Editor's note. A number of solvers indicated that as stated the problem is false. A correct statement would change the wording to read "... where $p_1, p_2 \dots p_n$ are the first n odd primes".

Solution by Albert J. Patsche, U.S. Army Weapons Command, Rock Island Arsenal, Illinois.

Let $N = (p_1p_2p_3 \dots p_n + 1)^{2^k} - 1$. Then because $a^{2^k} - 1 = (a-1)(a+1)(a^2+1)(a^{2^2}+1)\dots(a^{2^{k-1}}+1)$, $N = (p_1p_2 \dots p_n)[(p_1p_2 \dots p_n + 1) + 1][p_1p_2 \dots p_n + 1)^2 + 1] \dots [(p_1p_2 \dots p_n + 1)^{2^{k-1}} + 1]$. Next, denote $(p_1p_2 \dots p_n + 1)^r + 1$ by N_r. And since $(b+1)^r = \sum_{j=0}^{r} \binom{r}{j} b^{r-j}$, $N_r = \sum_{j=0}^{r-1} \binom{r}{j} (p_1p_2 \dots p_n)^{r-j} + 2$. Also, because $p_i (i = 1, 2, \dots n)$ divides $\sum_{j=0}^{r-1} \binom{r}{j} (p_1p_2 \dots p_n)^{r-j}$ but does not divide 2, p_i does not divide N_r. Hence, N_r must have at least one prime factor distinct from p_1 through p_n, call it P_r; and $P_r \neq 2$ since N_r is odd. Now consider $M = N_{2^{t+s}} + N_{2^t}$ where $t \geqq 0$ and $s \geqq 1$. Then, $M = (p_1p_2 \dots p_n + 1)^{2^{t+s}} + (p_1p_2 \dots p_n + 1)^{2^t} + 2 = (p_1p_2 \dots p_n + 1)^{2^t}\{(p_1p_2 \dots p_n + 1)^{(2^s-1)2^t} + 1\} + 2$. Upon factoring, $M = (p_1p_2 \dots p_n + 1)^{2^t}\{[(p_1p_2 \dots p_n + 1)^{2^t} + 1][\sum_{i=2}^{2^s}(-1)^i(p_1p_2 \dots p_n + 1)^{(2^s-i)2^t}]\} + 2 = (p_1p_2 \dots p_n + 1)^{2^t} N_{2^t} [\sum_{i=2}^{2^s}(-1)^i(p_1p_2 \dots p_n + 1)^{(2^s-1)2^t}] + 2$. Now, there exists an odd prime P_{2^t} distinct from p_1 through p_n such that $P_{2^t} \mid N_{2^t}$. Hence, $P_{2^t} \nmid M$; therefore $P_{2^t} \nmid N_{2^{t+s}}$. Thus, for each s (resp. t) $= 0, 1, \dots k-1$, N_{2^s} (resp. N_{2^t}) has a prime divisor P_{2^s} (resp. P_{2^t}) distinct from the p_i's. And, $P_{2^s} \neq P_{2^t}$ whenever $s \neq t$. Therefore the $n+k$ primes $p_1p_2 \dots p_n, P_1, P_2, P_{2^2}, \dots P_{2^{k-1}}$ are all divisors of $N = (\prod_{i=1}^{n} p_i)(\prod_{j=0}^{k-1} N_{2^j})$, and are all distinct.

Also solved by Eugen Peter Bauhoff, Frankfurt, Germany; David C. Brooks, Seattle Pacific College, Washington; William F. Fox, Moberly Junior College, Missouri; M. G. Greening, University of New South Wales, Australia; C. Bruce Myers, Austin Peay State University, Tennessee; Erik Laage-Petersen, Alborg, Denmark; S N. Rao, Western Michigan University; E. P. Starke, Plainfield, New Jersey; Burnett R. Toskey, Seattle University, Washington; Philip Tracy, Liverpool, New York; Kenneth M. Wilke, Topeka, Kansas; and the proposer.

A "Loot Begging" Cryptarithm

789. [March, 1971] *Proposed by Edwin P. McCravy, Midlands Technical Education Center, Columbia, South Carolina.*

Show that this "loot begging" cryptarithm

$$\begin{array}{r} S\ E\ N\ D \\ M\ O\ R\ E \\ \hline M\ O\ N\ E\ Y \end{array}$$

has exactly $\binom{b-8}{2}$ solutions in the base b.

Solution by V. S. Blanco, University of South Alabama.

From the cryptarithm we get immediately that $M = 1$, $O = 0$, $S = b - 1$, $E + 1 = N$, and $N + R + 1 = E + b$ or $N + R = E + b$, which by substitution of N lead to $R = b - 2$ or $R = b - 1$, but the latter is the value of S, therefore it is deleted and $R = b - 2$.

The cryptarithm changes to

$$\begin{array}{cccccc} & b-1 & E & E+1 & D \\ & & 1 & 0 & b-2 & E \\ \hline 1 & 0^* & E+1 & E^* & Y^* \end{array} \quad (\text{* carries one}).$$

From the rightest column we have (1) $D + E = Y + b$. Furthermore, since $S = b - 1$ and $R = b - 2$, then $D \leqq b - 3$, $N \leqq b - 3$, and (2) $E \leqq b - 4$. Subtracting (2) from (1) we have $D \geqq Y + 4 > Y$, then $Y \leqq 2$. Hence if $Y = 2$, then $D \geqq 6$ and D's range should be $b - 3 - 6 + 1 = b - 8$, but D cannot take an intermediate value between the consecutive numbers E and N, consequently the number of values of D, which is the number of solutions for $Y = 2$, is $b - 9$. For $Y = 3$, D has $b - 10$ values and will decrease in an arithmetic progression to 1 as Y increases.

Therefore, the total number of solutions is given by the sum of the A.P. $(b - 9) + (b - 10) + \cdots + 1 = (b - 9 + 1)(b - 9)/2 = \binom{b-8}{2}$.

In the base 10 the unique solution is

$$\begin{array}{ccccc} & 9 & 5 & 6 & 7 \\ & 1 & 0 & 8 & 5 \\ \hline 1 & 0 & 6 & 5 & 2 \end{array}$$

As another example, in base 12 there are $\binom{12-8}{2} = 6$ solutions with fixed values $O = 0$, $M = 1$, $R = \alpha$, $S = \beta$, and the table

Y	D	E	N
2	6	8	9
2	8	6	7
2	9	5	6
3	7	8	9
3	9	6	7
4	9	7	8

Also solved by David S. Gilliam, Idaho State University; M. G. Greening, University of New South Wales, Australia; John M. Howell, Littlerock, California; Burnett R. Toskey, Seattle University, Washington; Philip Tracy, Liverpool, New York; Charles W. Trigg, San Diego, California; and the proposer.

Trigg pointed out that the unique decimal solution appeared in the AMERICAN MATHEMATICAL MONTHLY, 40 (March, 1933) and elsewhere.

A Special Triangle

790. [March, 1971] *Proposed by Stanley Rabinowitz, Far Rockaway, New York.*

(1) Find all triangles ABC such that the median to side a, the bisector of angle B, and the altitude to side c are concurrent.

(2) Find all such triangles with integral sides.

I. *Solution by Leon Bankoff, Los Angeles, California.*

(1) Let D, E, F denote the feet of the concurrent cevians m_a, t_b, h_c issuing from A, B, C respectively. By Ceva's theorem, $AF \cdot BD \cdot CE = FB \cdot DC \cdot EA$, or $AF/FB = EA/CE = AB/CB$. The trigonometric equivalents of these identities are $\sin C/\sin A = \tan B/\tan A$ or briefly, $\sin C = \tan B \cos A$ (with absolute values applying throughout). This identity encompasses the necessary and sufficient conditions for the concurrence of the median to side a, the bisector of angle B, and the altitude to side c.

(2) It is known that all triangles having integer sides must contain one or another of the angles 60°, 90° or 120°. For example, a triangle with sides (3, 4, 5) has a 90° angle; one with sides (1, 1, 1) has all angles equal to 60°. A triangle with sides (3, 8, 7) has a 60° angle between side 3 and side 8 while a (3, 5, 7) triangle has a 120° angle between sides 3 and 5. (See problem 238, in the USSR Olympiad Problem Book, W. H. Freeman and Company, San Francisco, 1962, p. 347). Allowing A, B, C in turn to equal 60°, 90 or 120° subject to the restriction $\sin C/\sin B = \cos A/\cos B$ noted in part (1), we note the following results:

If either A, B, or C is equal to 60°, the triangle must be equilateral because the above restriction implies either $C \geqq B \geqq A$ or $C \leqq B \leqq A$. Thus, if either A or $C = 60°$, it is clear that the triangle is equilateral. If $B = 60°$, the only solution of $\sin C/\sin A = \sqrt{3}$ is $C = 60°$, $A = 60°$.

Suppose one of the three angles of triangle ABC is a right angle. If $A = 90°$, $\cos A = 0$, resulting in a degenerate triangle. If $B = 90°$, the order of magnitude designating B as the medial angle is violated. If $C = 90°$, the relation $\tan B = \sec A$ implies $\tan B \sin B = 1$, which leads easily to the solution $\cos B = 1/2(\sqrt{5} - 1)$. This result is impossible in a right triangle with integer sides.

If $A = 120°$ the Cevians m_a and t_b meet within the triangle while h_c lies outside the triangle, thus precluding concurrency. If $B = 120°$, the order of magnitude for C, B, A is violated. If $C = 120°$, the unique solution for $\sin A/\sin B$, subject to the constraint $\sin C = \tan B \cos A$, is irrational. Hence integer triangles with a 120° angle cannot meet the requirements for concurrency.

In conclusion, the equilateral triangle is the only integral triangle in which m_a, t_b and h_c are concurrent.

II. *Solution by Charles W. Trigg, San Diego, California.*

By Ceva's theorem and its converse, the necessary and sufficient condition that the three lines be concurrent is $(b \cos A)[ab/(a + c)](a/2) = (\pm a \cos B)[cb/(a + c)](a/2)$. Simplified, this becomes

$$\cos A = \pm \cos B\,(c/b) = \pm \cos B \sin C/\sin B.$$

Whereupon, $\pm \tan B = \sin C/\cos A$, depending upon whether B is acute or obtuse. In the latter case, the bisector is external [this MAGAZINE, 23 (May, 1950), 275–276].

When the cosines are replaced by their equivalents in terms of the sides, the necessary and sufficient condition takes the form

$$b^2 = a^2 + c^2 - 2ac^2/(a + c) = a^2 - c^2 + 2c^3/(a + c).$$

If the triangle has commensurable sides, and if the proper unit of measurement is chosen, a, b, c will be integers with no common factor. Moreover, a and c must be relatively prime, since any common factor of a and c would divide b. Hence $c^3/(a + c)$ cannot be an integer. The only way to make b an integer is to put $a = c = 1$, in which case the triangle is equilateral. In all other cases the three sides are incommensurable.

By putting $-c$ for c, the same argument applies to the case when the external bisector of angle B is used [AMERICAN MATHEMATICAL MONTHLY, 47 (March, 1940), 176].

The construction of the triangle when the bisector is internal is dealt with in the AMERICAN MATHEMATICAL MONTHLY, 44 (November, 1937), 599–600. For the construction when the bisector is external, see NATIONAL MATHEMATICS MAGAZINE, 15 (December, 1940), 149; 18 (November, 1943), 91.

Also solved by Michael Goldberg, Washington, D.C.; M.G. Greening, University of New South Wales, Australia; Ralph Jones, University of Massachusetts; Rina Rubenfeld, New York City Community College; E. P. Starke, Plainfield, New Jersey; and the proposer.

Solutions of a Cubic

791. [March, 1971] *Proposed by D. Rameswar Rao, Secunderabad, India.*

Show that the only solution in positive integers of $x^3 + y^3 + z^3 = u^3$ with x, y, z, u in arithmetic progression is $x = 3$, $y = 4$, $z = 5$, and $u = 6$.

Solution by David A. Rosen, Buffalo, New York.

Let $x = a - d$, $y = a$, $z = a + d$, and $u = a + 2d$.

Then $(a - d)^3 + a^3 + (a + d)^3 = (a + 2d)^3$. Expansion of the cubed binomials and combination of the like terms thus evolved yields the equation $2a^3 - 6a^2d - 6ad^2 - 8d^3 = 0$; or, upon division of the equation by 2: $a^3 - 3a^2d - 3ad^2 - 4d^3 = 0$.

Since a and d are integers, their quotient a/d must be rational. Let $r = a/d$, whence, $a = rd$. Substituting rd for a:

$$(rd)^3 - 3(rd)^2d - 3(rd)d^2 - 4d^3 = 0$$

$$r^3d^3 - 3r^2d^3 - 3rd^3 - 4d^3 = 0;$$

or, upon division of the equation by d^3 (since d does not equal zero):

$$r^3 - 3r^2 - 3r - 4 = 0$$

$$(r - 4)(r^2 + r + 1) = 0;$$

and, upon setting each factor equal to zero, we find that the roots of the cubic equation are $r = 4$ and $r = (-1 \pm i\sqrt{3})/2$. Rejection of the complex roots (since, as stated above, r must be rational) leaves $r = 4$ as the only pertinent root. Whence, because $a = rd$, $a = 4d$. So:

$$x = a - d = 3d$$

$$y = a = 4d$$

$$z = a + d = 5d$$

$$u = a + 2d = 6d.$$

That is, $(3d)^3 + (4d)^3 + (5d)^3 = (6d)^3$. Notice that the proposer only mentioned the case in which $d = 1$. In fact, d may be any positive integer (e.g., for $d = 2$, $6^3 + 8^3 + 10^3 = 12^3$).

Also solved by Leon Bankoff, Los Angeles, California (two solutions); Merrill Barnebey, Wisconsin State University, LaCrosse; David Bernklau, Brooklyn, New York; V. S. Blanco, University of South Alabama; Gerald Baisvert, Providence College, Rhode Island; Dermott A. Breault, Cyber, Inc., Cambridge, Massachusetts; Robert J. Bridgman, Mansfield State College, Pensylvannia; Gary L. Britton, University of Wisconsin, West Bend; David C. Brooks, Seattle Pacific College, Washington; Milo F. Bryn, Oklahoma State University; Edmund M. Clarke, Madison College, Virginia; Santo M. Diano, Havertown, Pennsylvania; William F. Fox, Moberly Junior College, Missouri; Herta T. Freitag, Hollins, Virginia; R. Garfield, College of Insurance, New York; Ira Gessel, Cambridge, Massachusetts; Michael Goldberg, Washington, D. C.; M. G. Greening, University of New South Wales, Australia; Philip Haverstick, Fort Belvoir, Virginia; John M. Howell, Littlerock, California; J. A. H. Hunter, Toronto, Canada; Erwin Just, Bronx Community College, New York; Václav Konecny, Jarvis Christian College, Texas; Lew Kowarski, Morgan State College, Maryland; Ralph Jones, University of Massachusetts; Eric Laage-Petersen, Alborg, Denmark; Gerald M. Leibowitz, University of Connecticut; H. R. Leifer, Pittsburgh, Pensylvania; Peter A. Lindstrom, Genesee Community College, New York; Paul T. Mielke, CUPM, Berkeley, California; Edward Moylan, Ford Motor Company, Dearborn, Michigan; George A. Novacky, Jr., University of Pittsburgh; Joseph O'Rourke, St. Joseph's College, Pensylvania; Albert J. Patsche, U. S. Army Weapons Command, Rock Island, Illinois; Walter J. Pavlick, St. Louis, Missouri; Bob Prielipp, Wisconsin State University, Oshkosh; Sally Ringland, Shippenville, Pensylvania; Rina Rubenfeld, New York City Community College; Robert S. Stacy, Albuquerque, New Mexico; E. P. Starke, Plainfield, New Jersey; Paul Sugarman, Massachusetts Institute of Technology; Burnett R. Toskey, Seattle University, Washington; Philip Tracy, Liverpool, New York; Charles W. Trigg, San Diego, California; Zalman Usiskin, University of Chicago; Michael Viernieri, Orange County Community College, New York; Edward H. Voorhees, Jr., University of Tennessee at Chattanooga; Kenneth M. Wilke, Topeka, Kansas; Wayne J. Zimmerman, St. Edwards University, Texas; and the proposer.

Most solvers pointed out that the uniqueness pertained to primitive solutions. Voorhees found that this problem is identical to Problem E1977 in the AMERICAN MATHEMATICAL MONTHLY, Vol. 74, page 438.

An Inscribed Tetrahedron

792. [March, 1971]. *Proposed by Murray S. Klamkin, Ford Scientific Laboratory.*

It is a known result that a necessary and sufficient condition for a triangle inscribed in an ellipse to have a maximum area is that its centroid coincide with the center of the ellipse. Show that the analogous result for a tetrahedron inscribed in an ellipsoid is not valid.

Solution by the proposer.

By means of an affine transformation, it suffices to consider a sphere instead of an ellipsoid. For a sphere, it is a known result that the inscribed regular tetrahedron has the maximum volume and for this case its centroid coincides with the center of the sphere. However, the converse result is not valid, i.e., if the centroid of an inscribed tetrahedron in a sphere coincides with the center of the sphere, the tetrahedron need not be regular but it must be isosceles (one whose pairs of opposite edges are congruent). This latter result can be obtained vectorially as follows:

Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ denote the four vertices on a unit sphere with center at O. Then if the two centroids coincide, we have $\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}+\boldsymbol{D}=0$ in addition to $\boldsymbol{A}^2=\boldsymbol{B}^2=\boldsymbol{C}^2=\boldsymbol{D}^2=1$. Whence, $(\boldsymbol{A}+\boldsymbol{B})^2=(\boldsymbol{C}+\boldsymbol{D})^2$ or $\boldsymbol{A}\cdot\boldsymbol{B}=\boldsymbol{C}\cdot\boldsymbol{D}$. Thus,

$$(\boldsymbol{A}-\boldsymbol{B})^2 = (\boldsymbol{C}-\boldsymbol{D})^2$$

and, similarly,

$$(\boldsymbol{A}-\boldsymbol{C})^2 = (\boldsymbol{D}-\boldsymbol{B})^2,$$

$$(\boldsymbol{A}-\boldsymbol{D})^2 = (\boldsymbol{B}-\boldsymbol{C})^2$$

and the tetrahedron is isocesceles.

Conversely, if the tetrahedron is isosceles, then $(\boldsymbol{A}-\boldsymbol{B})^2=(\boldsymbol{C}-\boldsymbol{D})^2$, $(\boldsymbol{A}-\boldsymbol{C})^2=(\boldsymbol{D}-\boldsymbol{B})^2$, $(\boldsymbol{A}-\boldsymbol{D})^2=(\boldsymbol{B}-\boldsymbol{C})^2$ in addition to $\boldsymbol{A}^2=\boldsymbol{B}^2=\boldsymbol{C}^2=\boldsymbol{D}^2=1$. Whence, $\boldsymbol{A}\cdot\boldsymbol{B}=\boldsymbol{C}\cdot\boldsymbol{D}$, $\boldsymbol{A}\cdot\boldsymbol{C}=\boldsymbol{D}\cdot\boldsymbol{B}$, $\boldsymbol{A}\cdot\boldsymbol{D}=\boldsymbol{B}\cdot\boldsymbol{C}$ and $(\boldsymbol{A}+\boldsymbol{B})^2=(\boldsymbol{C}+\boldsymbol{D})^2$. Then

$$(\boldsymbol{A}+\boldsymbol{B}-\boldsymbol{C}-\boldsymbol{D})\cdot(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}+\boldsymbol{D}) = 0$$

and, similarly,

$$(\boldsymbol{A}+\boldsymbol{D}-\boldsymbol{B}-\boldsymbol{C})\cdot(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}+\boldsymbol{D}) = 0.$$

Thus,

$$(\boldsymbol{A}-\boldsymbol{C})\cdot(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}+\boldsymbol{D}) = 0$$

and similarly,

$$(\boldsymbol{A}-\boldsymbol{B})\cdot(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}+\boldsymbol{D}) = 0,$$

$$(\boldsymbol{A}-\boldsymbol{D})\cdot(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}+\boldsymbol{D}) = 0.$$

Finally, the latter three equations imply $\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}+\boldsymbol{D}=0$ or that the two centroids coincide. (A geometric proof appears in N. Altshiller-Court, Modern Solid Geometry, MacMillan, New York, 1935, p. 95).

Lowest Order Determinant

793. [March, 1971] *Proposed by Gregory Wulczyn, Bucknell University.*

Find the determinant of lowest order with entries from the interval $-1 \leqq x \leqq +1$ whose value is 2^{38}.

Solution by J. R. Kuttler, Johns Hopkins Applied Physics Laboratory.

In the notation of [1], let $g(n)$ denote the maximum value of the determinant of an nth order matrix whose entries lie in the interval $-1 \leqq x \leqq +1$. By Hadamard's inequality [2], $g(n) \leqq n^{n/2}$. In particular,

$$g(18) \leqq 18^9 < 2^{38}.$$

On the other hand, it is known [3] that

$$g(20) = 20^{10},$$

and, by Theorem 11 of [1],

$$g(19) \leqq \frac{1}{20} g(20) = 20^9 > 2^{38}.$$

Therefore, 19 is the lowest order of a matrix with entries in the range $-1 \leqq x \leqq +1$ whose determinant is 2^{38}.

References

1. J. H. E. Cohn, On the value of determinants, Proc. Amer. Math. Soc., 14 (1963) 581–588.
2. J. Hadamard, Résolution d'une question relative aux déterminants, Bull. Sci. Math., Part 1, (2), 17 (1893), 240–246..
3. R. E. A. C. Paley, On orthogonal matrices, J. Mathematical Phys., 12 (1933) 311–320.

Also solved by Burnett R. Toskey, Seattle University, Washington. Two incorrect solutions were received.

Locus of the Incenter

794. [March, 1971]*Proposed by Leon Bankoff, Los Angeles, California.*

If I, O, H are the incenter, circumcenter and orthocenter, respectively, of a triangle ABC in which $C > B > A$, show that I must lie within the triangle BOH.

Solution by M. G. Greening, University of New South Wales, Australia.

R = circumradius; r = inradius; $x = \tan A/2$, $y = \tan B/2$, $z = \tan C/2$; ΔXYZ denotes the area of triangle XYZ.

Whether C be acute or obtuse, angle $OBH = C - A$ and BI bisects it internally, so that (i) I lies between BO and BH.

Also $BH = 2R\cos B$, $BI = r\operatorname{cosec}(B/2) = 4R\sin(A/2)\sin(C/2)$. Then $\Delta OBH = \frac{1}{2}\cdot 2R^2\cos B\sin(C-A) = m\alpha$.

$$\Delta BIO + \Delta BIH = \tfrac{1}{2}(1+\cos 2B)R\cdot BI\cdot \sin[\tfrac{1}{2}(C-A)]$$

$$= 2R^2(1+2\cos B)\sin(A/2)\sin(C/2)\sin[\tfrac{1}{2}(C-A)] = \beta$$

$$\alpha - \beta = 2R^2\sin[\tfrac{1}{2}(C-A)]\{\cos B\cos[\tfrac{1}{2}(C-A)] - (1+2\cos B)\sin(A/2)\sin(C/2)\}$$

$$= 2R^2\sin[\tfrac{1}{2}(C-A)]\cos B\{\cos(C/2)\cos(A/2) - (\sec B+1)\sin(A/2)\sin(C/2)\}$$

$$= 2R^2\sin[\tfrac{1}{2}(C-A)]\cos B\cos(C/2)\cos(A/2)\left\{1 - \frac{2xy}{1-y^2}\right\}.$$

As $B < \pi/2$, $1 - y^2 > 0$ and each of the other factors except the last is necessarily positive. So $\alpha - \beta$ has the same sign as $\gamma = 1 - y^2 - 2xz$. But $xy + yz + zx = 1$ so $\gamma = -[y^2 - (x+z)\,y + xz]$ which is positive for $x < y < z$, i.e., $\alpha > \beta$ which, with (i) implies I in interior of triangle OBH.

This also establishes that the point of tangency of the incircle and the nine point circle lies on the same side of OH as B.

Also solved by the proposer.

Bernoulli Numbers

795. [March, 1971] *Proposed by L. Carlitz, Duke University.*

Define $\{B_n\}$ by means of $B_0 = 1$ and $\sum_{k=0}^{n} \binom{n}{k} B_k = B_k$ for $n > 1$. Show that for arbitrary m, $n \geqq 0$,

$$(-1)^m \sum_{r=0}^{m} \binom{m}{r} B_{n+r} = (-1)^n \sum_{s=0}^{n} \binom{n}{s} B_{m+s}.$$

Solution by A. G. Shannon, New South Wales Institute of Technology, Australia.

We use induction on n and m, and the recurrence relation for combinatorial coefficients. For $n, m = 0, 1, 2$.

$$(1) \qquad (-1)^m \sum_{r=0}^{m} \binom{m}{r} B_{n+r} = (-1)^n \sum_{s=0}^{n} \binom{n}{s} B_{m+s}$$

can be readily shown if we use the fact that, for the Bernoulli numbers, $B_{2k+1} = 0$, for a positive integer k.

Assume (1) is true for $n = 0, 1, 2 \ldots N$ and for $m = 0, 1, 2 \ldots M$. Then (1) implies that

$$(2) \qquad (-1)^{M-1} \sum_{r=0}^{M} \binom{M}{r} B_{N+r} = (-1)^{N+1} \sum_{s=0}^{N} \binom{N}{s} B_{M+s}$$

and

$$(3) \qquad (-1)^{M-1} \sum_{r=0}^{M-1} \binom{M-1}{r} B_{N+r} = (-1)^{N} \sum_{s=0}^{N} \binom{N}{s} B_{M+s-1}$$

$$= (-1)^N \left\{ \sum_{s=0}^{N} \binom{N}{s+1} B_{M+s} + B_{M-1} \right\}.$$

Subtracting the corresponding parts of (3) from (2), we get

$$(-1)^{M-1} \sum_{r=1}^{M-1} \binom{M-1}{r-1} B_{N+r} = (-1)^{N+1} \left\{ \sum_{s=0}^{N+1} \binom{N+1}{s+1} B_{M+s} + B_{M-1} \right)$$

which can be rewritten as

$$(-1)^{M-1}\sum_{r=0}^{M-1}\binom{M-1}{r}B_{N+r+1}$$

$$= (-1)^{N+1}\sum_{s=0}^{N+1}\binom{N+1}{s}B_{M+s-1}$$

which justifies the induction on n. The induction on m follows from symmetry, and so (1) is proved.

Also solved by M. G. Greening, University of New South Wales, Australia; Wells Johnson, Bowdoin College, Maine; and the proposer.

ANSWERS

A532. The integral on the left is twice the area of the region between the semicircle $y = \sqrt{1-x^2}$, $-1 \leqq x \leqq 1$ and the x-axis, which is π. The integral on the right comes from the arc length formula for the arc length of the semicircle, which is π.

A533. No, because the field is not a group under multiplication. The set of nonzero elements is a group under multiplication.

A534. The team must win the last game played. Suppose that the team wins the $(n+1)$st game at the $(n+x+1)$. Of the preceding $(n+x)$ games the opposition must have won x games in some order. The probability of this is:

$$\binom{n-x}{x}(1-p)^x p^n.$$

So:

$$P = \sum_{x=0}^{n}\binom{n-x}{x}(1-p)^x p^{n+1}.$$

A535. $e^{2737} \gg e^{1610} \doteq 5^{1000} = (1+1+1+1+1)^{1000}$

$$= \sum_{(\Sigma n_i = 1000)} 1000!/n_1!\,n_2!\,n_3!\,n_4!\,n_5! \gg 1000!/70!\ 270!\ 300!\ 220!\ 140!$$

A536. If they were then

$$ar^{n_1} = \sqrt{p_1},\quad ar^{n_2} = \sqrt{p_2},\quad ar^{n_3} = \sqrt{p_3}\quad (n_1, n_2, n_3 \text{ distinct integers}).$$

Eliminating a and r yields

$$(p_1/p_2)^{n_2-n_1} = (p_2/p_3)^{n_1-n_3}$$

which is clearly impossible (by the unique factorization theorem). The result holds for any integral roots.

(Quickies on page 47)