MATHEMATICS MAGAZINE

INDEX TO VOLUME 36 JANUARY THROUGH DECEMBER, 1963

ARTICLES BY TITLES

ARTICLES BY AUTHORS

PROBLEMS AND SOLUTIONS

Contributors

Numbers refer to pages. Bold type indicates a problem solved and solution published; italics, a problem solved but solution not published; ordinary type, a problem proposed.

Solutions

The page on which a solution appears is in the parentheses following the number of the problem.
481 (71), 482 (72), 483 (73), 484 (73), 485 (74), 486 (75), 487 (76), 488 (134), 489 (135), 490 (135), 491 (136), 492 (137), 493 (138), 494 (139), 495 (198), 496 (200), 497 (201), 498 (201), 499 (202), 500 (203), 501 (204), 502 (265), 503 (266), 504 (266), 505 (267), 506 (269), 507 (269), 508 (270), 509 (321), 510 (322), 511 (323), 512 (323), 513 (324), 514 (324), 515 (326).

Quickies

306, 307, 308, 309 (77)
310, 311, 312 (142)
313, 314, 315, 316, 317 (206)
318, 319, 320, 321, 322 (270)
323, 324, 325, 326, (328)

The editors of the *Mathematics Magazine* wish to express their appreciation to the following who assisted us by reading and refereeing manuscripts.

H. L. Alder, Brother U. Alfred, Souren Babikian, R. D. Boswell, J. M. Calloway, R. Chalkley, J. I. Christopher, H. L. Dorwart, B. K. Gold, J. W. Green, A. E. Hallerberg, J. M. C. Hamilton, E. G. Harrell, V. E. Hoggatt, R. G. N. Hourston, J. M. Howell, A. Jaeger, G. D. James, P. W. Ketchum, J. M. Kingston, A. O. Lindstrom, Jr., C. Long, D. J. Merrill, F. D. Newman, W. H. Row, C. L. Seebeck, Jr., Anice Seybold, J. P. Schiller, H. E. Stelson, R. Suhr, K. W. Wegner, H. Wilson, W. B. Woolf, P. H. Yearout, H. Zuckerman.

MATHEMATICS MAGAZINE

CONTENTS

VOLUME 36 • NOVEMBER 1963 • NUMBER 5

MM

MATHEMATICS MAGAZINE

ROBERT E. HORTON, *Editor*

EDITORIAL CORRESPONDENCE should be sent to the Editor, ROBERT E. HORTON, Department of Mathematics, Los Angeles City College, 855 North Vermont Avenue, Los Angeles 29, California. Articles should be typewritten and double-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, H. M. GEHMAN, Mathematical Association of America, SUNY at Buffalo (University of Buffalo), Buffalo 14, New York.

ADVERTISING CORRESPONDENCE should be addressed to F. R. OLSON, Mathematical Association of America, SUNY at Buffalo (University of Buffalo), Buffalo 14, New York.

The MATHEMATICS MAGAZINE is published by the Mathematical Association of America at Buffalo, New York, bi-monthly except July-August. Ordinary subscriptions are: 1 year \$3.00; 2 years \$5.75; 3 years \$8.50; 4 years \$11.00; 5 years \$13.00. Members of the Mathematical Association of America may subscribe at the special rate of 2 years for \$5.00. Single copies are 65¢.

PUBLISHED WITH THE ASSISTANCE OF THE JACOB HOUCK MEMORIAL FUND OF THE MATHEMATICAL ASSOCIATION OF AMERICA

Second class postage paid at Buffalo, New York and additional mailing offices.

A STUDY OF CONIC SECTION ORBITS BY ELEMENTARY MATHEMATICS

RAPHAEL T. COFFMAN, Richland, Washington

The level of mathematics required for deriving the equations of planetary motion by the usual approach is sufficiently high to place the subject beyond the reach of most people. Because of the importance of the problem and the general interest in the subject, particularly in this era of earth satellites, a treatment is desirable at the lowest level of mathematics which will suffice. In this paper a derivation of the inverse square law and derivations of other equations relating to planetary motion are given, using only elementary mathematics. It is believed that the approach used is to some degree original, particularly in that limit theory is not used in any of the derivations. Also, the representation of acceleration as a vector originating at the tip of the velocity vector on the position-velocity diagram has not been elsewhere observed by the writer; however, it is related to the use of the acceleration vector on the hodograph. As used, the diagram makes it possible to write equations involving simultaneously radius vector, velocity and acceleration. The method used for finding tangents to the conics has been published in this magazine (Reference 1) and is included here because it is essential to the treatment.

Except for the use of the acceleration vector on the velocity vector and the derivation of the required properties of the conics there is little or no mathematics used which should be unfamiliar to a high school student. However, the treatment here given is not intended to be such as could readily be followed at this level. Considerably more exposition than has been given would probably be required in a high school level treatment. The nature of the problem limits the extent to which even an elementary solution can be simplified.

An acceleration which acts in the direction of motion of a point may be represented by a vector attached to the tip of the velocity vector to indicate the rate of change of the length of the latter, Fig. 1-a. If the direction of acceleration does not coincide with the direction of the velocity, Fig. 1-b, the acceleration may be represented by a vector originating at the tip of the velocity vector and parallel to the direction of the acceleration. That this is true may be shown with the aid of Fig. 1-c. Here the velocity, V, has been resolved into two components, V_1 and V_2, parallel and perpendicular, respectively, to the acceleration direction. Hence the acceleration acts only on V_1 and may be represented by a vector attached to the tip of V_1 and having the same direction. The side of the vector parallelogram opposite V_1 is equal to V_1 and its tip coincides with the tip of V. Its acceleration, which is the same as that of V_1, may be represented by a vector, a, with origin at the tip of V and parallel to V_1. This vector represents the acceleration of P and shows graphically the manner in which the acceleration affects V.

If the acceleration vector, hereafter designated by a, and shown as a broken line, is resolved into components parallel and perpendicular to V, the parallel component will determine the rate of change of magnitude of V and the per-

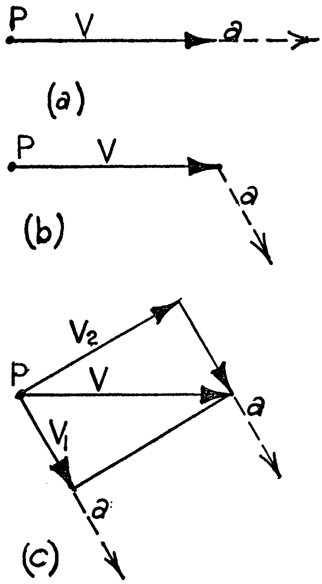


FIG. 1

pendicular component will determine the rate of change of direction, or angular velocity, of V.

In dealing with orbital motion, in which a body moves under the influence of an acceleration directed toward a fixed point, the approach used in this paper is to determine the acceleration which will keep the velocity vector always tangent to the curve which is the chosen path. The basic diagram is shown in Fig. 2, where S is the fixed point, P the orbiting body, V the velocity and a the acceleration. The acceleration vector is parallel to r, the radius vector. The component of a normal to V changes the direction of V at a rate which keeps V tangent to the curve representing the path of P.

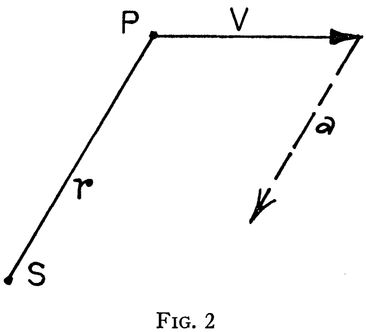


FIG. 2

To illustrate the method, consider the problem of finding an expression for the acceleration which will cause a point to move with constant velocity in a circular path. See Fig. 3. Since velocity is constant, the acceleration vector must be perpendicular to V. V and r have the same angular velocity since the angle

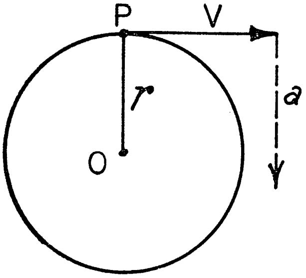


FIG. 3

between them remains 90°. Therefore, $V/r = a/V$ or, $a = V^2/r$, and is directed to the center of the circle.

Kepler's second law of planetary motion states that the line joining the sun and a planet sweeps out equal areas in equal times.

Newton proved that this is true for any body moving in any path because of a force directed to a fixed point. Since this theorem is essential in derivations to follow, a proof is given. In Fig. 4, P is a point moving with velocity V, S is a fixed point and r is the line joining S and P. If no acceleration acts on P it

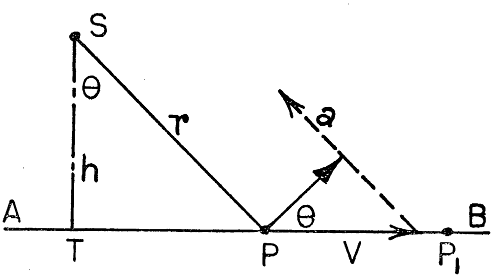


FIG. 4

moves with uniform velocity along line AB. In any equal time intervals the area swept out by r will be the area of a triangle whose base is the distance P moves in the interval and whose altitude is h, both of which are constant. Therefore, under these conditions equal areas are swept out in equal times. The rate of change of area, vA, is given by the equation: $vA = \frac{1}{2}rV\cos\theta$ where $V\cos\theta$ is the component of V perpendicular to r. If an acceleration directed to S acts on P its effect will be to move the tip of V along the line of the acceleration vector, a. This, however, does not produce a change in the vector $V\cos\theta$, hence does not affect the rate of change of area. Hence the rate of change of area is constant for any point in the path. Since $r\cos\theta = h$, $vA = \frac{1}{2}Vh$, or

$$Vh \text{ is constant,} \tag{1}$$

a corollary to the above theorem in Newton's *Principia*.

In Fig. 5 CDP is a parabola with focus at F and directrix AB. At any point, P, $FP = PE$. If V_1 and V_2 are vectors showing the rate of change of length of FP and PE, they are equal and the line drawn through the intersection of

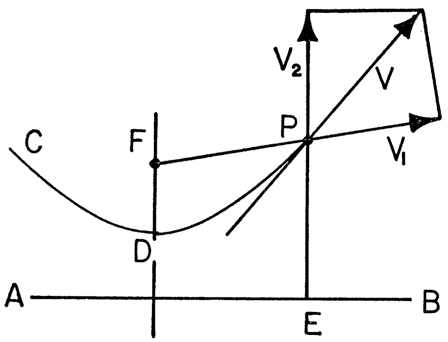


FIG. 5

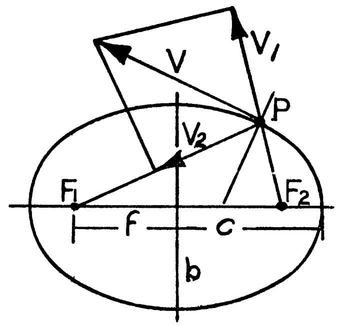


FIG. 6

perpendiculars to their tips bisects the angle FPE and determines V, the velocity of P.

Fig. 6 is an ellipse with foci F_1 and F_2. P is any point on the ellipse and $PF_1+PF_2=2c$, where c is the major semiaxis. If P is moving to the left PF_2 is increasing in length as shown by vector V_1 and PF_1 is decreasing at the same rate, indicated by vector V_2. The velocity, V, of P also has components perpendicular to both V_1 and V_2. The tip of V is at the point of intersection of perpendiculars drawn from the tips of V_1 and V_2. By congruence of triangles, V bisects the angle formed by the intersection of V_1 and V_2 and it follows that the perpendicular to V at P bisects angle F_1PF_2.

The direction of a point moving on a hyperbola is established in a similar manner, as shown in Fig. 7. Here F_1 and F_2 are the foci, P is any point on the curve and $F_1P-F_2P=2c$. Since F_1P and F_2P have a constant difference in length, V_1 and V_2 are equal, and the line containing V, the velocity of P, bisects the angle F_1PF_2.

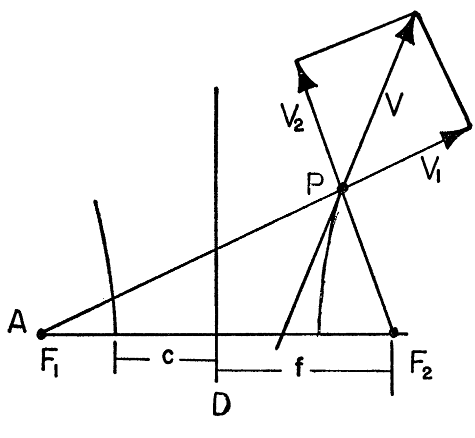


FIG. 7

The preceding material provides the basis for determining how the acceleration toward the focus must vary with distance if a point moves in a conic. The problem will be solved first for the parabola, since the derivation is simpler than for the other two conics.

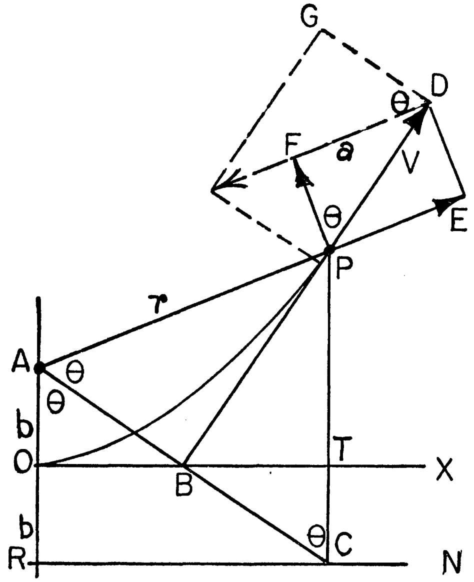


FIG. 8

In Fig. 8, P is a point on the parabola with focus A and directrix RN. OX is parallel to RN and $AO=OR=b$. The velocity vector V lies on PB, the bisector of angle APC. By plane geometry it is readily proved that the angles designated as θ are equal. Using the lower case v to indicate rate of change, or velocity, e.g. vx = rate of change of x, for the angular velocity of AP or r we have $v2\theta = V(\cos\theta)/r$, which, since $v2\theta=2v\theta$, becomes $v\theta = V(\cos\theta)/2r$. FP remains perpendicular to r, hence has the same angular velocity. The angular velocity of V is DG/V or $a(\cos\theta)/V$ where a is the acceleration vector, which is parallel to r. The angular velocity of angle FPD is the difference between the angular velocities of FP and V or $v\theta = V(\cos\theta)/r - a(\cos\theta)/V$. By giving $v\theta$ the previously obtained value, the above becomes $V(\cos\theta)/2r = V(\cos\theta)/r - a(\cos\theta)/V$, which reduces to:

$$V^2/2r = a. \tag{2}$$

From the law of equal area in equal time, $AB\cdot V=k$, where k is a constant, hence $V^2=k^2/\overline{AB}^2$. By similar triangles in Fig. 8, $AB: r = b: AB$, or $\overline{AB}^2 = rb$, hence $V^2=k^2/rb$. Substituting the right hand side of this equation for V^2 in the equation $V^2/2r=a$ we obtain:

$$a = k^2/(2r\cdot rb) = k^2/2br^2.$$

Representing $k^2/2b$, which is constant, by K we have:

$$a = K/r^2 \tag{3}$$

which shows that the acceleration toward the focus varies inversely as the square of the distance.

If equation (2) above is solved for V^2 and K/r^2 is substituted for a, the expression:

$$V^2 = K\cdot 2/r \tag{4}$$

is obtained. If the moving point has mass, this equation gives the relation of kinetic energy of the particle to its distance from the focus.

The acceleration to the focus of an ellipse is found with the aid of Fig. 9. Here, P is a point on the ellipse with major axis $2c=r_1+r_2$ and foci A and B. The velocity of P is V and the acceleration, a, is directed toward A. The perpendicular to V at P bisects angle APB, making angle $APB=2\theta$. It is apparent that the other angles so designated are equal to θ.

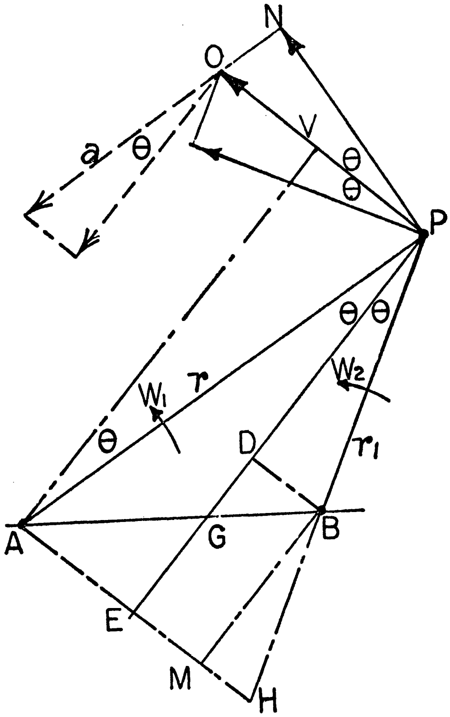


FIG. 9

The angular velocity, W_1, of r is: $W_1=V(\cos\theta)/r$, and of r_1 is: $W_2=V(\cos\theta)/r_1 = V(\cos\theta)/(2c-r)$, where $V\cos\theta$ is the component of V perpendicular to r and to r_1. The rate of change of angle APB, designated by $v2\theta$, is W_2-W_1. Since $v2\theta=2v\theta$ we have: $v\theta=\frac{1}{2}(V(\cos\theta)/2c-r-V(\cos\theta)/r)$ or

$$v\theta = V\cos\theta\,\frac{(r-c)}{r(2c-r)}\,\cdot$$

Considering the rate of change of θ, as represented by angle NPO, we have, since the angular velocity of PN is the same as that of r:

$$v\theta = a(\cos\theta)/V - V(\cos\theta)/r.$$

Substituting this expression for $v\theta$ in the above equation:

$$a(\cos\theta)/V - V(\cos\theta)/r = V\cos\theta(r-c))/r(2c-r).$$

When this is reduced and solved for a, we obtain the equation:

$$a = \frac{V^2}{r}\left(\frac{c}{2c - r}\right). \tag{5}$$

By the law of constant rate of change of area, $Vr \cos \theta = k$, where k is constant, hence, solving for V and squaring: $V^2 = k^2/(r^2 \cos^2 \theta)$. Using this value for V^2 in equation (5)

$$a = \frac{k^2 c}{r^3 \cos^2 \theta(2c - r)}. \tag{6}$$

This expresses a in terms of two variables, r and θ. The latter is a function of r and may be eliminated by the procedure given below.

In Fig. 9, AH is perpendicular to PE, BM is parallel to PE and BD is parallel to AH. Then, $\overline{AB}^2 = \overline{AM}^2 + \overline{MB}^2$.

$$\begin{aligned}
AE &= r \sin \theta \\
BD &= EM = (2c - r) \sin \theta \\
AM &= AE + EM = r \sin \theta + (2c - r) \sin \theta = 2c \sin \theta \\
MB &= DE = r \cos \theta - (2c - r) \cos \theta = 2(r - c) \cos \theta \\
\overline{AB}^2 &= [2c \sin \theta]^2 + [2(r - c) \cos \theta]^2 \\
&= 4c^2 \sin^2 \theta + 4c^2 \cos^2 \theta + 4r \cos^2 \theta(r - 2c) \\
\overline{AB}^2 &= 4c^2 + 4r \cos^2 \theta(r - 2c) \\
\overline{AB}^2 - 4c^2 &= 4r \cos^2 \theta(r - 2c) \\
\frac{4c^2 - \overline{AB}^2}{4} &= r \cos^2 \theta(2c - r) = C. \qquad (7)
\end{aligned}$$

The left side of this equation is a constant, hence, $r \cos^2 \theta(2c - r)$ is a constant, which may be called C. Returning to equation (6) above, it is apparent that the denominator is r^2C, hence, $a = k^2c/r^2C$, or, calling the three constants K,

$$a = K/r^2. \tag{8}$$

If $AB = 2f$, the expression $\overline{AB}^2 - 4c^2$ in (7) becomes $4f^2 - 4c^2$ and we have $4f^2 - 4c^2 = 4r \cos^2 \theta(r - 2c)$, or $c^2 - f^2 = r \cos^2 \theta(2c - r)$. Since $c^2 - f^2 = b^2$, where b is the minor semiaxis of the ellipse, we have $C = b^2$ and $a = k^2/r^2 \cdot c/b^2$.

If in the equation, $a = V^2/r \cdot c/(2c - r)$ we substitute K/r^2 for a and solve for V^2 we obtain:

$$V^2 = K(2/r - 1/c) \tag{9}$$

which will be discussed later.

The derivation of the acceleration law for the hyperbola follows, in general, the same method as used for the ellipse.

In Fig. 10, P is a point on the hyperbola with foci A and B. The velocity

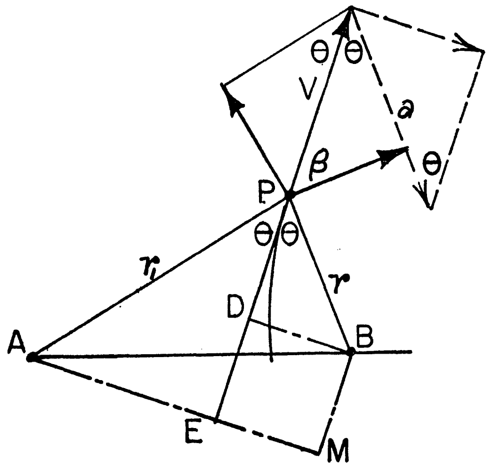


FIG. 10

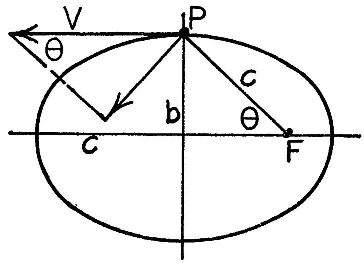


FIG. 11

of P is V and the acceleration, a, is directed toward B. The constant difference of the distance of P from A and B is:

$$r_1 - r = 2c.$$

If $V \sin\theta$ is the component of V perpendicular to r and to r_1 the rate of change of angle APB is:

$$\begin{aligned} v2\theta = 2v\theta &= -\left[V(\sin\theta)/r + V(\sin\theta)/r_1\right] \\ &= -V\sin\theta[1/r + 1/(2c+r)] \\ v\theta &= -V\sin\theta\,\frac{c+r}{r(2c+r)}\,. \end{aligned}$$

Also

$$\begin{aligned} v\theta = -v\beta &= -\left[V(\sin\theta)/r - a(\sin\theta)/V\right] \\ &= -V\sin\theta(1/r - a/V^2). \end{aligned}$$

Equating the two expressions for $v\theta$ and solving for a we obtain:

$$a = \frac{V^2 c}{r(2c+r)}\,. \tag{10}$$

Using the relation $Vr\sin\theta = k$ to eliminate V^2, the above becomes

$$a = \frac{k^2}{r^3}\cdot\frac{c}{(2c+r)\sin^2\theta}\,.$$

By the same steps as used for the ellipse it may be shown that $\frac{1}{4}\overline{AB}^2 - c^2 = r(2c+r)\sin^2\theta$ and that

$$a = K/r^2. \tag{11}$$

By substituting K/r^2 for a in the equation $a = V^2c/r(2c+r)$ and solving for V^2, the following is obtained:

$$V^2 = K(2/r + 1/c). \tag{12}$$

By the same procedure it may be proved that if the acceleration is directed away from A and the path remains a hyperbola the acceleration varies inversely as r_1^2.

It has now been shown that the law of acceleration, or attraction, to the focus of a conic is that of the inverse square. Also, it has been shown that for the ellipse $V^2=K(2/r-1/c)$, for the parabola $V^2=K\cdot 2/r$ and for the hyperbola $V^2=K(2/r+1/c)$. These three equations show that velocity determines which conic is the path. By inspection of the equation $V^2=K\cdot 2/r$, it is seen that as r increases V decreases, and when r is infinite V is zero. Hence the velocity of the particle in the parabolic path is just sufficient to allow the particle to travel to an infinite distance. In the same equation, if the quantity $2/r$ is decreased by $1/c$ the equation becomes that of an ellipse, $V^2=K(2/r-1/c)$. As stated earlier, c is the major semiaxis of the ellipse. If c is infinite the equation is that of a parabola, and as c decreases V decreases. Hence, in an ellipse the velocity is not sufficient to allow the particle to escape. If V is made zero the equation becomes $2/r-1/c=0$ or $r=2c$. This indicates that in an elliptical orbit the velocity at any point is sufficient to carry the particle to a distance $2c$ from the central point.

If $1/c$ is added to $2/r$ the equation becomes that of the hyperbola, $V^2=K(2/r+1/c)$. If r becomes infinite this becomes $V^2=K\cdot 1/c$, which shows the particle reaches infinity with a finite velocity.

Since kinetic energy is proportional to V^2, the three equations above are a measure of the kinetic energy of particles having mass. The difference in kinetic energy for any two distances r_1 and r_2 is the same for all three curves for which K has the same value, for the term $1/c$ is cancelled out:

$$V_1^2 - V_2^2 = K(1/r_1 - 1/r_2). \tag{13}$$

The term in parentheses is that obtained by integrating dr/r^2 between limits r_1 and r_2 and represents the change in potential energy as the distance changes by the amount r_1-r_2. Here, the expression has been obtained without integration.

Kepler's third law, that T^2 is proportional to r^3, where T is the time for one revolution and r is the mean distance of a planet from the sun, can be shown to be a consequence of the inverse square law. For a circular orbit $T=2\pi r/V$. Since $a=V^2/r$ and $a=k/r^2$, $V=k^{1/2}/r^{1/2}$ and $T=2\pi r^{3/2}/k^{1/2}$ or $T^2=4\pi^2r^3/k$. In Fig. 11 the rate of change of area swept out by the radius vector c at the point P, where c equals the major semiaxis, is:

$$vA_1 = \tfrac{1}{2}cV\sin\theta$$

or, since $\sin\theta=b/c$,

$$vA_1 = \tfrac{1}{2}Vb.$$

In a circle with radius c and velocity V the rate of change of area is $vA_2=\frac{1}{2}cV$. Combining the two gives

$$\frac{vA_1}{vA_2} = \frac{b}{c}\,.$$

The ratio of areas produced is proportional to the ratio of the rate of change of areas. When the point on the circle has made one revolution the area produced is πc^2. Therefore, in the same time:

$$\frac{vA_1}{vA_2} = \frac{A_1}{\pi c^2} = \frac{b}{c},$$

from which $A_1 = \pi bc$. This is the equation of an ellipse with semiaxes b and c. Hence, both the circular orbit and the elliptical orbit in which the major semiaxis is equal to the radius of the circle have the same period, and since Kepler's third law applies to the circle it applies also to the ellipse.

The equations which have been derived in this paper apply to particles. In order to make them applicable to physical bodies, as planets or earth satellites, it is necessary to know that the gravitational effect of a spherical body composed of homogeneous spherical shells is as though all the mass is at the center. An elementary proof of this has been derived by the writer, but because it is long and complex it has not been included here. The equation for the area of an ellipse is not generally derived below the level of integral calculus. It is, however, an easy matter to derive the equation from considerations of the relation of the ellipse to the circle.

Reference

1. The Reflection Property of the Conics. By R. T. Coffman and C. S. Ogilvy, MATHEMATICS MAGAZINE, **36**, No. 1, p. 11.

Answers

A323. Three. For three terms to be in A.P., we must have

$$2\binom{m}{n} = \binom{m}{n-1} + \binom{m}{n+1} \quad \text{or} \quad (m-2n)^2 = m+2$$

whence

$$m = a^2 - 2, \qquad 2n = a^2 \pm a - 2$$

In order to have four terms in A.P., $(a^2-a)/2 = (a^2+a-2)/2$ or $a=1$ and impossible. (See Note of Th. Motzkin, *Scripta Mathematica*, March, 1946, p. 14.)

A324. If A and B are both made equal to one, $(A+B)^{13} = 2^{13}$, which is the sum of the numerical coefficients, as all the other coefficients become unity. Hence $2^{13} = 8192$.

A325. Since $2z = 2xy/(x+y)$, $2xy - 2yz - 2zx = 0$, whereupon $(x+y-z)^2 = x^2+y^2 +z^2$.

A326. Assume $x \leqq y$. Then $y > 1$ and $x^x + y^y \leqq 2y^y \leqq y^{y+1} \leqq z^z$. Thus $x^x + y^y < z^z$ and there are no solutions.

(Quickies on page 328)

APPROXIMATIONS TO INCOMMENSURABLE NUMBERS BY RATIOS OF POSITIVE INTEGERS

PAUL D. THOMAS, U. S. Coast and Geodetic Survey

It is known that 355/113 gives π correct in the 6th decimal place ($355/113 = 3.141593$). Also 193/71 gives e correct in the 4th decimal place ($193/71 = 2.7183$). The purpose of this note is to explain how to find such rational approximations to π, e, etc. The basic methods are known and can be found, for instance, in a Textbook of Algebra, G. Chrystal, Part II, Ch. 32. (Although out of print for many years, it has been recently reprinted in paper by Dover.) Certain properties of the principal and intermediate convergents to the terminating simple continued fraction expansion of a rational number are involved.

Now every rational number x has a unique, terminating, simple continued fraction expansion of the form

$$x = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cdots}}} = a_1 + \frac{1}{a_2 +}\,\frac{1}{a_3 +}\,\frac{1}{a_4 +}\cdots \tag{1}$$

where a_1 is a positive integer or zero and $a_2, a_3, a_4, \cdots$ are positive integers. The set of integers $a_1, a_2, a_3, a_4, \cdots, a_n$ are called the *partial quotients* of the continued fraction.

The successive fractions formed from (1) as follows:

$$\frac{a_1}{1}, \quad a_1 + \frac{1}{a_2} = \frac{a_1a_2+1}{a_2}, \quad a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3}} = \frac{a_1a_2a_3 + a_1 + a_3}{a_2a_3+1},$$

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4}}} = \frac{a_1a_2a_3a_4 + a_1a_2 + a_1a_4 + a_3a_4 + 1}{a_2a_3a_4 + a_2 + a_4}, \cdots \tag{2}$$

are called the principal *convergents* of the continued fraction (1). (When $a_1 = 0$' the first convergent is written 0/1. Clearly the $(k+1)$th convergent may be derived from the kth by merely replacing a_k by $a_k + 1/a_{k+1}$. The terminal convergent is clearly the entire expansion (1).)

From (2) it is seen that $a_1 < a_1 + 1/a_2$,

$$a_1 + 1/a_2 > a_1 + \frac{1}{a_2 + 1/a_3}, \quad a_1 + \frac{1}{a_2 + 1/a_3} < a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + 1/a_4}}, \text{ etc.}$$

that is:

Each odd convergent is less and each even convergent is greater then every subsequent *convergent, and thus the value of the continued fraction* (1) *lies between the values of every two consecutive convergents.*

If $p_1, p_2, p_3, p_4, \cdots$ and $q_1, q_2, q_3, q_4 \cdots$ denote respectively the numerators and denominators of the principal convergents (2), then

$$\begin{aligned} &p_1 = a_1,\ p_2 = a_1a_2 + 1,\ p_3 = a_1a_2a_3 + a_1 + a_3, \\ &p_4 = a_1a_2a_3a_4 + a_1a_2 + a_1a_4 + a_3a_4 + 1, \\ &q_1 = 1,\ q_2 = a_2,\ q_3 = a_2a_3 + 1,\ q_4 = a_2a_3a_4 + a_2 + a_4, \cdots \end{aligned} \tag{4}$$

From (4) it is seen that $p_3 = a_3p_2 + p_1$, $q_3 = a_3q_2 + q_1$ and in general the numerator and denominator of any convergent are connected with those of the two preceding convergents by formulas

$$p_n = a_np_{n-1} + p_{n-2}, \qquad q_n = a_nq_{n-1} + q_{n-2}. \tag{5}$$

Clearly p_n, q_n form sequences of increasing positive integers. Again from (4) it is easily seen that $p_2q_1 - p_1q_2 = +1$, $p_3q_2 - p_2q_3 = -1$, $p_4q_3 - p_3q_4 = +1$ and in general

$$p_nq_{n-1} - p_{n-1}q_n = (-1)^n. \tag{6}$$

From (6) it is clear that every convergent is irreducible since if p_n, q_n had a common factor it would have to divide $(-1)^n$. Also from (4) we find $p_3q_1 - p_1q_3 = a_3$, $p_4q_2 - p_2q_4 = -a_4$, $\cdots$ and in general

$$p_nq_{n-2} - p_{n-2}q_n = (-1)^{n-1}a_n. \tag{7}$$

(The relations (5), (6), (7) are quite easily proved by induction). By dividing (6) by q_nq_{n-1} and (7) by q_nq_{n-2} one obtains the difference formulas for the principal convergents

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = (-1)^n/q_nq_{n-1}, \qquad \frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = (-1)^na_n/q_nq_{n-2}. \tag{8}$$

If p_n/q_n, p_{n+1}/q_{n+1}, p_{n+2}/q_{n+2} are three consecutive principal convergents of the continued fraction x (assuming that n is odd), then in the light of (3) one can write

$$\frac{p_n}{q_n} < \frac{p_{n+2}}{q_{n+2}} < x < \frac{p_{n+1}}{q_{n+1}}.$$

Subtracting p_n/q_n from each member of this inequality find

$$0 < \frac{p_{n+2}}{q_{n+2}} - \frac{p_n}{q_n} < x - \frac{p_n}{q_n} < \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}. \tag{9}$$

Now from (8)

$$\frac{p_{n+2}}{q_{n+2}} - \frac{p_n}{q_n} = (-1)^{n+1}a_{n+2}/q_nq_{n+2}, \qquad \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} = (-1)^{n+1}/q_nq_{n+1}$$

whence (9) becomes

$$\frac{(-1)^{n+1}a_{n+2}}{q_nq_{n+2}} < x - \frac{p_n}{q_n} < \frac{(-1)^{n+1}}{q_nq_{n+1}}$$

or since n was assumed odd,

$$\frac{a_{n+2}}{q_nq_{n+2}} < x - \frac{p_n}{q_n} < \frac{1}{q_nq_{n+1}}\,. \tag{10}$$

Since for distinct positive integers

$$\frac{1}{q_nq_{n+1}} < \frac{1}{q_n^{\,2}}$$

and from (5)

$$q_{n+2} = a_{n+2}q_{n+1} + q_n,$$

$$\frac{a_{n+2}}{q_nq_{n+2}} = \frac{a_{n+2}}{q_n(a_{n+2}q_{n+1} + q_n)} = \frac{1}{q_n(q_{n+1} + a_{n+2}^{-1}q_n)} \geqq \frac{1}{q_n(q_{n+1} + q_n)}$$

(equality when $a_{n+2}=1$) we may write (10) as

$$\frac{1}{q_n(q_n + q_{n+1})} \leqq \frac{a_{n+2}}{q_nq_{n+2}} < x - \frac{p_n}{q_n} < \frac{1}{q_nq_{n+1}} < \frac{1}{q_n^{\,2}}\,. \tag{11}$$

Clearly (11) provides *error limits* for the approximation to x by any convergent p_n/q_n. From (5) and (11) it is seen that if the next partial quotient is large then q_{n+1} will be large so that $1/q_nq_{n+1}$ will be small. Hence to obtain a good approximation to a continued fraction *it is advisable to take that convergent whose corresponding partial quotient immediately precedes a very much larger partial quotient.*

Consider now the two successive convergents

$$\frac{p_{n-1}}{q_{n-1}},\quad \frac{p_n}{q_n}$$

and write from (11)

$$\left| x - \frac{p_{n-1}}{q_{n-1}} \right| > \frac{a_{n+1}}{q_nq_{n+1}}, \qquad \left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_nq_{n+1}}\,. \tag{12}$$

But

$$\frac{1}{q_nq_{n+1}} < \frac{a_{n+1}}{q_{n-1}q_{n+1}},$$

since $q_{n-1} < a_{n+1}q_n$. ($q_n > q_{n-1}$ and a_{n+1} is a positive integer.) *This shows that each convergent is a closer approximation to the value of the fraction than is any preceding convergent.*

Now suppose that a/b is a closer approximation to the value of the fraction than p_n/q_n is. Then a/b must also be closer to the fraction than p_{n-1}/q_{n-1} is and by (3) a/b lies between p_n/q_n and p_{n-1}/q_{n-1}. Or, considering n as even,

$$\frac{p_{n-1}}{q_{n-1}} < \frac{a}{b} < \frac{p_n}{q_n},$$

and subtracting p_{n-1}/q_{n-1} from each member we get

$$0 < \frac{a}{b} - \frac{p_{n-1}}{q_{n-1}} < \frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}}. \tag{13}$$

From (8), with n even,

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = 1/q_nq_{n-1},$$

which with (12) gives

$$\frac{aq_{n-1} - bp_{n-1}}{bq_{n-1}} < \frac{1}{q_nq_{n-1}} \quad \text{or} \quad b > q_n(aq_{n-1} - bp_{n-1}).$$

But $aq_{n-1} - bp_{n-1} > 0$ since $a/b > p_{n-1}/q_{n-1}$. Hence $b > q_n$; that is: *If a/b is a closer approximation to the value of the continued fraction than p_n/q_n, is, its denominator b must be greater than q_n.*

If p_{n-2}/q_{n-2}, p_{n-1}/q_{n-1}, p_n/q_n are successive convergents to the continued fraction x we know by (3), if n is odd say, that p_{n-2}/q_{n-2}, p_n/q_n, x, p_{n-1}/q_{n-1} are in increasing order of magnitude. By (13) we know that no fraction whose denominator is less than q_{n-1} can lie in the interval p_{n-2}/q_{n-2}, p_{n-1}/q_{n-1}, and that no fraction whose denominator is less than q can lie in the interval p_n/q_n, p_{n-1}/q_{n-1}. But we have no assurance that a fraction whose denominator is less than q_n may not lie in the interval p_{n-2}/q_{n-2}, p_n/q_n for from (7), when n is odd, $p_nq_{n-2} - p_{n-2}q_n = a_n$ where $a_n \geqq 1$. If, when $a_n > 1$, we form from two successive principal convergents p_{n-2}/q_{n-2}, p_{n-1}/q_{n-1} the series of fractions

$$\frac{p_{n-2}}{q_{n-2}}, \frac{p_{n-2}+p_{n-1}}{q_{n-2}+q_{n-1}}, \frac{p_{n-2}+2p_{n-1}}{q_{n-2}+2q_{n-1}}, \cdots, \tag{14}$$
$$\frac{p_{n-2}+(a_{n-1})p_{n-1}}{q_{n-2}+(a_{n-1})q_{n-1}}, \frac{p_{n-2}+a_np_{n-1}}{q_{n-2}+a_nq_{n-1}} = \frac{p_n}{q_n}$$

form, according as n is odd or even, an increasing or decreasing series. The set of fractions (14) so generated and inserted between p_{n-2}/q_{n-2}, p_n/q_n are called *intermediate convergents*. Such a series may be formed for each partial quotient $a_n > 1$.

Now any two consecutive fractions of the series (14) may be represented by

$$\frac{P}{Q} = \frac{p_{n-2}+rp_{n-1}}{q_{n-2}+rq_{n-1}}, \qquad \frac{P'}{Q'} = \frac{p_{n-2}+(r+1)p_{n-1}}{q_{n-2}+(r+1)q_{n-1}}$$

where r is an integer such that $0 \leqq r \leqq a_n - 1$. Hence

$$\frac{P'}{Q'} - \frac{P}{Q} = \frac{P'Q - PQ'}{QQ'} = \frac{(q_{n-1} + rq_{n-1})[p_{n-2} + (r+1)p_{n-1}]}{QQ'}$$

$$\frac{-(p_{n-2} + rp_{n-1})[q_{n-2} + (r+1)q_{n-1}]}{QQ'}$$

$$= (p_{n-1}q_{n-2} - p_{n-2}q_{n-1})/QQ' = (-1)^{n-1}/QQ',$$

by (8), whence $P'Q - PQ' = (-1)^{n-1}$. That is neither pair P, Q; P', Q' have common factors since they would necessarily divide $(-1)^{n-1}$. And since $P' > P$, $Q' > Q$ when n is odd the series (14) is increasing. Also when n is odd, we have in increasing order of magnitude, where x is the value of (1),

$$p_{n-2}/q_{n-2},\ P/Q,\ p_n/q_n,\ x,\ p_{n-1}/q_{n-1}$$

or

$$\frac{P}{Q} < x < \frac{p_{n-1}}{q_{n-1}} \quad \text{whence} \quad x - \frac{P}{Q} < \frac{p_{n-1}}{q_{n-1}} - \frac{P}{Q}.$$

But

$$\frac{p_{n-1}}{q_{n-1}} - \frac{P}{Q} = \frac{Qp_{n-1} - Pq_{n-1}}{q_{n-1}Q} = \frac{p_{n-1}(q_{n-2} + rq_{n-1}) - q_{n-1}(p_{n-2} + rp_{n-1})}{q_{n-1}Q}$$

$$= \frac{p_{n-1}q_{n-2} - p_{n-2}q_{n-1}}{q_{n-1}Q} = \frac{(-1)^{n-1}}{q_{n-1}Q} = \frac{1}{q_{n-1}Q} \qquad \text{(since } n \text{ is odd)}$$

whence

$$x - \frac{P}{Q} < \frac{1}{q_{n-1}Q}$$

and by (11)

$$x - \frac{P}{Q} < \frac{1}{q_{n-1}Q} < \frac{1}{q_{n-1}^2}, \quad Q = q_{n-2} + rq_{n-1}. \tag{15}$$

Clearly (15) gives *error limits* for the approximation to the rational fraction x by an *intermediate convergent.*

From the above discussion it is seen that we may form for the continued fraction (1) two sequences of convergents using the *odd* principal convergents and their intermediate convergents for one (an increasing sequence), the *even* principal convergents and their intermediate convergents for the other (a decreasing sequence). These two sequences contain all the rational approximations to (1). That is one of these sequences will terminate, the last convergent being (1) itself. The other may be extended indefinitely for if p_{n-1}/q_{n-1}, p_n/q_n are the last two principal convergents, the series

$$\frac{p_{n-1}}{q_{n-1}},\ \frac{p_{n-1} + p_n}{q_{n-1} + q_n},\ \frac{p_{n-1} + 2p_n}{q_{n-1} + 2q_n},\ \cdots$$

forms either a continually increasing or a continually decreasing series in which no principal convergent occurs but whose individual terms approach more nearly the value p_n/q_n or x. (This is easily seen since we may consider the last partial quotient to be α and the last convergent $(p_{n-1}+\alpha p_n)/(q_{n-1}+\alpha q_n)$ whose limit as $\alpha \rightarrow \infty$ is p_n/q_n.)

Let us begin with a rational approximation to e, $e=2.7182818285$, and find the rational approximations to this number by the preceding methods. We need first to find the values of the partial quotients $a_1, a_2, \cdots, a_n$. We apply the method for finding the greatest common divisor of two integers:

$$\frac{27182818285}{10000000000} = 2 = a_1, \qquad \begin{array}{r} 27182818285 \\ 20000000000 \\ \hline 7182818285 \end{array} \Big| \; 10000000000 = 1 = a_2,$$

$$\begin{array}{rrl} 10000000000 & & \\ 7182818285 & & \\ \hline 2817181715 \,\big| & 7182818285 & = 2 = a_3 \\ & 5634363430 & \\ \cline{2-2} & 1548454855 \,\big| & 2817181715 = 1 = a_4, \text{ etc.,} \end{array}$$

and continuing in this way to termination of the division process find *the partial quotients* to be:

$$\begin{aligned} &a_1 = 2,\ a_2 = 1,\ a_3 = 2,\ a_4 = 1,\ a_5 = 1,\ a_6 = 4,\ a_7 = 1,\ a_8 = 1,\ a_9 = 6, \\ &a_{10} = 1,\ a_{11} = 1,\ a_{12} = 8,\ a_{13} = 1,\ a_{14} = 1,\ a_{15} = 12,\ a_{16} = 2,\ a_{17} = 3, \\ &a_{18} = 1,\ a_{19} = 1,\ a_{20} = 2,\ a_{21} = 1,\ a_{22} = 1,\ a_{23} = 1,\ a_{24} = 2,\ a_{25} = 3, \\ &a_{26} = 1,\ a_{27} = 4. \end{aligned} \tag{16}$$

By use of the formulas (5) the *principal convergents* corresponding to the *partial quotients* (16) are found to be:

$$\begin{aligned} &A_1 = 2,\ A_2 = 3,\ A_3 = 8/3,\ A_4 = 11/4,\ A_5 = 19/7,\ A_6 = 87/32, \\ &A_7 = 106/39,\ A_8 = 193/71,\ A_9 = 1264/465,\ A_{10} = 1457/536, \\ &A_{11} = 2721/1001,\ A_{12} = 23225/8544,\ A_{13} = 25946/9545, \\ &A_{14} = 49171/18089,\ A_{15} = 615998/226613,\ A_{16} = 1281167/471315, \\ &A_{17} = 4459499/1640558,\ A_{18} = 5740666/2111873, \\ &A_{19} = 10200165/3752431,\ A_{20} = 26140996/9616735 \\ &A_{21} = 36341161/13369166,\ A_{22} = 62482157/22985901 \\ &A_{23} = 98823318/36355067,\ A_{24} = 260128793/95696035, \\ &A_{25} = 879209697/323443172,\ A_{26} = 1139338490/419139207, \\ &A_{27} = 5436563657/2000000000 = 2.7182818285. \end{aligned} \tag{17}$$

For the $a_i > 1$ of (16) we compute the *intermediate convergents* from (17) by means of formulas (14):

$$\begin{aligned}
&\overline{A_1A_3}\ 5/2;\ \overline{A_4A_6}\ 30/11,\ 49/18,\ 68/25;\ \overline{A_7A_9}\ 299/110,\ 492/181,\\
&685/252,\ \underline{878/323,}\ 1071/394;\ \overline{A_{10}A_{12}}\ 4178/1537,\ 6899/2538,\\
&9620/3539,\ 12341/4540,\ 15062/5541,\ 17783/6542,\ 20504/7543;\\
&\overline{A_{13}A_{15}}\ 75117/27634,\ 124288/45723,\ 173459/63812,\ 222630/81901,\\
&271801/99990,\ 320972/118079,\ 370143/136168,\ 419314/154257,\\
&468485/172346,\ 517656/190435,\ 566827/208524;\\
&\overline{A_{14}A_{16}}\ 665169/244702;\ \overline{A_{15}A_{17}}\ 1897165/697928,\ 3178332/1169243;\\
&\overline{A_{18}A_{20}}\ 15940831/5864304;\ \overline{A_{22}A_{24}}\ 161305475/59340968;\\
&\overline{A_{23}A_{25}}\ 358952111/132051102,\ 619080904/227747137;\\
&\overline{A_{25}A_{27}}\ 2018548187/742582379,\ 3157886677/1161721586,\\
&4297225167/1580860793.
\end{aligned} \tag{18}$$

Finally we form our two sequences of odd and even convergents with intermediates from (17) and (18) which will be:

$$\begin{aligned}
&A_1,\ \overline{A_1A_3},\ A_3,\ A_5,\ A_7,\ \overline{A_7A_9},\ A_9,\ A_{11},\ A_{13},\ \overline{A_{13}A_{15}},\ A_{15},\ \overline{A_{15}A_{17}},\\
&A_{17},\ A_{19},\ A_{21},\ A_{23},\ \overline{A_{23}A_{25}},\ A_{25},\ \overline{A_{25}A_{27}},\ A_{27};
\end{aligned} \tag{O}$$

$$\begin{aligned}
&A_2,\ A_4,\ \overline{A_4A_6},\ A_6,\ A_8,\ A_{10},\ \overline{A_{10}A_{12}},\ A_{12},\ A_{14},\ \overline{A_{14}A_{16}},\ A_{16},\\
&A_{18},\ \overline{A_{18}A_{20}},\ A_{20},\ A_{22},\ \overline{A_{22}A_{24}},\ A_{24},\ A_{26},\ \cdots
\end{aligned} \tag{E}$$

or

$$\begin{aligned}
&2,\ 5/2,\ 8/3,\ 19/7,\ 106/39,\ 299/110,\ 492/181,\ 685/252,\ \underline{878/323,}\\
&1071/394,\ 1264/465,\ 2721/1001,\ 25946/9545,\ 75117/27634,\\
&124288/45723,\ 173459/63812,\ 222630/81901,\ 271801/99990,\\
&320972/118079,\ 370143/136138,\ 419314/154257,\ 468485/172346,\\
&517656/190435,\ 566827/208524,\ 615998/226613,\ 1897165/697928,\\
&3178332/1169243,\ 4459499/1640558,\ 10200165/3752431,\\
&36341161/13369166,\ 98823318/36355067,\ 358952111/132051102,\\
&619080904/227747137,\ 879209697/323443172,\ 2018548187/742582379,\\
&3157886677/1161721586,\ 4297225167/1580860793,\\
&5436563657/2000000000.
\end{aligned} \tag{O}$$

$$\begin{aligned}
&3,\ 11/4,\ 30/11,\ 49/18,\ 68/25,\ 87/32,\ 193/71,\ 1457/536,\\
&4178/1537,\ 6899/2538,\ 9620/3539,\ 12341/4540,\ 15062/5541,\\
&17783/6542,\ 20504/7543,\ 23225/8544,\ 49171/18089,\\
&665169/244702,\ 1281167/471315,\ 5740666/2111873,\\
&15940831/5864304,\ 26140996/9616735,\ 62482157/22985901,\\
&161305475/59340968,\ 260128793/95696035,\ 1139338490/419139207,\ \cdots
\end{aligned} \tag{E}$$

We are now in a position to find rational number approximations to e to fulfill certain conditions. For instance let us find the best rational approximation to e whose denominator does not exceed 100. Examining the series (O) and (E) we find from (E) the fraction with largest denominator less than 100 to be 193/71. Now examining (17) it is seen that 193/71 is the *principal convergent* A_8 and from (16) its corresponding *partial quotient* a_8 just precedes a larger $a_9=6$, so it is expected that 193/71 will be a good approximation. (See the discussion following (11).) From relations (11) and (17) we can compute the error limits $193/71=2.71830985$; which gives e correct when rounded to 4 decimals; an excess of 3 in the 5th decimal if e is rounded to 5 decimals; etc.

Consider now the best approximation to e by the ratio of two: 2-digit integers, 3-digit integers, 4-digit integers, 5-digit integers, etc.

Ratio of Two	*Series (E)*	*Series (O)*
2-digit integers	$87/32=2.71875$	
3-digit integers		$\dfrac{878}{323}=2.718266253$
4-digit integers	$9620/3539=2.718282$	$2721/1001=2.718281718$
5-digit integers	$49171/18089=2.7182818287$	$75117/27634=2.718281826$

Note that 49171/18089 *gives e correct to* 2 *in the* 10*th decimal place.* Examining (17) we see that 49171/18089 is the *principal convergent* A_{14} and its corresponding partial quotient a_{14} from (16) just precedes the large *partial quotient* $a_{15}=12$. Hence it would be expected that this ratio would give a good approximation to e. Similarly it is expected that 2721/1001 would give a good approximation to e since it is a *principal convergent*, A_{11}, and its corresponding partial quotient a_{11} just precedes a large partial quotient $a_{12}=8$.

If we take $\pi=3.1415926536$, and follow the same procedure as was done for e we find the *partial quotients:*

$$\begin{aligned} &a_1 = 3,\ a_2 = 7,\ a_3 = 15,\ a_4 = 1,\ a_5 = 292,\ a_6, = 1,\ a_7 = 1,\ a_8 = 1,\ a_9 = 4,\\ &a_{10} = 1,\ a_{11} = 1,\ a_{12} = 1,\ a_{13} = 45,\ a_{14} = 1,\ a_{15} = 1,\ a_{16} = 8. \end{aligned} \tag{19}$$

The principal convergents:

$$\begin{aligned} A_1 &= 3/1,\ A_2 = 22/7,\ A_3 = 333/106,\ A_4 = \underline{355/113},\ A_5 = 103993/33102,\\ A_6 &= 104348/33215,\ A_7 = 208341/66317,\ A_8 = 312689/99532,\\ A_9 &= 1459097/464445,\ A_{10} = 1771786/563977,\ A_{11} = 3230883/1028422,\\ A_{12} &= 5002669/1592399,\ A_{13} = 228350988/72686377,\\ A_{14} &= 233353657/74278776,\ A_{15} = 461704645/14695153,\\ A_{16} &= 3926990817/1250000000. \end{aligned} \tag{20}$$

We note from (19) and (20) that since a_5, a_9, a_{13} are relatively large *partial quotients* preceded by partial quotients which are unity, the approximations by

A_4, A_8, A_{12} should be very good. Actually from (20) find

$$A_4 = 355/113 = 3.141592\overset{\downarrow}{9}2035, \qquad A_8 = 3.141592653\overset{\downarrow}{6}18,$$

$$A_{12} = 3.141592653\overset{\downarrow}{6}00009.$$

Since there are 291 intermediate convergents between A_3 and A_5, we handle special problems as follows: Required the best representation of π by the ratio of two-4-digit numbers. From A_3 and A_4, (20), one has $333+r355<10{,}000$, whence $r<9667/355=27.2$. With $r=27$,

$$\frac{333+r355}{106+r113}=\frac{9918}{3157}=3.14159\overset{\downarrow}{0}117\cdots.$$

Similarly the best ratio of two-5 digit numbers is given from A_3 and A_4 where now $333+r355<100{,}000$, or $r<(99667)/335=280.75\cdots$, take $r=280$, whence

$$\frac{333+r355}{106+r113}=\frac{99733}{31746}=3.1415926\overset{\downarrow}{4}1.$$

Summary. Every rational number has a unique terminating simple continued fraction expansion (1) which is quite easily generated by applying the method for finding the greatest common divisor of two integers. The partial quotients obtained have corresponding principal convergents which are quite easily computed by (5). For those partial quotients greater than unity, intermediate convergents may be computed, and these with the principal convergents can be arranged into two series—one of odd and one of even convergents. Since these series will contain all the possible convergents to the continued fraction one can then choose from them the particular ones to satisfy any particular requirement as the best approximation by the ratio of two n-digit numbers, the best approximation where the denominator does not exceed some number n, etc. since, if no principal convergent satisfies the requirements, the intermediate convergent with the largest denominator satisfying the requirements gives the best approximation, etc.

SCIENCE AND LOGIC*

CECIL G. PHIPPS, Tennessee Polytechnic Institute

1. Introduction. The importance of science is apparent when one considers the use to be made of it. Science is the key with which man hopes to unlock the secrets of the Universe. By gaining an understanding of science, the Universe can be made to work for him instead of against him. No other force is working with comparable power to transform his life on this planet.

The world into which you who are reading this were born is gone beyond

* The preparation of this paper was sponsored by The Tennessee Academy of Science and the State Department of Education and supported by a grant from the National Science Foundation.

recall, except in fading memories. The world in which all of us are now living is changing rapidly. What the future holds we can only guess. At the present moment we are poised between a future with the same rapid development and a future ending in one grand holocaust made possible by the recent advances in science.

Under the pressure of these outside events, the curriculum of the schools has been forced to change. In an attempt to meet the changed conditions, some old subjects have received new emphasis, others have gained new contents, and new subjects have been added.

In the intensive study of science, logic has received great emphasis. It is the purpose of this note to examine how science and logic work together.

Often logic is applied to science through the medium of applied mathematics; it can even be applied to mathematics itself to develop a branch of pure mathematics for its own sake. However, the way in which logic is applied to any branch of science is much the same in every case.

2. Science. The question is often asked: What is Science? This question is easier to answer if we examine the separate branches of science rather than science as a whole. We examine the individual trees to gain a better understanding of the forest.

In the table below are listed some of the more important branches of science. Others could be added to the list and perhaps each of the ones given could be subdivided. The listing is for illustrative purposes only.

TABLE I

Mathematics	Psychology
Astronomy	Economics
Physics	Political Science
Chemistry	Social Science
Biology	Medicine

A *science*, or a *branch* of science, is any organized body of knowledge. The two key words in this definition are "organized" and "knowledge." Thus, social science, political science, and other organized bodies of knowledge are correctly designated as sciences under this definition, as well as the more familiar sciences, mathematics, physics and chemistry.

A distinction between the two groups of sciences as they appear above is in the manner of their organization. A deductive science is one which is organized entirely by the rules of logic. An empirical science is one which is organized as a result of observation and experiment.

Practically all sciences, as we know them, are mixtures of the two types. If, in popular jargon, one is considered more scientific than another, it is that we hold a body of knowledge to be more scientific the more its facts are derived by logical reasoning. In effect, we tend to classify the sciences by the amount of deductive reasoning which they contain.

On this basis, logic itself and several branches of mathematics would head the list as the most scientific subjects. At the bottom of the list would appear the social sciences and, perhaps surprisingly, much in the practice of medicine.

Further inspection yields another conclusion. When an ordering is made in the sciences according to the amount of logic used in drawing conclusions, it is almost identical with an ordering based on the amount of mathematics used to reach these conclusions.

This discovery points up the importance of mathematics and the increasing emphasis placed upon its study. One's ability to work in many fields of science is often limited by the kind and amount of mathematics he does not know.

For those sciences which make the least direct use of mathematics a substitute has been devised in the methods of statistics based on the laws of probability. For example, medicine and other science cannot predict, with certainty, how long a given individual will live. But mortality tables can be constructed which show the probability of his living any given length of time. These figures are so dependable that, based upon a large number of individuals, life insurance can be sold on long contracts as a sound financial investment.

3. Logic: Analogy. Having taken this introductory look at science our attention is turned next to logic. A first impulse is to think that logic applies only to mathematics and philosophy. This is not strictly true; logic applies to all scientific endeavor. In fact, machines can be built which solve simple problems in logic.

There are three methods of logical reasoning. In order of increasing logical rigor, they are *analogy*, *induction* and *deduction*. All three methods are used in science.

Logic applies to statements about objects and the relations between them. These statements can be divided rather precisely into two types: specific or particular statements about particular objects and general statements about a class of objects.

For example, "this is a right angle" is a particular statement; "all right angles are equal" is a general statement. Again, "this animal is a frog" is a particular statement; "all frogs begin life as pollywogs" is a general statement.

The different methods of reasoning apply to the different ways one can pass from one type of statement to another statement of the same or a different type. Numerous examples will make the meaning clear.

In using analogy, one is reasoning from one particular statement to another particular statement. For example, if you know a family in which the grandfather's name is John and the father's name is John, then if you expect the name of the eldest son to be John also, you are reasoning by analogy.

Since the name of this son might not be John, it is possible to reach a false conclusion from true premises by the use of analogy. Therefore, the method must be used with caution.

However, there are numerous cases where the method is useful. Aristotle was using analogy when he classified whales and dolphins among the fishes. It is only because we have more knowledge that we use analogy to classify them among the mammals.

The use of analogy is an indispensable method of teaching. The parables of Christ were His most effective teaching device. Whenever we give an illustration of a general principle, we are using analogy. In English composition, similes and metaphors are examples of brief analogies. In fact, it would be almost impossible to learn anything beyond our own individual experience if we could not reason by analogy; there would be no transfer of training.

In mathematics, an analogy can be set up between the real numbers and the points of a line. Further, an analogy can be set up between all the complex numbers and the points of a plane. These two examples are, so far as we know, perfect analogies; that is, conclusions reached in reasoning by analogy are always true.

In a more elaborate example from the field of physics, the analogy is not perfect beyond a certain point. Consider the following table comparing the flow of water in pipes to the flow of electricity in wires.

TABLE II

water	electricity
pipe	wire
valve	switch
reservoir	storage battery
friction to flow	resistance
etc.	etc.
?	Magnetic field about a conductor carrying a current

In the last item, the analogy breaks down: there is no counterpart in the hydraulic system for the magnetic field surrounding a wire carrying a current.

From this example and many similar ones it is possible to draw some general conclusions. When the analogy holds, the same mathematics applies in both cases; when the analogy fails, the mathematics must be different. For example, Maxwell's equations describe the action of the magnetic field; these equations do not apply to a hydraulic system. For each new situation there must be new mathematics leading to the infinite variety of mathematics.

The ability to recognize analogies where they exist is often considered a mark of intelligence and scholastic ability. In this connection they are used extensively in tests administered by the Armed Forces and many schools.

4. Logic: Induction. The next method of reasoning to be considered is induction. In this method one reasons from a set of particular examples to a general conclusion. Use has been made of it frequently in the above discussion. The water-electricity example is an illustration.

However, like analogy, induction is not always a sure method of reasoning. Suppose for example that one writes down the first four odd numbers, 1, 3, 5, 7. They are prime numbers; but one cannot safely conclude from this that all odd numbers are prime. A single example such as 9 or 15 would show the general conclusion to be false.

But just as analogy was indispensable in the teaching and learning process, induction with all its shortcoming is an indispensable part of the process known as "the scientific method." This process explains the answer to the question:

how does one go about the scientific study of something? The method is basic to all attempts to give a scientific explanation to the Universe and its various phenomena. It involves a succession of steps in which there is not necessarily a last one.

The first step in all cases involves the collection of all pertinent data and observations. In astronomy, for instance, these observations may extend over several centuries.

With the data and observations before him, the observer is ready for the second important step. In a flash of comprehension or inspiration, he perceives a general principle behind all these. It may be a trivial one such as a child discovering that we use the week-day names over and over again. Or it may be a profound one such as a concept for the solar system or a theory of evolution.

With a general principle stated, the next step is to test it. Is it true; i.e., does it conform to the known facts better than any other similar hypothesis? The various theories of the solar system are examples of this method in action.

After the tests of whether the general principle explains, the next step is to test whether it predicts. It is of little use if it does not. If it does predict, then the predicted results are tested by an experiment. The prediction of the bending of light in the theory of relativity is an example of such a situation.

This process of revision, refinement, or even replacement, is an endless process. This is the scientific method.

The ability to modify or reject a favorite hypothesis is the mark of a loyal and true scientist. In some cases the Communists find this hard to do; they attempt to make the facts conform to their theories instead of the other way round.

5. Logic: Deduction. At one step in the scientific method, conclusions were drawn from the hypotheses by the use of deductive reasoning. This method leads from the general to the general or from the general to the particular. When one speaks of logic, he usually refers to the methods of deductive logic.

A purely deductive science, or the deductive part of any science, is composed of four parts: undefined terms, defined terms, assumptions and theorems. The purpose of the first three parts is to produce the fourth.

Just as one cannot raise a crop of wheat without some wheat as seed, so one cannot construct a deductive system without some undefined terms with which to begin the discussion. There must be something to talk about.

A common understanding of the undefined terms is gained by having common experiences or by analogies from other experiences. Sometimes it may take a long time and much study before one gains a complete understanding of these terms. For this reason the general public may fail to grasp the full meaning of some scientific situations.

A measure of one's intellectual maturity is in part a measure of one's supply of undefined terms with which he is familiar. If one who is equipped with this understanding knows the rules of logic and is willing to abide by them, he is ready to begin the study of a deductive system.

Examples of undefined terms can be taken from many fields. In geometry, point, line and plane are three such terms. In physics, electron is an undefined term: some of the things it will do are known; what it is is unknown. In chemistry, the names of the elements are usually taken as undefined terms. In biology, life is an undefined term: it can be described but, until it is created in the laboratory, it must remain undefined.

Other terms used in a deductive system may be defined by relating them to the undefined terms or other terms previously defined. There are standard procedures and tests for the construction of a good definition. However, this is not the place to discuss them.

In passing, we note that defined terms are not a logical necessity for ultimately each defined term is expressible by means of undefined terms alone. Definitions therefore are merely a human convenience; they save time and are an almost indispensable aid to human thought. If some animals can reason, as some psychologists believe, the animals do not use definitions since they lack the power of speech.

Terms, both defined and undefined, may be any part of speech except interjections and pronouns. Some terms which are verbs or verb phrases are of special importance: they express conditions called relations. Some prepositions, such as "in" and "on," may do the same.

Assumptions form the third group of parts in a deductive system; they express relations between the terms of the system. For example, "two points determine a line" is a statement which expresses the relation between two points and the line passing through them. In biology, until the present time at least, it is assumed that "all life comes from a living thing."

Now, most of us have heard or been engaged in arguments. But, arguments are possible only if two or more people are reasoning from different premises or assumptions. The argument is certain to be inconclusive as long as the differing assumptions remain unstated. However, once they are stated, then there may be disagreement, but there can be no argument.

Hence to develop a deductive system it is necessary to have complete agreement on the assumptions to be used. Among the several conditions which these assumptions must satisfy, the most important is that they be consistent, that is, it must not be possible to draw contradictory conclusions from them.

6. Logic: Theorems. The fourth group of parts of a deductive system is composed of theorems. These are additional relations between the terms derived from the assumptions by the rules of logic. The development of these theorems is the sole reason for the existence of the system. In a large system, the number of theorems is as endless as the varieties of one's experience or one's imagination.

There is no certain way to find all the theorems of a system. And, if a statement is suspected of being a theorem, there is no certain way of discovering a logical proof of this fact. This is the part of science which remains an art depending, as it does, upon the skill, imagination and experience of the worker.

The discovery of new facts, the discovery of a logical proof or the verification of a fact by an experiment is research. To a practitioner engaged in this work this discovery is a thrill that repays him for the hours and days of drudgery which went into the project.

SIMPLE REGULAR SPHERE PACKINGS IN THREE DIMENSIONS

IAN SMALLEY, Northampton College, London

Introduction. Rogers (1) has defined a packing as a system of equal closed spheres in n-dimensional space in which no two spheres have any inner point in common. If classified in terms of density, regular packings in three dimensions range from $0.056 \cdots$ to $0.740 \cdots$. These limits have been established empirically, they are not the proven bounds of density. In the highest density packing each sphere touches twelve others. There is an infinite number of variations of this packing in which the spheres have slightly differing neighbour relationships. There are only two in which the spheres are all equivalent as Barlow discovered in 1883. Interest in the loosest packings came much later; the loosest packings known were discovered by Heesch and Laves in 1933. The packings of extreme density have been fully discussed by Melmore (2).

There exists a range of regular packings having densities intermediate between the two extremes. These may be distinguished by density; by coordination number (CN), that is the number of spheres touching every sphere in the pack; and by the shape of the Voronoi polyhedron associated with the lattice formed by the centres of the spheres. The Voronoi polyhedron (VP) is a convex polyhedron associated with each point in the lattice, and these polyhedra pack together to fill space. The planes which form the faces of the polyhedron are the perpendicular bisectors of the lines joining the centre of the reference sphere to the centres of the nearest spheres. Rogers uses this terminology but Coxeter (3) prefers to call this figure the Dirichlet region. He defines the two dimensional case as a polygon whose interior consists of all the points in the plane which are nearer to a particular lattice point than any other lattice point.

Regular packings can be represented by point lattices formed by the centres of the spheres and may be divided into two types, simple and non-simple. The lattice is represented by a unit cell, this is the simplest convenient figure which by means of suitable translatory movements will describe the whole lattice. The unit cell for the simple lattice is the simplest possible: a parallelepiped with lattice points only at the corners. The cell for a non-simple lattice will have lattice points in the interior or on the boundary. It is possible for different parallelepipeds to give the same lattice provided they have the same volume. One particular cell is chosen for the derivation of the range of simple packs, the conditions for this are described below. The unit cell of a simple pack can be defined by the centres of eight adjacent spheres and each of these spheres contributes to eight cells. Each cell contains in effect one sphere.

Derivation of Packings. The packings consist of spheres which are:

(1) of unit diameter,
(2) of unit density, the external space has zero density,
(3) of equal coordination and with the same spatial orientation with respect to contact points.

The separate packings are derived by variation of the possible shape of the unit cell of the centre point lattice. If the cell is defined by the shape of all six sides it is unnecessary to also define the angles. One simple cell is equivalent to the unit lattice as defined by Hilbert and Cohn-Vossen (4), this is the cube of unit edge. This gives a packing for which the VP is a cube of equal size. The unit lattice has unit volume and generates the fundamental point lattice as defined by Hardy and Wright (5). It can be defined by three vectors making unit intercepts on three orthogonal axes. By varying the angles between these vectors within the limits prescribed by simple packing other packings can be derived.

Each side of the lattice generating parallelepiped represents a layer of spheres and may vary between a square and a rhombus with angles of 60° and 120° which represents the closest packing of equal spheres in the plane. The square faces can be represented by the symbol A; the 60° rhombus faces by C and the intermediate rhombus by B. Cell variation can be shown by considering the movement of one layer of four spheres over the other. This movement must be translatory only, no rotation is allowed. Movement parallel to a layer edge is represented by the symbol a; movement parallel to a layer face diagonal provided that the two diagonals are equal is represented by b.

The cubic cell already mentioned represents a packing with coordination number 6 and a density of 0.513. This density is obtained directly from the volume of the cell since each cell in a simple packing contains the equivalent of one sphere. The cubic cell has six A faces and two of these must remain constant throughout all the movements. Because there are no B or C faces the cell symbol is 600. The multiplication table below shows the group of complete movements from 600.

	I	aAp	bAp
I	I	aAp	bAp
aAp	aAp	I	aAp
bAp	bAp	aAp	I

The symbol p means that the move is complete. The move is considered complete when the top layer rests in the interstices in the lower layer. Movement of a layer face will cause the distance between cell layer faces to change. A complete move produces the maximum possible change in the layer separation.

$$\boxed{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \end{matrix}} \qquad \boxed{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{\sqrt{2}}{2} \end{matrix}}$$

$$\begin{aligned} aAp &= 600 \rightarrow 402 \\ &\quad CN + 2 \\ bAp &= 600 \rightarrow 204 \\ &\quad CN + 6 \end{aligned}$$

aAp $\qquad$ bAp

Any limited intermediate move can be symbolized by l, thus; $aAl = 600 \rightarrow 420$, $CN+0$. $bAl = 600 \rightarrow 240$, $CN+0$. These limited moves produce cells with B sides which have a range of possible angles. In the movement matrices these are defined by x, y and z. These have an allowable range as follows; $90° > x, y, z > 60°$. It is sufficient to define the acute angles; three are required because three pairs of sides make up each cell. All cells having no C sides have $CN=6$.

$$\underset{aAl}{\begin{array}{|ccc|}1 & 0 & 0\\ 0 & 1 & 0\\ N & 0 & M\end{array}} \qquad \underset{bAl}{\begin{array}{|ccc|}1 & 0 & 0\\ 0 & 1 & 0\\ N & N & \sqrt{M^2+N^2}\end{array}} \qquad \underset{cAl}{\begin{array}{|ccc|}1 & 0 & 0\\ 0 & 1 & 0\\ N & K & \sqrt{M^2-K^2}\end{array}} \qquad \begin{array}{l}M = \sin x\\ N = \cos x\\ K = \cos y\end{array}$$

The c move is in any direction other than a or b. Move cAp leads to the 222 structure; this structure with two A, two B and two C faces is only produced by this move. Any move short of this (i.e. cAl) gives the same structure as bAl, that is 240.

$$cAp = 600 \rightarrow 222,\ CN+2. \qquad cAl = 600 \rightarrow 240,\ CN+0.$$

The 240 structure produced by bAl is a special case of the general 240 structure as produced by cAl. Since the b move is parallel to an A face diagonal the four B faces will be equal. The move matrix for cAl is shown above; it represents the general matrix for moves of an A face. While the coordination number is still six the Voronoi polyhedron is a distorted cube. When C sides are produced the CN increases and the VP is a more complex figure. For 402 it is a hexagonal prism of unit height and for other $8CN$ packings it is a distorted version of this. With the 204 structure the CN rises to twelve and the VP is a rhombic dodecahedron. The production of this solid from a $12CN$ pack is described by Steinhaus (6) and Cundy and Rollett (7).

When a B or C face is being moved the symbol b will not suffice for diagonal movement since the diagonals are unequal. The symbol b' will be used to represent movement parallel to the short diagonal and b'' for moves parallel to the long diagonal. The 420 cell makes a convenient starting point for variations of a B layer. If a coordinate system based on this cell is used certain move matrices are similar to those describing moves from 600. The table of moves is shown below.

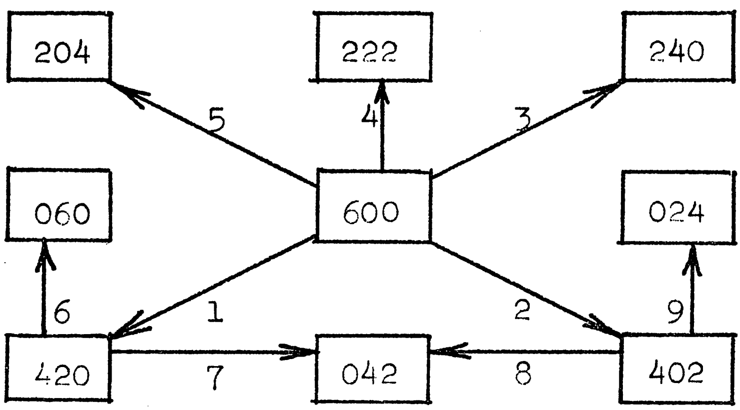
	I	aBp	$b'Bp$
I	I	aBp	$b'Bp$
aBp	aBp	I	aBp
$b'Bp$	$b'Bp$	aBp	I

B moves tend to produce 060 structures with six B faces and a minimum of regularity. Of the six possible moves (ignoring b'' moves) of a B face four lead to 060. These are shown in the table below. A C face is a special B face with the angles fixed. The a; b' group of moves is the same as for a B face, simple substitution of C for B in the table above will give the C move table. Five of the possible six C moves give 042, the other one produces the only $10CN$ structure, 024. If the VP for this packing is constructed by drawing the planes bisecting the lines connecting the centre of one sphere to the centres of the ten nearest spheres a ten sided solid is produced which does not pack to fill space and thus cannot be the true VP. With this packing it is necessary to consider not only the nearest spheres but also the next nearest, this gives a distorted truncated octahedron, a figure with fourteen sides. There appears to be an inequality

$$R_2 < \sqrt{2}R_1$$

which determines the nature of the Voronoi polyhedron; in this R_1 is the distance of the nearest spheres and R_2 is the distance of the next nearest. This is derived from the 600 pack in which the edge of the VP, which is a cube, lies in the plane at $R_2/2$ from the reference sphere and this distance is $\sqrt{2}/2$. If this relationship applies then the VP will have more than n faces where n is the CN. For the 024 pack $R_2 = 1.225$; therefore the VP has more than ten sides. For all other packs the VP has n sides.

The 006 cell is obtained from 402 by a b'' move; this has the same VP as the 204 cell and is just another way of representing the one $12CN$ pack. It is preferable to use 204 then the b'' move is not required at all. The complete range of



moves and simple structures produced can be illustrated by means of the figure above. The numbers represent the moves as follows:

1, aAl. 2, aAp. 3, cAl, bAl. 4, cAp. 5, bAp. 6, $b'Bp$, aBl, $b'Bl$, cBl. 7, aBp, cBp. 8, $b'Cp$, cCp, aCl, $b'Cl$, cCl. 9, aCp.

The pack symmetries are based on crystallographic criteria which are used to assign crystals to one of seven classes. Only the monoclinic system is not represented by a simple sphere packing. Certain apparent anomalies occur due to choice of unit cell. The twelve coordinated packing has a very high symmetry but is represented in the table by a cell of the lowest symmetry, i.e., triclinic. If a non-simple cell is chosen cubic symmetry can be shown.

Some details of the packings produced are given in the table below.

TABLE OF POSSIBLE CELLS REPRESENTING SIMPLE PACKINGS

Symbol	Sides	*CN*	Angles	Symmetry
600	$6A$	6	90°, 90°, 90°	cubic
420	$4A2B$	6	90°, 90°, x	orthorhombic
240	$2A4B$	6	90°, x, y	triclinic
060	$6B$	6	x, y, z	triclinic
402	$4A2C$	8	90°, 90°, 60°	hexagonal
204	$2A4C$	12	90°, 60°, 60°	triclinic
006	$6C$	12	60°, 60°, 60°	rhombohedral
042	$4B2C$	8	60°, x, y	triclinic
024	$2B4C$	10	60°, 60°, 75°31′	tetragonal
222	$2A2B2C$	8	90°, 60°, x	triclinic

Non-simple packings are of several varieties: (a) high density, these are variants of 204, (b) low density, with coordination numbers of three and four and (c) intermediate density, based on the $8CN$ pack for which the VP is a truncated octahedron. This figure is the so-called Kelvin body, of great importance in space filling studies. If every sphere in a packing is necessarily equivalent then there is only one non-simple $12CN$ packing; if two types of spheres are allowed (same size and CN but different VP) then there is an infinite number of $12CN$ non-simple packings. These have been discussed by Wells (8) and it is interesting to note that their existence has been categorically denied by Graton and Fraser (9).

References

1. C. A. Rogers, Proc. London Math. Soc., (3) **8** (1958) pp. 609–620.
2. S. Melmore, Mineral Mag., **28** (1949) pp. 479–483.
3. H. S. M. Coxeter, Introduction to Geometry, John Wiley, New York (1961) p. 53.
4. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination (English translation of Anschauliche Geometrie). Chelsea, New York (1956) p. 44.
5. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford, London, 3rd ed. (1954) p. 26.
6. H. Steinhaus, Mathematical Snapshots, Oxford, New York, 2nd ed. (1950) p. 172.
7. H. M. Cundy and A. P. Rollett, Mathematical Models, Oxford, London, 2nd ed. (1961) p. 196.
8. A. F. Wells, Solid State Physics, **7,** Academic Press, New York (1958) pp. 425–503.
9. L. Grafton and H. J. Fraser, J. Geol., **43** (1935) pp. 785–909.

VECTOR SPACE AXIOMS FOR GEOMETRY

CURTIS M. FULTON, University of California at Davis

We want to show that the axioms and definitions of the theory of real vector spaces, when slightly modified, suffice to establish plane hyperbolic, elliptic, and Euclidean geometry.

Notations used in [1] will apply throughout. In this paper the elements of the vector space will be called straight lines or simply lines and we agree to let x

and αx represent the same line. Also, we restrict ourselves to three-dimensional vector spaces. In the definition of inner product [1, p. 121] we effect one important change: positiveness is not included. We postulate, however, that there exist elements x, called proper lines, for which $(x, x)>0$.

The three geometries arise quite naturally from the trichotomy which we find at the start. Let x_1, x_2, x_3 form a basis. If y_i, $i=1, 2, 3$ is a triple of lines, we have [1, p. 10], $y_i=\sum_{r=1}^{3}\alpha_{ir}x_r$. For the inner products we obtain $(y_i, y_j) = \sum_{r,s}\alpha_{ir}\alpha_{js}(x_r, x_s)$. Hence, using obvious notation for determinants, $|(y_i, y_j)| = |\alpha|^2\cdot|(x_r, x_s)|$. As is easily seen, the y's are dependent if and only if $|\alpha|=0$. If the y's are independent, the determinant $|(y_i, y_j)|$ has the same sign as $|(x_r, x_s)|$ and is zero if the latter determinant vanishes. Designating by Δ the determinant pertaining to such a basis, we distinguish between the following cases:

$\Delta > 0$ elliptic geometry

$\Delta = 0$ Euclidean geometry

$\Delta < 0$ hyperbolic geometry.

Using the simpler notation x, y, z for a basis, we write the determinant

$$\Delta = \begin{vmatrix} (x, x) & (x, y) & (x, z) \\ (y, x) & (y, y) & (y, z) \\ (z, x) & (z, y) & (z, z) \end{vmatrix}. \tag{1}$$

Employing a convenient notation for the cofactors of this determinant, we note that

$$\Delta(z, z) = -\Delta_{12}^{2} + \Delta_{11}\Delta_{22}. \tag{2}$$

Other identities, to be used later, are

$$\Delta_{11}\Delta_{22}\Delta_{33} - \Delta_{12}\Delta_{23}\Delta_{31} = \Delta[(x, x)(y, y)(z, z) - (x, y)(y, z)(z, x)], \tag{3}$$

$$\Delta = \Delta_{11}(x, x) + \Delta_{22}(y, y) - \Delta_{33}(z, z) + 2\Delta_{12}(x, y). \tag{4}$$

In what follows, all lines being introduced are assumed to be proper lines unless the contrary is explicitly stated. If the cofactor Δ_{33} in (1) is positive, we say that the lines x, y intersect and define a point. The existence of intersecting lines is postulated. In the case of intersecting lines a linear combination, say $u=\alpha x+\beta y$, $\alpha\beta\neq 0$, is easily shown to be a proper line which intersects x and y. We agree to consider three such lines as concurrent. The lines of the pencil are divided into two classes or sets according as $\alpha\beta>0$ or $\alpha\beta<0$. Now the absolute value of $(x, y)/\sqrt{(x, x)(y, y)}$ is less than unity and is not changed when either element is multiplied by a nonzero scalar. It seems natural then to let the expression under discussion represent the cosine of one of the "angles" determined by the sign of $\alpha\beta$. This in itself is not satisfactory unless the measure of the angle, denoted by xy, has the property that $xy=ux+uy$.

To fix our ideas let us make $(x, x)=(y, y)=1$. Since the line u is completely determined by the ratio $\alpha:\beta$, we may also reduce (u, u) to unity. We can make $\alpha>0$, $\beta>0$ for one of the angles and define its measure by $\cos xy=(x, y)$. If now u is a line pertaining to this angle, we can show that u and x determine two sets of lines, one of which is a subset of the set defined by $\alpha>0$, $\beta>0$. In fact, consider $\gamma u+\delta x=(\gamma\alpha+\delta)x+\gamma\beta y$, where the factors of x and y are positive when we make $\gamma>0$ and $\delta>0$. But then, using the above definition, $\cos ux=(u, x)$. We finally justify our definition showing by a simple manipulation that $\cos xy =\cos ux \cos uy-\sin ux \sin uy$. Clearly, the cosine of the other angle, $\alpha\beta<0$, would be defined as $-(x, y)$.

To discuss a basic difference between our geometries we consider a basis x, y, z. We assume that these lines intersect in pairs. Letting $u=\alpha x-\beta y$, $\alpha\beta\neq 0$, we find

$$(u, u)(z, z) - (u, z)^2 = \alpha^2\Delta_{22} + 2\alpha\beta\Delta_{12} + \beta^2\Delta_{11}. \tag{5}$$

Because of (2) this expression is positive for all α, β if $\Delta>0$. In other words, in elliptic geometry all lines u intersect z. In Euclidean geometry there is one ratio $\alpha:\beta$ for which (5) is zero corresponding to the "parallel" line. Eventually, in the hyperbolic case, (5) can also become negative for lines that are neither intersecting nor parallel. Now on account of (2), $\Delta_{12}\neq 0$ in Euclidean and hyperbolic geometry. Hence in these geometries the lines u for which $\alpha\beta\Delta_{12}>0$ all intersect z. On the other hand, in the set characterized by the opposite sign of $\alpha\beta$ there are lines that do not intersect z. The lines u of the former set give rise to a set of points of intersection on z. By definition, these points form the "segment" whose endpoints are the intersections of x, z and y, z. If, for example, $\Delta_{12}>0$, then the lines $u=\alpha x-\beta y$ for which $\alpha\beta>0$ belong to the angle "opposite" the segment. Denoting the angle measure by xy and making $(x, x)=(y, y)=1$, we have $\cos xy=-(x, y)$. In elliptic geometry there are, of course, two segments determined by a pair of points.

Next, we want to define the "length" of a segment. For this purpose we consider two unit lines through its endpoints, perpendicular to z, namely

$$x' = \Delta_{22}^{-1/2}[x - (x, z)z], \qquad y' = \Delta_{11}^{-1/2}[(y, z)z - y], \tag{6}$$

where we assume that $(x, x)=(y, y)=(z, z)=1$. It is found that $(x', z)=(y', z) =0$, $(x', x')=(y', y')=1$, and

$$(x', y') = \Delta_{12}/\sqrt{\Delta_{11}\Delta_{22}}. \tag{7}$$

This expression can reasonably be expected to define "distance." For simplicity of argument let us single out hyperbolic geometry. Using, if necessary, $-x$ instead of x we make $(x', y')>1$ in accordance with (2). The lines of the pencil $\alpha'x'+\beta'y'$ may be written in the form $\alpha x-\beta y+\gamma z$, where $\alpha'\beta'=\alpha\beta(\Delta_{11}\Delta_{22})^{1/2}$. Geometrically speaking, a line $\alpha'x'+\beta'y'$ and the corresponding line $\alpha x-\beta y$ intersect on z, as long as the former is a proper line. If $\alpha'\beta'>0$, then $\alpha\beta>0$ and we are dealing with lines passing through points of the segment. Thus we define (7) as representing the hyperbolic cosine of the distance between x, z and y, z if $\Delta_{12}>0$.

We are now ready to prove the law of cosines for angles. Let x, y, z be three independent lines intersecting in pairs. In the identity (3) the factor of Δ is positive since $\sqrt{(x, x)(y, y)}>(x, y)$. Hence in hyperbolic geometry as well as in

(Continued on page 311)

as interpretations of $\sin \theta_s$ and $\cos \theta_s$ as functions of angle measure θ_s. For example, for $s=\pi$ (radian measure), we make the identifications

$$\sin \theta_\pi = \sin x \quad \text{and} \quad \cos \theta_\pi = \cos x,$$

and for $s=180$ (degree measure)

$$\sin \theta_{180} = \sin\left(\frac{\pi}{180}\, x\right) \quad \text{and} \quad \cos \theta_{180} = \cos\left(\frac{\pi}{180}\, x\right)$$

for all values of x. Consequently, when we define the sine and cosine functions of angle measure θ_s in trigonometry, these functions are unique for each unit of angle measure but each unit gives rise to a different pair of functions of arc length which we have shown can be identified with one of the pairs of numerical functions (subsets of $R \times R$) in the family of pairs $\sin (kx)$ and $\cos (kx)$ determined axiomatically by properties P_1, P_2, and P_3'. Thus, radian measure is not the only unit of angle measure which gives rise to numerical functions. However, the radian is the only unit of angle measure for which the measure of the angle and all the trigonometric functions of that measure can be represented by lengths employing the same unit of length, the radius of the unit circle. Therefore, radian measure of angles takes on a fundamental role in classical mathematics involving the trigonometric functions.

With these identifications, we can interpret the arithmetic operations on the trigonometric functions and their variables. For example, we have

$$\theta_s + \sin \theta_s = T + \sin\left(\frac{\pi}{s}\, T\right) = x + \sin\left(\frac{\pi}{s}\, x\right)$$

for all values of x for each value of s. Therefore, in particular

$$1^\circ + \sin 1^\circ = 1 + \sin\left(\frac{\pi}{180}\right) \doteq 1.0175, \text{ degree units,}$$

and $1_r + \sin 1_r = 1 + \sin 1 \doteq 1.8415$, radian units.

References

1. J. D. Mancill, *Modern Analytical Trigonometry*, New York: Dodd, Mead, 1960.

(Continued from page 301)

Euclidean geometry, $\Delta_{12}\Delta_{23}\Delta_{31} > 0$. Without loss of generality we can make $\Delta_{12} > 0$, $\Delta_{23} > 0$, $\Delta_{31} > 0$. Then (7) is the hyperbolic cosine of the "side" of the triangle whose endpoints are x, z and y, z. If we again assume that $(x, x) = (y, y) = (z, z) = 1$, we have for the interior angles opposite the sides $\Delta_{12} = (x, z)(y, z) - (x, y) = \cos xz \cos yz + \cos xy$ and $(\Delta_{11}\Delta_{22})^{1/2} = \sin xz \sin yz$. This shows the validity of the law of cosines for hyperbolic geometry.

In elliptic geometry the lines (6) intersect on account of (7) and (2). Thi- makes the definition of distance quite obvious. Also, the existence of four tris angles can be ascertained without difficulty. We simply consider the concurrent

(Continued on page 318)

phism and isomorphism which are fundamental to contemporary mathematics.

Chapter VIII is devoted to the intrinsic study of permutation of sets, finite or infinite.

In Chapter IX, the introduction to the notion of operator groups widens considerably the scope of the theory of groups. Due to this notion, the theory of vector spaces and the theory of ideals are found directly attached to the main body of the theory of groups.

The last chapter is devoted to the concept of dimension theory of operator groups. Here we find the theories of Jordan-Hölder, Schreier and Zassenhaus. We know that this theory of group operators is essentially due to Emmy Noether and Krull.

There are ample exercises in each chapter, and abundant review exercises at the end of each chapter to test the understanding of the reader.

A new feature of the book is the colored plates, illustrating abstract theories in terms of colored figures. These figures help the reader from a pedagogic point of view in understanding more deeply the theories. They suggest methods of instruction of the fundamental concepts of the theory of groups.

This is an excellent book on the elementary theory of groups and is recommended highly for those who begin the study of groups for the first time.

SOUREN BABIKIAN
Los Angeles City College

(Continued from page 311)

lines $\alpha x-\beta y$, $\beta y-\gamma z$, $\gamma z-\alpha x$, $\alpha\beta\gamma\neq 0$. Each such line determines a side and an angle of a triangle. There are four possible combinations of signs of the ratios $\alpha:\beta$, $\beta:\gamma$, $\gamma:\alpha$.

The Euclidean case requires different treatment since $(x', y')=1$ if again $\Delta_{12}>0$. But then $(x'-y', x'-y')=0$ and $x'-y'$ is not a proper line. It is expedient to show that there is just one such "ideal" line i. Since our dimension number is three, we can set $i=\alpha x'+\beta y'+\gamma z$ and $(i, i)=0$ implies $\alpha+\beta=0$, $\gamma=0$. We normalize i by taking $\alpha=(\Delta_{11}\Delta_{22})^{1/2}$. Hence

$$i = (\Delta_{11}\Delta_{22})^{1/2}(x' - y'). \tag{8}$$

If we further assume that $\Delta_{23}>0$, $\Delta_{31}>0$, it is easily shown that we may write the symmetric expression

$$i = \Delta_{11}^{1/2} x + \Delta_{22}^{1/2} y + \Delta_{33}^{1/2} z. \tag{9}$$

We now define the distance between x, z and y, z as $(\Delta_{11}\Delta_{22})^{-1/2}$, which is the scalar factor of i in the expression (8) for $x'-y'$. Similar definitions apply to the other sides of the triangle due to the symmetry of (9). Our normalization process indicates that there cannot exist an absolute unit of length in Euclidean geometry. From the identity (4) the law of cosines is now immediately derived, taking into account that $(x, y)=-\cos xy$.

Reference

1. P. R. Halmos, Finite-dimensional vector spaces, 2d ed., D. Van Nostrand, 1958.

TEACHING OF MATHEMATICS

EDITED BY ROTHWELL STEPHENS, Knox College

This department is devoted to the teaching of mathematics. Thus, articles of methodology, exposition, curriculum, tests and measurements, and any other topic related to teaching, are invited. Papers on any subject in which you, as a teacher, are interested, or questions which you would like others to discuss, should be sent to Rothwell Stephens, Mathematics Department, Knox College, Galesburg, Illinois.

THE SINE AND COSINE FUNCTIONS

JULIAN D. MANCILL, University of Alabama

1. Introduction. Traditionally, the trigonometric functions have been characterized, essentially, as functions of angle measure in Euclidean geometry. This is the natural and historical development of the trigonometric functions relating to their application to triangle measurement and perhaps this should be the method used to introduce trigonometry. However, many recent writers have attempted to free the trigonometric functions from the concept of angle measure by defining them as functions of the set of real numbers representing the arc length along the unit circle measured from a fixed point on the circle. Such a procedure merely shifts the emphasis from angle measure in radians to the measure of arc length along the unit circle, both measures being represented by the same set of real numbers. However, it is desirable for many of the applied sciences, as well as for "pure" analysis, to have the trigonometric functions defined as functions of a real variable, free of any prescribed metric. This we propose to do in a very elementary way.

We shall employ a postulational development of these functions by assuming the existence (for the moment) of a pair of real valued functions $\sin x$ and $\cos x$ with the following three properties:

$$P_1:\quad \sin(x-y) \equiv \sin x \cos y - \cos x \sin y,$$

$$P_2:\quad \cos(x-y) \equiv \cos x \cos y + \sin x \sin y,$$

$$P_3:\quad \lim_{x\to 0^+} \frac{\sin x}{x} = 1,$$

where x and y are real variables with domains all real numbers. It will be shown that these postulates define a unique pair of functions $\sin x$ and $\cos x$. Then it is obvious that the one-parameter family of pairs of functions, $\sin(kx)$ and $\cos(kx)$, have the properties P_1 and P_2, and

$$P_3':\quad \lim_{x\to 0^+} \frac{\sin(kx)}{x} = k \lim_{x\to 0^+} \frac{\sin(kx)}{kx} = k > 0,$$

for each real positive value of k. Thus, we shall show the existence of a one-parameter family of pairs of sine and cosine functions with domains all real numbers which have all the properties of the classical sine and cosine functions of angle measure in trigonometry. Then we shall identify $\sin(kx)$ and $\cos(kx)$, for each value of k, with their circular function counterparts in trigonometry, where the value of k is determined by the unit of angle (or circular arc) measure employed in defining the circular functions.

2. Some Derived Properties. We shall derive some of the familiar properties of $\sin x$ and $\cos x$ from postulates P_1, P_2, and P_3 without further definitions. Then it follows easily from P_1, P_2, and P_3' that the functions $\sin(kx)$ and $\cos(kx)$ have corresponding properties for each value of k.

(I) $$\sin 0 = 0.$$

This property follows from P_1 for $x=y$.

(II) $$\cos 0 = 1.$$

If we set $y=0$ in P_1, we have

$$\sin x = \sin x \cos 0,$$

or

$$\sin x(1 - \cos 0) = 0$$

for all values of x. However, it follows from P_3 that there are values of x for which $\sin x \neq 0$. Therefore, $1-\cos 0$ must be zero and

$$\cos 0 = 1.$$

(III) $$\sin^2 x + \cos^2 x \equiv 1.$$

(IV) $$\sin(-y) \equiv -\sin y.$$

If we set $x=0$ in P_1, we obtain

$$\sin(-y) \equiv \sin 0 \cos y - \cos 0 \sin y = -\sin y.$$

In the same way we can prove that

(V) $$\cos(-y) \equiv \cos y.$$

(VI) $$\sin(x+y) \equiv \sin x \cos y + \cos x \sin y.$$

This property follows from

$$\sin(x+y) \equiv \sin[x-(-y)]$$

and the preceding properties. In a similar way we can prove

(VII) $$\cos(x+y) \equiv \cos x \cos y - \sin x \sin y.$$

Now, all the other well-known identities of trigonometry can be easily derived for $\sin x$ and $\cos x$ in terms of the basic ones derived here.

(VIII) $$\lim_{x\to 0} \frac{\sin x}{x} = 1 \quad \text{and} \quad \lim_{x\to 0} \frac{\sin(kx)}{x} = k.$$

The first of these limits follows from P_3 and (IV), and the second from P_3' and (IV).

(IX) Sin x *is continuous at* $x = 0$.

From the definition of limit and (VIII), we note that

$$\frac{\sin x}{x} = 1 + \epsilon,$$

where $\epsilon \to 0$ as $x \to 0$, from which it follows that

$$\sin x = x(1 + \epsilon).$$

Therefore, we have

$$\lim_{x \to 0} \sin x = 0 = \sin 0,$$

and the property is proved.

(X) Sin x *and* cos x *are continuous everywhere.*

To prove this we shall make use of the relation

$$\left| \sin x_1 - \sin x_2 \right| \equiv 2 \left| \sin \frac{x_1 - x_2}{2} \right| \cdot \left| \cos \frac{x_1 - x_2}{2} \right|.$$

Since

$$\left| \cos x \right| \leqq 1,$$

we have

$$\left| \sin x_1 - \sin x_2 \right| \leqq M \left| \sin \frac{x_1 - x_2}{2} \right|,$$

where M is a positive constant, which proves our property in view of (IX). The continuity of cos x follows from a similar argument.

(XI) Sin x and cos x *are differentiable* with

$$\frac{d}{dx}(\sin x) = \cos x \quad \textit{and} \quad \frac{d}{dx}(\cos x) = -\sin x.$$

We have

$$\frac{d}{dx}(\sin x) = \lim_{h \to 0} \frac{\sin(x + h) - \sin x}{h} = \lim_{h \to 0} \frac{2 \sin \frac{h}{2} \cos \left(x + \frac{h}{2} \right)}{h}$$

$$= \lim_{h \to 0} \frac{\sin \frac{h}{2}}{\frac{h}{2}} \cdot \lim_{h \to 0} \cos \left(x + \frac{h}{2} \right)$$

$$= \cos x,$$

as follows from P_3 and the continuity of cos x. Proof of the second part can be made in a similar way.

We can now prove the uniqueness property

(XII) *The functions* $\sin x$ *and* $\cos x$ *satisfying the three postulates* P_1, P_2, *and* P_3 *are unique.*

To prove this property we suppose there are two pairs of functions $\sin_1 x$, $\cos_1 x$ and $\sin_2 x$, $\cos_2 x$ satisfying our postulates for all values of x. Now we define the function

$$f(x) \equiv (\sin_1 x - \sin_2 x)^2 + (\cos_1 x - \cos_2 x)^2,$$

which is differentiable for all values of x, and

$$\begin{aligned} f'(x) &= 2(\sin_1 x - \sin_2 x)(\cos_1 x - \cos_2 x) + 2(\cos_1 x - \cos_2 x)(-\sin_1 x + \sin {}_2x) \\ &\equiv 0. \end{aligned}$$

Therefore, $f(x)$ is a constant and if we evaluate it at $x=0$, we obtain

$$f(0) = f(x) \equiv 0.$$

Thus, we have

$$\sin_1 x \equiv \sin_2 x \quad \text{and} \quad \cos_1 x \equiv \cos_2 x.$$

Up to this point we have not faced the problem of showing there exists a pair of functions $\sin x$ and $\cos x$ satisfying our postulates. To accomplish this we shall show that the functions

$$S(x) = x - \frac{x^3}{3!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \cdots$$

and

$$C(x) = 1 - \frac{x^2}{2!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots,$$

satisfy our postulates and, therefore, must be the functions we are seeking, since there is only one pair by the uniqueness property. That these infinite series define functions of x for all real values of x follows from well known properties of such series.

Postulate P_3 is satisfied by $S(x)$ because

$$\lim_{x\to 0} \frac{S(x)}{x} = \lim_{x\to 0} \left(1 - \frac{x^2}{3!} + \cdots\right) = 1.$$

We could show that postulates P_1 and P_2 are satisfied by this pair of functions by direct substitution and operations on infinite series but we shall follow a method similar to that in the proof of (XIII). First we note that

$$\frac{dS(x)}{dx} = 1 - \frac{x^2}{2!} + \cdots = C(x)$$

and

$$\frac{dC(x)}{dx} = -x + \frac{x^3}{3!} - \cdots = -S(x)$$

from known properties of series.

Now we define the function

$$g(x) \equiv \{S(x-y) - [S(x)C(y) - C(x)S(y)]\}^2 + \{C(x-y) - [C(x)C(y) + S(x)S(y)]\}^2$$
$$\equiv g_1^2(x) + g_2^2(x),$$

where we consider y to be a parameter. Differentiating $g(x)$ with respect to x, we have

$$\begin{aligned} g'(x) &= 2g_1(x)\{C(x-y) - [C(x)C(y) + S(x)S(y)]\} \\ &\quad + 2g_2(x)\{-S(x-y) - [-S(x)C(y) + C(x)S(y)]\} \\ &= 2g_1(x)g_2(x) + 2g_2(x)[-g_1(x)] \equiv 0, \end{aligned}$$

and it follows that $g(x)$ is a constant. Evaluating it at $x=0$, we find

$$g(0) = g(x) \equiv 0,$$

from which it follows that $g_1(x) \equiv g_2(x) \equiv 0$ and postulates P_1 and P_2 are satisfied.

Thus, we have exhibited the unique functions

$$\sin x \equiv S(x) \quad \text{and} \quad \cos x \equiv C(x),$$

which satisfy our postulates.

To complete our formulation of the properties of the functions $\sin x$ and $\cos x$ from our postulates we need to derive their periodic and other numerical properties. We shall not accomplish this in its entirety here because we shall identify these functions with the circular sine and cosine functions of radian measure of angles developed in trigonometry and their numerical properties will follow because of the uniqueness property. However, we shall derive their periodic property.

First, we shall locate the first positive zero of $\cos x$. To do this we shall make use of the fact that the terms in the series expression for $\sin x$ and $\cos x$ may be rearranged in any manner without affecting their properties. Thus, we see that

$$\cos x = \left(1 - \frac{x^2}{2!}\right) + \frac{x^4}{4!}\left(1 - \frac{x^2}{30}\right) + \cdots$$

in which for each value of x the terms are all positive after a certain point on. Therefore, we have

$$\cos x > 1 - \frac{x^2}{2!}, \qquad 0 < x^2 < 30.$$

For example, these conditions are fulfilled if $x = \sqrt{2}$ and, therefore,

$$\cos \sqrt{2} > 0.$$

Also, there is no zero of $\cos x$ between 0 and $\sqrt{2}$.

Next let us regroup the terms of the series as follows

$$\cos x = 1 - \left(\frac{x^2}{2!} - \frac{x^4}{4!}\right) - \left(\frac{x^6}{6!} - \frac{x^8}{8!}\right) - \cdots,$$

from which it follows that

$$\cos x < 1 - \frac{x^2}{2!}\left(1 - \frac{x^2}{12}\right), \qquad 0 < x^2 < 56,$$

and in particular

$$\cos \sqrt{3} < -1/8 < 0.$$

Thus, we have shown that $\cos x$ has at least one zero on the interval $\sqrt{2} < x < \sqrt{3}$ because it is continuous and changes sign on this interval. Now we wish to show that there is just one zero on this interval.

First, we see that $\sin x > 0$ for $\sqrt{2} \leqq x \leqq \sqrt{3}$ from

$$\begin{aligned} \sin x &= x\left(1 - \frac{x^2}{3!}\right) + \frac{x^5}{5!}\left(1 - \frac{x^2}{42}\right) + \cdots \\ &> x\left(1 - \frac{x^2}{6}\right), \qquad 0 < x^2 < 42. \end{aligned}$$

Therefore,

$$\frac{d \cos x}{dx} = -\sin x < 0, \qquad \sqrt{2} < x < \sqrt{3},$$

and there is exactly one zero of $\cos x$ on this interval which we shall momentarily denote by z. It follows that

$$\cos z = 0 \quad \text{and} \quad \sin z = 1.$$

Now, from (VI) and (VII), we can write

$$\begin{aligned} \sin(x + z) &= \sin x \cos z + \cos x \sin z = \cos x, \\ \sin(x + 2z) &= \sin[(x + z) + z] = \cos(x + z) = \cos x \cos z - \sin x \sin z \\ &= -\sin x, \\ \sin(x + 3z) &= \sin[(x + z) + 2z] = -\sin(x + z) = -\cos x, \\ \sin(x + 4z) &= \sin[(x + 2z) + 2z] = -\sin(x + 2z) = \sin x, \end{aligned}$$

for all values of x. In a similar way we can show that

$$\cos(x + 4z) = \cos x.$$

From these relations we see that

$$\sin 2z = 0, \qquad \cos 2z = -1, \qquad \sin 3z = -1, \qquad \cos 3z = 0.$$

It has just been shown that $\sin x$ and $\cos x$ are periodic with a period $4z$. Since these functions are continuous, they have a smallest positive period, and it follows from $\sin(x+z)=\cos x$ that they have the same smallest period, say p. We shall now show that $p=4z$. Since

$$\sin(x+p) = \sin x$$
$$\cos(x+p) = \cos x$$

for all values of x, we have $\sin p=0$ and $\cos p=1$, for $x=0$. Therefore,

$$\cos p = \cos^2\frac{p}{2} - \sin^2\frac{p}{2} = 1 \quad \text{and} \quad \cos^2\frac{p}{2} + \sin^2\frac{p}{2} = 1,$$

from which we obtain

$$\cos^2\frac{p}{2} = 1 \quad \text{and} \quad \sin\frac{p}{2} = 0.$$

But $\cos p/2$ cannot be $+1$ with $\sin p/2=0$, for then $p/2$ would be a period and p would not be the minimum period. Consequently, we have

$$\sin\frac{p}{2} = 0 \quad \text{and} \quad \cos\frac{p}{2} = -1.$$

Now, with these properties, we have

$$\cos\frac{p}{2} = \cos^2\frac{p}{4} - \sin^2\frac{p}{4} = -1$$
$$\cos^2\frac{p}{4} + \sin^2\frac{p}{4} = 1,$$

or

$$\cos\frac{p}{4} = 0.$$

Therefore, we have shown that $p/4=z$ since z is the first positive zero of $\cos x$.

We have shown that the functions $\sin x$ and $\cos x$ oscillate periodically between $+1$ and -1, and it is easily seen that between these extreme values they are monotone, because the sign of the derivative in each case is constant between these points.

The value of z may be approximated more precisely than $\sqrt{2}<z<\sqrt{3}$ by several means. For example, if $y=\sin x$, $-z\leqq x\leqq z$, it follows from properties (III) and (XI) that

$$y'^2 + y^2 = 1,$$

and we define

$$x = \int_0^y \frac{dt}{\sqrt{1-t^2}} \equiv \operatorname{Arc}\sin y, \quad -1 < y < 1.$$

But it is easily shown that

$$\sin\frac{z}{2} = \cos\frac{z}{2} = \frac{1}{\sqrt{2}}\cdot$$

Consequently, we see that

$$\frac{z}{2} = \int_0^{1/\sqrt{2}} \frac{dt}{\sqrt{1-t^2}} = \text{Arc}\sin\frac{1}{\sqrt{2}},$$

from which $z/2$ may be approximated directly, resulting in

$$2z = 3.14159\cdots.$$

In the next section we shall show that $2z=\pi$, where π is the constant ratio of the circumference of a circle to its diameter in Euclidean geometry. Therefore, we could have taken $2z$ as the definition of π in the present context.

3. Geometric Interpretation. In order to identify the functions $\sin(kx)$ and $\cos(kx)$ with their circular function counterparts in trigonometry, we shall construct an interpretation in Euclidean geometry. To this end let us consider the equation

$$C:\quad x^2+y^2 = 1/k^2, \qquad k>0$$

of the circle with center at the origin of the coordinate system and radius $1/k$. We shall make use of the following parametric representation of this circle:

$$C:\quad x = \frac{1}{k}\cos(kt), \qquad y = \frac{1}{k}\sin(kt),$$

for each positive value of k, where $\sin(kt)$ and $\cos(kt)$ are the functions defined by P_1, P_2, and P_3' of Section 1. Now we note the interesting property

(XIII) *The length of arc on the circle C with the radius $1/k$ from the point $\left(\frac{1}{k},\ 0\right)$ to the point $\left[\frac{1}{k}\cos(kT), \frac{1}{k}\sin(kT)\right]$ is T.*

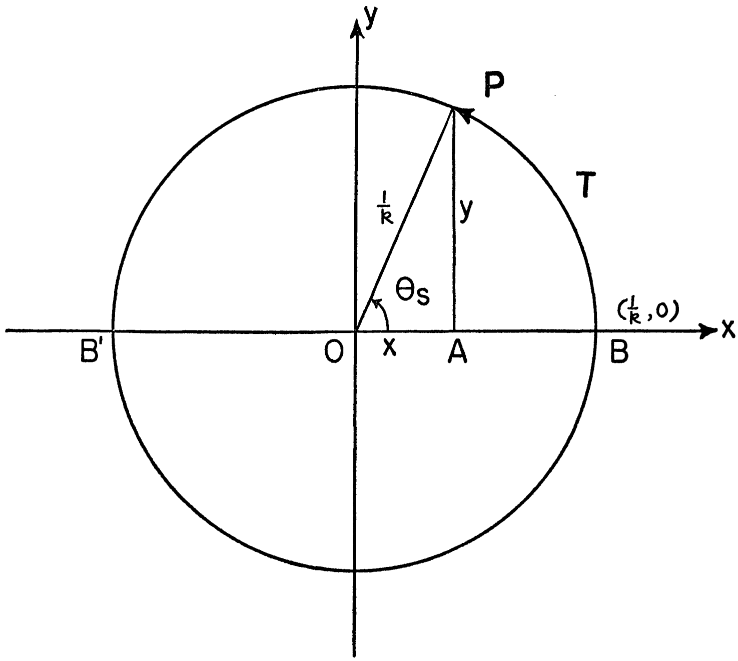
This follows easily from the formula for arc length in calculus or it may be shown directly from the definition of arc length and the properties of the functions $\sin(kt)$ and $\cos(kt)$. Therefore, if on the circle C a point P is chosen with abscissa OA equal to $(1/k)\cos(kT)$, then the ordinate AP is equal to $(1/k)\sin(kT)$ and the length of the circular arc BP is T. Consequently, if we let s denote the measure of the straight angle BOB' with respect to some arbitrary unit of angle measure (for an elementary discussion of angle measure, see [1], Chapter 1), then from well-known properties of Euclidean geometry

$$\frac{\theta_s}{T} = \frac{s}{\pi/k} \quad\text{or}\quad \theta_s = \frac{ks}{\pi}T = \frac{s}{\pi}\frac{T}{r}, \qquad r = 1/k,$$

where θ_s denotes the measure (in s units) of the angle at the center of C subtended by the arc BP. Now if we choose

$$k = \pi/s,$$

then $\theta_s = T$ and the measure of the arc T equals that of the subtended angle for the circle with radius $r=1/k=s/\pi$; that is, *the number of linear units* (π/s*-th* part of the radius) *in the length of the arc BP equals the number of units of angle measure* (s units) *of the subtended angle for the circle with radius* $r=s/\pi$.



It follows, therefore, that we may take the length of arc T on this circle as the measure of the subtended angle in s units of angle measure. Thus, we have exhibited a measure function for angles in s units, simply the arc length function on the circle of radius $r=s/\pi$, and our three properties P_1, P_2, and P_3' insure the properties

$$\sin\left(\frac{\pi}{s}\,T\right) = \sin\theta_s = \frac{\pi}{s}\,y, \qquad \cos\left(\frac{\pi}{s}\,T\right) = \cos\theta_s = \frac{\pi}{s}\,x.$$

At the same time we have an identification of the functions $\sin(kx)$ and $\cos(kx)$ with the circular functions $\sin\theta_s$ and $\cos\theta_s$ for each value of $k=\pi/s$, all values of s, simply by interpreting the real variable x as the arc length T on the circle with radius $r=1/k=s/\pi$. Therefore, we have all the numerical properties of $\sin(kx)$ and $\cos(kx)$ for each value of k, including their periodicity, from their circular function counterparts as follows from the uniqueness property. In this connection, we note the correspondence between the completion of the periods of $\sin x$ and $\cos x$ and the fact that at the same time the point P describes the circumference of the unit circle ($k=1$). Therefore, we conclude that $p=4z=2\pi$ or $z=\pi/2$.

Conversely, we may consider the identifications

$$\sin\theta_s = \sin\left(\frac{\pi}{s}\,x\right) \quad \text{and} \quad \cos\theta_s = \cos\left(\frac{\pi}{s}\,x\right)$$

as interpretations of $\sin\theta_s$ and $\cos\theta_s$ as functions of angle measure θ_s. For example, for $s=\pi$ (radian measure), we make the identifications

$$\sin\theta_\pi = \sin x \quad\text{and}\quad \cos\theta_\pi = \cos x,$$

and for $s=180$ (degree measure)

$$\sin\theta_{180} = \sin\left(\frac{\pi}{180}x\right) \quad\text{and}\quad \cos\theta_{180} = \cos\left(\frac{\pi}{180}x\right)$$

for all values of x. Consequently, when we define the sine and cosine functions of angle measure θ_s in trigonometry, these functions are unique for each unit of angle measure but each unit gives rise to a different pair of functions of arc length which we have shown can be identified with one of the pairs of numerical functions (subsets of $R\times R$) in the family of pairs $\sin(kx)$ and $\cos(kx)$ determined axiomatically by properties P_1, P_2, and P_3'. Thus, radian measure is not the only unit of angle measure which gives rise to numerical functions. However, the radian is the only unit of angle measure for which the measure of the angle and all the trigonometric functions of that measure can be represented by lengths employing the same unit of length, the radius of the unit circle. Therefore, radian measure of angles takes on a fundamental role in classical mathematics involving the trigonometric functions.

With these identifications, we can interpret the arithmetic operations on the trigonometric functions and their variables. For example, we have

$$\theta_s + \sin\theta_s = T + \sin\left(\frac{\pi}{s}T\right) = x + \sin\left(\frac{\pi}{s}x\right)$$

for all values of x for each value of s. Therefore, in particular

$$1^\circ + \sin 1^\circ = 1 + \sin\left(\frac{\pi}{180}\right) \doteq 1.0175, \text{ degree units,}$$

and $1_r+\sin 1_r = 1+\sin 1 \doteq 1.8415$, radian units.

References

1. J. D. Mancill, *Modern Analytical Trigonometry*, New York: Dodd, Mead, 1960.

(Continued from page 301)

Euclidean geometry, $\Delta_{12}\Delta_{23}\Delta_{31}>0$. Without loss of generality we can make $\Delta_{12}>0$, $\Delta_{23}>0$, $\Delta_{31}>0$. Then (7) is the hyperbolic cosine of the "side" of the triangle whose endpoints are x, z and y, z. If we again assume that $(x, x)=(y, y)=(z, z)=1$, we have for the interior angles opposite the sides $\Delta_{12}=(x, z)(y, z)-(x, y)=\cos xz \cos yz+\cos xy$ and $(\Delta_{11}\Delta_{22})^{1/2}=\sin xz \sin yz$. This shows the validity of the law of cosines for hyperbolic geometry.

In elliptic geometry the lines (6) intersect on account of (7) and (2). Thi- makes the definition of distance quite obvious. Also, the existence of four tris angles can be ascertained without difficulty. We simply consider the concurrent

(Continued on page 318)

COMMENTS ON PAPERS AND BOOKS

EDITED BY HOLBROOK M. MACNEILLE, Case Institute of Technology

This department will present comments on papers published in the MATHEMATICS MAGAZINE, lists of new books, and reviews.

In order that errors may be corrected, results extended, and interesting aspects further illuminated, comments on published papers in all departments are invited.

Communications intended for this department should be sent to Holbrook M. MacNeille, Department of Mathematics, Case Institute of Technology, Cleveland 6, Ohio

THE INTERSECTION POINTS OF PERIMETER BISECTORS

SISTER M. STEPHANIE SLOYAN, Georgian Court College, Lakewood, New Jersey

In the September-October 1962 issue of the MATHEMATICS MAGAZINE (p. 251) Charles Pinzka has proved that the three perimeter bisecting lines, each of which passes through a corresponding midpoint of a side of a triangle, meet in a point S. If we consider the triangle ABC as inscribed in a unit circle, and change to complex coordinates ($A=t_1$, $B=t_2$, $C=t_3$), the line through vertex A and the incenter I is:

$$z(1-t_1\bar{I})-\bar{z}t_1(t_1-I)+t_1^2\bar{I}-I=0.$$

Pinzka has shown that a perimeter bisecting line is parallel to this and passes through the midpoint of the side opposite vertex A. Such a line is

$$z(1-t_1\bar{I})-\bar{z}t_1(t_1-I)=\frac{\sigma_1\sigma_3-\sigma_1\sigma_3t_1\bar{I}+t_1^2\bar{I}\sigma_3-t_1^2\sigma_2+t_1I\sigma_2-\sigma_3I}{2\sigma_3}.$$

Three such lines intersect in S which may be expressed as

$$S=\frac{\sigma_1-I}{2}$$

where σ_1 is the coordinate of the vector to the orthocenter. Therefore S can be found for a triangle by bisecting the segment between the orthocenter and the symmetric of the incenter with respect to the center of the circumscribed circle.

Since the determinant

$$\begin{vmatrix} I & \bar{I} & 1 \\ \dfrac{\sigma_1-I}{2} & \dfrac{\sigma_2-\bar{I}\sigma_3}{2\sigma_3} & 1 \\ \dfrac{\sigma_1}{3} & \dfrac{\sigma_2}{3\sigma_3} & 1 \end{vmatrix}=0,$$

the centroid $G(\sigma_1/3)$ lies on line IS. Moreover, because $(I+2S)/3=G$, the centroid divides the segment from I to S in the ratio 2:1. Pinzka has proved that S is the midpoint of I and N (N the intersection of perimeter bisecting lines through the vertices), so N may be represented as

$$N=\sigma_1-2I.$$

Because $(2I+N)/3=G$, G divides the segment IN in the ratio 1:2. Since $IG/GS=2$ and $IN/NS=-2$, the segment IS is harmonically separated by G and N.

Now let us consider the poristic system of triangles circumscribed to a circle of center I and inscribed to a unit circle with center O. Triangle ABC is one of these triangles as is a triangle with vertices, say, $A'B'C'$. Ramler [1] has shown that the orthocenters of all such triangles lie on a circle with center $2I$, and radius equal to $OI \cdot OI/R$ where OI is the distance from the circumcenter to the in-center, and R is the radius of the circumscribing circle.

It can be shown that if, from a fixed point outside a circle P, secants are drawn through the circle, and the midpoints found between each intersection of the secants with the circle and the external point, the midpoints lie on a circle Q. The center of circle Q is a point midway between the center of circle P and the external point, and the radius of circle Q is equal to one-half the radius of circle P. For all triangles of the poristic system under discussion the point $-I$ is fixed, and the orthocenter σ_1 (or σ_1', σ_1'', etc.), lies on a circle of center $2I$. Since S is the midpoint between the orthocenter of any triangle and $-I$, S lies on a circle of center $I/2$ and radius equal to $\frac{1}{2}OI \cdot OI/R$ or radius equal to one-half the orthocentric radius.

The coordinates of the vector to N, σ_1-2I, can be written as $(2\sigma_1-4I)/2$. If σ_1, the orthocenter, lies on a circle of center $2I$ then $2\sigma_1$ lies on a circle with center $4I$ and radius twice the orthocentric radius. Then N will be the midpoint between each "$2\sigma_1$" and $-4I$. Therefore each N of a triangle of the poristic system lies on a circle with center O (the circumcenter) and radius equal to the orthocentric radius.

Reference

1. O. J. Ramler, A poristic system of triangles, Duke Mathematical Journal, **17** (1950), 446.

FURTHER RESULTS WITH N^2-N+A

J. A. H. Hunter, Toronto, Ontario, Canada

The intriguing article on this well-known polynomial, by S. Kravitz in the May 1962 issue, prompted me to continue further along the lines suggested by him. My findings may be of interest.

All of what follows has resulted from applications of a method based on elementary theory of quadratic congruences. Having no computer available, it was necessary to devise procedures that would not entail an undue amount of tedious calculation. Maybe a brief mention of this method will serve to show the principle on which the results were obtained.

In the case of n^2-n+17, say we seek the minimum value of n that will generate a product of four distinct factors, not necessarily different. In the notation used, then, we seek to evaluate $h_{17}(4)$.

Set $n^2-n+17=abcd$, $a\leqq b\leqq c\leqq d$.

The initial test was carried out on $abcd=17^3\cdot d$.

Then, by congruences, we find that $d=4913k^2\pm 611k+19$.

Here we obviously have the minimum value from $d=19$, so $f\{h_{17}(4)\} = f_{17}(306) = 17\cdot 17\cdot 17\cdot 19$, with $h_{17}(4) = 306$.

This was very simple, but it illustrates the principle.

Referring to Mr. Kravitz's article, I have established the value $h_{41}(5) = 14{,}145$, giving $f(n) = 41\cdot 47^4$. I have also confirmed $h_{41}(6) = 139{,}564$, giving $f(n) = 41^4\cdot 61\cdot 113$: this value was announced by Mr. Kravitz in the June 1961 issue of *Recreational Mathematics Magazine.*

The values of $h_A(m)$, for $A=3$, 5, 11, 17, 41, and for $m=3$, 4, 5, 6, are now established and/or confirmed as shown in the following tabulation. I also tabulate the corresponding prime factorizations of $f\{h_A(m)\}$.

A	3	5	11	17	41
$h_A(3)$	7	16	54	113	421
$h_A(4)$	12	30	132	306	1722
$h_A(5)$	16	146	561	2643	14145
$h_A(6)$	97	280	2794	19347	139564

A	3	5	11	17	41
$f\{h_A(3)\}$	$3^2\cdot 5$	$5\cdot 7^2$	$13^2\cdot 17$	$19\cdot 23\cdot 29$	$47\cdot 53\cdot 71$
$f\{h_A(4)\}$	$3^3\cdot 5$	$5^3\cdot 7$	$11^3\cdot 13$	$17^3\cdot 19$	$41^3\cdot 43$
$f\{h_A(5)\}$	3^5	$5^2\cdot 7\cdot 11^2$	$11\cdot 13^4$	$19^2\cdot 23\cdot 29^2$	$41\cdot 47^4$
$\{h_A(6)\}$	$3^4\cdot 5\cdot 23$	$*5^7$	$11^4\cdot 13\cdot 41$	$17^3\cdot 29\cdot 37\cdot 71$	$41^4\cdot 61\cdot 113$

* Oddly enough, $h_5(6) = h_5(7)$.

I now venture to define an even more interesting function which, for lack o a better name, I suggest as the Hunter function. $H(m)$ is defined as being the value, of n in n^2-n+A, which generates the smallest product of m distinct and *different* prime factors.

Establishing values of $H(m)$ depends on the same basic principles, but involves far more working. Without a computer I found it impracticable to evaluate for higher orders than $H(5)$. The values for $H(3)$, $H(4)$, and $H(5)$ are shown in the following tabulation:

A	3	5	11	17	41
$H_A(3)$	19	20	66	113	421
$H_A(4)$	72	86	275	664	2912
$H_A(5)$	787	531	1936	6443	38914

A	3	5	11	17	41
$f\{H_A(3)\}$	$3\cdot 5\cdot 23$	$5\cdot 7\cdot 11$	$11\cdot 17\cdot 23$	$19\cdot 23\cdot 29$	$47\cdot 53\cdot 71$
$f\{H_A(4)\}$	$3\cdot 5\cdot 11\cdot 31$	$5\cdot 7\cdot 11\cdot 19$	$11\cdot 13\cdot 17\cdot 31$	$17\cdot 19\cdot 29\cdot 47$	$41\cdot 47\cdot 53\cdot 83$
$f\{H_A(5)\}$	$3\cdot 5\cdot 11\cdot 23\cdot 163$	$5\cdot 7\cdot 11\cdot 17\cdot 43$	$11\cdot 13\cdot 17\cdot 23\cdot 67$	$19\cdot 23\cdot 29\cdot 37\cdot 47$	$43\cdot 47\cdot 61\cdot 71\cdot 173$

In conclusion I should like to acknowledge with gratitude the help I have had from Mr. Kravitz in checking some of my results.

A COMMENT ON "FINDING THE NTH ROOT OF A NUMBER BY ITERATION"

E. M. ROMER, Air Force Institute of Technology

This note serves to simplify the work done in reference (1) and to correlate this work with earlier sources in the literature. In reference (1), a recursion formula for determining the nth root of a number is obtained:

$$x = \frac{1}{n}\left[(n-1)x_0 + \frac{N}{x_0^{n-1}}\right], \tag{1}$$

where

N = number whose nth root is desired
x_0 = present estimate of required root
x = subsequent estimate of required root.

Equation (1) as well as equations obtained in references (2) and (3) are special cases of a generalized recursion sequence obtained in reference (4). It is shown in reference (4) that these equations are among the m useful sequences that can be obtained; where m is an integer which satisfies the inequality

$$2^{m-1} < n \leqq 2^m.$$

The conclusions regarding the magnitude of the error in x as a function of the error in x_0 can be obtained in a somewhat simpler fashion than is shown in reference (1) if the *fractional* error, i.e., the ratio of the absolute error to the desired root, is analyzed. One can define the fractional error of the present and subsequent estimates (α_0, and α, respectively) by equations (2) and (3):

$$x_0 = N^{1/n}(1 + \alpha_0), \tag{2}$$

$$x = N^{1/n}(1 + \alpha), \tag{3}$$

and the combination of (2) and (3) with (1) gives:

$$1 + \alpha = \frac{1}{n}\left[\frac{(n-1)(1+\alpha_0)^n + 1}{(1+\alpha_0)^{n-1}}\right]. \tag{4}$$

Solving (4) for α and rearranging:

$$\alpha = \left(\frac{n-1}{n}\right)\alpha_0 + \frac{1}{n}\left[\frac{1}{(1+\alpha_0)^{n-1}} - 1\right]. \tag{5}$$

From (5), it is clear that for positive values of α_0, the second term is negative and

$$\alpha < \left(\frac{n-1}{n}\right)\alpha_0 \tag{6}$$

which corresponds to

$$\epsilon' < \left(\frac{n-1}{n}\right)\epsilon$$

as was shown in reference (1). (ϵ' and ϵ are the subsequent and present *absolute* errors, respectively.) For α_0 with a magnitude less than one, a binomial expansion of (5) gives

$$\alpha = \left(\frac{n-1}{2}\right)\alpha_0^2 + R, \tag{7}$$

where the remainder, R, contains only terms in α_0 of third and higher order, and for $\alpha_0 \ll 1$, α is approximately given by

$$\alpha \approx \left(\frac{n-1}{2}\right)\alpha_0^2, \tag{8}$$

which corresponds to

$$\epsilon' \approx (n-1)\epsilon^2/2N^{1/n}$$

as was shown in reference (1).

References

1. Henry Laufer, "Finding the Nth Root of a Number by Iteration," MATHEMATICS MAGAZINE, **36**, May 1963, p. 153.

2. G. R. Taylor, "An Approximate for Any Positive Integral Root," MATHEMATICS MAGAZINE, **35**, March 1962, pp. 107–8.

3. E. M. Romer, "An Iterative Procedure for Obtaining Fractional Roots of Real Numbers," Wright Air Development Center, WADC TN-59-116, Wright-Patterson AFB, Ohio, April 1959.

4. E. M. Romer, "Recursion Formulae to Obtain Integral Roots of Real Numbers," Air Force Institute of Technology Technical Report 61-3, Wright-Patterson AFB, Ohio, March 1963.

BOOK REVIEWS

The Lore of Large Numbers. By Philip J. Davis. Random House, Inc., New York, 1961, x+165 pp., paper back, $1.95.

This is volume 6 of the New Mathematical Library which is designed to appeal to high school students and laymen. It does, indeed, contain much stimulating material. Considering the space devoted to e, π, and recurring decimals, a better title might have been *Long Numbers*. In the course of discussion of large numbers, some topics dealt with in down-to-earth fashion are: scales of notation, division by zero, exponential notation, the personality of numbers, theory of residues, and simultaneous linear equations. Critical editing could have polished a few rough spots without reducing the readability. Consider, for example: page 8, "$\sqrt{2}$ cannot be a fraction"; page 9, "complex numbers . . . have the peculiar property that . . . we cannot tell which is the larger"; page 13, "An animal who must find food on the high branches must grow a long neck"; page 19, "If 2 were multiplied by itself 11 times, the resulting number would be called the 11th power of 2." An index would have been helpful. Answers are given to selected members of the problem sets which are an integral part of the book. The excellent short bibliography should encourage wider reading.

CHARLES W. TRIGG
Los Angeles City College

Apollonius of Perga: Treatise on Conic Sections. Edited by T. L. Heath. Barnes and Noble, Inc., New York, 1961, clxx+254 pp., hard cover, $9.00.

This is an authentic reprint of Sir Thomas Heath's scholarly treatment which had not previously been available for many years. The work of Apollonius has been put into modern notation without loss of the spirit and continuity of the original. Prefacing this rendition and occupying over one-third of the volume is an earlier history of the conic sections among the Greeks and an introduction to the *Conics* of Apollonius. This extensive preamble, written with clarity and enviable style, greatly enhances the value of the book.

CHARLES W. TRIGG
Los Angeles City College

Groups. By G. Papy. Dunod, Paris. 1961. 249 pages.

The role of the concept of "group" being fundamental to mathematics, it is desirable that the elements of the theory of groups be taught on the secondary level, not as a supplementary chapter, but rather as a motivating element in the construction of the mathematical structure. This is the purpose of the author in writing this book.

In spite of the modernization in the mathematical instruction on the secondary level, nearly all the students leave their secondary education without having an idea of this major element of mathematical thought.

The purpose of this book is to contribute to the indispensable reform of modernizing secondary school mathematics by presenting the fundamental elements of the theory of groups.

In this transition period, the book is written to be used by secondary school teachers, students intending to specialize in mathematics and physical science, as well as for those students intending to teach in secondary schools.

It is the author's wish that the great part of the contents of this book be integrated in the curriculum of secondary school mathematics.

Chapter I presents the definitions and fundamental properties of groups.

Chapter II illustrates the concept of group by examples and counter-examples which are important in the development of group idea.

Chapter III is particularly useful for its application to mathematical instruction on the secondary level, and illustrates clearly the problems relative to negative and fractional exponents as well as to primitive roots of unity.

In Chapter IV, the theory is elevated to higher levels by the consideration of subgroups and of stable parts of groups. Experience reveals that the study of those parts of groups already known, interest most the students.

Chapter V reviews some supplementary results of commutative groups.

Chapter VI shows how the application of group concepts enriches the exposition of elementary arithmetic, and lead naturally to concepts of rings and lattices. The teaching of arithmetic along these lines leads to the determination of the theory of groups. The methods used are basic and can be applied frequently in more advanced studies of mathematics.

Chapter VII studies the important relations which could exist between different types of groups. It expounds in particular the theories of homomorphism and isomorphism which are fundamental to contemporary mathematics.

Chapter VIII is devoted to the intrinsic study of permutation of sets, finite or infinite.

In Chapter IX, the introduction to the notion of operator groups widens considerably the scope of the theory of groups. Due to this notion, the theory of vector spaces and the theory of ideals are found directly attached to the main body of the theory of groups.

The last chapter is devoted to the concept of dimension theory of operator groups. Here we find the theories of Jordan-Hölder, Schreier and Zassenhaus. We know that this theory of group operators is essentially due to Emmy Noether and Krull.

There are ample exercises in each chapter, and abundant review exercises at the end of each chapter to test the understanding of the reader.

A new feature of the book is the colored plates, illustrating abstract theories in terms of colored figures. These figures help the reader from a pedagogic point of view in understanding more deeply the theories. They suggest methods of instruction of the fundamental concepts of the theory of groups.

This is an excellent book on the elementary theory of groups and is recommended highly for those who begin the study of groups for the first time.

SOUREN BABIKIAN
Los Angeles City College

(Continued from page 311)

lines $\alpha x-\beta y$, $\beta y-\gamma z$, $\gamma z-\alpha x$, $\alpha\beta\gamma\neq 0$. Each such line determines a side and an angle of a triangle. There are four possible combinations of signs of the ratios $\alpha:\beta$, $\beta:\gamma$, $\gamma:\alpha$.

The Euclidean case requires different treatment since $(x', y')=1$ if again $\Delta_{12}>0$. But then $(x'-y', x'-y')=0$ and $x'-y'$ is not a proper line. It is expedient to show that there is just one such "ideal" line i. Since our dimension number is three, we can set $i=\alpha x'+\beta y'+\gamma z$ and $(i, i)=0$ implies $\alpha+\beta=0$, $\gamma=0$. We normalize i by taking $\alpha=(\Delta_{11}\Delta_{22})^{1/2}$. Hence

$$i = (\Delta_{11}\Delta_{22})^{1/2}(x' - y'). \tag{8}$$

If we further assume that $\Delta_{23}>0$, $\Delta_{31}>0$, it is easily shown that we may write the symmetric expression

$$i = \Delta_{11}^{1/2} x + \Delta_{22}^{1/2} y + \Delta_{33}^{1/2} z. \tag{9}$$

We now define the distance between x, z and y, z as $(\Delta_{11}\Delta_{22})^{-1/2}$, which is the scalar factor of i in the expression (8) for $x'-y'$. Similar definitions apply to the other sides of the triangle due to the symmetry of (9). Our normalization process indicates that there cannot exist an absolute unit of length in Euclidean geometry. From the identity (4) the law of cosines is now immediately derived, taking into account that $(x, y)=-\cos xy$.

Reference

1. P. R. Halmos, Finite-dimensional vector spaces, 2d ed., D. Van Nostrand, 1958.

PROBLEMS AND SOLUTIONS

EDITED BY ROBERT E. HORTON, Los Angeles City College

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. Proposals should be accompanied by solutions, when available, and by any information that will asist the editor. Ordinarily, problems in well-known textbooks should not be submitted. Solutions should be submitted on spearate, 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 City College, 855 North Vermont Avenue, Los Angeles 29, California.

PROPOSALS

530. *Proposed by Maxey Brooke, Sweeny, Texas.*

"Here is something interesting," said Bobby's father. "The product of my age and Bobby's age remains the same even when the two digits in each age are reversed. And our ages are not divisible by 11."

"That's nothing," said Bobby's grandfather. "The product of my age and Bobby's also remains the same when the digits in each age are reversed."

"You've got nothing on me," said Bobby's great-grandfather. "The product of my age and Bobby's age also remains the same when the two digits in each age are reversed."

How old is Bobby?

531. *Proposed by Ben B. Bowen, Vallejo Junior College, California.*

Given the differential equation

$$\sqrt{1 - y^2}\, dx = \sqrt{1 - x^2}\, dy,$$

a student immediately wrote the solution

$$x\sqrt{1 - y^2} = y\sqrt{1 - x^2} + c.$$

His text book gave the solution $\sin^{-1} x - \sin^{-1} y = c$.

Given the differential equation $f(y)dx = f(x)dy$, find f such that $xf(y) - yf(x) = c$ is a solution.

532. *Proposed by Josef Andersson, Vaxholm, Sweden.*

The direction of the axis of a parabola is given. Construct the parabola if three points on the curve are also given.

533. *Proposed by David L. Silverman, Beverly Hills, California.*

What is the largest integer which cannot be partitioned into distinct squares?

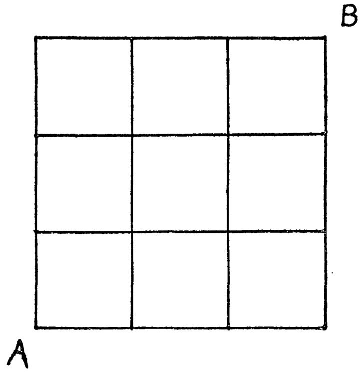
XENOPHOBES

How often is the hostile critic,
Of a process analytic,
A stranger to the palaces
Of classical analysis.

MARLOW SHOLANDER

534. *Proposed by Brother U. Alfred, St. Mary's College, California.*

Given a three-by-three grid as shown



Determine the number of paths from A to B under the limitation that no point of the grid is covered twice by any one path.

535. *Proposed by Murray S. Klamkin, State University of New York at Buffalo.*

It is known that if the family of integral curves of the linear differential equation $y'+P(x)y=Q(x)$ is cut by the line $x=a$, then the tangents at the points of intersection are concurrent. Prove, conversely, that if for a first order equation $y'=F(x, y)$ the tangents (as above) are concurrent, then $F(x, y)$ is linear in y.

536. *Proposed by D. Moody Bailey, Princeton, West Virginia.*

A line through the incenter I of triangle ABC meets sides BC, CA, and AB at the respective points M, N, O. Points D, E, and F are chosen on sides BC, CA, and AB so that

$$\frac{BD}{DC} = -\frac{c}{b}\cdot\frac{BM}{MC}, \qquad \frac{CE}{EA} = -\frac{a}{c}\cdot\frac{CN}{NA},$$

and

$$\frac{AF}{FB} = -\frac{b}{a}\cdot\frac{AO}{OB}$$

where a, b, and c are the sides opposite vertices A, B, and C of the triangle. Show that rays AD, BE, and CF are concurrent at a point P which lies on the circumcircle of triangle ABC.

SOLUTIONS

Late Solutions

489, 490. *Thomas R. Hamrick, San Quentin, California.*
495. *Sam Sesskin, Hempstead, New York.*
496. *Joseph B. Bohac, St. Louis, Missouri.*

502. *Thomas R. Hamrick, San Quentin, California, and P. S. Pappas, Whitestone, New York.*

An American Alphametic

509. [March 1963] *Proposed by Huseyin Demir, Middle East Technical University, Ankara, Turkey.*

Solve the cryptarithm

$$\begin{array}{ccccccc} & U & N & I & T & E & D \\ & S & T & A & T & E & S \\ \hline A & M & E & R & I & C & A \end{array}$$

in the base 11, introducing the digit α.

Solution by Anton Glaser, Ogontz Campus, Pennsylvania State University.

My solution was obtained as follows:

(1) $A=1$ [In any numeration system, adding two digits can result in "carrying" at most unity.]

(2) $D\neq 0$ [Suppose $D=0$, then we get contradiction of $S=A=1$ vs. $S\neq A$]

(3) $S\neq 0$ [Similar to (2)]

(4) $E\neq 0$ [Suppose $E=0$, then $C=A=1$ contradicting $C\neq A$]

(5) $T\neq 0$ [Suppose $T=0$, then either $I=A=1$ vs. $I\neq A$ or $T=I=0$ vs. $T\neq I$]

(6) $U\neq 0$ and $S\neq 0$ by usual rules of cryptarithms

(7) $E\neq\alpha$ [Suppose $E=\alpha$, then $C=E=\alpha$ vs. $C\neq E$]

(8) $U+S>9$

(9) $U+S>\alpha$ if nothing was "carried" from previous column

(10) $D+S=11_{\text{(eleven)}}=12_{\text{ten}}=\text{twelve}$

Only the digits shown in table at right are possible for D and S, and only in the combinations shown.

[Since $A=1$ neither D nor S can be 1]

[Neither D nor S can be 6, since either would imply $D=S=6$ vs. $D\neq S$]

D	
2	α
3	9
4	8
5	7
7	5
8	4
9	3
α	2

(11) $T\neq 2$, $T\neq 3$, $T\neq 4$, and $T\neq 5$ [For $T=2$, $T=3$, $T=4$, and $T=5$ and the seven possible values of E that go with each of these four values of T, there resulted in each case a contradiction of some sort.]

(12) For $T=6$ and $E=5$, the remaining letters could be assigned a one-to-one correspondence with the remaining digits that would satisfy the cryptarithm.

$$\begin{array}{ccccccc} & 8 & \alpha & 2 & 6 & 5 & 3 \\ & 9 & 6 & 1 & 6 & 5 & 9 \\ \hline 1 & 7 & 5 & 4 & 2 & 0 & 1 \end{array}$$

Also solved by Josef Andersson, Vaxholm, Sweden; Merrill Barneby, University of North Dakota; Maxey Brooke, Sweeny, Texas; Harry M. Gehman, State University of New York at Buffalo; Wahin Ng, San Francisco, California; Norman Harelik, Mather High School, Chicago, Illinois; J. A. H. Hunter, Toronto, Ontario, Canada; Robert Sandling, Columbia University; Anita Skelton, Watervliet Arsenal, New York; David L. Silverman, Beverly Hills, California; Orvan Sommers, West Bend High School, Wisconsin; C. W. Trigg, Los Angeles City College; Hazel S. Wilson, Jacksonville University, Florida; Brother Louis F. Zirkel, Archbishop Molloy High School, Jamaica, New York; and the proposer.

A Power of Two

510. [March 1963] *Proposed by Miltiades S. Demos, Drexel Institute of Technology.*

Evaluate

$$\prod_{k=1}^{n} \sin\left(\frac{2k-1}{2n}\cdot\frac{\pi}{2}\right).$$

Solution by L. Carlitz, Duke University.

In the identity

$$x^{2n}+1=\prod_{k=1}^{2n}\left(x-e^{(2k-1)\pi i/2n}\right)$$

take $x=1$. Then

$$\begin{aligned} 2 &= \prod_{k=1}^{2n}\left(1-e^{(2k-1)\pi i/2n}\right) \\ &= \prod_{k=1}^{2n} e^{(2k-1)\pi i/4n}\left(e^{(2k-1)\pi i/4n}-e^{-(2k-1)\pi i/4n}\right) \\ &= 2^{2n}\prod_{k=1}^{2n}\sin\frac{(2k-1)\pi}{4n} \\ &= 2^{2n}\left(\prod_{k=1}^{n}\sin\frac{(2k-1)\pi}{4n}\right)^2 \end{aligned}$$

so that

$$\prod_{k=1}^{n}\sin\frac{(2k-1)\pi}{4n}=\frac{\sqrt{2}}{2^n}.$$

Also solved by Josef Andersson, Vaxholm, Sweden; J. L. Brown, Jr., Pennsylvania State University; Daniel I. A. Cohen, Midwood High School, Brooklyn, New York; Eldon Hansen, Lockheed Missiles and Space Company, Palo Alto, California; Paul D. Thomas, U. S. Coast and Geodetic Survey, Washington, D. C.; and the proposer.

Dollar Tangents

511. [March 1963]. *Proposed by Maxey Brooke, Sweeny, Texas.*

With a silver dollar, trace a circle. Choose a point on this circle. Using only the silver dollar and a pencil, construct a circle through the point tangent to the first circle. Assume that you can trace a circle through any two points less than a dollar's diameter apart.

Solution by the proposer. We proceed using the following steps. Let A be the given circle and a be the given point on it. Draw circle B through a intersecting A again at b. Determine c diametrically opposite b on circle B. Draw circle C through points a and c. C will be the required circle tangent to A at a.

To determine a point diametrically opposite a given point on a circle, let A be the given circle and a be the given point. Then draw circle B through a intersecting A at b. Draw circle C through b intersecting B at c. Draw circle D through c intersecting A at d and C at e. Draw E through d and e intersecting A at f. Point f is diametrically opposite a in circle A.

Joseph Konhauser pointed out that this problem appeared with solution in the October, 1954, issue of *Eureka*, the journal of the Cambridge University Mathematical Society, Junior Branch of the Mathematical Association.

A Difference Operation

512. [March 1963]. *Proposed by R. N. Karnawat and J. M. Gandhi, Government College, Bhilwara, Rajasthan, India.*

Prove the following identity:

$$\binom{n}{2}1^x - \binom{n}{3}2^x + \cdots (-1)^n\binom{n}{n}(n-1)^x = (-1)^{x+1}, \text{ for } n > x,$$

or

$$= (-1)^x[n! - 1], \text{ for } n = x.$$

Solution by James C. Hickman, University of Iowa.

The identity is only true for x an integer. For example if we set $x=1/2$ and $n=2$ we have

$$\binom{2}{2}1^{1/2} = 1 \neq (-1)^{3/2}.$$

For x an integer the series may be represented by

$$\begin{aligned}(1-E_y)^n y^x \big|_{y=-1} - y^x \big|_{y=-1} &= (-1)^n \Delta_y^n y^x \big|_{y=-1} - y^x \big|_{y=-1}\\ &= (-1)^{x+1}, \qquad n > x,\\ &= (-1)^x(n!-1), \qquad n = x,\end{aligned}$$

where E_y and Δ_y are the shift and difference operators operating on y and where the results are obtained by using well-known results concerning operating on polynomials with the difference operator.

Also solved by Joseph Andersson, Vaxholm, Sweden; J. L. Brown, Jr., Penn-

sylvania State University; Eldon Hansen, Lockheed Missiles and Space Company, Palo Alto, California; Dmtri Thoro, San Jose State College, California; Brother Louis F. Zirkel, Archbishop Molloy High School, Jamaica, New York; and the proposers.

Brown noted that the problem is a special case of Problem E1253 in the *American Mathematical Monthly*, February, 1957, p. 109.

A Collinearity

513. [March 1963]. *Proposed by Leon Bankoff, Los Angeles, California.*

DE is a chord of a circle O perpendicular to the diameter AB at C. A circle J is inscribed in the space bounded by AC, CE and the arc AE. Show that DJ cuts AC at the point where AC touches the incircle I of triangle ACD.

Solution by Gilbert Labelle, Université de Montréal, Canada.

Let the radius of circle O be 1, θ be $<AOD$ $(0 \leqq \theta \leqq \pi)$, P the common point to O and J, r and R the radii of I and J respectively, M and N the points of contact of I and J with AB. Then we must prove that:

$$\frac{r}{CD} = \frac{R}{CD + R}, \tag{1}$$

where $CD = \sin\theta$.

The relation $pr = S_{ADC}$ gives:

$$r = \frac{2\sin^2\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)}{1 + \sin\left(\frac{\theta}{2}\right) + \cos\left(\frac{\theta}{2}\right)},$$

The relation $(\overline{OP} - R)^2 - R^2 = \overline{ON}^2$ gives $(1-R)^2 - R^2 = (\cos\theta + R)^2$ then: $R = 2\cos\theta/2(1-\cos\theta/2)$,

And (1) reduces to

$$\frac{\sin\frac{\theta}{2}}{1 + \sin\frac{\theta}{2} + \cos\frac{\theta}{2}} = \frac{1 - \cos\frac{\theta}{2}}{1 + \sin\frac{\theta}{2} - \cos\frac{\theta}{2}},$$

a true relation.

Also solved by Josef Andersson, Vaxholm, Sweden; P. R. Nolan, Department of Education, Dublin, Ireland; Hazel S. Wilson, Jacksonville University, Florida; and the proposer.

Integral Simplification

514. [March 1963]. *Proposed by Joseph W. Andrushkiw, Seton Hall University.*

Show that if $f(z)$ is an odd function integrable on $[-1, 1]$, then

$$\int_0^{2k\pi} x^2 f(\sin x)dx = -2k^2\pi^2 \int_0^{\pi} f(\sin x)dx$$

and apply it to evaluate

$$\int_0^{2\pi} \frac{x^2 \sin x}{1+\cos^2 x}\, dx.$$

Solution by J. L. Brown, Jr., Pennsylvania State University.

Since $(x-k\pi)^2 f(\sin x)$ is antisymmetric about $x=k\pi$, we have

$$\int_0^{2k\pi} (x-k\pi)^2 f(\sin x)dx = 0,$$

or

$$\int_0^{2k\pi} x^2 f(\sin x)dx - 2k\pi \int_0^{2k\pi} xf(\sin x)dx + k^2\pi^2 \int_0^{2k\pi} f(\sin x)dx = 0.$$

From the odd symmetry,

$$\int_0^{2kx} f(\sin x)dx = 0,$$

so that

$$\int_0^{2k\pi} x^2 f(\sin x)dx = 2k\pi \int_0^{2k\pi} xf(\sin x)dx.$$

But

$$\begin{aligned}\int_0^{2k\pi} xf(\sin x)dx &= \sum_{i=0}^{2k-1} \int_{i\pi}^{(i+1)\pi} xf(\sin x)dx \\ &= \sum_{i=0}^{2k-1} \int_0^{\pi} (y+i\pi)f[(-1)^i \sin y]dy = \sum_{i=0}^{2k-1} \int_0^{\pi} (y+i\pi)(-1)^i f(\sin y)dy \\ &= \int_0^{\pi} yf(\sin y)dy\left(\sum_{i=0}^{2k-1}(-1)^i\right) + \pi\left(\sum_{i=0}^{2k-1}(-1)^i i\right)\int_0^{\pi} f(\sin y)dy \\ &= 0 - k\pi\int_0^{\pi} f(\sin y)dy = -k\pi \int_0^{\pi} f(\sin y)dy.\end{aligned}$$

Thus

$$\int_0^{2k\pi} x^2 f(\sin x)dx = -2k^2\pi^2 \int_0^{\pi} f(\sin y)dy$$

as required. Using this result,

$$\int_0^{2\pi} \frac{x^2 \sin x}{1+\cos^2 x}\, dx = -2\pi^2 \int_0^{\pi} \frac{\sin x}{1+\cos^2 x}\, dx = -4\pi^2 \tan^{-1}(1) = -\pi^3$$

from standard integral tables.

Also solved by Josef Andersson, Vaxholm, Sweden; Mark S. Fineman and Martin van Buren (jointly), Floral Park, New York; and the proposer.

A Complex Sequence

515. [March 1963]. *Proposed by R. J. Mansfield, Research Council of Alberta, Canada.*

Find an explicit expression for the nth term in the sequence:

$$1,\ 1,\ 1+i,\ 1+2i,\ 3i,\ \cdots$$

Solution by R. G. Buschman, Oregon State University.

Consider the terms indexed $0, 1, 2, \cdots, n, \cdots$ so that with s_0 through s_5 given we have that, if $\{s_n\}$ is a sequence starting with the given numbers, then also $\{s_n + \binom{n}{6} a_n\}$ is a sequence starting with the same numbers, for an arbitrary sequence $\{a_n\}$.

Now if the nth term of the sequence were defined by

$$s_0 = s_1 = 1,\ s_n = s_{n-1} + is_{n-2} \qquad (n \geqq 2),$$

then we would have a unique sequence generated. The characteristic equation for this difference equation is $z^2 - z - i = 0$, which has discriminant $1+4i$. Hence, using one of the known representations for the solution, we have, for example, the explicit expression

$$2^n s_n = \sum_{k=0}^{[n/2]} \binom{n}{2k} (1+4i)^k + \sum_{k=0}^{[(n-1)/2]} \binom{n}{2k+1} (1+4i)^k.$$

Also solved by Josef Andersson, Vaxholm, Sweden; Dermot A. Breault, Sylvania Applied Research Laboratory, Waltham, Massachusetts; Brother T. Brendan, St. Mary's College, California; Maxey Brooke, Sweeny, Texas; J. L. Brown, Jr., Pennsylvania State University; L. Carlitz, Duke University; Daniel I. A. Cohen, Midwood High School, Brooklyn, New York; Dewey C. Duncan, East Los Angeles College; Harry M. Gehman, State University of New York at Buffalo; H. W. Gould, West Virginia University; Gilbert Labelle, Université de Montréal (two solutions); Barry Litvack, University of Michigan; John W. Milsom, Texas College of Arts and Industries; Michael J. Pascual, Watervliet Arsenal, New York; Brother Louis F. Zirkel, Archbishop Molloy High School, Jamaica, New York; and the proposer.

Duncan, Brown, Gould, Milsom, and Zirkel pointed out the possibility of many different sequences arising after the five terms given unless some further hypothesis about the sequence is stated.

Comment on Problem 486

486. [May 1962 and January 1963]. *Proposed by Max Rumney, London, England.*

Solution by the proposer.

It was the intention of the proposer that all the a_{ij} be different, and the problem should have the additional words "all a_{ij} being different" between "a_{33}" and "so."

Notation. Since all the N are different we shall order them thus $N_1 > N_2 > N_3$ and their partitions respectively

$$(a_{11},\ a_{12},\ a_{13}),\ (a_{21},\ a_{32},\ a_{23}) \quad \text{and} \quad (a_{31},\ a_{32},\ a_{33}).$$

Denote the determinant of these partitions by D, and if a different partition say $(a'_{11}, a'_{12}, a'_{13})$ is used then the determinant will be D'.

N minima. The solver must also investigate the smallest N_1, N_2, N_3 (N_{mn}) which are capable of representation as $D=0$. The proposer found that for $N_1=6$, $N_2=17$, 18, $N_3=25$, 22. $D=0$ with (1, 3, 2) with (4, 5, 8), (6, 7, 12) and (4, 6, 8) (5, 7, 10).

For even N, N_{mn} are 6, 18, 26, $D=0$ for (1, 2, 3) (4, 5, 9) (6, 7, 13).

If the N are consecutive then $N_1 \geqq 16$, e.g., for $N_1=16$

$$\begin{vmatrix} 1 & 13 & 2 \\ 4 & 5 & 8 \\ 3 & 9 & 6 \end{vmatrix} = 0$$

and, of course, there may be many $D=0$.

Limits. We now say that $N_1 \geqq 6$, $N_2 \geqq 17$, $N_3 \geqq 22$ will give an unlimited number of N, but for consecutive N, $N \geqq 16$.

Method. If

$$a_{11}x + a_{12}y = a_{13}, \qquad a_{21}x + a_{22}y = a_{23}, \qquad a_{31}x + a_{32}y = a_{33} \tag{I}$$

then $D=0$.

Let $x=k$, $N_1 > a_{11}(k+1)$, $y=0$, then as in (I), we have

$$a_{11}k + a_{12}y = a_{13}, \qquad a_{12} = N_1 - a_{11} - a_{13}$$

and

$$D = k\begin{vmatrix} a_{11} & a_{12} & a_{11} \\ a_{21} & a_{22} & a_{21} \\ a_{31} & a_{32} & a_{31} \end{vmatrix} = 0$$

since $a_{13} = ka_{11}$.

The value of k can generally be obtained by inspection and its values be such that all a_{ij} be different. It is easy to see that with the exceptions mentioned (and there may be some more), representations will not be unique and these will be the many D's. If all N *are even*, an alternative and simpler method is given. The number of representations will depend on the number of partitions of $N/2$ into two parts.

Let $x=y=1$.

Then

$$a_{11}x + \left(\frac{N}{2} - a_{11}\right)y = \frac{N}{2}$$

and similarly for the other a_{ij}, and, of course,

$$D = \begin{vmatrix} a_{11} & \dfrac{N_1}{2} - a_{11} & \dfrac{N_1}{2} \\ a_{21} & \dfrac{N_2}{2} - a_{21} & \dfrac{N_2}{2} \\ a_{31} & \dfrac{N_3}{2} - a_{31} & \dfrac{N_3}{2} \end{vmatrix} = 0$$

since col 1+col 2=col 3.

We give two examples to illustrate the method.

(a) N's are 37, 60, 89. Choose $x = 4, y = 0$

then

$$3 \times (4) + \times (0) = 12, \; 7 \times (4) \times 25 \times (0) = 28, \; 8 \times (4) + 49(0) = 32$$
$$37 = 3 + 22 + 12, \; 60 = 7 + 25 + 28, \; 89 = 8 + 49 + 32,$$

and $D=0$.

(b) N's are 16, 30, 808.

$$\begin{array}{ll} \dfrac{16}{2} = 8, \text{ partitions are say } 3, 5 & \\ \dfrac{40}{2} = 20, \text{ partitions are say } 7, 13 & \text{and} \\ \dfrac{808}{2} = 404, \text{ partitions are say } 107, 297 & \end{array} \quad \begin{vmatrix} 3 & 5 & 8 \\ 7 & 13 & 20 \\ 107 & 297 & 404 \end{vmatrix} = 0.$$

QUICKIES

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

Q 323. Determine the maximum number of consecutive terms of the coefficients of a binomial expansion which are in arithmetic progression. [*Submitted by Murray S. Klamkin.*]

Q 324. The sum of the numerical coefficients of $(A+B)^{13}$ is which of the following: 2304, 1127, 8192? [*Submitted by Raphael T. Coffman.*]

Q 325. If $2z$ is the harmonic mean of two rational numbers x and y, then $x^2+y^2+z^2$ is the square of a rational number. [*Submitted by C. W. Trigg.*]

Q 326. Solve in positive integers the equation $x^x+y^y=z_z$. [*Submitted by David L. Silverman.*]

(Answers on page 280)

The ratio of areas produced is proportional to the ratio of the rate of change of areas. When the point on the circle has made one revolution the area produced is πc^2. Therefore, in the same time:

$$\frac{vA_1}{vA_2}=\frac{A_1}{\pi c^2}=\frac{b}{c},$$

from which $A_1=\pi bc$. This is the equation of an ellipse with semiaxes b and c. Hence, both the circular orbit and the elliptical orbit in which the major semi-axis is equal to the radius of the circle have the same period, and since Kepler's third law applies to the circle it applies also to the ellipse.

The equations which have been derived in this paper apply to particles. In order to make them applicable to physical bodies, as planets or earth satellites, it is necessary to know that the gravitational effect of a spherical body composed of homogeneous spherical shells is as though all the mass is at the center. An elementary proof of this has been derived by the writer, but because it is long and complex it has not been included here. The equation for the area of an ellipse is not generally derived below the level of integral calculus. It is, however, an easy matter to derive the equation from considerations of the relation of the ellipse to the circle.

Reference

1. The Reflection Property of the Conics. By R. T. Coffman and C. S. Ogilvy, MATHEMATICS MAGAZINE, **36**, No. 1, p. 11.

Answers

A323. Three. For three terms to be in A.P., we must have

$$2\binom{m}{n}=\binom{m}{n-1}+\binom{m}{n+1}\quad\text{or}\quad(m-2n)^2=m+2$$

whence

$$m=a^2-2,\qquad 2n=a^2\pm a-2$$

In order to have four terms in A.P., $(a^2-a)/2=(a^2+a-2)/2$ or $a=1$ and impossible. (See Note of Th. Motzkin, *Scripta Mathematica*, March, 1946, p. 14.)

A324. If A and B are both made equal to one, $(A+B)^{13}=2^{13}$, which is the sum of the numerical coefficients, as all the other coefficients become unity. Hence $2^{13}=8192$.

A325. Since $2z=2xy/(x+y)$, $2xy-2yz-2zx=0$, whereupon $(x+y-z)^2=x^2+y^2+z^2$.

A326. Assume $x\leqq y$. Then $y>1$ and $x^x+y^y\leqq 2y^y\leqq y^{y+1}\leqq z^z$. Thus $x^x+y^y<z^z$ and there are no solutions.

(Quickies on page 328)

MATHEMATICS MAGAZINE VOL. 36, NO. 5, NOV.-DEC., 1963