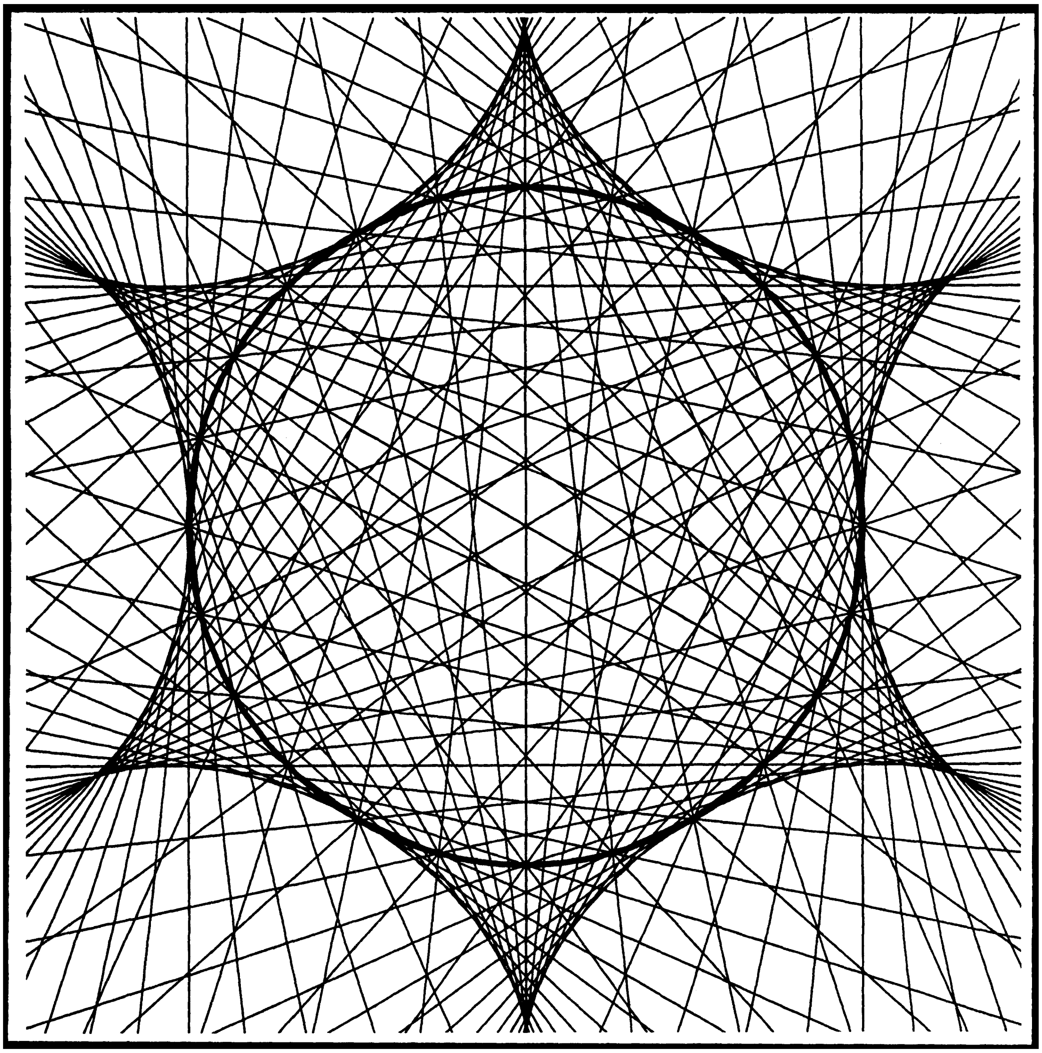
Vol. 73, No. 3, June 2000

MATHEMATICS MAGAZINE



- Trochoids and Bungee Cords
- Possible and Impossible Pyramids
- A Bifurcation Problem in DE

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. The *Magazine* is not a research journal, so the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Manuscripts on history are especially welcome, as are those showing relationships among various branches of mathematics and between mathematics and other disciplines.

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 71, pp. 76–78, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Send new manuscripts to Frank Farris, Editor-Elect, Department of Mathematics and Computer Science, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053-0290. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

AUTHORS

Tom Sibley received his B.A. from St. Olaf College in 1973. His Peace Corps experience teaching high school mathematics in the Congo led him to choose teaching as a career. After finishing his Ph.D. at Boston University in 1980, he taught college mathematics in Liberia. He has been at St. John's University since 1984. His research interests have focused on symmetry and finite geometries. He stumbled upon the idea of pyramid completeness while writing problems for his geometry textbook. Because at that time he couldn't solve this question for the taxicab metric, he decided not to include it in the textbook.

Andrew Simoson received his Ph.D. from the University of Wyoming in 1979 and has been chair of the mathematics department at King College ever since. While he was a Fulbright professor at the University of Dar es Salaam in Tanzania on a 1997–98 sabbatical, the student body there struck on a haphazard basis, effectively nullifying twenty teaching days. During the periods of waiting for classes to begin again, he dabbled in depicting envelopes using *Mathematica*, and so stumbled upon the insights presented in this paper.

Duff Campbell received his Ph.D. from Boston University in 1997, and since then has taught at the United States Military Academy at West Point, NY. **Samuel R. Kaplan** received his Ph.D. from Boston University in 1996. He joined the faculty at the University of North Carolina at Asheville in 1999 after teaching three years at Bowdoin College. Their collaboration began when Campbell and Kaplan were in graduate school, and is based on their experiences assistant teaching differential equations.

Vol. 73, No. 3, June 2000

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, the Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

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

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036-1385.

Printed in the United States of America

ARTICLES

The Trochoid as a Tack in a Bungee Cord

ANDREW J. SIMOSON
King College
Bristol, TN 37620

1. Introduction

Perhaps the most beautiful curves studied in calculus are the *trochoid* family. In the mathematical zoo of the typical calculus text, these trochoids are presented as the curves traced out by a tack embedded somewhere in a wheel, which in turn is rolling around a circle. In FIGURE 1(a), the small *railroad wheel* has a tack in its rim which overlaps the larger circular road; the loopy curve is the trochoid, the path traced out by the tack. In FIGURE 1(b), the wheel is rolling on the inside of the circle. In this

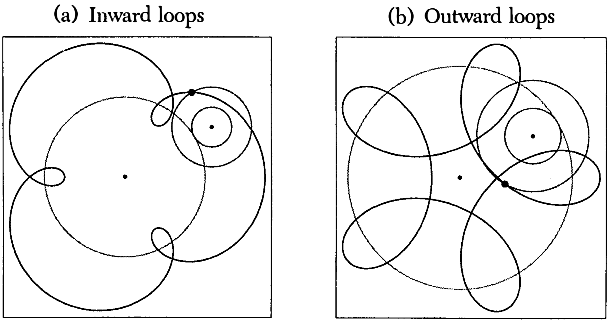


FIGURE 1
The trochoids

paper, we present these trochoids in a more wild setting where they are both hidden and defined by a family of lines, giving an alternate model of the trochoid as the curve traced out by a tack in a bungee cord whose ends are held by two runners proceeding around a circular track. If we imagine this bungee cord as being made of glow-stick material, then a time-lapse overhead, stroboscopic photo of the action leads us to wrestle with the classic notion of the envelope to a curve and then on to fractional derivatives and harmonic envelopes as we explore the nature of this alternative model of the trochoid. We conclude with a generalization and several problems, some of which could serve as open-ended student projects.

2. The trochoid equations

Let us derive, via the classic wheel model, parametric equations for a trochoid in terms of a variable t referred to as time. Take a tack of glow-stick material and embed it in a "railroad" wheel which is rolling in the xy-plane around a fixed circle whose center is the origin and whose radius R is a positive real number. By custom we disallow the road to be a single point and we stipulate that the distance from the center of the wheel to the road is nonzero. To allow for some normalized standard equations, let us say that the tack's initial position is on the x-axis at a point other than the origin. Let P and Q be the positions of the center of the wheel and of the tack, respectively. Let the nonzero number r be the "radius" of the wheel so that the center of the wheel is initially at the point $P_o = (R + r, 0)$ and where the tack is s units from the wheel's center so that the tack is initially at $Q_o = (R + r + s, 0)$, where s is any real number with $R + r + s \neq 0$. We interpret the wheel as rolling around the outside of the circle of radius R if $r > 0$, and as rolling around the inside if $r < 0$; furthermore, we interpret the tack as initially being on the right hand side of the wheel if $s > 0$, and on the left hand side if $s < 0$.

Now a time-lapse photo of the glowing tack traced out by Q is called a trochoid when the tack's motion is periodic in time. In order to obtain periodic motion, $\frac{R}{r}$ must be some non-zero rational number m. Let θ be the angle between the positive x-axis and a ray from the origin through P. Let ϕ be the angle between the direction P_oQ_o and the direction PQ. So Q is given by the sum of the position of the wheel's center and the vector from this center to the tack; that is, the sum $(R + r)(\cos\theta, \sin\theta) + s(\cos\phi, \sin\phi)$, where initially $\theta = 0 = \phi$. See FIGURE 2 for the case when $R > 0$ and

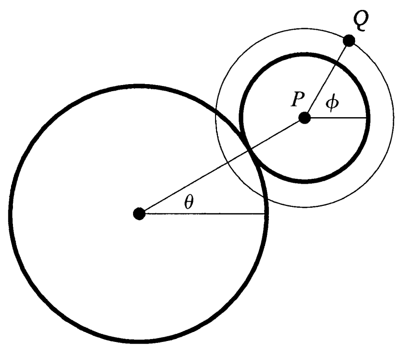


FIGURE 2
The wheel setup

$s > r > 0$. By the arclength formula, we know that $R\theta = r\phi$. Hence Q can be written as $(R + r)(\cos\theta, \sin\theta) + s(\cos m\theta, \sin m\theta)$. Next, choose any non-zero rational number p and let $q = mp$. Using the reparameterization $pt = \theta$ results in Q as $r(1 + \frac{q}{p})(\cos pt, \sin pt) + s(\cos qt, \sin qt)$. Note that if $\frac{q}{p} = -1$, then Q's path is a circle of radius s. Otherwise, since any non-zero scalar multiple of this curve is simply a rescaling (along with a mirror image in the case of a negative scalar), then any curve of the form $\alpha(\cos pt, \sin pt) + \beta(\cos qt, \sin qt)$ is a trochoid as well, where α and β are real numbers with $\alpha + \beta \neq 0$. Let us standardize these equations by dividing by $\alpha + \beta$ so that the tack is initially at $(1, 0)$. Now let $\lambda = \frac{\alpha}{\alpha+\beta}$ so that $1 - \lambda = \frac{\beta}{\alpha+\beta}$. Hence the standard *classic* equations for a trochoid are given by

$$(x, y) = (\lambda\cos pt + (1-\lambda)\cos qt, \lambda\sin pt + (1-\lambda)\sin qt), \qquad (1)$$

where p and q are rational numbers, $p \neq -q$, and λ is any real number. Notice that the path defined by (1) is symmetric with respect to the x-axis and that $(x(0), y(0)) = (1, 0)$.

To present an alternative to the wheel model, the mathematical *bush* in which we hope to spot the trochoids, consider the following problem: Two people A and B

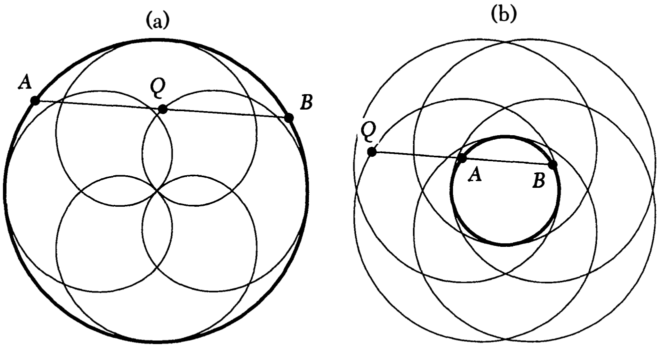


FIGURE 3
Following the tack

walk along holding a taut bungee cord between them. We shall assume that this bungee cord stretches uniformly and can collapse to a point. Embed a glow-stick tack in the bungee cord; a time-lapse photo will trace out a curve as A and B walk along. For example FIGURE 3(a) depicts A and B walking counterclockwise around the unit circle, where A walks five times as fast as B and where the tack is in the middle of the bungee cord; the line in the figure represents the bungee cord at a particular time, while the path within the unit circle is the time-lapse photo of the tack. In general if A makes p circuits of the unit circle while B makes q circuits, where p and q are rational numbers, then at time t, A's position is $(\cos pt, \sin pt)$ and B's position is $(\cos qt, \sin qt)$. Now embed the tack somewhere in the bungee cord between A and B, and let λ be the ratio of the length between the tack and B and of the length between A and B, (where $\lambda = 1$ if A and B have the same position). Then the tack's position at time t is given by $(x, y) = (\lambda \cos pt + (1 - \lambda)\cos qt, \lambda \sin pt + (1 - \lambda)\sin qt)$, where λ is any real number between 0 and 1; this is exactly the trochoid equation we found before.

To encompass the more general case, we envision the two runners as each holding onto a point of a cosmic bungee line (rather than just a cord) which stretches linearly as the distance between the two runners varies. In this way, the tack can be positioned anywhere along the bungee line, not just between the two runners. For example, FIGURE 3(b) gives the trochoid for $p = 5$, $q = 1$, $\lambda = 2$. Here the time-lapse photo of the tack is a trochoid outside the unit circle.

As a concession to simplicity, we do not exclude degeneracies as we did before. One of these corresponds to the two runners remaining motionless at $(1, 0)$ for all time. Hence we say that the standard bungee cord trochoid parameterizations are given by

$$T(p, q, \lambda) = (x, y) = (\lambda \cos pt + (1 - \lambda)\cos qt,\ \lambda \sin pt + (1 - \lambda)\sin qt), \quad (2)$$

where p and q are any rational numbers and λ is any real number. Note that if we allow $q = -p$ and $\lambda = \frac{1}{2}$, then the path generated by (2) is a line segment. The basic difference between (1) and (2) is that (2) allows any ellipse to be considered as a trochoid.

A more stunning way to perceive the trochoids is to imagine that A and B travel about separate unit circle tracks, one a unit distance above the other; if the cords themselves are made of glow-stick material, a time-lapse photo of the cord in motion yields a surface in $\mathbb{R}^3$, and each horizontal cross section is a trochoid. A two-dimensional parameterization of this surface is

$$G(t, \lambda) = (\lambda \cos pt + (1 - \lambda)\cos qt, \lambda \sin pt + (1 - \lambda)\sin qt, \lambda) \quad (3)$$

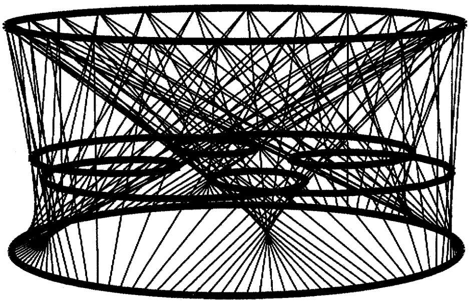


FIGURE 4
Layers of trochoids

where $0 \leq t \leq 2\pi$ and λ is a real number. Note that $G(t, 1)$ and $G(t, 0)$ are the two unit circle tracks on which A and B are walking. FIGURE 4 is a stroboscopic time-lapse photo of this surface wherein a few of the bungee cords are illuminated and the solid curve is the cross section for $\lambda = \frac{9}{16}$.

3. Envelopes

Overhead shots of FIGURE 4 yield striking views, as in FIGURE 5(a, b, c). That is, when "squashed back" into the plane, the bungee cords at any time t appear to be tangent to a curve we shall call C as in FIGURE 5(d). Intuitively, there are two natural ways to view C. From a global perspective, C is where the surface of FIGURE 5 would *fold* upon itself when squashed down into the plane. From a local perspective, (x, y) is a point on C if (x, y) is on the bungee cord at some time t and (x, y) always stays on the *same* side of the bungee cord near time t.

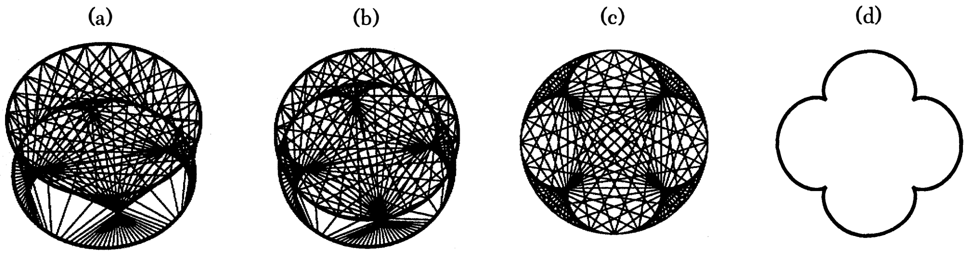


FIGURE 5
Moving overhead

Let's consider the global interpretation first. Observe that if we squash or *project* this surface of FIGURE 5 onto the xy-plane so that each $G(t, \lambda) = (x, y, \lambda)$ is projected to $(x, y, 0)$, then the surface folds upon itself creating creases or curves much like a heap of linen sheets pressed flat onto the floor develops creases. If a point (x, y) is on such a fold then there should be a normal vector to this surface at the point (x, y, t) which is parallel to the xy-plane; that is, the third component of this normal vector should be 0. Recall from vector calculus that a normal to the surface as given by G is obtained by crossing G_t and G_λ of (3), where G_t and G_λ are the gradients of G with respect to t and λ. Now

$$G_t = (-\lambda p \sin pt - (1-\lambda) q \sin qt, \lambda p \cos pt + (1-\lambda) q \cos qt, 0)$$

and

$$G_\lambda = (\cos pt - \cos qt, \sin pt - \sin qt, 1).$$

Therefore the third component of the cross product $G_t \times G_\lambda$ is

$$-\lambda p(\sin^2 pt + \cos^2 pt) + (1-\lambda)q(\sin^2 qt + \cos^2 qt) \\ + (\lambda p - (1-\lambda)q)(\cos pt \cos qt + \sin pt \sin qt).$$

Setting the above expression to 0, using some trigonometric identities, and solving for λ gives

$$\lambda = \frac{q}{p+q}.$$

Substituting this expression into (2) gives the parametric equations for C as

$$(x(t), y(t)) = \left(\frac{p\cos(qt) + q\cos(pt)}{p+q}, \frac{p\sin(qt) + q\sin(pt)}{p+q}\right). \tag{4}$$

Observe that (4) is a very special case of (1); when p and q are both positive then (4) is a parameterization of the epicycloid (i.e., when $s = r$), and when p and q are of opposite signs then (4) is a parameterization of the hypocycloid (i.e., when $r < 0$ and $s = |r|$). See FIGURE 6 for the hypocycloid corresponding to $p = 5$, $q = -1$; note that this curve lies outside the unit circle.

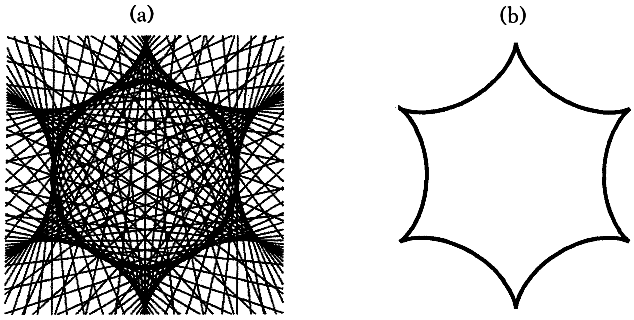


FIGURE 6
A hypocycloid

The classic way to achieve this same result is to use a representation involving the points (x, y) on the bungee cord lines, rather than the convex combination representation of (1) as used above. To obtain this representation, note that a point (x, y) is on the bungee line at time t if and only if

$$\frac{y - \sin pt}{x - \cos pt} = \frac{\sin qt - \sin pt}{\cos qt - \cos pt}.$$

After rearranging terms and using a trigonometric addition identity in the above expression, we observe that if (x, y) is on the bungee cord at time t then $F(x, y, t) = 0$ where

$$F(x, y, t) = x(\sin pt - \sin qt) + y(\cos qt - \cos pt) + \sin(qt - pt). \tag{5}$$

Note that the set of points (x, y, t) for which $F = 0$ is a surface in $\mathbb{R}^3$, as typified by FIGURE 7; observe that for each t-value pictured on the vertical axis, there is exactly one line, except for those times t for which the runners meet, in which case there is an entire plane of lines.

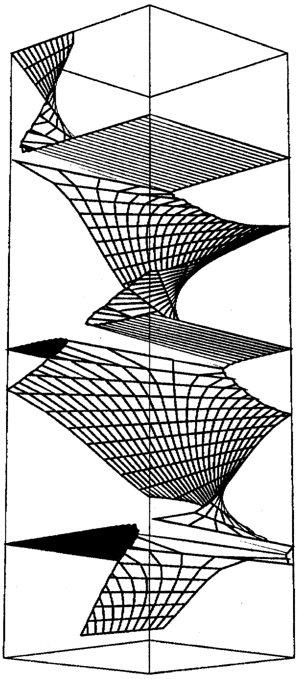


FIGURE 7
The surface $F = 0$ for $p = 2$, $q = 1$

As before, we want to determine the fold set of this surface when it is projected onto the xy-plane. This time, observe that a point (x, y) is on the fold set whenever there is a tangent plane to the surface at (x, y, t) which is perpendicular to the xy-plane. Via vector calculus, this condition translates to $F_t(x, y, t) = 0$, where F_t is the partial derivative of F with respect to t. Hence the fold curve C, classically called the *envelope* of the family of lines $F(x, y, t) = 0$ indexed by t, is obtained by solving the system

$$F(x, y, t) = 0 \quad \text{and} \quad F_t(x, y, t) = 0. \tag{6}$$

After a good deal of straightforward simplification, the parametric solutions to these equations is that given in (4). See [**8**] for details; Example 3 of [**2**] gives the same result using an altogether different approach. See [**3**, p.76] for a more formal presentation of this definition of the envelope.

Now let's consider the local perspective. Observe that if t is not a time at which A and B occupy the same place, then for each such t and for each real number c, $F(x, y, t) = c$ is a line parallel to the line $F(x, y, t) = 0$ in the xy-plane. For such times t, we say that a point (x, y) is *above* the bungee line if $F(x, y, t) > 0$, *below* the line if $F(x, y, t) < 0$ and *on* the line if $F(x, y, t) = 0$. Therefore, to find those points (x, y, t) for which (x, y) is both on the bungee cord at time t and for which (x, y) always stays on the same side of the bungee cord near time t, we need to identify those points (x, y, t) for which both $F(x, y, t) = 0$ and for which $F(x, y, s)$ is either solely nonnegative or solely nonpositive for s-values near t; that is, t is both a root of $F(x, y, s)$ and a local extreme point of $F(x, y, s)$ with respect to s. Such points will be contained in the solution set to the system of equations (6) above. Both the global and local perspectives lead us to solving the same system of equations!

4. Which trochoids are bungee envelopes?

Since the epicycloid and hypocycloid appear ever so nicely as envelopes of the bungee cords, it is natural to ask whether the general trochoid can be perceived as the envelope of a similar family of lines. With this in mind we pose the following problem, where f and g are functions from $\mathbb{R}$ into $\mathbb{R}$.

> **A bungee cord characterization:** Given a curve Γ in the xy-plane, find a function pair (f, g) such that the envelope of the family of lines $\{\mathscr{L}_t | t \in \mathbb{R}\}$ is Γ, where $\mathscr{L}_t$ is the line through $(\cos f(t), \sin f(t))$ and $(\cos g(t), \sin g(t))$.

For example, for the epicycloid and hypocycloid classes, the function pairs are the linear functions (pt, qt) where p and q are different nonzero rational numbers with $p \neq -q$. In general, if A and B proceed at varying rates about the unit circle then we have the following result, where S_h and C_h represent $\sin h(t)$ and $\cos h(t)$, respectively, where h is a function in terms of the variable t.

THEOREM 1: A PARAMETERIZED ENVELOPE. *The envelope formed by the family of lines whose endpoints are* $(\cos f(t), \sin f(t))$ *and* $(\cos g(t), \sin g(t))$, *where* f *and* g *are different differentiable functions, is the parametric set of equations*:

$$(x(t), y(t)) = \left(\frac{f'C_g + g'C_f}{f' + g'}, \frac{f'S_g + g'S_f}{f' + g'} \right). \tag{7}$$

To derive these equations note that a point (x, y) on any of these lines must satisfy the equation

$$y - \sin f(t) = \frac{\sin g(t) - \sin f(t)}{\cos g(t) - \cos f(t)} (x - \cos f(t))$$

for some t. Rearranging terms, we can represent the set of points (x, y, t) satisfying such equations as the zero level set of the function $F(x, y, t) = x(S_f - S_g) + y(C_g - C_f) + S_{g-f}$. Hence $F_t = x(C_f f' - C_g g') + y(S_f f' - S_g g') + C_{g-f}(g' - f')$. After much straightforward simplification, similar to the steps used in solving (6), the solution set to the equations $\{F = 0, F_t = 0\}$ can be written in the desired form.

In terms of the bungee cord characterization problem, we would like to find that function pair (f, g) which yields a given trochoid curve Γ as the envelope. Since these trochoid curves are so strongly related to the epicycloid and hypocycloid, one might guess that finding the bungee cord characterizations for the trochoids ought to be easy. A poor first guess to match a translate Γ of the polar curve $r = \frac{1}{2} + \cos \theta$ appears in FIGURE 8(a) where $f(t) = 2t + 2\sin 2t$ and $g(t) = t$; Γ is the lighter curve and the envelope is the heavier curve. One "near miss" for the bungee cord

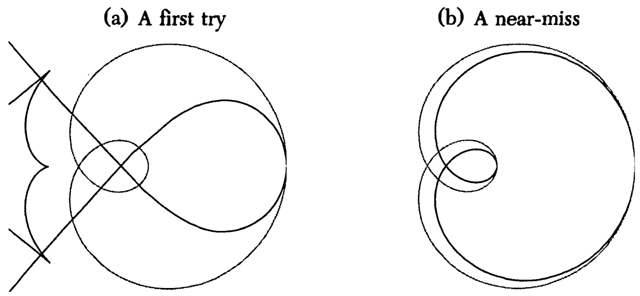


FIGURE 8
Bogus trochoids

characterization of Γ is the function pair $(f, g) = (2t + \frac{\pi}{2}\sin\frac{t}{2}, 2t - \frac{\pi}{2}\sin\frac{t}{2})$; the envelope for this function pair in FIGURE 8(b) surely looks like a trochoid; but with a fair amount of work one can show that it matches no horizontal translate of any polar curve of the form $r = \alpha + \beta\cos\theta$ for any positive numbers α and β. As far as we know, finding the bungee cord characterization for the trochoids remains an open problem.

5. Harmonic envelopes

Although we are unable to produce the general trochoid as an envelope of a family of bungee cord lines held by two runners moving about the unit circle, we can flush some trochoids from their hiding places in the family of lines given in (5) by employing the higher order partial derivatives of F. Observe that if $F = 0$ can be perceived as a surface in $\mathbb{R}^3$, so can $\frac{\partial^u F}{\partial t^u} = 0$. In fact, such surfaces will look very much like FIGURE 7. The intersection of any two of these surfaces will then be some path in $\mathbb{R}^3$. With this idea in mind, we define $\mathscr{F}_u$, called the u^{th} *harmonic* envelope, as the projection of the intersection of the surface $F = 0$ and the surface $\frac{\partial^u F}{\partial t^u} = 0$ where u is any positive integer. These harmonic envelopes suggest some natural conjectures. That is, from FIGURE 9(2), the curve $\mathscr{F}_2$ looks like a trochoid; and sifting through the equations yields the following result, where we use the word "harmonic" in the sense of obtaining similar patterns at "deeper" levels.

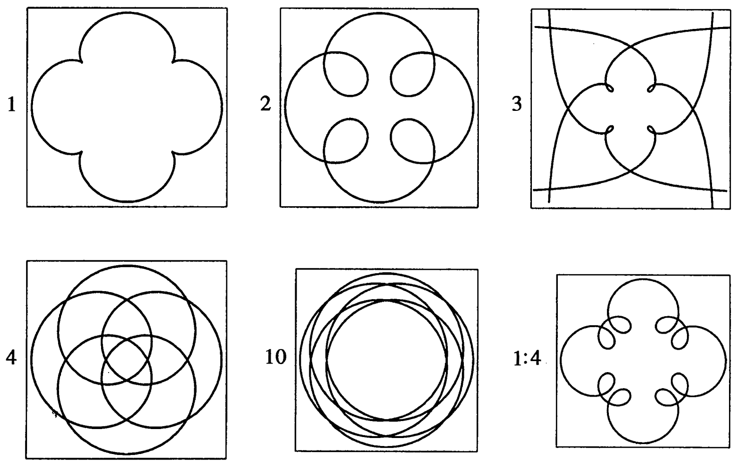


FIGURE 9
Harmonic envelopes for $p = 5$, $q = 1$

THEOREM 2: HARMONIC ENVELOPES. *Let n be a positive integer. Let F be the function in* (5). *The curve $\mathscr{F}_{2n}$ is a trochoid whose parametric representation is*

$$(x, y) = \left(\frac{\left(p^{2n} - (q-p)^{2n}\right)C_p + \left((q-p)^{2n} - q^{2n}\right)C_q}{p^{2n} - q^{2n}}, \right.$$

$$\left. \frac{\left(p^{2n} - (q-p)^{2n}\right)S_p + \left((q-p)^{2n} - q^{2n}\right)S_q}{p^{2n} - q^{2n}} \right). \qquad (8)$$

To generate (8) note that $\frac{\partial^{2n}F}{\partial t^{2n}} = 0$ is equivalent to the following:

$$x\left(p^{2n}S_p - q^{2n}S_q\right) + y\left(q^{2n}C_q - p^{2n}C_p\right) + (q-p)^{2n}S_{q-p} = 0.$$

From $F = 0$, we have $x = \frac{S_{p-q} + y(C_p - C_q)}{S_p - S_q}$. Substituting this expression into $F_t = 0$ along with much straightforward simplification, we have the desired equation for y. The formula for x is similarly derived. Then observe that (8) is a special case of (1).

FIGURES 9(4) and 9(10) are the trochoids $\mathscr{F}_4$ and $\mathscr{F}_{10}$. The odd orders fail to yield trochoids however; for example Figure 9(3) gives $\mathscr{F}_3$. Experimenting with other combinations gives interesting curves as well; for example, Figure 9(1 : 4) is an exotic curve which is the projection of the intersection of the surfaces $F_t = 0$ and $F_{tttt} = 0$.

6. Interpreting the second harmonic

To understand the significance of the second harmonic as a trochoid, let us recall from linear algebra the formula for the distance between a point $X = (x, y)$ and a line $\mathscr{L}$ through the points A and B. The projection vector P of the vector AX onto $\mathscr{L}$ is

$$P = \frac{(X-A)\cdot(B-A)}{(B-A)\cdot(B-A)}(B-A).$$

Hence the Euclidean distance from X to $\mathscr{L}$ is the square root of the dot product of $X - A - P$ with itself, which upon simplification is

$$H(x,y,t) = \sqrt{(X-A)\cdot(X-A) - \frac{((X-A)\cdot(B-A))^2}{(B-A)\cdot(B-A)}}\,, \tag{9}$$

where we identify A and B with the position of the two runners, namely, $A = (\cos pt, \sin pt)$ and $B = (\cos qt, \sin qt)$. We see that in some ways F is a shadow of H in the sense that, except when the two runners A and B are near the same location, when H is far from 0 then so is F. That is, for almost all t-values, the solution set in the xy-plane for $F(x, y, t) = 0$ is the same as that for $H(x, y, t) = 0$. The reason for the disparity is that when the runners meet at time t, the solution set to $F(x, y, t) = 0$ is the entire xy-plane whereas the solution set to $H(x, y, t) = 0$ is the empty set. FIGURE 10 shows the typical relation between F and H; in particular, it compares the functions $F(0, .85, t)$ and $H(0, .85, t)$ (as a lighter curve) for $p = 2$, $q = 1$; note that the roots of the two curves agree except at the time $t = 0$ and $t = 2\pi$, the place where the runners meet.

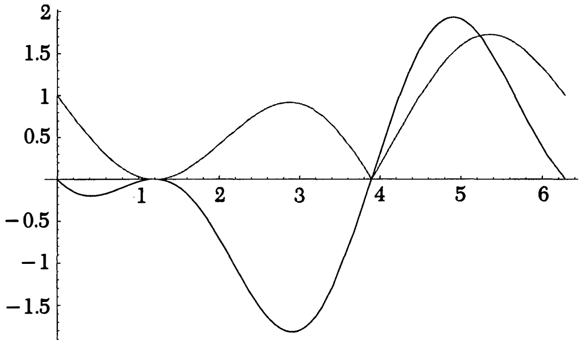


FIGURE 10
Distance functions of (x, y) to the bungee cord

Thus if we interpret F as a "signed distance function" of the point (x, y) from the bungee cord at time t, the solution set of the condition $\frac{\partial^2 F}{\partial t^2} = F_{tt} = 0$ includes the points at which $F(x, y, t)$ is changing most rapidly with respect to t; that is, intuitively, this solution set to $F = 0 = F_{tt}$ includes those points for which the bungee cords are sweeping over the point (x, y) at time t most rapidly. To use a baseball analogy, if we think of the point (x, y) as a baseball, the cord $F(x, y, t) = 0$ at time t as a baseball bat, and $F_t(x, y, t)$ as the speed of a baseball bat at time t, then the condition $F = 0 = F_{tt}$ translates into hitting only those baseballs which can be hit as hard as is possible; that is, intuitively, the second harmonic are those places of optimal impact. To summarize more simply, with respect to time, the first harmonic is basically the set of extremes as well as the zeroes of F, whereas the second harmonic is basically the set of inflections as well as the zeroes of F.

7. Some geometry of the higher harmonics

The intuitive nature of the higher harmonic trochoids is rather unclear to this writer. However, besides having shown that a purely mechanical manipulation of the derivative gives the observation that the even orders yield trochoids and the odd orders greater than one fail to do so, we can give some geometrical insight into the warping and rotating influence by the u^{th} derivative acting on the original family of lines. Doing so involves fractional derivatives, unfamiliar to many, which we explain below.

For any real number u, let the curve $\mathscr{F}_u$, the u^{th} *harmonic* envelope, be defined as the x-y projection of the solution set of $\{F = 0, \frac{\partial^u F}{\partial t^u} = 0\}$. In particular, with this operator it is helpful to see how the epicycloid $\mathscr{F}_1$ in Figure 9(1) *morphs* into the trochoid $\mathscr{F}_2$ of Figure 9(2) as u increases from 1 to 2. As will be shown, the derivative metamorphosis of $\mathscr{F}_1$ becoming $\mathscr{F}_2$ encounters a stage $\mathscr{F}_{1.35}$ which is similar in kind to the stage $\mathscr{F}_3$ in the metamorphosis of $\mathscr{F}_2$ into $\mathscr{F}_4$.

The fractional derivative (see **[6]** and **[7]**) dates back to 1867. Integral to its definition is the *gamma* function: $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$, where $\alpha > 0$; extend this function to all real numbers by using the recursion relation $\Gamma(\alpha) = \frac{\Gamma(\alpha+1)}{\alpha}$ for all α. See Figure 11(a). Then the somewhat intimidating definition due to Grünwald (1867) of the u^{th} *fractional derivative* expanded about a where u and a are any real numbers is

$$\frac{d^u f(x)}{[d(x-a)]^u} = \lim_{N\to\infty} \left\{ \frac{\left[\frac{x-a}{N}\right]^{-u}}{\Gamma(-u)} \sum_{j=0}^{N-1} \frac{\Gamma(j-u)}{\Gamma(j+1)} f\left(x - j\frac{x-a}{N}\right) \right\}. \tag{10}$$

With a few appropriate tricks, we can find $\mathscr{F}_u$. First of all, we take $a = 0$. Two familiar properties of the ordinary derivative operator remain, namely, for any functions f and

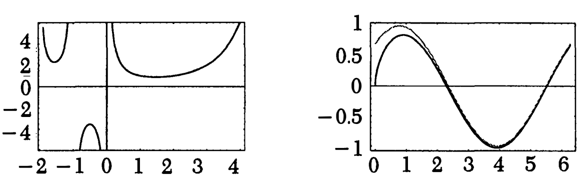


FIGURE 11
The Γ function and the semiderivative

g and constants α and β we have

$$\textbf{Linearity}: \quad \frac{d^u(\alpha f(x)+\beta g(x))}{dx^u} = \alpha\frac{d^u f(x)}{dx^u} + \beta\frac{d^u g(x)}{dx^u}. \tag{11}$$

$$\textbf{A limited chain rule}: \quad \frac{d^u f(\alpha x)}{dx^u} = \alpha^u \frac{d^u f(\alpha x)}{d(\alpha x)^u} \quad \text{where } \alpha > 0. \tag{12}$$

Some familiar properties of the ordinary derivative are hopelessly absent. In particular, it turns out that the u^{th} derivative of a periodic function fails in general to be periodic. To demonstrate this phenomenon, let's consider the *semi-derivative*, a special case of the general definition as presented by Courant and Hilbert [**6**, p. 50]; applying their definition to the sine function gives

$$\frac{d^{\frac{1}{2}}\sin x}{dx^{\frac{1}{2}}} = \sqrt{2}\int_0^{\sqrt{\frac{2x}{\pi}}} \cos\left(x - \frac{\pi}{2}t^2\right)dt. \tag{13}$$

Fresnel integrals of this kind are found in the standard reference [**1**, p. 300]. The light curve of FIGURE 11(b) depicts the graph of the semiderivative of $\sin x$ on $[0, 2\pi]$. As x increases, the semi-derivative of $\sin x$ converges to $\sin(x+\frac{\pi}{4})$, the heavier curve of FIGURE 11(b). And in general for large x it can be shown [**6**, p. 110] that

$$\frac{d^u \sin x}{dx^u} = \sin\left(x + \frac{\pi}{2}u\right) + \frac{x^{-1-u}}{\Gamma(-u)} - \frac{x^{-3-u}}{\Gamma(-u-2)} + \cdots$$

and

$$\frac{d^u \cos x}{dx^u} = \cos\left(x + \frac{\pi}{2}u\right) + \frac{x^{-2-u}}{\Gamma(-u-1)} - \frac{x^{-4-u}}{\Gamma(-u-3)} + \cdots.$$

Thus, for $u > 0$ and for x sufficiently large,

$$\frac{d^u \sin x}{dx^u} \approx \sin\left(x + \frac{\pi}{2}u\right) \quad \text{and} \quad \frac{d^u \cos x}{dx^u} \approx \cos\left(x + \frac{\pi}{2}u\right). \tag{14}$$

By (11), (12) and (14), for $0 < p \le q$, $\frac{\partial^u F}{\partial t^u} = 0$ is asymptotically equal to

$$x\left(p^u \sin\left(pt + \frac{\pi u}{2}\right) - q^u \sin\left(qt + \frac{\pi u}{2}\right)\right) + y\left(q^u \cos\left(qt + \frac{\pi u}{2}\right) - p^u \cos\left(pt + \frac{\pi u}{2}\right)\right)$$
$$+ (q-p)^u \sin\left((q-p)t + \frac{\pi u}{2}\right) = 0. \tag{15}$$

If $p > q$, then the last term on the left in (15) can be replaced with

$$-(p-q)^u \sin\left((p-q)t + \frac{\pi u}{2}\right).$$

Similarly, the reader can adapt (15) appropriately if either p or q is negative by taking advantage of the odd and even property of the sine and cosine functions, respectively. For each u, the family of lines $\frac{\partial^u F}{\partial t^u} = 0$ parameterized by t has an envelope. FIGURE 12, where $p = 4$ and $q = 1$, illustrates this phenomenon for the u-values $\{1, 1.3, 1.6, 1.9\}$, suggesting that the u^{th} partial derivative acting on the family of lines as given by $F = 0$ tends to constrict and rotate them into another family of lines in a somewhat exotic fashion. Let's define $\mathscr{G}_u$ to be the projection onto the xy-plane of the intersection of the surface $F = 0$ with the surface given by (15). Note that as t gets large, $\mathscr{F}_u$ approaches $\mathscr{G}_u$. Thus we can view $\mathscr{F}_u$ as being asymptotically periodic.

The graphs of these asymptotic harmonic envelopes as in FIGURE 13, show how the epicycloid of FIGURE 9(1) changes with the fractional derivative into the trochoid of FIGURE 9(2). To describe this transformation, we imagine a surgical team of four simultaneously snipping FIGURE 9(1) at the four points furthest from the origin, so

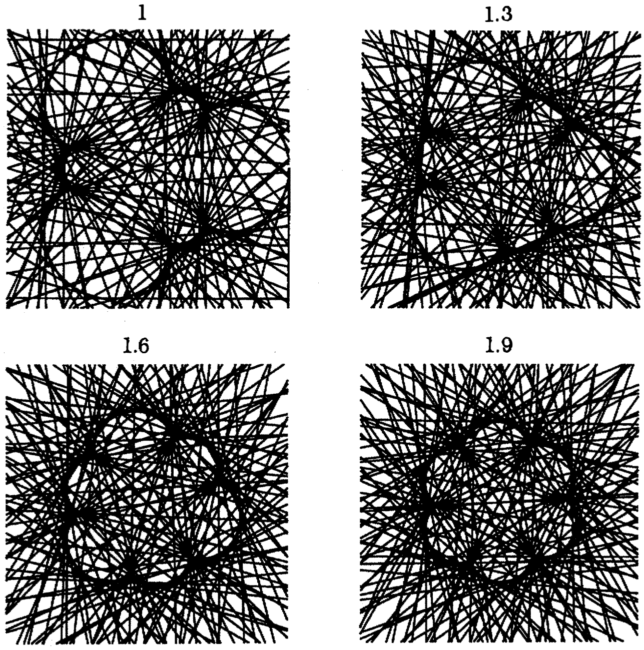


FIGURE 12

Envelopes of the family of lines $\frac{\partial^u F}{\partial t^u} = 0$ for $p = 4$, $q = 1$

creating eight thrashing strands which explode outwards to the point at infinity somewhat like the severed ends of a piano string erupting under great tension; each surgeon grapples with the two strands he or she created, wrestling their communal patient clockwise, causing the curve to cross at its cusps so forming loops, as they push the strands back into themselves, and ultimately meld the ends at the four points nearest the origin. FIGURE 14 shows how the trochoid emerges from "nothing" into an epicycloid by examining the fractional derivative homotopy as u ranges from 0 to 1.

8. A generalization and some questions

Just as the wheel model of the trochoid generalizes to multiple wheels in **[4]**, so the bungee cord model generalizes. That is, if λ_i are real numbers with $\Sigma_{i=1}^{n} \lambda_i = 1$ and a_i

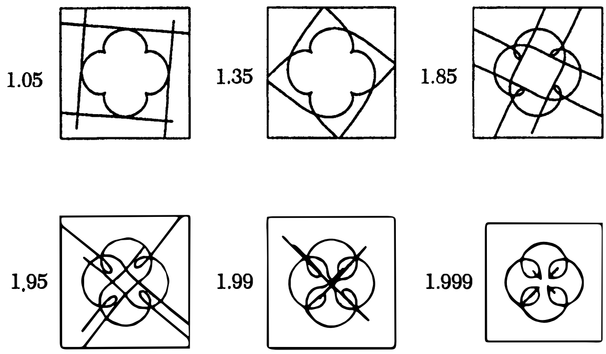


FIGURE 13

An epicycloid morphing into the trochoid $\mathscr{F}_2$

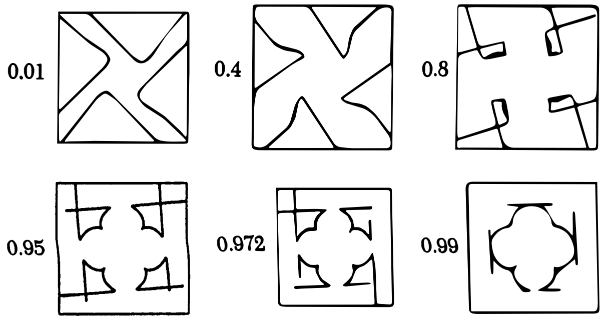


FIGURE 14
An epicycloid emerging from nothing

are rational numbers then any curve of the form

$$\left(\sum_{i=1}^{n}\lambda_i\cos(a_it),\ \sum_{i=1}^{n}\lambda_i\sin(a_it)\right)$$

is a generalized trochoid. Such a generalized trochoid is generated by following the path of a tack in a stretchable rubber sheet whose corners are held by n people where person i moves around the unit circle with speed a_i. Actually, we view the entire xy-plane as a cosmic rubber sheet which can fold in upon itself, wherein the tack moves around in the plane as a fixed linear combination of the position of the runners. FIGURE 15 depicts the path of a tack in a triangular sheet whose corners are positioned at $(\cos t, \sin t)$, $(\cos 5t, \sin 5t)$ and $(\cos 17t, \sin 17t)$ and $\lambda_1 = \frac{1}{2}$, $\lambda_2 = \frac{1}{3}$ and $\lambda_3 = \frac{1}{6}$, respectively. Alternatively, we could think of the tack as being the hub of n bungee cords, each one of different elasticity, whose separate ends are held by the n runners.

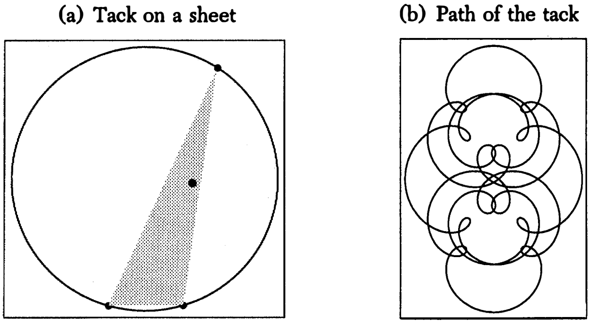


FIGURE 15
A "three-person" trochoid

Finally, we pose some problems for the reader.

1. The ellipses $(x, y) = (\cos t, a\sin t)$ are the trochoids $T\left(1, -1, \frac{a+1}{2}\right)$, where a is any number.
2. The polar flowers $r = \cos\left(\frac{m}{n}\theta\right)$, where m and n are integers, are the trochoids $T\left(n+m, n-m, \frac{1}{2}\right)$.
3. The polar cardioid $r = 1 + a\cos\theta$, where a is any positive number, is the translated trochoid $\frac{a+2}{2}T\left(1, 2, \frac{a}{a+2}\right) + \left(\frac{a}{2}, 0\right)$.
4. The epicycloid and hypocycloid of p cusps are $T\left(p+1, 1, \frac{1}{p+2}\right)$ for $p \geq 0$ and $T\left(p-1, -1, \frac{1}{2-p}\right)$ for $p \geq 3$, respectively, for $p \in \mathbb{Z}$.
5. Some trochoids of p non-intersecting "inner" and "outer" loops are $T\left(p+1, 1, \frac{1+2p}{p^2+2p}\right)$ and $T\left(p-1, -1, \frac{1-2p}{p^2-2p}\right)$ for $p \geq 3$ with $p \in \mathbb{Z}$.

6. Find a bungee cord characterization of the general trochoid. Find other familiar or unusual curves which can be generated from elementary function pairs (f, g).
7. Classify the curves in the xy-plane as given by the solution set to the system $\frac{\partial^u F}{\partial t^u} = \frac{\partial^v F}{\partial t^v}$, where u and v are positive integers. Is the case for $u = 1$ and $v = 4$ as given in FIGURE 9 the most exotic of these curves?
8. Each family of lines given by $\frac{\partial^u F}{\partial t^u} = 0$ as indexed by t has an envelope where u is a real number. See FIGURE 12. Some of these are epicycloids and hypocycloids. How are these envelopes related (as u changes)?
9. Let $\mathscr{H}_1$ be the projection of the intersection of the surfaces $H(x, y, t) = 0$ and $H_t(x, y, t) = 0$, where H is as given in (9). Note that both $F = 0$ and $H = 0$ are simply different representations of the same family of lines, and since the envelope of a family of lines should intuitively be independent of the representation, verify that the first harmonic $\mathscr{F}_1$ and $\mathscr{H}_1$ are the same almost everywhere. The reader should be forewarned that the computer-generated parametric solution for $\mathscr{H}_1$ runs on for numerous pages. Contrast the second harmonics as well, that is, $\mathscr{F}_2$ and $\mathscr{H}_2$.
   As an altogether different problem with respect to these harmonics, try this elusive question: does $\lim_{u \to 0} \mathscr{F}_u$ exist, and if so, is it a path? See FIGURE 14 (0.01).
10. In this paper we allowed two runners to proceed about the unit circle connected by a bungee cord. Now allow them to proceed about any trochoid. In particular, try the polar curve $r = \cos p\theta$, following Maurer **[5]**, where p is a positive integer with runners A and B positioned on the track at the polar angles $\theta + \delta$ and θ, respectively, for a fixed δ, while θ ranges from 0 to 2π, so that the radial positions of A and B are $\cos p(\theta + \delta)$ and $\cos p\theta$, respectively. What kinds of paths are traced out by tacks embedded in such bungee cords? FIGURE 16 shows the cords and the envelope corresponding to $p = 5$ and $\delta = \frac{97\pi}{180}$. Are any of these familiar curves?

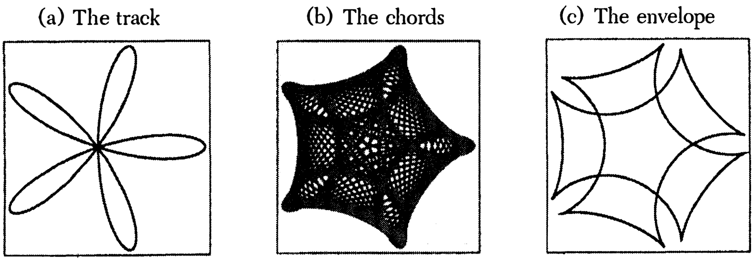


FIGURE 16
An envelope for Maurer

REFERENCES

1. Milton Abramowitz and Irene A. Stegun, editors, *Handbook of Mathematical Functions*, Dover, New York, NY, 1968.
→ Irl Bivens and Andrew Simoson, Beholding a rotating beacon, this MAGAZINE 71 (1998), 83–104.
3. J. W. Bruce and P. J. Giblin, *Curves and Singularities*, Cambridge University Press, Cambridge, UK, 1984.
→ Frank A. Farris, Wheels on wheels on wheels—surprising symmetry, this MAGAZINE 69 (1996), 185–189.
5. Peter M. Maurer, A rose is a rose..., *Amer. Math. Monthly* 94 (1987), 631–645.
6. Keith B. Oldham and Jerome Spanier, *The Fractional Calculus*, Academic Press, New York, NY, 1974.
7. B. Ross, A brief history and exposition of the fractional calculus, *Lecture Notes in Mathematics* vol. 457, Springer-Verlag, New York, NY, 1975, 1–36.
8. Andrew Simoson, An envelope for a spirograph, *College Math. J.* 28:2 (1997), 134–139.

The Possibility of Impossible Pyramids

TOM SIBLEY
St. John's University
Collegeville, MN 56321

Introduction

When can we form a triangle or a pyramid from edges with given lengths? Euclid found that for three segments to be the sides of a triangle, it is necessary and sufficient for the length of each segment to be shorter than the sum of the lengths of the other two. Thus, for any three lengths that could potentially be sides of a triangle, there actually is a Euclidean triangle whose sides have those lengths. In short, Euclidean plane geometry is "triangle complete," as defined and shown analytically below. It seems natural to expect that Euclidean space is "pyramid complete"; that is, given six lengths with appropriate restrictions we can find a Euclidean pyramid whose edges have those lengths. But could there be "impossible pyramids"? Might there be lengths for the sides of a pyramid that form four triangular faces even though no Euclidean pyramid exists with those lengths?

As part of this investigation we shall describe a family of metric spaces which includes that of Euclidean geometry.

DEFINITIONS. *A metric space* (X, d) *is a non-empty set* X *together with a metric (or distance)* d *such that for any elements* P, Q, $R \in X$,

(i) $d(P,Q) \geq 0$;
(ii) $d(P,Q) = d(Q,P)$;
(iii) $d(P,Q) = 0 \Leftrightarrow P = Q$;
(iv) $d(P,Q) + d(Q,R) \geq d(P,R)$.

Property (iv) is the *triangle inequality*. A metric space is *triangle complete* if, for any three non-negative real numbers a, b, and c satisfying the triangle inequality in any order (i.e., $a \leq b + c$, $b \leq a + c$, and $c \leq a + b$), there are three points P, Q, and R in X such that $d(P,Q) = a$, $d(P,R) = b$, and $d(Q,R) = c$.

Let's show that Euclidean plane geometry (i.e., $\mathbb{R}^2$ with the usual Euclidean metric) is triangle complete. Given three non-negative reals a, b, and c satisfying the three triangle inequalities, let $P = (0,0)$ and $Q = (a,0)$. We need to find a point $R = (x,y)$ such that $d(P,R) = b$ and $d(Q,R) = c$ or, equivalently, $x^2 + y^2 = b^2$ and $(a-x)^2 + y^2 = c^2$. These equations' solutions are $x = (a^2 + b^2 - c^2)/2a$ and $y = \pm\sqrt{b^2 - x^2} = \pm\sqrt{(b-x)(b+x)}$. We will establish triangle completeness once we show that $b^2 - x^2 \geq 0$. Now $b + x = ((a+b)^2 - c^2)/2a \geq 0$ because $a + b \geq c$. The two other triangle inequalities show that $|a - b| \leq c$ and so $b - x = (c^2 - (a-b)^2)/2a \geq 0$.

The hyperbolic plane is also triangle complete since Euclid's Proposition 22 from Book I holds in hyperbolic geometry. However, not every metric space is triangle complete. For instance, a sphere in $\mathbb{R}^3$ of radius r with distances measured along its surface cannot be triangle complete because we cannot find points arbitrarily far apart on a sphere. In fact, even if we restrict a, b, and c to distances that occur on the sphere and satisfy the triangle inequality, they still need not determine a triangle whose sides are arcs of great circles. For example, first consider an equilateral triangle on a sphere whose three vertices lie on the equator. Note that no equilateral triangle on the sphere can have longer sides. Let $d = 2\pi r/3$ be the distance between two of these points and let a, b, and c be equal lengths slightly larger than d. Then there is no spherical triangle with sides a, b, and c.

Euclidean pyramids

Let's turn to the problem of constructing a pyramid in three-dimensional Euclidean space from six given lengths. First of all, appropriate triples of lengths must form four triangles to make the faces of the pyramid, say $\triangle PQR$, $\triangle PQS$, $\triangle PRS$, and $\triangle QRS$. Thus the triangle inequalities must hold for these triples. We want to glue the corresponding edges of the triangles together to make a pyramid. We formalize this intuitive idea as follows:

DEFINITION. *A metric space* (X, d) *is* pyramid complete *if, for any sextuple of non-negative real numbers* (a, b, c, d, e, f) *such that each of the triples* $\{a, b, c\}$, $\{a, d, e\}$, $\{b, d, f\}$, *and* $\{c, e, f\}$ *satisfies the triangle inequalities in any order, there are four points* P, Q, R, *and* S *in* X *so that* $d(P, Q) = a$, $d(P, R) = b$, $d(Q, R) = c$, $d(P, S) = d$, $d(Q, S) = e$, *and* $d(R, S) = f$.

(See FIGURE 1; the triple $\{a, b, c\}$ corresponds to $\triangle PQR$.)

Is Euclidean space pyramid complete? The following example shows that the answer is no.

Example 1. Consider the sextuple $(14, 8, 8, 8, 8, 8)$. We need $d(P, Q) = 14$, and all the others to be 8. We can readily make $d(P, R) = d(P, S) = d(Q, R) = d(Q, S) = 8$ as well, but we will see that these distances restrict the possible values of $d(R, S)$. If we fix P, Q, and R and rotate triangle $\triangle PSQ$ around the side $\overline{PQ}$, we get the largest distance between R and S when all four points are in the same plane; see FIGURE 2. For M, the midpoint of $\overline{PQ}$, the Pythagorean theorem applied to $\triangle PMS$ and $\triangle PMR$ gives $d(S, M) = d(R, M) = \sqrt{d(P, R)^2 - d(P, M)^2} = \sqrt{8^2 - 7^2} \approx 3.873$. Thus, no matter how we rotate $\triangle PSQ$, $d(R, S) \le d(R, M) + d(M, S) < 7.75$, which is less than the required 8.

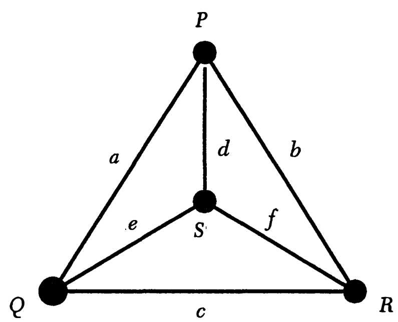


FIGURE 1

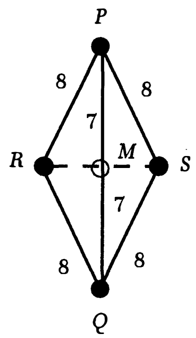


FIGURE 2

Remarks A scaled down version of Example 1 shows that no 3-dimensional space that is "locally Euclidean" (technically a 3-dimensional manifold), such as hyperbolic space, can be pyramid complete. Several mathematicians, starting with Niccolo Tartaglia in 1560, have published versions of the following formula relating the volume V of a pyramid to the lengths of its six sides:

$$V^2 = \tfrac{1}{6}\big(a^2f^2(b^2+c^2+e^2+d^2) + b^2e^2(a^2+c^2+f^2+d^2) + c^2d^2(a^2+b^2+f^2+e^2) \\ - a^4f^2 - a^2f^4 - b^4e^2 - b^2e^4 - c^4d^2 - c^2d^4 \\ - a^2b^2c^2 - a^2e^2d^2 - b^2f^2d^2 - c^2f^2e^2\big).$$

The triangle inequalities for the faces together with the requirement that V^2 non-negative give necessary and sufficient conditions on the lengths for a Euclidean pyramid to exist (see **[4]**).

The taxicab metric

Are any metric spaces pyramid complete? Yes! We show that $\mathbb{R}^3$ with the taxicab metric is pyramid complete. The taxicab metric provides an easily-explored geometry, often called taxicab geometry. (See, e.g., **[2]**, **[3]**, **[5]**, and **[7]**.) The "taxicab" name comes, in $\mathbb{R}^2$, from the distances travelled by cars if all streets run either North-South or East-West. (See FIGURE 3.) Cars that stay on the roads cannot benefit from the Pythagorean theorem, so the total distance between any two points is the sum of the distances in the principal directions. We consider the taxicab metric only in $\mathbb{R}^3$, although it is readily defined in $\mathbb{R}^n$.

DEFINITION. *For two points* (x_1, y_1, z_1) *and* (x_2, y_2, z_2) *in* $\mathbb{R}^3$,

$$d_T((x_1, y_1, z_1), (x_2, y_2, z_2)) = |x_2 - x_1| + |y_2 - y_1| + |z_2 - z_1|.$$

See, e.g., [**6**, pp. 214–219] for a proof that $\mathbb{R}^3$ with the taxicab metric d_T is a metric space. Moreover, we can show that $\mathbb{R}^3$ with the taxicab metric is triangle complete. Let a, b, and c satisfy the three triangle inequalities and, without loss of generality, assume that a is the largest. Let $P = (0,0,0)$, $Q = (a,0,0)$, and $R = (x, y, 0)$, where $x = (a+b-c)/2$ and $y = (b+c-a)/2$. From the triangle inequalities $0 \le x$ and $0 \le y$. It is easy to check that $d_T(P,R) = x + y = b$ and $d_T(Q,R) = (a-x) + y = c$, as in FIGURE 4. Note that this argument makes no use of the third coordinate, so the "taxicab plane" is also triangle complete.

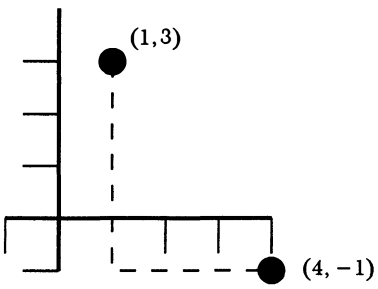


FIGURE 3
In the taxicab metric
$d((1,3),(4,-1)) = 3 + 4 = 7$

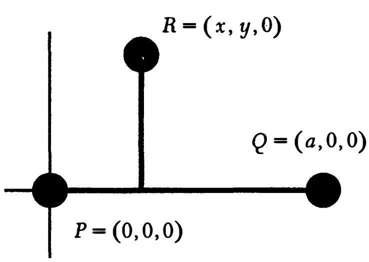


FIGURE 4

Let's reconsider the six lengths from Example 1 using the taxicab metric. Although no Euclidean pyramid exists with these edge lengths, the following example shows that there is such a "taxicab" pyramid in $\mathbb{R}^3$.

EXAMPLE 2. Let $P = (0, 0, 0)$, $Q = (7, 7, 0)$, $R = (7, 0, 1)$, and $S = (3, 4, 1)$ as in FIGURE 5. Then $d_T(P, Q) = 14$, $d_T(P, R) = 8$, $d_T(Q, R) = 8$, $d_T(P, S) = 8$, $d_T(Q, S) = 8$, and $d_T(R, S) = 8$.

FIGURE 5 suggests a general approach to proving that taxicab geometry is pyramid complete. Note that R is above one corner of the rectangle in the xy-plane with opposite corners P and Q, and that the amount R is raised above that rectangle depends on how much the sum of the distances $d_T(P, R)$ and $d_T(Q, R)$ exceeds $d_T(P, Q)$ in triangle $\triangle PQR$. That is, $d_T(P, R) + d_T(Q, R) - d_T(P, Q) = 2$, twice the height of R above the plane of that rectangle. If we were given just the first five distances $d_T(P, Q) = 14$, $d_T(P, R) = 8$, $d_T(P, S) = 8$, $d_T(Q, R) = 8$, and $d_T(Q, S) = 8$, the triangle inequalities for triangle $\triangle PRS$ (or $\triangle QRS$) limit $d_T(R, S)$ to any number from 0 to 16. Thus pyramid completeness requires us to obtain pyramids for all of the values for $d_T(R, S)$ from 0 to 16. In FIGURE 5 we can obtain the values between 0 and 14 by sliding S along the diagonal between $R = S'$ and S''. (The special conditions of this example force R and S' to be the same point.) To see this note that all points $(x, y, 0)$ on the line $x + y = 7$ with $0 \le x,\ y \le 7$ are a distance of 7 from both P and Q. Hence the points $(x, y, 1)$ will be a distance of 8 from both P and Q. However, to stretch the distance $d_T(R, S)$ beyond 14, we need a different tactic. Note that the two points labeled S^* are the maximum distance of 16 from R and that the segments connecting S'' to them provide a way to vary $d_T(R, S)$ continuously from 14 to the maximum distance 16 while keeping S the correct distance from P and Q.

THEOREM 1. *Taxicab geometry on* $\mathbb{R}^3$ *is pyramid complete.*

Proof. Suppose that we are given a sextuple of non-negative numbers (a, b, c, d, e, f) such that all triangle inequalities hold on the following triples: $\{a, b, c\}$, $\{a, d, e\}$, $\{b, d, f\}$, and $\{c, e, f\}$. (Refer to FIGURE 1 for the relationship between these lengths and the points P, Q, R, and S.) For ease, we assume $b + c \ge d + e$. (If $b + c < d + e$, we switch R and S in what follows.) We follow Example 2, choosing $P = (0, 0, 0)$, $Q = (t, u, 0)$, and $R = (t, 0, v)$, where $d_T(P, Q) = a$, $d_T(P, R) = b$, and $d_T(Q, R) = c$. (See FIGURE 6.)

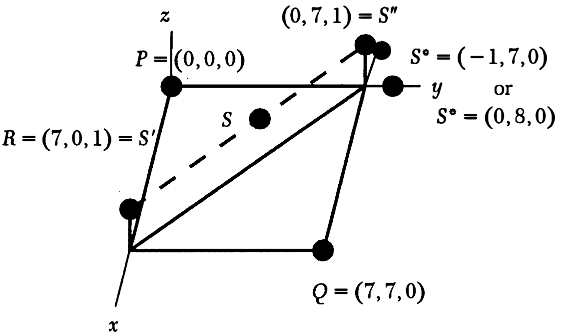


FIGURE 5
S = (3,4,1)

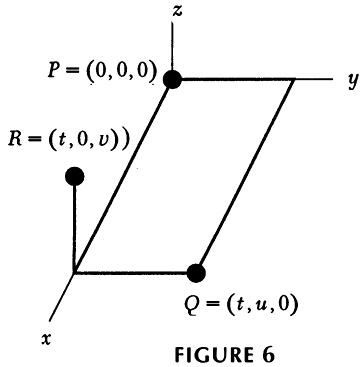


FIGURE 6

From the definition of the taxicab metric we get the equations

$$t + u = a, \quad t + v = b, \quad \text{and} \quad u + v = c.$$

The solutions

$$t = .5(a + b - c), \quad u = .5(a + c - b), \quad \text{and} \quad v = .5(b + c - a)$$

are all non-negative because of the triangle inequalities satisfied by a, b, and c.

Next we place S. We could solve three additional equations for S with various cases, but a geometric approach provides more insight. As in the discussion following Example 2 we treat $d_T(R, S)$ as a variable. We use the fourth and fifth numbers, d and e, to determine the range of possible locations for S relative to points P and Q, which is the range of values for $d_T(R, S)$. We will show that this range includes f. Figures 7 through 10 illustrate our strategy. The dashed lines in FIGURES 7, 8, and 9 describe the possibilities for S when $b \geq c$. If $b < c$, only FIGURE 8 changes significantly, as shown in FIGURE 10. In all cases, we will show that S' is the candidate for S as close to R as possible and S^* is the candidate for S as far from R as possible.

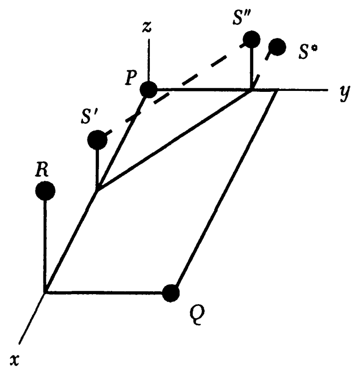


FIGURE 7

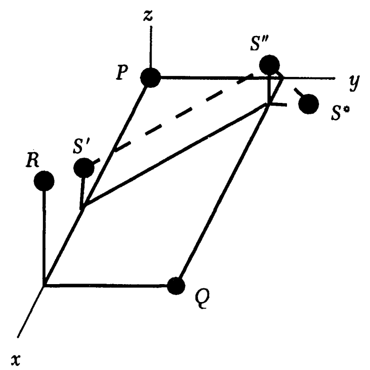


FIGURE 8

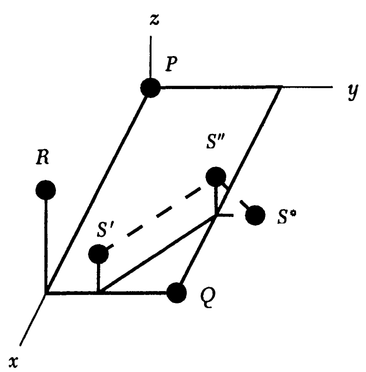


FIGURE 9

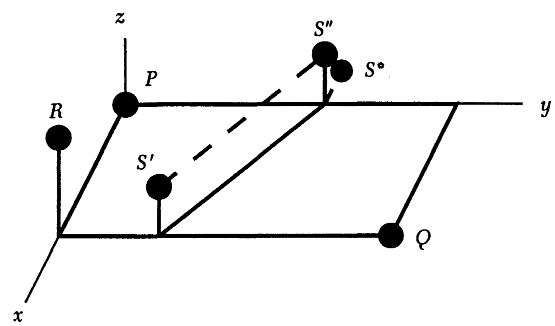


FIGURE 10

We consider now the possibilities for S, showing that these points satisfy $d = d_T(P, S)$ and $e = d_T(Q, S)$. Let $k = 0.5(d + e - a)$. Because a, d, and e satisfy the triangle inequality in any order, we see that $k \geq 0$. As in Example 2, k will be the height above the xy-plane of the points on the dashed line between S' and S''. Consider the points $X = (x, y, 0)$ satisfying $0 \leq x,\ y \leq d - k = 0.5(a + d - e)$ and $x + y = d - k$. These points X form the solid diagonal line. Then $d_T(P, X) = d - k$ and $d_T(Q, X) = e - k = 0.5(a + e - d)$, since the triangle inequalities for a, d, and e ensure that both $d - k \geq 0$ and $e - k \geq 0$. Given $X = (x, y, 0)$ as above, the points $S = (x, y, k)$ satisfy $d_T(P, S) = d - k + k = d$ and $d_T(Q, S) = e$. The tables below give the coordinates for S', S'', S^*, and S between S'' and S^* for the cases corresponding to Figure 7 through 10, which depend on how $d - k$ compares with t and u. For the column for S, $0 \leq w \leq k$.

Figure	$d - k$	S'	S''
7	$d - k \leq t, u$	$(d - k, 0, k)$	$(0, d - k, k)$
8	$u < d - k \leq t$	$(d - k, 0, k)$	$(d - k - u, u, k)$
9	$t, u < d - k$	$(t, d - k - t, k)$	$(d - k - u, u, k)$
10	$t < d - k \leq u$	$(t, d - k - t, k)$	$(0, d - k, k)$

Figure	S^*	*S between S'' and S^**
7	$(-k, d - k. 0)$	$(-w, d - k, k - w)$
8	$(d - k - u, u + k, 0)$	$(d - k - u, u + w, k - w)$
9	$(d - k - u, u + k, 0)$	$(d - k - u, u + w, k - w)$
10	$(-k, d - k, 0)$	$(-w, d - k, k - w)$

We can readily check that, throughout the table,

$$d = d_T(P, S') = d_T(P, S'') = d_T(P, S^*) = d_T(P, S).$$

Similarly, we find that, throughout the table,

$$d_T(Q, S') = d_T(Q, S'') = d_T(Q, S^*) = d_T(Q, S) = t + u + 2k - d,$$

which equals e once we rewrite t, u, and k in terms of a, b, c, d, and e. Thus in all cases the points between S' and S'' and between S'' and S^* are the correct distances from P and Q.

Finally, we show that the points S' and S^* in each case give, respectively, the minimum and maximum distances between R and S compatible with the relevant triangle inequalities. We now need the assumption $b + c \geq d + e$, which implies that $d_T(P, R) + d_T(Q, R)$ exceeds $d_T(P, Q)$ by more than $d_T(P, S) + d_T(Q, S)$ exceeds $d_T(P, Q)$. Hence the z-coordinate of R must be greater than the z-coordinate of S. In Figures 7 and 8, where $S' = (d - k, 0, k)$, S' is "between" P and R. That is, $d_T(P, R) = d_T(P, S') + d_T(S', R)$. Because $d_T(P, R) = b$, $d_T(P, S') = d$ and $b \leq d + f$, we see that $d_T(S', R) \leq f$. In Figures 9 and 10, $S' = (d - k - u, u, k)$ is "between" Q and R and a similar argument shows that $d_T(S', R) \leq f$. In the same way we see that $f \leq b + d = d_T(R, S^*)$ in Figures 7 and 10 and $f \leq c + e = d_T(R, S^*)$ in Figures 8 and 9. In short, $d_T(R, S') \leq f \leq d_T(R, S^*)$. This completes the proof. ■

Related metrics

The Euclidean and taxicab metrics are special cases of the so-called p-metrics (for $1 \leq p < \infty$), a family of metrics defined as follows for $\mathbb{R}^3$.

DEFINITION. *On* $\mathbb{R}^3$, *define* d_p *for any real number* $p \geq 1$ *by*

$$d_p((x, y, z), (k, l, m)) = \left(|x-k|^p + |y-l|^p + |z-m|^p\right)^{1/p}.$$

The Euclidean metric is the case $p = 2$ and the taxicab metric is the case $p = 1$. See, e.g., [**6**, pp. 214–219] for a proof that d_p is a metric on $\mathbb{R}^n$.

It can be shown that $\mathbb{R}^n$ with any of the p-metrics is triangle complete. Since the proof requires more advanced ideas from real analysis, we omit it. Interestingly, however, we can settle the question of pyramid completeness without advanced mathematics. FIGURE 11 illustrates in two dimensions the shape of the "p-circles" of the same radius for different values of $p < \infty$. Note that the taxicab "circle" has straight sides, while all the others are curved. This flatness is the key to pyramid completeness for the taxicab metric. Each of the other p-circles is *strictly convex*: Each point on these circles has a tangent line that intersects the circle in only that point. Similarly, for $p > 1$ the three-dimensional "p-spheres" are strictly convex: Each point has a tangent plane that intersects the p-sphere at only one point.

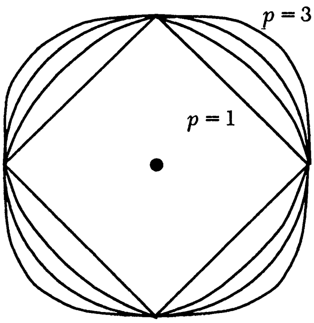


FIGURE 11
p circles for
$p = 1, p = 1.5, p = 2, p = 3.$

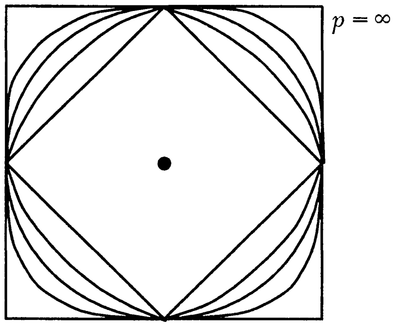


FIGURE 12
p circles
including $p = \infty$

EXAMPLE 3. The metric space $\mathbb{R}^3$ with the p-metric for $1 < p < \infty$ is not pyramid complete. Consider the following values for the sides of the pyramid: $a = 10$, $b = c = d = e = 5$ and $f > 0$. Let P and Q be any two points in $\mathbb{R}^3$ with $d_p(P, Q) = 10$. Consider the p-spheres of radius 5 centered at P and Q. By the strict convexity of these p-spheres, they have exactly one point of intersection. This means that the only possibility for a pyramid with the given sides $a, b, c, d,$ and e is for $f = 0$. Hence $\mathbb{R}^3$ with the p-metric is not pyramid complete for $1 < p < \infty$.

REMARK. A more sophisticated argument shows that for any particular value of $p > 1$ and $a > 0$ there are values for $b, c, d, e,$ and f all the same and slightly bigger than $a/2$ giving actual triangles that cannot be made into pyramids.

The family of p-metrics has one more member—the ∞-metric. The p-metrics arise in functional analysis, an area unrelated to triangle and pyramid completeness. (See [**1**] and [**6**].)

DEFINITION. *On* $\mathbb{R}^3$, *define* d_∞ *by*

$$d_\infty((x,y,z),(k,l,m)) = \max\{|x-k|,|y-l|,|z-m|\}.$$

FIGURE 12 adds the "∞-circle" to the p-circles of FIGURE 11. As p increases, the shape of a p-circle approaches the shape of the ∞-circle; this indicates why the notation d_∞ is natural. More precisely, in $\mathbb{R}^n$, $d_\infty(P,Q) = \lim_{p\to\infty} d_p(P,Q)$. (See, e.g., **[6]** for more on d_∞.) Because circles in both the d_1 and d_∞ metrics have straight sides, we conjecture that $\mathbb{R}^3$ with the d_∞ metric is pyramid complete.

THEOREM 2. *The metric space* $\mathbb{R}^3$ *with* d_∞, *the* ∞-*metric, is pyramid complete.*

Proof. Suppose that for the sextuple of non-negative numbers (a,b,c,d,e,f), all triangle inequalities hold for the triples $\{a,b,c\}$, $\{a,d,e\}$, $\{b,d,f\}$, and $\{c,e,f\}$. (See FIGURE 1 for the relationship between these lengths and the points P, Q, R, and S.) For ease we consider the case where a is the largest value and e is the smallest among b, c, d, and e. (The other cases require only a relabeling of points.) Let $P = (b,0,0)$, $Q = (c,a,e)$, $R = (0,b,0)$, and $S = (f,d,0)$. (See FIGURE 13.)

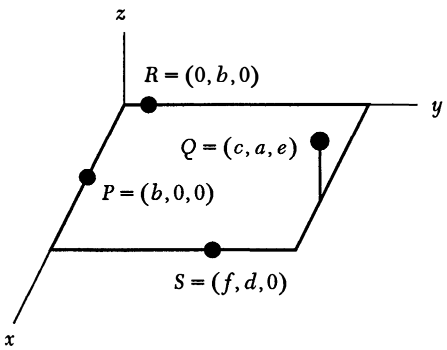


FIGURE 13

Then $d_\infty(P,R) = b$. Also, $d_\infty(P,Q) = a$ and $d_\infty(Q,R) = c$ because $0 \le e \le c \le a$ and the triangle inequality $a \le b + c$ guarantees $a - b \le c$. The other triangle inequalities similarly ensure that $d_\infty(P,S) = d$, $d_\infty(Q,S) = e$ and $d_\infty(R,S) = f$. ■

EXAMPLE 4. An unordered set of six lengths may form a Euclidean pyramid in some orders, but not in others. Consider the sextuple $(10,7,6,6,6,6)$. In this order there is a Euclidean pyramid with its vertices having approximate coordinates $P = (10,0,0)$, $Q = (0,0,0)$, $R = (4.35, 4.132, 0)$, and $S = (5, -0.907, 3.190)$. However, the reasoning in Example 1 shows that the sextuple $(10,6,6,6,6,7)$ fails to have a corresponding Euclidean pyramid. The definition of pyramid completeness we have chosen leads to simpler proofs than a weaker one allowing reordering of lengths. Examples 1 and 3 show that the p-metrics for $1 < p < \infty$ are not pyramid complete, even allowing such reordering.

Interested readers can explore higher-dimensional or other variations of this problem, such as the following suggestion from Walter Sizer, of Moorhead State University:

> The areas of the four faces of a pyramid satisfy an inequality similar to the triangle inequality: The sum of any three of these areas is not less than the fourth area. Given four numbers that satisfy this area inequality in all arrangements, is there a pyramid in Euclidean geometry (or other metric spaces) whose faces have these numbers as their areas?

In summary, only the metric spaces with the extreme metrics, $p = 1$ and $p = \infty$, are pyramid complete.

Acknowledgment. I thank the referees for Example 4 and for other helpful suggestions.

REFERENCES

1. Edwin Beckenbach and Richard Bellman, *An Introduction to Inequalities*, Random House, New York, NY, 1961.
2. Eugene F. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*, Dover Publications, New York, NY, 1986.
3. Fernand Prevost, The conic sections in taxicab geometry, *The Math. Teacher* 91 (April, 1998), 304–317.
4. J. F. Rigby, The edge-lengths of a tetrahedron, *Facta Universitatis Ser. Math. Inform.* #5 (1990), 85–87.
5. Thomas Q. Sibley, *The Geometric Viewpoint: A Survey of Geometries*, Addison-Wesley Longman, Reading, MA, 1998.
6. George F. Simmons, *An Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, NY, 1963.

→ Katye O. Sowell, Taxicab geometry—a new slant, this MAGAZINE 62 (1989), 238–248.

A Web-Searchable Database for *Mathematics Magazine*

We encourage readers (and prospective contributors) to visit the *Magazine*'s Web-searchable database of Articles, Notes, Proofs Without Words, and more. The database can be found through MAA Online, at

`http://www.maa.org`

(click on Journals and then on Mathematics Magazine), or directly from its "home" at Harvey Mudd College, at

`http://www.math.hmc.edu/MathMag/`

The database now contains information going back to 1927, the *Magazine*'s inception. A similar database allows readers to search for information on the *College Mathematics Journal.*

A Bifurcation Problem in Differential Equations

DUFF CAMPBELL
U.S. Military Academy
West Point, NY 10996

SAMUEL R. KAPLAN
University of North Carolina at Asheville
Asheville, NC 28804

A parameter may be defined as "a constant that varies." As such a constant is varied, the qualitative behavior of whatever is being studied may change dramatically. An example from differential equations is the equation

$$\frac{d^2y}{dt^2} = ky, \quad \text{where } k \text{ is a constant.}$$

The qualitative behavior of solutions to this differential equation, and the nature of the equilibrium solution $y \equiv 0$, are each dramatically different for $k > 0$ and $k < 0$. We say that $k = 0$ is a *bifurcation value* for the parameter k in this differential equation. Bifurcation theory is an important area of research in dynamical systems and the basics of this theory can be introduced in a sophomore-level differential equations course **[1]**. Here we look at a simple bifurcation problem which arose out of a lab project. This problem allows the instructor to bring into the classroom an interesting applied problem, numerical experimentation, qualitative analysis, and the geometry of a three-dimensional phase space.

An early model in a differential equations course is the logistic population model:

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{N}\right),$$

where P is the population, t is time, k is the intrinsic growth rate for small populations and N is the carrying capacity.

With a small alteration, we have an interesting, applied bifurcation problem [**1**, Lab 1.1, p. 132].

1. A harvesting problem

Suppose we model the population of a particular species of fish in a lake with a logistic differential equation,

$$\frac{dP}{dt} = 0.25P\left(1 - \frac{P}{4}\right),$$

where $P(t)$ is the number of fish in tens of thousands at time t in years, $k = 0.25$ is the intrinsic growth rate parameter and $N = 4$ (that is, 40,000 fish) is the carrying capacity of the lake. Suppose it has been decided that the lake will be opened to controlled fishing and the student must decide how many fishing licenses can safely be issued. Each license allows for 20 fish to be caught in one year's time. A modification of the model which takes fishing into account is

$$\frac{dP}{dt} = 0.25P\left(1 - \frac{P}{4}\right) - 0.002L, \tag{1}$$

where L is the number for fishing licenses issued. How many licenses can be issued safely?

2. A geometric approach to the harvesting problem

The problem can easily be solved by a geometric analysis, based on studying a plot of dP/dt as a function of P for several values of L. The power of this qualitative analysis is that we can understand the model and predict the long-range behavior without having to solve the differential equation explicitly.

For $L = 0$, the right-hand side of (1) has two roots, one at $P = 0$ and the other at $P = 4$ (see FIGURE 1). These two roots represent the equilibrium solutions for the original differential equation. At $P = 0$ we have no fish, and thus no change. At $P = 4$ we have stasis at the carrying capacity; the death rate and birth rate are the same. For populations between 0 and 4, dP/dt is positive, so the population is increasing. For populations larger than 4, dP/dt is negative so the population is decreasing. This information can be portrayed in a phase line (see FIGURE 1).

FIGURE 2 shows the graphs for several values of L, from which the qualitative behavior of the solution is clear. The right-hand side of (1) has a maximum of $1/4 - 0.002L$ when $P = 2$. When the term $0.002L$ is greater than $1/4$, that is, when L is larger than 125, the parabola is completely below the P-axis so dP/dt is always negative. This means that no matter what the initial population, if more than 125 licenses are issued the fish population will die out.

At this point one can introduce the idea of a threshold population by studying what happens to the equilibria for values of L between 0 and 125 (see FIGURE 2). From FIGURE 2 we see that for L between 0 and 125 there are two equilibrium populations $P_1 < P_2$. The larger one is a *sink* and the smaller a *source*. If $P(0)$ is greater than P_2 then the population decreases with limit P_2, and if $P(0)$ is between P_1 and P_2 the population increases with limit P_2. However if $P(0) < P_1$ the population decreases,

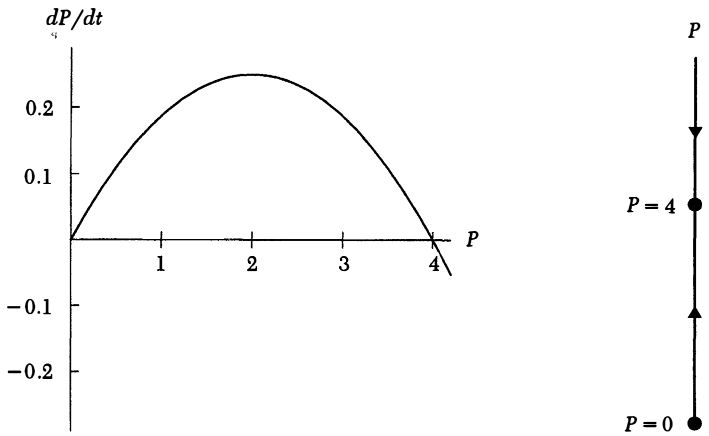


FIGURE 1
A plot of dP/dt versus P and its associated phase line.

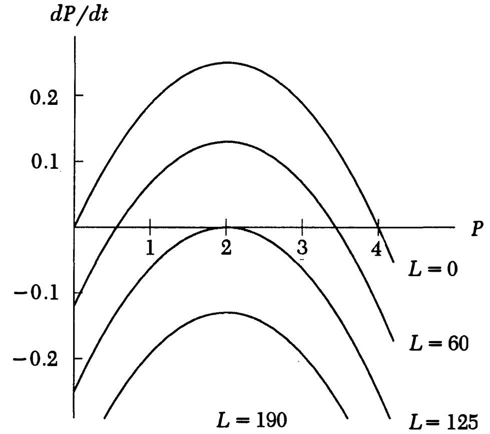


FIGURE 2
Graphs for several values of L, from which the qualitative behavior of the solution is deduced.

reaching 0 in finite time. This is shown by the graphs for the case $L = 60$ in FIGURE 3 (b). Thus P_1 is a "threshold" level below which the fish population must not fall if the population is not to be wiped out.

When $L = 125$ the two equilibrium populations coalesce to form a single "semi-stable" equilibrium at $P = 2$, which is also a threshold population. When the initial population $P(0) > 2$ then the population decreases with limit 2, while initial populations below 2 go extinct.

If $L > 125$, for example in the case $L = 190$ in FIGURE 3, all solutions decrease to zero regardless of the initial population $P(0)$. The value $L = 125$ is called a *bifurcation* value for the parameter L in our model, because the behavior of the solutions, and the number and type of equilibria, are dramatically different for $L < 125$ and $L > 125$. One can expect students to perform a qualitative analysis like the one above for many autonomous first-order differential equations.

3. The lab project

Now suppose the amount of fishing is periodic over the course of a year. This leads to the following differential equation:

$$\frac{dP}{dt} = 0.25P\left(1 - \frac{P}{4}\right) - 0.002\,L(1 + \sin 2\pi t). \qquad (2)$$

The student's task is to propose a procedure for determining the number L of fishing licenses to issue, and to justify that choice. Since this is a *non-autonomous* differential equation (that is, the right hand side of (2) depends explicitly on the independent variable t), the geometric technique demonstrated in Section 2 will not work.

When students at Boston University examined this non-autonomous equation numerically, they found the following: for parameter values below the old bifurcation value of $L = 125$, they observed, not two equilibrium solutions, but what appeared to be two periodic solutions. That is, they saw two solutions which oscillated about the old equilibrium values, with period one year, as expected from the differential equation. Between the two periodic solutions, there are solutions which leave the

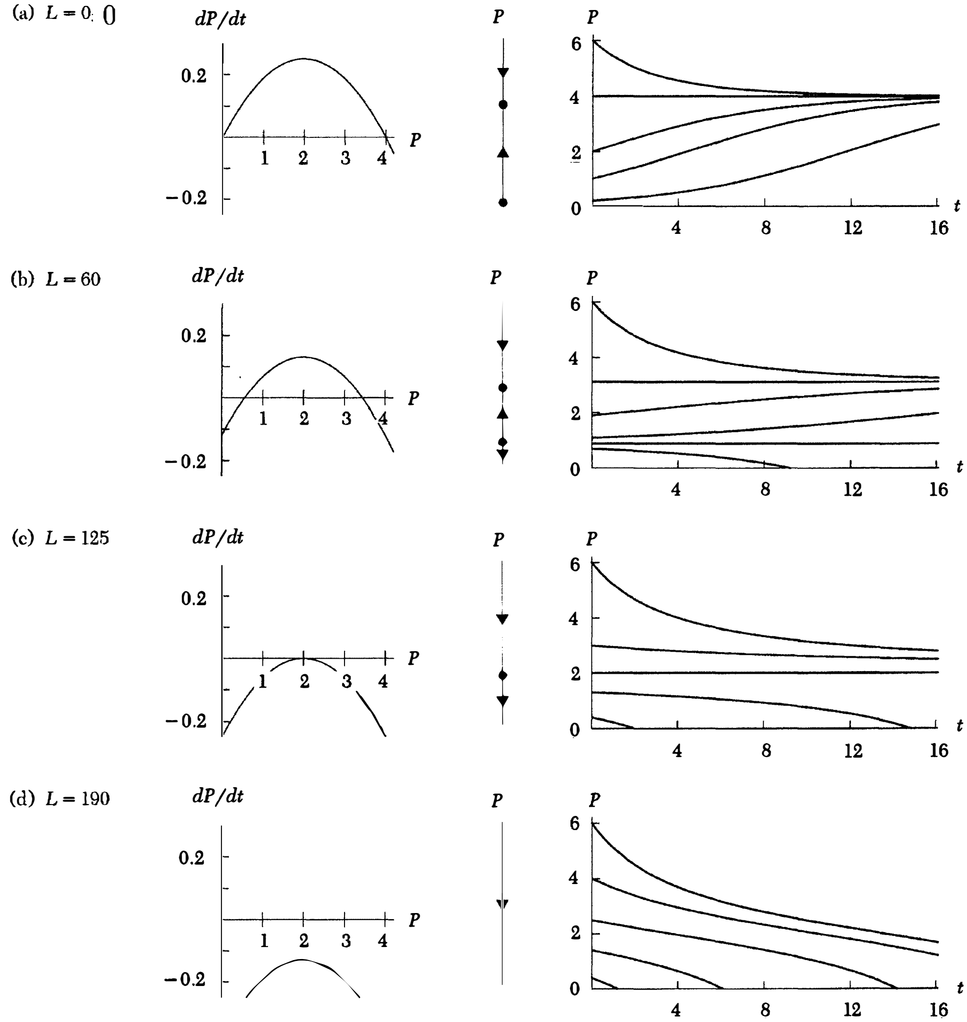


FIGURE 3
Phase line and solution sketch for several values of L.

lower periodic solution and approach the upper one, but they too oscillate up and down (see FIGURE 4). Outside the periodic solutions, just as in the autonomous situation, solutions "decreased." So the situation was much as before; any initial population above a certain threshold was attracted to a stable sink, and initial populations below that threshold eventually died out. However, all solutions, including the sink and the source, were now oscillatory (compare solutions in FIGURE 4).

As the value of L increased, the two periodic solutions got closer and closer together, and eventually coalesced, giving a scenario much like the earlier bifurcation behavior of (1): one periodic solution, populations above that approaching the periodic solution, and populations below that dying off. Increasing L further, the periodic solution disappeared, and all solutions decreased in an oscillatory manner. The situation matched what happened in the autonomous case, with the oscillatory behavior as an additional facet.

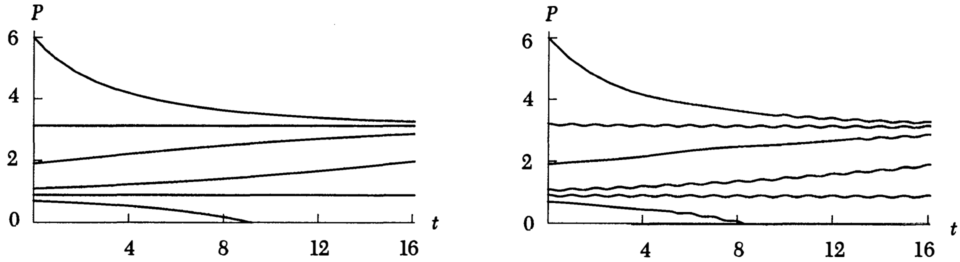


FIGURE 4
Several solutions to (1) and (2) (left and right respectively) for $L = 60$.

An important question about this bifurcation arises: When does the periodic solution disappear? Is it again at the old bifurcation value of $L = 125$, or has the periodic harvesting changed this? To examine this question, some students looked at longer and longer time scales (see Figure 5). Their results suggest that the bifurcation value occurs at a lower value than $L = 125$, but since we have no analytic methods available to deal with this situation (at least not in a sophomore-level course), it's hard to tell. At the time the first author had to grade this, there was some dispute as to what actually was going on—the course decided on (always a useful technique) was to raise doubt in the students' minds, to encourage further exploration (and buy some time). Those students who had asserted that the bifurcation value for (2) was still $L = 125$ were shown Figure 5, right; those who asserted that the bifurcation must have occurred earlier were encouraged to think about computer round-off as well as the inherent error in Euler's method (or any numerical technique).

This parallels the real-world situation: if fishing is allowed at exactly the bifurcation value, the population lives on the brink of disaster. If a sudden fluctuation in the population causes the population to rise a bit (a fisherman takes a day off, a shark dies or gets indigestion, twins are born!), the effect will be flattened out by the downward tendency of all solutions above the unique periodic solution. But any fluctuation (illegal over-harvesting, a sudden epidemic, poor breeding conditions, etc.) that pushes the population below the periodic solution will result in extinction. Similarly, this is how computer error has an effect: an overestimate of the population will eventually disappear, but an underestimate will get magnified immediately.

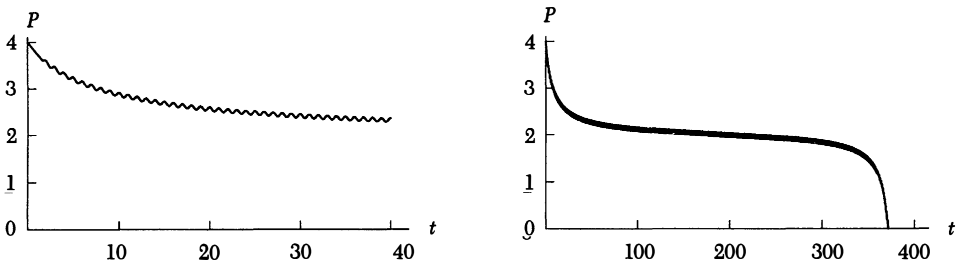


FIGURE 5
Solution at L = 125 shown with two time scales.

4. Increasing the dimensions; decreasing the complexity

We now show how this question may be analyzed further. A key observation is that the non-autonomous term in (2) is $x(t) = \sin 2\pi t$, which is (part of) a solution to the first-order system

$$\begin{cases} \dfrac{dx}{dt} = 2\pi y \\ \dfrac{dy}{dt} = -2\pi x. \end{cases} \tag{3}$$

We may therefore transform the one-dimensional non-autonomous system (2) into the **three**-dimensional **autonomous** system

$$\begin{cases} \dfrac{dx}{dt} = 2\pi y \\ \dfrac{dy}{dt} = -2\pi x \\ \dfrac{dP}{dt} = 0.25P(1 - P/4) - 0.002L(1 + x). \end{cases} \tag{4}$$

A plot of one solution (with $x(0) = 0$, $y(0) = 1$, and $P(0) = 4$) is shown in FIGURE 6.

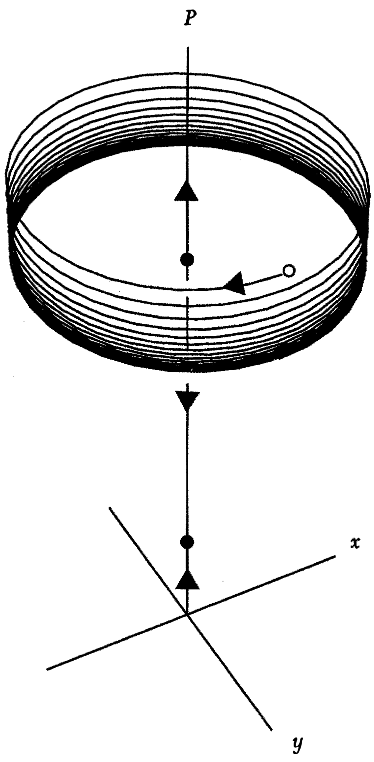


FIGURE 6
Solutions to (4) for $L = 75$. The initial condition $x(0) = 0$, $y(0) = 1$, $P(0) = 4$, is marked by an open circle. The equilibria are designated by the black dots.

The goal of making the system autonomous is to see in what way the equilibria guide the dynamics. To see how this occurs, we solve for the equilibria. For L between 0 and 125 there are two equilibria at $x = 0$, $y = 0$, and $P = 2(1 \pm \sqrt{1 - 0.008L})$. When $L = 125$ there is only one equilibrium at $x = 0$, $y = 0$ and $P = 2$. For $L > 125$ there are no equilibria. [The usual next step is to linearize the system to see what sort of qualitative behavior one may expect near each equilibrium. Unfortunately in this case two of the eigenvalues are pure imaginary, so the equilibria are non-hyperbolic. Thus the technique of local linearization is of limited use here.]

Students can explore real-world facets to this problem. For example, the population should always be positive. Since the limiting solution is oscillatory, one should check at least numerically that the attracting periodic solution is always above $P = 0$ (see FIGURE 7, left). For L beyond the bifurcation value one can ask how long it will take for the fish to die out (see FIGURE 7, right).

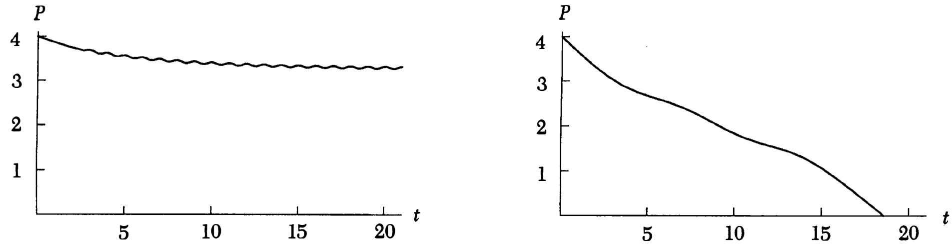


FIGURE 7

Solutions to (2) with $L = 75$ (left) and with $L = 200$ (right).

One can also use this problem to introduce the idea of an *invariant set* for a differential equation. If we let $C(x, y) = x^2 + y^2$ then it is easy to check that for $\{x(t), y(t)\}$, a solution to (4), $dC/dt = 0$. This means that solutions of (4) are partitioned by the level sets of C. Level sets of C are cylinders centered on the P-axis. The level set where $C = 0$ is the P-axis. When L is between 0 and 125, the equilibria lie on this invariant line (see FIGURE 6). Solutions along this line conform to the original model given by (1), and in fact the P-axis in this phase space is exactly the phase line for (1) analyzed earlier. That is, the original problem of non-periodic harvesting (1) is embedded in the generalized periodic harvesting problem (4).

This raises the question of what the other invariant cylinders represent. To see, we explicitly introduce a second parameter, b, to (2) to form:

$$\frac{dP}{dt} = 0.25P\left(1 - \frac{P}{4}\right) - 0.002L(1 + b\sin 2\pi t), \tag{5}$$

then different values of b correspond to solutions on different invariant cylinders. When $b = 0$, (5) yields our original harvesting problem (1). When $b = 1$, (5) yields our periodic harvesting problem (2). One can raise the question of how to interpret, in real-life terms, values of b between zero and one, or values of b strictly larger than one.

When L is between 0 and the bifurcation value, solutions (with $P(0)$ large enough) asymptotically approach some periodic solution. This limiting periodic function appears as an attracting *limit cycle* in (x, y, P) coordinates. The presence of invariant sets as well as attracting and repelling cycles opens the door to discussion and exploration of the geometry of solutions to differential equations.

5. Bifurcation in the general periodic harvesting problem

To explore the actual bifurcation value for (2) let's look at a more general periodic harvesting problem,

$$\frac{dP}{dt} = kP(1-P) - H(1+\sin 2\pi t). \tag{6}$$

Here we have eliminated the carrying capacity N by letting P represent a scaled population, measuring the proportion of the carrying capacity which is present at time t. The constant k is still the intrinsic growth rate for small populations and H is the harvesting constant, representing a proportion of the carrying capacity. Thus $H = 0.1$ means that the harvesting rate varies between 0% and 20% of the carrying capacity. Fix the values of these parameters and suppose $P(t)$ is a solution to (6). We then calculate $P(t+1) - P(t)$:

$$\begin{aligned} P(t+1) - P(t) &= \int_t^{t+1} \frac{dP}{ds}\, ds \\ &= \int_t^{t+1} [kP(s)(1-P(s)) - H(1+\sin 2\pi s)]\, ds \\ &= \int_t^{t+1} kP(s)(1-P(s))\, ds - H\int_t^{t+1} (1+\sin 2\pi s)\, ds \\ &= \int_t^{t+1} kP(s)(1-P(s))\, ds - H. \end{aligned}$$

We know that the expression $kP(1-P) \leq k/4$ since $k/4$ is the maximum value of this function, achieved when $P = 1/2$. For a non-constant solution to (6) (there are clearly no constant solutions) we therefore have the following strict inequality:

$$\begin{aligned} P(t+1) - P(t) &= \int_t^{t+1} kP(s)(1-P(s))\, ds - H \\ &< k/4 - H. \end{aligned}$$

For a periodic solution to exist ($P(t+1) - P(t) \equiv 0$) we therefore have the restriction that $H < k/4$. Recall that for a constant harvest the bifurcation value is $kN/4$ (thus $H = k/4$ after scaling). The bifurcation value for the periodic harvesting problem is therefore strictly smaller than that for the constant harvesting problem.

To determine a lower bound on the value of the bifurcation we consider the geometry of the solution space. In FIGURES 8, left and center, we plot the curve of zero slope (i.e., $dP/dt = 0$), called a *nullcline*. For clarity, the region where $dP/dt > 0$ is shaded; the nullcline is the boundary between the shaded and the unshaded regions. Notice also that the region $1/2 \leq P \leq 1$ constitutes a "trapping region" for solutions; once in, a solution cannot pass out of this region. This is because $\left.\frac{dP}{dt}\right|_{P=1} \leq 0$, while $\left.\frac{dP}{dt}\right|_{P=1/2} \geq 0$ (when $H \leq k/8$). Now consider the portion of the nullcline $dP/dt = 0$ from the local maximum at $A = (3/4, 1)$ to the local minimum at $B = (5/4, 1/2(1+\sqrt{1-8H/k}))$. Any solution that starts on this portion of the nullcline must decrease until it intersects the rising nullcline $dP/dt = 0$ again, between points B and $C = (7/4, 1)$. The solution must then cross the nullcline horizontally and increase until it intersects the nullcline again, somewhere between the points C and the next minimum, $D = (9/4, 1/2(1+\sqrt{1-8H/k}))$. In fact, as the portion of the nullcline connecting C and D is a translate of the portion connecting A and B, we may use the evolution of solutions to construct a continuous mapping of this portion of the nullcline to itself. (This is the *Poincaré return map*.) As the endpoints of the domain map into the interior of the domain, we see there that exists at least one fixed point under this mapping; in fact, a more detailed argument shows that this is a strict contraction mapping, and thus there exists a unique, attracting, fixed point. This fixed point corresponds to a periodic solution. Thus an attracting periodic solution exists, with $P(t) \geq \frac{1}{2}$, for values of $H \leq k/8$. A similar argument (with time reversed) shows that for $H < k/8$ there exists a unique, repelling, periodic solution with $P(t) \leq \frac{1}{2}$. Thus the bifurcation must occur for an attracting $k/8 < H < k/4$.

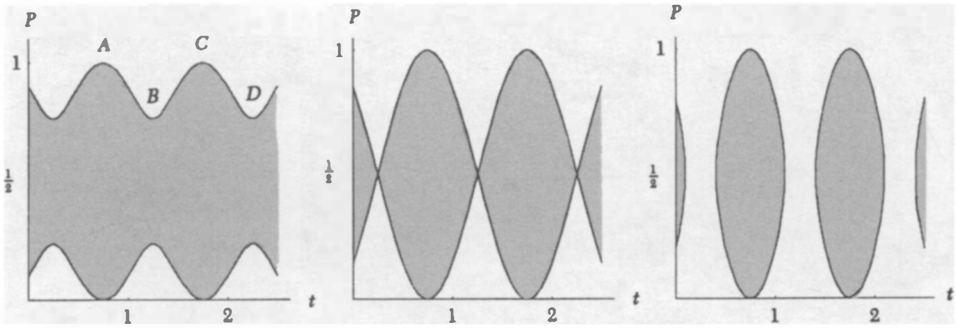


FIGURE 8

Regions of positive slope are shown in gray for $0 < H < k/8$ (left), $H = k/8$ (center), $H > k/8$ (right).

When k is relatively small, the bifurcation value of H is close to $k/4$. For $H > k/8$ it seems possible for a solution to pass between two regions of positive slope and lead to extinction. For sufficiently large k this is exactly what does occur (see FIGURE 9). A more detailed analysis could explore this dependence on k.

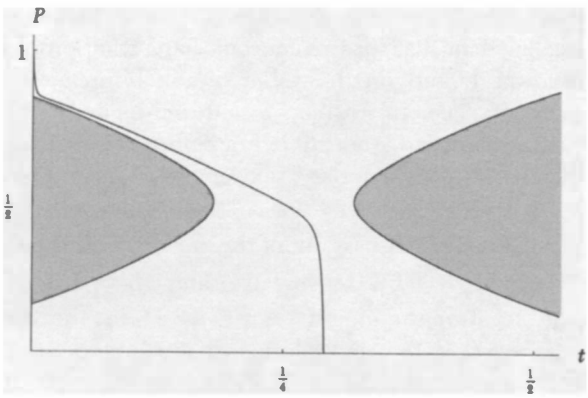


FIGURE 9

For sufficiently large k a solution can pass between two regions of positive slope.

Acknowledgment. We would like to thank Paul Blanchard of Boston University for writing the original lab project and for many helpful suggestions. Thanks also to Keith Ramsay (formerly of Boston College), who shared the proof for the upper bound of the bifurcation value; and to the referees, who substantially improved this article with their constructive criticisms and comments.

REFERENCE

1. Blanchard, Devaney, and Hall, *Differential Equations*, Brooks-Cole Publishing, Boston, MA, 1998.

Resources on the web

Dynamical Systems and Technology Project: math.bu.edu/DYSYS/dysys.html

BU Differential Equations project: math.bu.edu/odes/

Nebraska Wesleyan University DE Resources: brillig.nebrwesleyan.edu/~glarose/delabs/

Books for further reference

Differential Equations and Their Applications, 4th ed., Martin Braun, Springer-Verlag, New York, NY, 1993.

Higher-dimensional thinking

From an advertisement for two monographs, *Multifoliate Numbers* and *Multifoliate Cyclic Equations*, which appeared appeared on the back cover of the March 1937 issue of the National Mathematics Magazine:

> A one dimensional universe of numbers furnishes a mathematics which is adequate for one dimensional thought. In how many dimensions does a man think? Thought of time is one dimensional; thought of space is three dimensional; simultaneous thought of time and space is four dimensional. Thought of color is three dimensional; add color to time and space and then thought becomes seven dimensional. Thought of sound is x dimensional. Determine x for yourself; you will find some difficulties. Thought of time, color, space, and sound is $7+x$ dimensional. And this is not the end. ...
>
> Evolution, that is, the extraction of roots, unfolds a universe of true, commutative, algebraic numbers of an unlimited number of dimensions. ... The explicit recognition of the fact has been delayed because the symbol of evolution, the radical sign, taken by itself, furnishes only a vague and ambiguous symbolization of the whole truth. The underlying truth only needs clarification by reclothing evolution with an adequate code of symbols.

Price: One dollar per volume.

NOTES

Rectangles in Rectangles

JOHN E. WETZEL
University of Illinois at Urbana-Champaign
Urbana, IL 61801-2975

Introduction When does one rectangle fit into another? In 1956, in the *Monthly*, L. Ford asked for a necessary and sufficient condition for a $p \times q$ rug to fit on an $a \times b$ floor. Necessary conditions are easy to find: If a $p \times q$ rectangle fits into an $a \times b$ rectangle (FIGURE 1), then

$pq \le ab$ (the area condition)
$p^2 + q^2 \le a^2 + b^2$ (the diameter condition)
$p + q \le a + b$ (the perimeter condition)
$\min\{p, q\} \le \min\{a, b\}$ (the thickness condition).

But none of these necessary conditions is sufficient. For example, a rectangle with sides 9 and 4 does not fit into a rectangle with sides 8 and 6, but all four necessary conditions are satisfied.

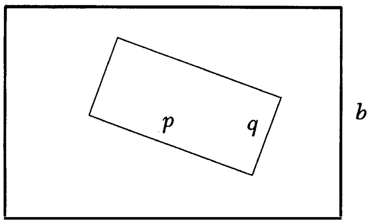


FIGURE 1

Sufficient conditions seem scarcer. One can show that if the thickness condition $\min\{p, q\} \le \min\{a, b\}$ is satisfied and if either of the conditions

$$\max\{p, q\} \le \min\{a, b\} \quad \text{and} \quad p + q \le \sqrt{a^2 + b^2}$$

holds, then a $p \times q$ rectangle fits into an $a \times b$ rectangle. But neither of these sufficient conditions is necessary; an 88×13 rectangle fits into an 81×59 rectangle (see below), but both of these two conditions are false.

In answer to Ford's question, W. Carver **[3]** gave the following mysterious-looking necessary and sufficient condition: a $p \times q$ rectangle $(p \ge q)$ fits into an $a \times b$ rectangle $(a \ge b)$ if and only if

$$p \le a \quad \text{and} \quad q \le b$$

or

$$p > a \quad \text{and} \quad b \ge \frac{2pqa + (p^2 - q^2)\sqrt{p^2 + q^2 - a^2}}{p^2 + q^2}.$$

Carver's ingenious elementary argument is entirely geometric. In this note we give a straightforward development of Carver's conditions that is quite different from the argument he suggested, and we prove:

THEOREM 1. (Carver **[3]**) *Suppose an $a \times b$ rectangle T is given, with the notation arranged so that $a \geq b$. Then a $p \times q$ rectangle R with $p \geq q$ fits into T if and only if*

(*a*) $p \leq a$ *and* $q \leq b$, *or*

(*b*) $p > a, q \leq b$, *and* $\left(\frac{a+b}{p+q}\right)^2 + \left(\frac{a-b}{p-q}\right)^2 \geq 2.$

In Section 4 we use this result to find the smallest rectangle of given shape that can accommodate first every rectangle of perimeter two, and then every rectangle of diameter one. In Section 5 we pose some related unsolved problems of undetermined difficulty.

2. Rectangles in rectangles In 1964 H. Steinhaus **[16]** asked for a necessary and sufficient condition on the six sides for one triangle to fit into another. Nearly thirty years later, K. A. Post **[12]** gave a set of 18 inequalities relating the sides of the triangles that are necessary and sufficient in the sense that if one of the inequalities is correct, the first triangle fits in the second; and if the first triangle fits in the second, then one (at least) of the inequalities is correct. Post's argument relies on a fitting lemma for triangles that is of interest in its own right: if one triangle fits in another, then it fits in such a way that two of its vertices lie on the same side of the containing triangle.

Our argument for rectangles also relies on a fitting lemma: The largest rectangle similar to a given rectangle that lies in a target rectangle T has its vertices on the sides of T.

LEMMA 2. (Mok-Kong Shen)[1] *Suppose rectangles T and R_0 are given. Then there is a largest rectangle R similar to R_0 that fits in T, and R fits in T with all four of its vertices on the sides of T.*

Proof. The existence of R is a routine consequence of compactness, but for completeness we include a short argument using sequences. Suppose that R_0 has sides u_0 and v_0. Let $\lambda = \sup\{t > 0 : \text{a } tu_0 \times tv_0 \text{ rectangle fits in } T\}$, and let R be the rectangle with sides $u = \lambda u_0$ and $v = \lambda v_0$. To show that R fits in T, take a sequence $\{\lambda_k\}$ of positive reals that increase to λ, and let R_k be the rectangle with sides $u_k = \lambda_k u_0$ and $v_k = \lambda_k v_0$. For each index k there are points K_k, L_k, M_k, N_k in T so that the rectangle $K_k L_k M_k N_k$ is congruent to R_k. Passing successively to convergent subsequences (using the Bolzano-Weierstrass theorem), we arrange for $\{K_k\}$, $\{L_k\}$, $\{M_k\}$, and $\{N_k\}$ all to converge, say to K, L, M, and N, respectively. Then the rectangle $R = KLMN$ is congruent to R_0 and lies in T.

How is such a maximal rectangle R situated in T? At least two vertices of R must lie on the sides of T, or a larger rectangle similar to R_0 could be found in T. We examine the possibilities.

If two vertices of R lie on the same side of T, then the sides of R must be parallel to those of T; and the maximality forces the other two vertices to lie on the sides of T as well.

[1]In response to my May 7, 1998 post to the newsgroup `sci.math.research` concerning this question, Mok-Kong Shen posted on May 11, 1998 a partial solution that includes a version of this fitting lemma. Mr. Shen declined my invitation to be coauthor of this note.

If two vertices of R lie on opposite sides of T and if at least one of the two other vertices of R is inside, a small translation of R moves both inside T; and then a suitable small rotation about its center moves R completely inside T, contrary to the maximality. So if two vertices of R lie on opposite sides of T, then all four vertices must lie on T.

If two vertices of R lie on adjacent sides of T and either of the other two vertices is on a side, we have two opposite vertices of R on the sides of T, and the previous argument applies. If both the other vertices are inside T, a small translation moves R completely inside T, contrary to the maximality. So in every case, all four vertices of R must lie on T. ∎

We say that a rectangle R is *inscribed* in a rectangle T if each (closed) side of T contains a vertex of R (see FIGURE 2). When a rectangle with sides p and q is inscribed in a rectangle with sides a and b, the lengths p, q, a, and b are related by the quartic equation

$$q^4 - (a^2 + b^2 + 2p^2)q^2 + 4abpq - p^2(a^2 + b^2 - p^2) = 0, \tag{1}$$

an equation that over the years has appeared frequently in the problem literature. The earliest references of which I am aware are in 1896, although (1) surely must have been known many years earlier. In a problem posed in the *Monthly*, M. Priest asked for the "...length of a piece of carpet that is a yard wide with square ends, that can be placed diagonally in a room 40 feet long and 30 feet wide, the corners of the carpet just touching the walls of the room." A solution **[14]** published the following year by C. D. Schmitt, et al., gives a trigonometric derivation of (1). Equation (1) also appears as an example in a discussion of quartic equations in Merriman and Woodward **[10**, p. 20**]**, an 1896 mathematics reference book for engineers, with the curious footnote: "This example is known by civil engineers as the problem of finding the length of a strut in a panel of the Howe truss."[2]

The question was posed again in 1914 by C. E. Flanagan. The solution by Otto Dunkel **[4]**, published in 1920, includes a geometric derivation of (1) in the elegant symmetric form

$$2 - \left(\frac{a+b}{p+q}\right)^2 - \left(\frac{a-b}{p-q}\right)^2 = 0$$

and, regarding this as an equation in q for given a, b, and p, an elaborate investigation of the location of the roots.

3. The fitting region Suppose the given target rectangle T has sides of length a and b, where $a \geq b$. By the *fitting region* $F = F(T)$ we mean the set of points $\{p, q\}$ so that the rectangle with sides p and q fits in T. We find it convenient *not* to assume that $p \geq q$. It will simplify the language to regard a point as a 0×0 rectangle and a line segment of length $\ell > 0$ as an $\ell \times 0$ rectangle. With this agreement we see that F is a closed subset of the closed quarter-disk of radius $\sqrt{a^2 + b^2}$ that is symmetric in the 45° line.

The fitting region $F(T)$ is star-shaped with respect to the origin, that is to say, if $P \in F(T)$ and if Q lies on the segment OP (where O is the origin), then $Q \in F(T)$.

[2] The Howe truss, a linear assemblage of cross-braced rectangular structural "panels," was invented by one William Howe (1803–1852) and patented in 1840 and 1842. It was the most popular bridge truss design in the United States in the latter half of the 19th century. See **[6]**.

More generally, if $(p, q) \in F(T)$ and if $0 \le r \le p$ and $0 \le s \le q$ (or if $0 \le r \le q$ and $0 \le s \le p$), then $(r, s) \in F(T)$.

An $x \times y$ rectangle R fits in T with its sides parallel to those of T precisely when $x \le a$ and $y \le b$, or $x \le b$ and $y \le a$. Writing $S_{uv} = \{(x, y): 0 \le x \le u, 0 \le y \le v\}$, we conclude that $S_{ab} \cup S_{ba} \subseteq F(T)$. We will find that $F(T)$ is formed from $S_{ab} \cup S_{ba}$ by adding two curvilinear triangular tabs T_1 and T_2 symmetric in the 45° line (see $T_1 = UVW$ in FIGURE 4).

It remains to analyze the situation in which the largest fitting rectangle $R = KLMN$ similar to a given rectangle R_0 fits in T with exactly one vertex on each side of $T = ABCD$. Suppose the vertices are labeled as in FIGURE 2, with $AB = CD = a$,

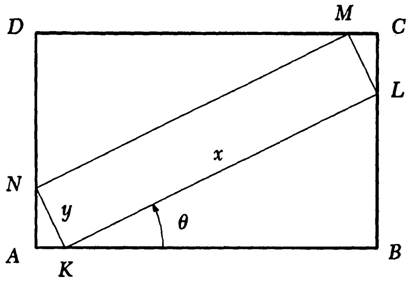


FIGURE 2

$BC = DA = b$, $K \in AB$, $L \in BC$, $M \in CD$, and $N \in DA$. Write $x = KL = NM$ and $y = KN = LM$. Taking the directed angle $\theta = \angle BKL$ as parameter, we see that

$$\begin{cases} x \cos\theta + y \sin\theta = a, \\ x \sin\theta + y \cos\theta = b \end{cases}$$

(cf. [14]). Solving for x and y gives parametric equations for part of the boundary curve of the fitting region:

$$\begin{cases} x = \dfrac{a\cos\theta - b\sin\theta}{\cos 2\theta}, \\ y = \dfrac{b\cos\theta - a\sin\theta}{\cos 2\theta}. \end{cases}$$

It follows that

$$\begin{cases} x + y = \dfrac{a+b}{\sqrt{2}} \dfrac{1}{\sin\left(\theta + \frac{\pi}{4}\right)}, \\ x - y = \dfrac{a-b}{\sqrt{2}} \dfrac{1}{\cos\left(\theta + \frac{\pi}{4}\right)}. \end{cases} \tag{2}$$

Rotating the coordinate axes through the angle $-\pi/4$ using the usual coordinate transformation $x = (\xi + \eta)/\sqrt{2}$, $y = (-\xi + \eta)/\sqrt{2}$, we arrive at the parametric equations

$$\begin{cases} \xi = \dfrac{a-b}{2} \dfrac{1}{\cos\left(\theta + \frac{\pi}{4}\right)}, \\ \eta = \dfrac{a+b}{2} \dfrac{1}{\sin\left(\theta + \frac{\pi}{4}\right)}. \end{cases}$$

To eliminate the parameter, we solve for $\sin(\theta + \frac{\pi}{4})$ and $\cos(\theta + \frac{\pi}{4})$ and square and add, obtaining the equation $\xi^2\eta^2 - c^2\xi^2 - d^2\eta^2 = 0$, i.e., (since $\eta = (x+y)/\sqrt{2} > 0$),

$$\eta = \frac{c|\xi|}{\sqrt{\xi^2 - d^2}},$$

where $c = \frac{1}{2}(a+b)$ and $d = \frac{1}{2}(a-b)$. Suppose $a \neq b$. Then $d \neq 0$, and the curve has two branches symmetric in the η-axis with asymptotes $\eta = c$ and $\xi = \pm d$ (see FIGURE 3). Write $\eta = \Gamma(\xi)$ for the branch with $\xi > 0$ and $\eta > 0$. Plainly Γ is a decreasing function whose graph is concave upwards.

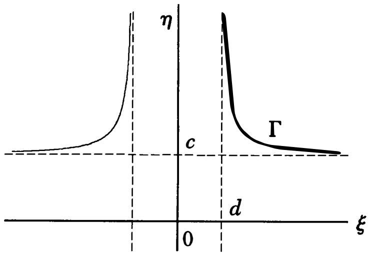


FIGURE 3

In terms of (x, y), Γ can be written (from (2)) in the form $f(x, y; a, b) = 2$, where

$$f(x, y; a, b) = \left(\frac{a+b}{x+y}\right)^2 + \left(\frac{a-b}{x-y}\right)^2.$$

The curve Γ is asymptotic to the lines $x + y = (a+b)/\sqrt{2}$ and $y - x = \pm(a-b)/\sqrt{2}$. For (x, y) in the quadrant $x > |y|$, we see that $f(x, y; a, b) < 2$ on the concave side of Γ and $f(x, y; a, b) > 2$ on the side of Γ that contains the asymptotes. The curve Γ meets the x-axis at $x = \sqrt{a^2 + b^2}$ with slope $-(a^2+b^2)/2ab < -1$, and it passes through the point (a, b) with slope a/b. Since Γ is concave, it meets the line segment $x = a$, $0 \leq y \leq b$ at (a, b) and again at a point U whose y-coordinate b_0 is the root of the cubic

$$y^3 + by^2 - 3a^2y + a^2b = 0$$

in the interval $0 < y < b$. Let U, V, W be the points with coordinates (a, b_0), $(a, 0)$, $(\sqrt{a^2+b^2}, 0)$, respectively, and let T_1 be the curvilinear triangular region bounded by the line segments UV and VW and the arc of Γ with endpoints U and W (see FIGURE 4). Let T_2 be the mirror image region of T_1 in the 45° line through the origin. We shall see that the fitting region $F(T)$ is precisely $S_{ab} \cup S_{ba} \cup T_1 \cup T_2$.

In the square case $a = b$ the graph Γ collapses to the line $\eta = c$, i.e., $x + y = a\sqrt{2}$; and T_1 and T_2 become ordinary triangles.

So finally we come to our result.

THEOREM 3. *A rectangle R with sides p and q fits in a rectangle T with sides a and b (with $a \geq b$) if and only if the point (p, q) lies in the set $S_{ab} \cup S_{ba} \cup T_1 \cup T_2$.*

Proof. If $(p, q) \in S_{ab} \cup S_{ba}$, then $p \leq a$ and $q \leq b$ or vice versa, and in either case the rectangle with sides p, q fits in T with its sides parallel to those of T. If

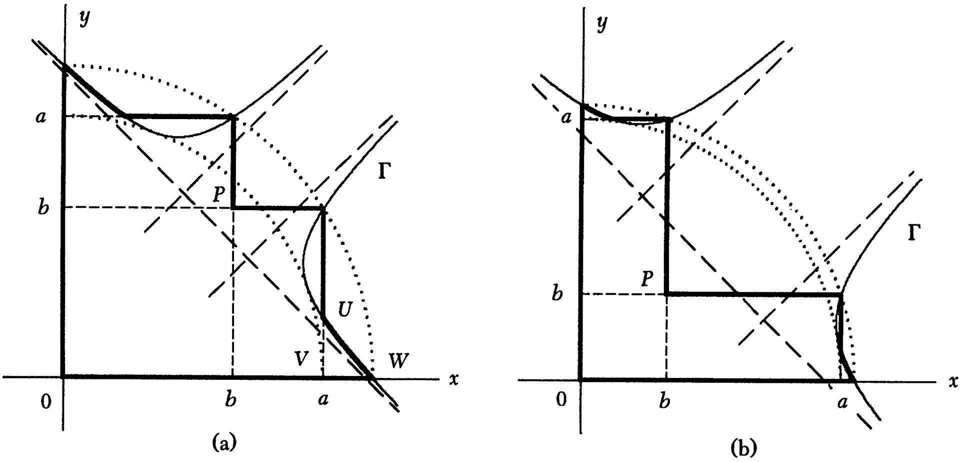


FIGURE 4

$(p, q) \in T_1$, the ray from the origin through (p, q) meets Γ at a point (x, y) whose coordinates are the sides of the largest rectangle R' in T similar to a $p \times q$ rectangle R; and R fits in R' and hence in T. If $(p, q) \in T_2$, then $(q, p) \in T_1$, and the above argument applies. So if (p, q) is any point in $S_{ab} \cup S_{ba} \cup T_1 \cup T_2$, the rectangle with sides p and q fits in T.

Conversely, if a rectangle with sides p and q fits into T, the largest similar rectangle R', with sides $x \geq p$ and $y \geq q$, fits in T in one of only two possible ways. If two vertices of R' lie on the same side of T, then either $x \leq a$ and $y \leq b$ or vice versa, and $(x, y) \in S_{ab} \cup S_{ba}$; so $(p, q) \in S_{ab} \cup S_{ba}$. Otherwise, R' is inscribed in T, (x, y) lies on Γ, and (p, q) lies on the radius to (x, y) and so lies in $S_{ab} \cup S_{ba} \cup T_1 \cup T_2$. ■

It follows, as asserted, that $F(T) = S_{ab} \cup S_{ba} \cup T_1 \cup T_2$. Theorem 1 follows easily: if $p \geq q$, conditions (a) are equivalent to $(p, q) \in S_{ab}$; otherwise, $(p, q) \in T_1$ if and only if $p \geq a, q \leq b$, and $f(a, b; p, q) \geq 2$; and these are conditions (b).

The assertion in the Introduction that the conditions

$$\min\{p, q\} \leq \min\{a, b\} \quad \text{and} \quad p + q \leq \sqrt{a^2 + b^2}$$

are sufficient can be read off FIGURE 4, because if $p \geq q$ and $a \geq b$, then

$$\left\{(p, q) : q \leq b \text{ and } p + q \leq \sqrt{a^2 + b^2}\right\} \subseteq F(T).$$

It is worth noting that in the square case, a $p \times q$ rectangle fits into a square of side a if and only if $\max\{p, q\} \leq a$ or $p + q \leq a\sqrt{2}$ (cf. **[15]**).

4. Two rectangle covering problems Geometric covering problems have attracted considerable interest in recent years. We have recently described the smallest equilateral triangle that can cover any triangle of perimeter two (see Wetzel **[18]**), and in the same article we reported an easily verified claim by one J. Smith that the smallest equilateral triangle that can cover every triangle of diameter one (i.e., every triangle whose longest side has length one) has side $(2\cos 10°)/\sqrt{3}$. Here we use Theorem 3 to solve two analogous problems for rectangles.

First we determine the smallest rectangle R similar to a given rectangle R_0 that can accommodate every rectangle of perimeter two. Evidently the width of R must be at least $1/2$, and if R_0 is long and thin, taking the width of R to be $1/2$ suffices. However, if the sides of R_0 are more nearly equal, we need further to arrange for the diagonal of R to be at least one. More precisely, we prove the following:

THEOREM 4. *An $a \times b$ rectangle R_0 is given, with $a \geq b$. The smallest rectangle R similar to R_0 that can accommodate every rectangle of perimeter two is $a/2b \times 1/2$ if $a \geq b\sqrt{3}$ and $a/d \times b/d$ if $a \leq b\sqrt{3}$, where $d = \sqrt{a^2+b^2}$ is the diagonal of R_0.*

Proof. The dual question is more natural in the present context: given a rectangle R with sides of lengths a and b, determine the largest possible ℓ so that R can accommodate every rectangle of perimeter ℓ. In geometric terms, we seek the largest ℓ so that the line segment $\sigma(\ell)$ with endpoints $(\ell/2, 0)$ and $(0, \ell/2)$ lies in the fitting region $F(R)$.

Since Γ meets the x-axis at the point $d = \sqrt{a^2+b^2}$ with slope at most -1, as ℓ increases from 0, $\sigma(\ell)$ remains in $F(R)$ until it encounters either the point $P(b,b)$ or the intercept $(d, 0)$ (see FIGURE 4). The segment $\sigma(\ell)$ reaches the point P before the point $(d,0)$ precisely when $2b \leq d$, i.e., precisely when $a \geq b\sqrt{3}$. So when $a \geq b\sqrt{3}$ the largest possible ℓ is $4b$, and when $a \leq b\sqrt{3}$ the largest possible ℓ is $2d$.

Dually, the smallest rectangle similar to a given $a \times b$ rectangle (with $a \geq b$) that can accommodate every rectangle of perimeter ℓ is $\ell a/4b \times \ell/4$ when $a \geq b\sqrt{3}$ and $\ell a/2d \times \ell b/2d$ when $a \leq b\sqrt{3}$. The assertion follows when $\ell = 2$. ■

COROLLARY 5. *The rectangle of least area that can accommodate every rectangle of perimeter two has sides $1/2$ and $\sqrt{3}/2$.*

As a second example, we determine the smallest rectangle similar to a given rectangle that can accommodate every rectangle of diameter one.

THEOREM 6. *An $a \times b$ rectangle R_0 is given, with $a \geq b$. The smallest rectangle R similar to R_0 that can accommodate every rectangle of diameter one is $a/(b\sqrt{2}) \times 1/\sqrt{2}$ if $a \geq b\sqrt{2}$ and $1 \times b/a$ if $a \leq b\sqrt{2}$.*

Proof. Direct arguments are easy to give, but a dual argument that parallels the proof of Theorem 4 is straightforward. For the given rectangle R_0 with fitting region $F(R_0)$, we seek the largest d so that every rectangle with diagonal d fits in R_0. Write $\Phi(\rho)$ for the quarter circular arc $x^2 + y^2 = d^2$, $0 \leq x \leq d$. In geometric terms, we seek the largest d so that $\Phi(d) \subseteq F(R_0)$.

It is a calculus exercise to show that Γ (as a function of x and y) has slope $-x/y$ at precisely one point, the point (x_1, y_1) where

$$\begin{cases} x_1 = \sqrt{\frac{a}{2}\left(a + \sqrt{a^2 - b^2}\right)}, \\ y_1 = \sqrt{\frac{a}{2}\left(a - \sqrt{a^2 - b^2}\right)}. \end{cases} \tag{3}$$

Since $x_1^2 + y_1^2 = a^2$, it follows that Γ and the quarter circle $\Phi(a)$ are tangent at the point (x_1, y_1) given by (3) (see FIGURE 4). Consequently, the largest radius d so that $\Phi(d) \subseteq F(R_0)$ is just $\min\{a, b\sqrt{2}\}$, depending on whether for increasing d, $\Phi(d)$ meets P or Γ first.

Dually, the smallest rectangle similar to a given $a \times b$ rectangle (with $a \geq b$) that can accommodate every rectangle of diameter d is $ad/b\sqrt{2} \times d/\sqrt{2}$ when $a \geq b\sqrt{2}$ and $d \times bd/a$ when $a \leq b\sqrt{2}$. The assertion follows when $d = 1$. ■

COROLLARY 7. *The rectangle of least area that can accommodate every rectangle of diameter one has sides 1 and $1/\sqrt{2}$.*

5. Some further questions Here are a few interesting related problems, of undetermined difficulty.

Find necessary and sufficient conditions on a, b, c, s for a triangle with sides a, b, c to fit in a square with side s; for a square with side s to fit in a triangle with sides a, b, c.

More generally, find necessary and sufficient conditions on a, b, c, p, q for a triangle with sides a, b, c to fit in a rectangle with sides p, q; for a rectangle with sides p, q to fit in a triangle with sides a, b, c.

Each of these fitting problems has associated covering problems. How small a square (or rectangle of prescribed shape) can accommodate every triangle of perimeter two? Of diameter one? How small a triangle of prescribed shape can accommodate every rectangle of perimeter two? Of diameter one?

Find necessary and sufficient conditions on the lengths a, b, c and p, q, r for a box (i.e., a rectangular parallelepiped) with edges p, q, r to fit in a box with edges a, b, c. (F.M. Garnett asked this question in 1923, and W.B. Carver **[2]** supplied a fragmentary answer in 1925.) What are the analogous results for orthotopes in $\mathbb{E}^d$? Again, there are related covering problems.

REFERENCES

1. Robert Blasch, Solution to Problem 30, *Pentagon* 10 (1950), 42–43. (Posed by John K. Osborn, *Pentagon* 9 (1950), 109.)
2. W. B. Carver, Solution to Problem 3036, *Amer. Math. Monthly* 32 (1925), 47–49. (Posed by F. M. Garnett, *Amer. Math. Monthly* 30 (1923), 337.)
3. W. B. Carver, Solution to Problem E1225, *Amer. Math. Monthly* 64 (1957), 114–6. (Posed by L. R. Ford, *Amer. Math. Monthly* 63 (1956), 421.)
4. Otto Dunkel, Solution I to Problem 416, *Amer. Math. Monthly* 27 (1920), 327–30. (Posed by C. E. Flanagan, *Amer. Math. Monthly* 21 (1914), 156.)
5. Robert C. Fisher and Alan D. Ziebur, *Integrated Algebra and Trigonometry*, Prentice-Hall, Englewood Cliffs, NJ, 1957. (See p. 259.)
6. Howe, William, *Britannica Online* (accessed June 30, 1998).
7. James F. Hurley, *Litton's Problematical Recreations*, Van Nostrand Reinhold, New York, NY, 1971. (See Problem 41, p. 205.)
8. W. W. Johnson, Solution to Problem 473, *Amer. Math. Monthly* 23 (1916), 173. (Posed by Frank B. Morris, *Amer. Math. Monthly* 22 (1915), 267.)
9. Samuel I. Jones, *Mathematical Nuts for Lovers of Mathematics*, rev. ed., S. I. Jones, Nashville, TN, 1936. (See problem 14, p. 178.)
10. Mansfield Merriman and Robert S. Woodward, *Higher Mathematics*, John Wiley & Sons, New York, NY, 1896.
11. Leroy F. Meyers, Solution to Problem 244, *Crux Math.* 4 (1978), 19. (Posed by Steven R. Conrad, *Crux Math.* 3 (1977), 130.)
12. K. A. Post, Triangle in a triangle: on a problem of Steinhaus, *Geom. Dedicata* 45 (1993), 115–20.
13. William R. Ranson, *One Hundred Curious Mathematical Problems*, J. Weston Walch, Portland, ME, 1955. (See pp. 60–63.)
14. Cooper D. Schmitt, W. H. Harvey, and B. F. Finkel, Solution II to Problem 67, *Amer. Math. Monthly* 4 (1897), 87–88. (Posed by F. M. Priest, *Amer. Math. Monthly* 3 (1896), 278.)
15. D. L. Silverman, Solution to Problem E1584, *Amer. Math. Monthly* 71 (1964), 209–10. (Posed by D. J. Newman, *Amer. Math Monthly* 70 (1963), 438.)
16. Hugo Steinhaus, *One Hundred Problems in Elementary Mathematics*, Basic Books, New York, NY, 1964.
17. Edmund H. Umberger, Solution to Problem 1275, *School Sci. Math.* 33 (1933), 678. (Posed by Paul Veigle, *School Sci. Math.* 33 (1933), 332.)

→ John E. Wetzel, The smallest equilateral cover for triangles of perimeter two, this MAGAZINE 70 (1997), 125–30.

Bernoulli Trials and Family Planning

EUGENE F. KRAUSE
University of Michigan
Ann Arbor, MI 48104

Introduction A sports-minded young couple resolves to continue having babies until they have enough boys for a touch football team (7) and enough girls for a basketball team (5). How many babies should they plan on having? Assuming boys and girls are equally probable, we shall see that they had better plan on having $15\frac{81}{512} \approx 15.158$ babies. If we use the statistically more realistic probabilities of 0.514 for a boy and 0.486 for a girl, then (we shall see) they should expect to have about 14.745 babies.

We have two goals in this paper. One, of course, is to pose and solve the general problem of which the family-planning question above is a special case. The other is to provide some insight into the process of problem-solving by describing two quite different strategies that we employed. Before we can look at the strategies, though, we need to formulate the general problem.

We are concerned with Bernoulli trials: repeated performances of a single random experiment such that on each performance there are just two possible outcomes, success with probability p or failure with probability $q = 1 - p$. We suppose that these Bernoulli trials continue until the (first) moment when the number of successes has reached s and the number of failures has reached r. The question is this: On average, how many trials (performances) are required? We shall refer to this expected number of trials as the waiting time for s successes and r failures, and we shall denote it by $w(s, r)$.

A front-end, recursive-inductive strategy On the *first* trial, one of two things will happen. Either we will have a success, with probability p, and then expect to wait an additional $w(s-1, r)$ trials; or we will have a failure, with probability $q = 1 - p$, and then expect to wait an additional $w(s, r-1)$ trials. That is,

$$w(s, r) = p[1 + w(s-1, r)] + q[1 + w(s, r-1)]$$

which simplifies to a fundamental recurrence relation

$$w(s, r) = 1 + qw(s, r-1) + pw(s-1, r) \qquad (r, s = 1, 2, \dots) \qquad (1)$$

Our plan is to use this recurrence relation to generate a list of formulas for $w(s, 1)$, $w(s, 2), \dots$ in which we will look for a pattern. To get started we need two initial conditions. The initial condition

$$w(s, 0) = s/p \qquad (s = 1, 2, \dots) \qquad (2)$$

follows from the simple observation that $w(s, 0) = s \cdot w(1, 0)$ and the well-known fact [**1**, p. 189] that the waiting time for *one* success in Bernoulli trials is the reciprocal of the probability of success on a single trial. (For example, the waiting time for a "4" to turn up when one rolls a die repeatedly is 6.) The other initial condition, of course, is

$$w(0, r) = r/q \qquad (r = 1, 2, \dots) \qquad (3)$$

Fixing $r = 1$ in (1) yields a recurrence relation for the function of one variable, $w(s, 1)$

$$w(s, 1) = pw(s-1, 1) + 1 + qs/p$$

Iterating this recurrence relation produces a formula for $w(s,1)$

$$w(s,1)=\sum_{i=0}^{s-1}p^i+q\sum_{i=0}^{s-1}p^{i-1}(s-i)+p^s/q$$

which, upon use of the easily derived identities (4) and (5),

$$\sum_{i=0}^{s-1}p^i=\frac{1-p^s}{1-p}=\frac{1-p^s}{q} \tag{4}$$

$$\sum_{i=0}^{s-1}p^{i-1}(s-i)=\frac{1}{q}\left(\frac{s}{p}-p^{s-1}\right)+\frac{1}{q^2}(p^{s-1}-1) \tag{5}$$

becomes the first formula in (6) below.

Next we fix $r=2$ in (1) and use the formula we just found for $w(s,1)$ to arrive at a recurrence relation for the function of one variable, $w(s,2)$:

$$w(s,2)=pw(s-1,2)+1+qs/p+p^s.$$

Iterating this recurrence relation produces a formula for $w(s,2)$:

$$w(s,2)=\sum_{i=0}^{s-1}p^i+sp^s+q\sum_{i=0}^{s-1}p^{i-1}(s-i)+2p^s/q,$$

which, on using identities (4) and (5), simplifies to the second formula in (6) below.

Fixing r at 3, then 4 in (1), and repeating twice the process that we carried out in cases $r=1$ and $r=2$, leads us to the final two formulas in (6):

$$\begin{aligned}
w(s,1)&=\frac{s}{p}+p^s\left[\frac{1}{q}\binom{s-1}{0}\right]\\
w(s,2)&=\frac{s}{p}+p^s\left[\frac{2}{q}\binom{s-1}{0}+\binom{s}{1}\right]\\
w(s,3)&=\frac{s}{p}+p^s\left[\frac{3}{q}\binom{s-1}{0}+2\binom{s}{1}+q\binom{s+1}{2}\right]\\
w(s,4)&=\frac{s}{p}+p^s\left[\frac{4}{q}\binom{s-1}{0}+3\binom{s}{1}+2q\binom{s+1}{2}+q^2\binom{s+2}{3}\right]
\end{aligned} \tag{6}$$

The decision to express the polynomials in s (the expressions inside the brackets in (6)) in terms of the basis $\binom{s-1}{0}$, $\binom{s}{1}$, $\binom{s+1}{2}$, ... instead of the standard basis $s^0, s^1, s^2, \ldots$ was based on exploratory work with the special case $p=q=\frac{1}{2}$ where the standard basis gave a chaotic picture while the basis of binomial coefficients revealed a pattern.

Inside the brackets in the formula for $w(s,4)$ in (6), patterns finally reveal themselves in the exponents of q, in the numerical coefficients, and in the upper and lower entries in the binomial coefficients. Since these patterns persist in the formulas for $w(s,3)$, $w(s,2)$, and $w(s,1)$, it is a good bet that they hold in general. That is the content of our first theorem. And we call it a theorem rather than a conjecture because it can be proved by double induction. We omit that proof, which relies on equations (1)–(3).

THEOREM 1. *If $w(s,r)$ denotes the waiting time for s successes and r failures in Bernoulli trials, where the probability of success is p and the probability of failure is*

$q = 1 - p$, then

$$w(s,r) = \frac{s}{p} + p^s \sum_{k=1}^{r} kq^{r-1-k}\binom{s+r-1-k}{r-k} \tag{7}$$

where we agree that $\binom{-1}{0} = 1$ *so that* $w(0,r) = r/q$.

Note. When we apply this theorem to the case of the sports-minded couple, we have two choices. We can think of having a son as success and a daughter as failure, and calculate $w(7,5)$ with $p = q = \frac{1}{2}$ to arrive at $15\frac{81}{512}$. With $p = 0.514$ and $q = 0.486$, we get $w(7,5) \approx 14.745$. Or we can view sons as failures and daughters as successes and calculate $w(5,7)$. Again we get $15\frac{81}{512}$ when $p = q = \frac{1}{2}$, and we get (approximately) 14.745 when $p = 0.486$ and $q = 0.514$.

A back-end, deductive strategy For the front-end strategy we looked at what happens on the *first* trial, and this led us to a recurrence relation for $w(s,r)$. For the back-end strategy we look at what happens on the *last* trial; that is, on the trial that completes the project of attaining (at least) s successes and (at least) r failures. This will lead us to a probability distribution and then the expected value for an appropriate random variable.

Let K be the waiting time (random variable) until s successes and r failures have been attained. Now s successes and r failures can be attained in k trials in two ways: either $s-1$ successes occur in the first $k-1$ trials, and $(k-1)-(s-1) \geq r$, and the k^{th} trial results in a success; or $r-1$ failures occur in the first $k-1$ trials, and $(k-1)-(r-1) \geq s$, and the k^{th} trial results in a failure. Thus

$$P(K=k) = \binom{k-1}{s-1}p^{s-1}q^{k-s}p + \binom{k-1}{r-1}q^{r-1}p^{k-r}q \qquad (k \geq r+s)$$

That is, the probability distribution for K is

$$P(K=k) = \binom{k-1}{s-1}p^s q^{k-s} + \binom{k-1}{r-1}p^{k-r}q^r \qquad (k = r+s, r+s+1, \dots) \tag{8}$$

Thus the expected value of K, which we called $w(s,r)$ before, is

$$w(s,r) = \sum_{k=r+s}^{\infty} k\left[\binom{k-1}{s-1}p^s q^{k-s} + \binom{k-1}{r-1}p^{k-r}q^r\right] \tag{9}$$

In order to deduce a finite formula for $w(r,s)$ from (9), we need the following identity, which is just our old equation (2) in different notation.

$$\sum_{k=s}^{\infty} k\binom{k-1}{s-1}p^s q^{k-s} = \frac{s}{p} \tag{10}$$

Note that the righthand side, s/p, is the waiting time for s successes, i.e., $w(s,0)$. But the lefthand side is also $w(s,0)$ because it is just (9) in the special case of $r = 0$. The deduction of a finite formula proceeds in a straightforward way:

$$\begin{aligned}
w(s,r) &= \sum_{k=r+s}^{\infty} k\binom{k-1}{s-1}p^s q^{k-s} + \sum_{k=r+s}^{\infty} k\binom{k-1}{r-1}p^{k-r}q^r \\
&= \sum_{k=s}^{\infty} k\binom{k-1}{s-1}p^s q^{k-s} - \sum_{k=s}^{r+s-1} k\binom{k-1}{s-1}p^s q^{k-s} \\
&\quad + \sum_{k=r}^{\infty} k\binom{k-1}{r-1}p^{k-r}q^r - \sum_{k=r}^{r+s-1} k\binom{k-1}{r-1}p^{k-r}q^r \\
&= \frac{s}{p} - p^s \sum_{k=s}^{r+s-1} k\binom{k-1}{s-1}q^{k-s} + \frac{r}{q} - q^r \sum_{k=r}^{r+s-1} k\binom{k-1}{r-1}p^{k-r} \\
&= \frac{s}{p} - p^s \sum_{k=s}^{r+s-1} s\cdot\frac{k}{s}\binom{k-1}{s-1}q^{k-s} + \frac{r}{q} - q^r \sum_{k=r}^{r+s-1} r\cdot\frac{k}{r}\binom{k-1}{r-1}p^{k-r} \\
&= \frac{s}{p} + \frac{r}{q} - sp^s \sum_{k=s}^{r+s-1} \binom{k}{s}q^{k-s} - rq^r \sum_{k=r}^{r+s-1} \binom{k}{r}p^{k-r}.
\end{aligned}$$

THEOREM 2. *If $w(s,r)$ denotes the waiting time for s successes and r failures in Bernoulli trials, where the probability of success is p and the probability of failure is $q = 1 - p$, then*

$$w(s,r) = \frac{s}{p} + \frac{r}{q} - sp^s \sum_{k=s}^{r+s-1} \binom{k}{s}q^{k-s} - rq^r \sum_{k=r}^{r+s-1} \binom{k}{r}p^{k-r}. \tag{11}$$

When we apply (11) in the special case of $s = 7, r = 5$ with $p = q = \frac{1}{2}$, we again get $15\frac{81}{512}$. For $p = 0.514$ and $q = 0.486$, we get (approximately) 14.745. And note that formula (11), unlike formula (7), is *obviously* symmetric in the sense that (11) is unchanged if we interchange both s with r and p with q.

Conclusion Two different strategies led to two different looking but equivalent solutions to the same problem. The front-end strategy is simple and straightforward, but soon leads one into rather heavy algebraic manipulation to establish a pattern. Induction confirms *that* the pattern holds in general, but still leaves one wondering *why*.

The back-end strategy, while requiring a stronger conceptual background in probability, leads to a simpler derivation of a qualitatively more satisfying solution: formula (11) is obviously symmetric. Quantitatively, however, formula (7) has an edge. The summation in formula (7) has r terms, while the two summations in formula (11) produce $r + s$ terms. The computational advantage of formula (7) suggests that it would be worthwhile to deduce (7) using methods similar to the ones used to deduce (11). Here is such a deduction.

The waiting time for s successes and r failures is equal to the waiting time for s successes plus the weighted average of the waiting times for k failures ($k = 0, 1, 2, \dots, r$), assuming that $r - k$ failures occurred while we waited for the s^{th} success. The waiting time for s successes is, of course, s/p, and the waiting time for k failures is k/q. The probability (weight) that $r - k$ failures occur among the first

$s-1$ successes, and that then the s^{th} success occurs, is $\binom{s-1+r-k}{r-k}p^{s-1}q^{r-k}p$. Thus

$$w(s,r)=\frac{s}{p}+\sum_{k=0}^{r}\binom{s-1+r-k}{r-k}p^{s}q^{r-k}\frac{k}{q}$$

which reduces to (7).

Acknowledgment. I would like to acknowledge a heavy debt of gratitude to two anonymous referees and to my longtime friend and colleague, Chung-Tuo Shih, for their insights that shaped this paper.

REFERENCES

1. F. Mosteller, R. Rourke, and G. Thomas, Jr., *Probability with Statistical Applications*, 2nd edition, Addison Wesley, Reading, MA, 1970.

Sailing Around Binary Trees and Krusemeyer's Bijection

YUJIN SHEN
The Richard Stockton College of New Jersey
Pomona, NJ 08240

JIANG T. SHULT
Kansas State University
Manhattan, KS 66506

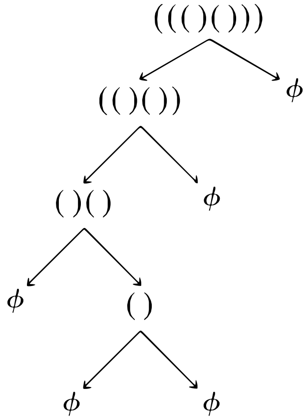
Introduction The Catalan numbers, defined numerically by the equation

$$c_n=\frac{1}{n+1}\binom{2n}{n},$$

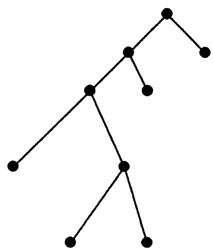
can be used to describe many mathematical models. For example, Mark Krusemeyer presented two different types of parenthesizations in [3]. The first type of parenthesization is a sequence of n pairs of matching parentheses such that, up to any point in the sequence, the number of left parentheses is greater than or equal to the number of right parentheses. Such a sequence of n pairs of parentheses is called *well-defined*. For example, (( )) is a well-defined sequence, while ( )) (is not. The second type of parenthesization is associated with possible ways to multiply $n+1$ different things if the multiplication is non-associative and non-commutative. For example, $(ab)c$ and $a(bc)$ are two different products of three letters. We use Krusemeyer's terminology to call a product of $n+1$ different things with n pairs of parentheses, or $n-1$ pairs if you omit the outer pair, a *bracketing*. Let A_n be the set of well-defined sequences of n pairs of parentheses and B_n the set of bracketings with n multiplications. It is known (see, e.g., [4] and [5]) that $|A_n|=|B_n|=c_n$.

Richard Guy remarked in [2] that "an examination of the symmetries in the two cases makes it unlikely you'll find a direct combinatorial comparison" between the two types of parenthesizations. Interestingly, Krusemeyer [3] constructed a direct recursive correspondence from A_n to B_n, and proved, by induction, that the correspondence is a bijection. Krusemeyer's bijection is the first direct recursive bijection from A_n to B_n. Then David Callan [1] obtained a direct combinatorial bijection from A_n to B_n and proved that his bijection agreed with Krusemeyer's. In this paper, we will present bijections to both A_n and B_n from the set C_n of *full binary trees* with $2n+1$ vertices. (A full binary tree is a tree such that every vertex, except the leaves, has exactly two sons.) Two bijections from A_n to C_n will be constructed, one recursively and one by a direct combinatorial method. It can be shown easily by induction that these two bijections are equivalent, although they appear completely different.

Full binary trees and Krusemeyer's bijection We will obtain first a bijection from A_n to C_n, and then a bijection from C_n to B_n; both bijections are based on Krusemeyer's bijection [3]. Starting with a well-defined sequence of parentheses, Krusemeyer recursively breaks the sequence down into smaller, well-defined sequences at the breaking point, which is the first point in the sequence where the number of left parentheses is equal to the number of right parentheses, until no parentheses are left. For example, the sequence ((( ) ( ))) is broken down as shown below, where ϕ is the empty set of no parentheses:

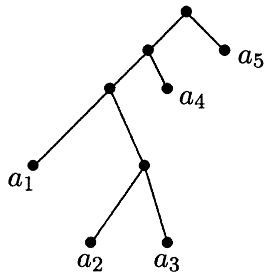


This process of breaking down the sequence gives rise naturally to a full binary tree with $2n+1$ vertices. In the preceding example, we get the following tree:

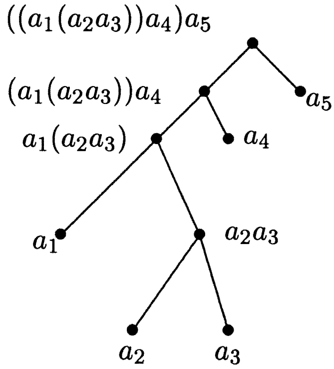


The process is reversible, in the following manner. Assign ϕ to each leaf of the full binary tree. Starting at the lowest leaves, we put a pair of parentheses around whatever the left son has and put them before whatever the right son has and then put the whole thing at the vertex of their father. Continuing this process until the root is reached, we will get ((() ())). The idea illustrated in the example extends to give the desired bijection from A_n to C_n.

Next, we describe a correspondence from C_n to B_n using the full binary tree above as an example. Label leaves from left to right using letters a_1, a_2, a_3, a_4, and a_5, as follows:



Starting with the lowest leaves (a_2 and a_3 in our example), we multiply the labels for the sons of each vertex and use the product as a label for that vertex. As we move up the tree, parentheses become necessary, and when we reach the root, we will have a bracketing of $a_1, a_2, \ldots a_n$. In the example, the labeled tree will look like this:

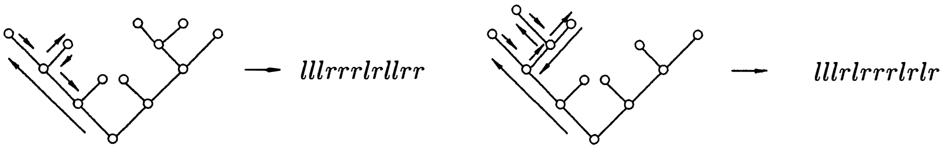


We started at ((() ())) and ended at $((a_1(a_2a_3))a_4)a_5$; this is the fourth example of Krusemeyer's note **[3]** on page 260. Notice that the process is also reversible, and that the process extends to a correspondence from C_n to B_n. The proof that the correspondences from A_n to C_n and from C_n to B_n are actually bijections is very similar to the proof outlined in **[3]**.

The full binary tree representation makes Krusemeyer's bijection more natural and visible. Stanton and White **[5]** give an indirect bijection from A_n to C_n through the set of binary trees on n vertices and the set of ordered trees on $n + 1$ vertices.

Sailing around the tree We now present a second bijection from A_n to C_n. For convenience, full binary trees are oriented in the upper half-plane, with a unique root vertex on the x-axis and every non-leaf vertex being a fork with two upward edges, one to the left and one to the right, labeled "l" or "r" respectively. Two edges are said to be matched if they are partners in a fork.

Sir Francis Drake circumnavigates this tree in a clockwise direction, starting at the records in his log the type of each edge *that he has not previously encountered*, marking an "l" for a left edge and an "r" for a right edge. The left-to-right order of Drake's entries reflects the chronological order in which he encounters edges. (During the periods when he passes along an edge whose other side was previously recorded, he makes no entry in the logbook.) Here are two sample voyages and their corresponding logbooks:

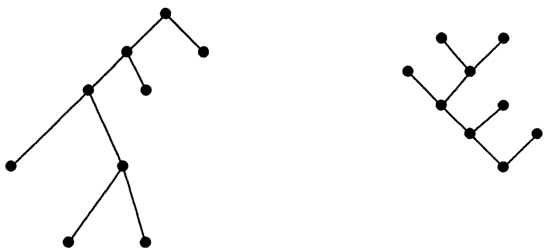


Each logbook corresponds to a well-defined sequence of parentheses, since each right edge possesses a previously recorded matching left edge and each edge belongs to a matching pair. Each tree, moreover, is easily reconstructed from Drake's logbook. One begins by fixing a root vertex; upon scanning the log-word, one carries out the following directions:

1. If an "l" is read, construct an edge reaching upward and to the left of the present vertex, and then climb that edge to its end vertex.
2. If an "r" is read, slide directly down the tree built so far until you first encounter a vertex not already equipped with a right edge. At that vertex, sprout a branch upward to the right and then climb it.
3. When all letters are used, slide back to the root vertex.

This recipe produces one edge for each letter in the log. It also requires a tree builder to clamber all over this tree, coming to rest at the end of each step, at each of the vertices of the constructed tree, exactly once. Each leaf is created exactly when one reads an "r" or has run out of letters. Note also that one is never asked to sprout a left edge or a right edge where an edge has already existed. Thus each non-leaf vertex has sprouted both a left and a right edge. The result is a full binary tree of C_n. One can easily show by induction that such a correspondence is a bijection.

The two bijections are equivalent An easy inductive proof shows that the two bijections from A_n to C_n presented above are equivalent. We omit the proof, but we will use the sequence ((()())), which is the same as *lllrlrrr*, as an example. By the first bijection, we have already generated the full binary tree below at left:



Now, by the second bijection, we generate the full binary tree above at right. Turning either tree upside down produces the other.

Acknowledgment. We thank the referee for pointing out Callan's letter [**1**] and for detailed suggestions on exposition.

REFERENCES

1. David Callan, Letter to the editor, this MAGAZINE 71 (1998), 240.
→ R.K. Guy, The second strong law of small numbers, this MAGAZINE, 63 (1990), 3–20.
→ Mark Krusemeyer, A parenthetical note, this MAGAZINE, 69 (1996), 257–260.
4. Ivan Niven, *Mathematics of Choice*, Random House, New York, NY, 1965, 141–152.
5. Dennis Stanton and Dennis White, *Constructive Combinatorics*, Springer-Verlag, New York, NY, 1986, 59–64.

The £450 Question

J. H. E. COHN
Royal Holloway University of London
Egham, Surrey TW20 0EX
United Kingdom

1. Introduction Only rarely does Pure Mathematics make an appearance in the general press, and so it is perhaps noteworthy that in its 1995 New Year's edition the *Sunday Telegraph* in England offered a prize of £450, over $700, for a solution to the following problem:

PROBLEM. *Firstly, find positive integers* A, B, C, D *such that* $\frac{A^3}{B^3} + \frac{C^3}{D^3} = 6$ *where each of* A, B, C, D *is less than* 100, *and secondly, either give a second solution to the above equation where all the variables are integers exceeding* 100 *with no common factors, or demonstrate that no such second solution can exist.*

The first part is quite easy, although according to [**2**, p. 572], Legendre thought that no such integers existed. Assuming the rational numbers A/B and C/D to be in lowest terms, as we obviously may, multiplying by B^3 we see that B^3C^3/D^3 must be an integer and so $D \mid B$ and similarly $B \mid D$, whence $B = D$. Thus the question really concerns pairwise coprime positive integer solutions to the equation $X^3 + Y^3 = 6Z^3$. Since the cube of any integer is congruent to one of 0, ± 1 modulo 9, it follows that for any solution $3 \mid Z$ and hence that $81 \mid X^3 + Y^3 = (X+Y)(X^2 - XY + Y^2)$. Since 3 does not divide both X and Y, the second factor is not divisible by 9, and hence 27 divides $X + Y$. Also, $X + Y$ must be even, and writing $X + Y = 54k$, $X - Y = 2l$ and $Z = 3Z'$ we obtain $k(l^2 + 243k^2) = Z'^3$. Clearly $k = 1, l = 10, Z' = 7$ satisfy this, whence $A = 37, B = D = 21, C = 17$ is the required first solution.

Writing $y = l/k$ and $x = Z'/k$, we see that finding pairwise coprime non-zero integers satisfying $X^3 + Y^3 = 6Z^3$ is equivalent to finding rational points on the elliptic curve $y^2 = x^3 - 243$; a similar argument shows that for any positive integer a, $X^3 + Y^3 = aZ^3$ leads to $y^2 = x^3 - d$ for suitable d. These curves have been extensively studied, and it is known that unless $d = 432$ if there is even one rational point then there are infinitely many, that the tangent-secant method described below in §2 defines a group operation, and that the group is finitely generated. In our case, we have already seen that there is a rational point, and it transpires that there is just one generator defined by the solution (37, 17, 21) to the original problem. (See [**1**, §82] for further details.) In this note we shall however avoid the use of these methods and use elementary methods as far as possible, in accordance with the spirit of the question.

This would appear to solve the given problem, since we now know that there are infinitely many pairwise coprime integer solutions to the equation $X^3 + Y^3 = 6Z^3$. The trouble is that the variables are required to exceed 100, and in particular to be *positive*. Most of the methods for finding further solutions seem to produce values of X and Y of opposite signs, and the numerical values rapidly increase. Another way of looking at the problem is to seek rational points on the curve $x^3 + y^3 = 6$ depicted in FIGURE 1, with some features we shall need. It has infinitely many rational points, and there appears no reason to suppose that there are not infinitely many in the first quadrant. Finding them, or even proving their existence, is another matter.

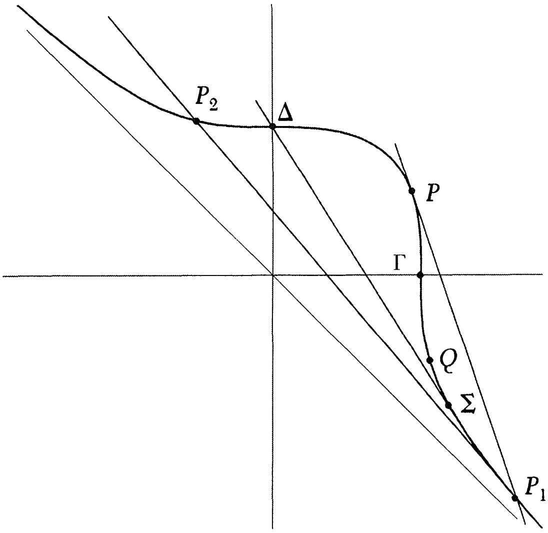


FIGURE 1

Our plan is as follows. In the following section we describe the tangent-secant method, and use it to prove that there are infinitely many pairwise coprime positive integer solutions to the equation $X^3 + Y^3 = 6Z^3$, and use repeated applications of the tangent method to find a solution to the original problem. In §3 we show that the solution obtained in this way is not the smallest such solution, by finding a smaller one; §4 demonstrates that the solution found in §3 is the smallest solution, and finally in §5 we tabulate the first two positive solutions of the equation $X^3 + Y^3 = aZ^3$ for the next eight values of a for which such solutions exist.

2. The tangent–secant method Since any straight line meets a cubic curve with rational coefficients in three points, it follows that if two of these are rational the third must also be so. Thus the tangent at any rational point will intersect the curve at a new rational point—neither of the points of inflection on the curve, nor the point at which the tangent is parallel to the asymptote are rational points. Similarly, the chord joining two rational points that are not reflections of each other in the axis of symmetry, $x = y$, will meet the curve in a third rational point.

Given any rational point P in homogeneous coordinates (X, Y, Z), where X and Y have no common factor, the tangent at P intersects the curve again at P_1, called the *tangential* of P in [4], where

$$X_1 = X(X^3 + 2Y^3),\ Y_1 = -Y(2X^3 + Y^3),\ Z_1 = Z(X^3 - Y^3),$$

and it is easily verified that X_1 and Y_1 also have no common factor. Since $|Z_1| > |Z|$, from any given starting point we obtain by successive applications of this process infinitely many distinct rational points having steadily increasing denominators, which without loss of generality we always take to be positive. From $P(37, 17, 21)$ we obtain $P_1(2237723, -1805723, 960540)$ followed by P_2 with coordinates

$$-1276454530530789553459441,$$

$$2983517129811443394501441,$$

$$1641849890114429433751236 0.$$

Although we may construct infinitely many rational points in this way from any starting point, not every rational point is a tangential, since for any tangential it is easily seen that $4 \mid Z_1$. There are certainly points on the curve for which this is false, e.g., P itself.

We have not yet proved that there are infinitely many *positive* rational points, but this presents little difficulty. Denote by Γ and Δ the points where the curve meets the axes, and by Σ the point whose tangent passes through Δ; these points are *not* rational points. Let the reflection of P_2 in the axis of symmetry be Q. We already know that there are infinitely many rational points, and by symmetry, there are infinitely many on the section to the right of Δ. If there are infinitely many on the section between Γ and Δ, the proof is complete. Otherwise there must be infinitely many on the part of the curve to the right of Γ. This section of the curve is convex, as is easily verified. Thus the tangent at any rational point between Γ and Σ will meet the curve again at a positive rational point, and so if there are infinitely many rational points between Γ and Σ, the proof will be complete. On the other hand, if there were not, then there would have to be infinitely many rational points to the right of Σ. For any point to the right of P_1 the secant joining this point to P will meet the curve again between Δ and P, and finally for any point between Σ and P_1 its join to Q will meet the curve again in the positive quadrant.

Although we have proved the existence of infinitely many positive rational solutions, we have not yet demonstrated even one, but repeating the above tangent process yields P_3 with coordinates

X = 67,795,980,510,382,142,472,326,399,266,506,183,877,357,337,513,870,737,934, 706,199,386,093,375,292,356,829,747,318,557,796,585,767,361;
Y = 792,220,572,662,549,608,190,252,926,112,121,617,686,087,939,438,245,665,806, 051,608,621,113,641,830,336,450,448,115,419,524,772,568,639;
Z = 436,066,841,882,071,117,095,002,459,324,085,167,366,543,342,937,477,344,818, 646,196,279,385,305,441,506,861,017,701,946,929,489,111,120,

the solution produced by Kevin Brown which won the £450 prize. Whether an answer involving numbers with 102 digits was expected is questionable. Although this does indeed give a second solution as required by the newspaper, it is not the next smallest solution. It is actually the third smallest solution, not the second solution as we shall see presently. P_5 gives a further solution in which Z has over 1600 digits. I shall not trouble the reader, or the printer, by recording its coordinates in full.

3. A second method We show next that Brown's solution is not the smallest. As is shown in [**3**, p.129] any solution of $AR^3 + BS^3 + CT^3$ where $ABC = 6$ gives one of $X^3 + Y^3 = 6Z^3$ on substituting $X = (U+V)^3 - 9UV^2$, $Y = 9U^2V - (U+V)^3$, $Z = 3RST(U^2 - UV + V^2)$ where $U = AR^3$ and $V = -BS^3$. There are only the two essentially different cases $A = 1, B = 2, C = 3$ and $A = B = 1, C = 6$ and a preliminary calculation found four small solutions:

A	B	C	R	S	T
1	2	3	1	1	−1
1	2	3	5	−4	1
1	2	3	655	−488	−253
1	1	6	37	17	−21

The result of the above process on the first solution of this table gives the last, equivalent to the point P mentioned above. Starting with the second we obtain the solution P_1, or rather its reflection in the line $x = y$, and one further iteration gives the solution

$$X = 1{,}498{,}088{,}000{,}358{,}117{,}387{,}964{,}077{,}872{,}464{,}225{,}368{,}637{,}808{,}093{,}957{,}571{,}271{,}237;$$
$$Y = 1{,}659{,}187{,}585{,}671{,}832{,}817{,}045{,}260{,}251{,}600{,}163{,}696{,}204{,}266{,}708{,}036{,}135{,}112{,}763;$$
$$Z = 1{,}097{,}408{,}669{,}115{,}641{,}639{,}274{,}297{,}227{,}729{,}214{,}734{,}500{,}292{,}503{,}382{,}977{,}739{,}220,$$

which we shall prove to be the solution of the problem with smallest Z.

Repeated iterations on the last two do not produce a positive solution below $10^{20{,}000}$.

4. The Diophantine equation $X^3 + Y^3 = 6Z^3$ We now seek to prove that the solution just obtained is the smallest; obviously brute force computation is out of the question, and we develop parametric forms for the variables X, Y, and Z in which the bounds on the parameters are considerably smaller than the Z we have above, which exceeds 10^{57}. As above we stipulate that $(X, Y) = 1$; it then follows that $Z = 3Z'$ and, since X and Y cannot both be divisible by 3, that

$$\frac{X+Y}{54} \cdot \frac{X^2 - XY + Y^2}{3} = Z'^3$$

where the two factors on the left have no common factor. Thus $X + Y = 54\sigma^3$, say, and then if $X - Y = 2\rho$ we obtain

$$\rho^2 + 243\sigma^6 = \tau^3, \tag{1}$$

where $Z' = \sigma\tau$. Since we supposed X and Y to have no common factor, the same is now true of ρ and 3σ, and so ρ, 3σ, and τ are pairwise coprime. Also ρ and σ have opposite parity and so τ must be odd. Conversely, any such solution of (1) in integers ρ, σ, and τ leads to one of the given equations $(\rho + 27\sigma^3, -\rho + 27\sigma^3, 3\sigma\tau)$. To avoid confusion we use square brackets to denote a solution $[\rho, \sigma, \tau]$ of (1). The condition for the point to be a positive rational point now takes the form $\rho < 27\sigma^3$, assuming both ρ and σ to be positive, and the tangential of $[\rho, \sigma, \tau]$ is $[8\rho^4 - 36\rho^2\tau^3 + 27\tau^6, 2\rho\sigma, \tau(\rho^2 + 2187\sigma^6)]$.

We now consider the equation (1) in its own right. We first factorize in the field $\mathbb{Q}[\sqrt{-3}\,]$. The two factors on the left of the equation $(\rho + 9\sigma^2\sqrt{-3}\,)(\rho + 9\sigma^2\sqrt{-3}\,) = \tau^3$ are coprime in the field and so each must be a cube times a unit. The only units in the field are ± 1, $\pm\frac{1}{2}(1 + \sqrt{-3}\,)$ and $\pm\frac{1}{2}(1 - \sqrt{-3}\,)$, and so it follows by absorbing the $\pm$ sign into c and d and taking conjugates if necessary that there are only two cases to consider, with $\rho + 9\sigma^2\sqrt{-3}$ equaling either $\frac{1}{2}(1 + \sqrt{-3}\,)(\frac{1}{2}(c + d\sqrt{-3}\,))^3$ or $(\frac{1}{2}(c + d\sqrt{-3}\,))^3$, where in each case c and d are rational integers of like parity. There is no loss of generality in assuming c and d to be even. For if they were odd and $c \equiv d \pmod 4$, then we may replace $\frac{1}{2}(c + d\sqrt{-3})$ by

$$\tfrac{1}{2}(c + d\sqrt{-3}) \cdot \tfrac{1}{2}(-1 + \sqrt{-3}) = -\frac{c + 3d}{4} + \frac{c - d}{4}\sqrt{-3},$$

and similarly if $c \equiv -d \pmod 4$, on multiplying by $-\frac{1}{2}(1 + \sqrt{-3})$. Thus let $c = 2a$ and $d = 2b$. Then $\rho + 9\sigma^2\sqrt{-3}$ equals either $\frac{1}{2}(1 + \sqrt{-3})(a + b\sqrt{-3})^3$ or $(a + b\sqrt{-3})^3$, with $\tau = a^2 + 3b^2$ and since τ is odd, a and b have opposite parity; furthermore, since $3 \nmid \tau$, we have $3 \nmid a$.

The first case would imply that $18\sigma^3 = a^3 + 3a^2b - 9ab^2 - 3b^3$ and hence $3 \mid a$ which is impossible. The second gives $9\sigma^3 = 3a^2b - 3b^3$ and hence $3\sigma^3 = b(a + b)(a - b)$. Here the three factors on the right are pairwise coprime, for any common factor of any two would divide both a and b and hence both ρ and σ. The cases that can now arise are as follows:

1. $b = L^3$, $a \pm b = M^3$, $a \mp b = 3N^3$, whence $\pm 2L^3 - M^3 + 3N^3 = 0$;
2. $b = 3L^3$, $a + b = M^3$, $a - b = N^3$, whence $6L^3 - M^3 + N^3 = 0$.

Thus making the appropriate sign changes, it is easily verified that the process outlined in §3, unlike that of §2, has a converse; not only can solutions of $X^3 + Y^3 = 6Z^3$ be found from known solutions of $AR^3 + BS^3 + CT^3 = 0$ where $ABC = 6$, but for each solution of the former, there is a solution of an appropriate equation of the latter type. Thus for example the solution P_3 yields the solution $L = 160841972528$, $M = 120418942015$, $N = 129900299507$ of the equation $2L^3 = M^3 + 3N^3$.

We use this method to prove that the solution found above in §3 is the smallest solution, by which we mean the positive solution with the smallest Z, apart from P. We find that in either case $\sigma = LMN$, $Z = 3\sigma\tau$, and $\tau > 3^{\frac{5}{3}}\sigma^2$. Thus $|LMN| \leq 3^{-\frac{8}{9}}Z^{\frac{1}{3}}$. In particular, for any smaller solution than the one we have, we find that we should have to have $|LMN| < 3.9 \cdot 10^{18}$. While this number is a considerable improvement on the previous bound for XYZ, it is still too large for effective computation and we need a further reduction.

The method of effecting this depends on which of the two cases above arises. In the second case, we have returned to our original equation, and then the process can be repeated. Omitting the details, we find that at the next stage we must obtain a solution of one of the above equations with $|LMN| < 500$. These were tabulated above without finding a smaller iterated solution, and so this completes the discussion of the second case.

In the first case, we consider the equation $M^3 - 2L^3 = 3N^3$, where we seek solutions in integers of either sign having no common factor, and factorize in the field $\mathbb{Q}[\varphi]$ where $\varphi^3 = 2$. This field has fundamental unit $\varepsilon = \varphi - 1$, class number $h = 1$, and the rational prime 3 factors completely as $\varepsilon(\varphi + 1)^3$. We thus obtain $N = c^3 + 2d^3 + 4e^3 - 6cde$ with $M - \varphi L = \varepsilon^r(\varphi + 1)(c + d\varphi + e\varphi^2)^3$ for some rational integers c, d, e having no common factor, and $r = 0$, 1 *or* 2. The cases $r = 1$ *or* 2 are easily dismissed since each would imply that $3 \mid M$ and so we obtain

$$M = c^3 + 2d^3 + 4e^3 + 12cde + 6(ec^2 + cd^2 + 2de^2)$$

$$-L = c^3 + 2d^3 + 4e^3 + 12cde + 3(c^2d + 2d^2e + 2e^2c)$$

$$0 = (c^2d + 2d^2e + 2e^2c) + (ec^2 + cd^2 + 2de^2)$$

whence $M-2L-3N=54cde$. Then we find $\tau=a^2+3b^2=2L^6+\frac{1}{2}(M^6+9N^6)$ and so

$$\frac{2\tau}{3}=\frac{M^6+4L^6+9N^6}{3}\geq\frac{1}{81}\frac{M^6+2^6L^6+3^6N^6}{3}$$
$$\geq\frac{1}{81}\left(\frac{|M|+2|L|+3|N|}{3}\right)^6\geq\frac{1}{81}(18cde)^6$$

or $\tau\geq\frac{1}{54}(18cde)^6$. For a positive solution of the original problem $\rho<27\sigma^3$, whence $\tau^3<972\sigma^6$, and so since $Z=3\sigma\tau$ we obtain $Z^6=3^6\sigma^6\tau^6>\frac{3}{4}\tau^9>\frac{3}{4}(\frac{1}{54}(18cde)^6)^9$ and so for any smaller solution than the one we have, we need $cde<237{,}000$. We also have $(4e(c+d)+c^2+2d^2)^2=c^4-8c^3d-12c^3d^2-8cd^3+4d^4$, and so the form on the right hand side must be a square. Clearly c must be odd because otherwise L, M, and N would all be even, and a short calculation reveals that the only such solutions in the range are

c	d	e	L	M	N
1	0	0	−1	1	1
1	−1	0	4	5	−1
5	−1	−4	−488	655	−253
5	20	−4	−3449	26309	18269

Of these, the first gives the solution $37, 17, 21$ and the others do not give positive solutions, concluding the proof.

5. Other values of a Adapting the argument of §2 slightly will show that for any a for which $x^3+y^3=a$ has infinitely many rational solutions there will be infinitely many positive ones. We list below the two smallest positive solutions of $X^3+Y^3=aZ^3$ for the next eight values of a for which there are infinitely many integer solutions. It is seen that the size of these is rather unpredictable.

a	Smallest positive solution	Second positive solution
7	(5, 4, 3)	(9226981, 4381019, 4989780)
9	(2, 1, 1)	(415280564497, 676702467503, 348671682660)
12	(89, 19, 39)	(14243182991376555689930151122399, 13563910882413198071001945741601, 7655979840040760645368155264060)
13	(7, 2, 3)	(26441619018689, 40343602894936, 18636783082845)
15	(397, 683, 294)	Z has 104 digits
17	(11663, 104940, 40831)	(6686709776467673063832642547079, 23985426834891954919065940763238, 9395089174626523412779432507327)
19	(5, 3, 2)	(8, 1, 3)
20	(19, 1, 7)	(3556849302231767279, 749958645746307601, 1314436091995106388)

Acknowledgment. The author wishes to express his sincere thanks to the referee for some very helpful suggestions.

REFERENCES

1. B. N. Delone and D. K. Faddeev, *The Theory of Irrationalities of the Third Degree*, American Mathematical Society, Providence, RI, 1964.
2. L. E. Dickson, *History of the Theory of Numbers, Vol. II*, Carnegie Institution, Washington, DC, 1921 (reprint Chelsea, New York, NY, 1952).
3. L. J. Mordell, *Diophantine Equations*, Academic Press, London, UK, 1966.
4. J. J. Sylvester, On certain ternary cubic-form equations, *Amer. J. Math.*, 2 (1879) 280–285.

How to Determine Your Gas Mileage

ALEXANDER J. IZZO
Bowling Green State University
Bowling Green, OH 43403

Introduction Suppose we want to determine the gas mileage of a car. By definition,

$$\text{gas mileage} = \frac{\text{miles driven}}{\text{gallons of gas used}}.$$

So all we have to do is fill up the tank, drive around a bit, refill the tank, and then divide the miles driven by the amount of gas we put into the tank the second time. The trouble with this approach is that gas mileage varies a great deal with driving conditions. In particular, gas mileage in the city is generally much lower than that on the highway. We really want to determine two gas mileages: one for the city and one for the highway. If each time between filling the tank, the car is driven only in the city or only on the highway, then there isn't any problem, but generally between fuelings we do both city and highway driving. We want a method that can handle this situation. We will present such a method below using least squares. This method is suitable for inclusion in a linear algebra course and is quite satisfactory in practice. However, it is not optimal. Improving the method will lead us naturally to the introduction of weighted least squares. Good references for the elementary linear algebra used in this paper include **[1]** and **[2]**.

To obtain the data needed to apply the method described below, the fuel tank must sometimes be filled completely so that we know how much gas has been used. In addition, the driver must record:

(i) the amount of gas put into the gas tank at each fueling;
(ii) the odometer reading each time the tank is filled;
(iii) the odometer reading before and after each highway trip.

From these records we obtain the required data, namely, the amount of gas used between the $(j-1)^{th}$ and j^{th} times the tank is filled and the distances driven in the city and on the highway during that interval. We will use the following notations:

g_j = gallons of gas used between the $(j-1)^{th}$ and j^{th} fillings of the tank;
c_j = miles driven in the city during that interval;
h_j = miles driven on the highway during that interval;
y_1 = gas mileage in the city;
y_2 = gas mileage on the highway.

Our goal is to determine the values of the unknowns y_1 and y_2 from the data g_j, c_j, and h_j.

The method By the definition of gas mileage, the amount of gas used between the $(j-1)^{th}$ and j^{th} times the tank is filled should be $(c_j/y_1)+(h_j/y_2)$. Thus we obtain one equation for each time the tank is refilled. If the tank is refilled m times, we obtain the following system of m equations in the unknowns y_1 and y_2

$$\frac{c_1}{y_1}+\frac{h_1}{y_2}=g_1, \quad \dots, \quad \frac{c_m}{y_1}+\frac{h_m}{y_2}=g_m.$$

Letting $x_1=1/y_1$ and $x_2=1/y_2$, we obtain the linear system

$$\begin{gathered} c_1x_1+h_1x_2=g_1 \\ \vdots \\ c_mx_1+h_mx_2=g_m. \end{gathered} \tag{1}$$

Setting

$$A=\begin{pmatrix} c_1 & h_1 \\ \vdots & \vdots \\ c_m & h_m \end{pmatrix}, \quad x=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{and} \quad g=\begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix},$$

we can rewrite the system in matrix form as $Ax=g$.

As every good linear algebra student knows, the vectors that can be written in the form Ax are precisely those in the column space of A, so our system has a solution if and only if g is in the column space of A. But the column space of A is at most two-dimensional, while g lies in m-space, so the vector g is almost certain to lie outside the column space of A, as shown in FIGURE 1 below.

Given a system of equations with no solution, it is natural to look for a value of x that makes the difference between the right and left sides as small as possible. That is, we look for a vector $\bar{x}$ that minimizes the (euclidean) length of $A\bar{x}-g$. This is the method of *least squares*. As shown in FIGURE 1, the length of $A\bar{x}-g$ is minimized when $A\bar{x}-g$ is perpendicular to the column space of A.

It is easy to see that a vector is perpendicular to the column space of A if and only if it lies in the null space of A^t. Therefore, we have

$$A^t(A\bar{x}-g)=0 \quad \text{or equivalently} \quad A^tA\bar{x}=A^tg. \tag{2}$$

It is not difficult to see that the matrix A^tA has the same null space as A. So if the columns of A are linearly independent, then the null spaces of A and A^tA each contain only the zero vector, and hence the square matrix A^tA is invertible. Thus,

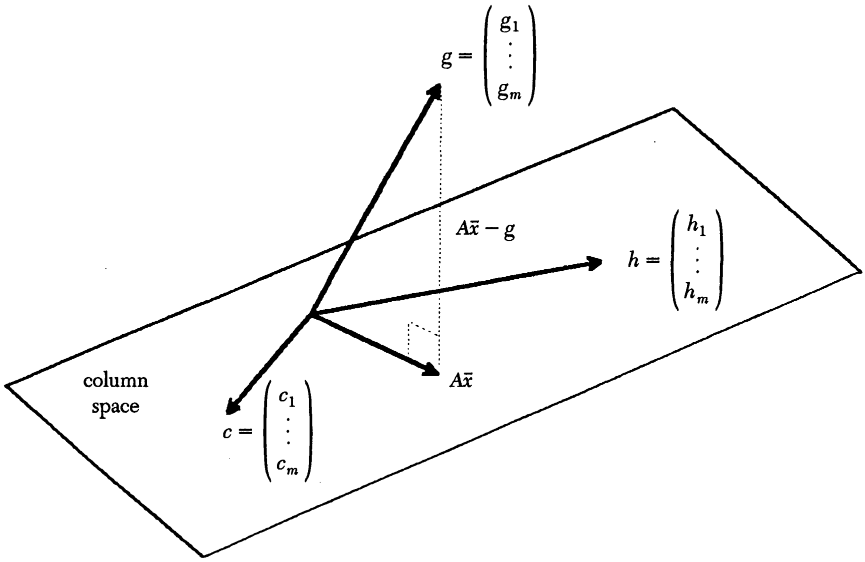


FIGURE 1

assuming that the columns of A are linearly independent, we arrive at our solution:

$$\bar{x} = (A^tA)^{-1}A^tg.$$

If we set $c = (c_1, \dots, c_m)$, $h = (h_1, \dots, h_m)$, and $g = (g_1, \dots, g_m)$, then we can write the solution in terms of dot products as

$$\bar{x} = \begin{pmatrix} c\cdot c & c\cdot h \\ c\cdot h & h\cdot h \end{pmatrix}^{-1} \begin{pmatrix} c\cdot g \\ h\cdot g \end{pmatrix} = \begin{pmatrix} \dfrac{(h\cdot h)(c\cdot g) - (c\cdot h)(h\cdot g)}{(c\cdot c)(h\cdot h) - (c\cdot h)^2} \\ \dfrac{(c\cdot c)(h\cdot g) - (c\cdot h)(c\cdot g)}{(c\cdot c)(h\cdot h) - (c\cdot h)^2} \end{pmatrix}.$$

Recalling that the gas mileages y_1 and y_2 were the reciprocals of x_1 and x_2 and writing out the dot products as sums, we obtain completely explicit formulas for the gas mileages:

$$\begin{aligned} y_1 &= \frac{1}{x_1} = \frac{\Sigma c_j^2 \Sigma h_j^2 - (\Sigma c_j h_j)^2}{\Sigma h_j^2 \Sigma c_j g_j - \Sigma c_j h_j \Sigma h_j g_j} \\ y_2 &= \frac{1}{x_2} = \frac{\Sigma c_j^2 \Sigma h_j^2 - (\Sigma c_j h_j)^2}{\Sigma c_j^2 \Sigma h_j g_j - \Sigma c_j h_j \Sigma c_j g_j}. \end{aligned} \tag{3}$$

We assumed above that the columns of A were linearly independent. If the columns of A are linearly *dependent*, then $\bar{x}$ is not uniquely determined by equation (2), and in fact the set of solutions to equation (2) forms a line, so it is not possible to obtain useful values for the city and highway gas mileages. Of course if m is much bigger than 1, then the columns of A are very unlikely to be linearly dependent.

I presented the problem of determining gas mileage in a linear algebra class at the University of Michigan, and it was very successful. I covered the method of least

TABLE 1 GAS MILEAGE DATA

j	c_j	h_j	g_j	j	c_j	h_j	g_j	j	c_j	h_j	g_j
1	15	283	7.017	18	134	224	10.592	35	207	83	9.074
2	116	171	8.051	19	199	182	10.246	36	52	312	9.519
3	2	367	10.062	20	284	0	10.007	37	76	558	14.841
4	2	300	7.430	21	178	0	6.458	38	0	132	3.148
5	221	69	9.645	22	153	146	9.476	39	11	539	12.356
6	93	215	9.907	23	49	180	6.161	40	194	138	9.832
7	130	87	8.599	24	338	0	13.042	41	334	0	9.872
8	1	360	10.100	25	148	87	7.471	42	304	45	10.786
9	49	219	8.65	26	28	282	9.004	43	147	225	11.234
10	0	273	7.700	27	144	0	5.855	44	0	359	10.200
11	41	268	9.810	28	0	298	6.929	45	129	97	7.824
12	17	321	9.310	29	35	528	14.912	46	193	132	12.032
13	76	613	17.120	30	1	363	8.490	47	256	0	10.762
14	193	0	9.088	31	26	173	5.954	48	194	0	8.507
15	79	178	9.208	32	290	0	14.341	49	41	327	10.310
16	228	0	10.107	33	125	412	15.922	50	233	84	11.113
17	157	191	10.711	34	129	138	8.577				

TABLE 2 GAS MILEAGE RESULTS

Method	City (mi/gal)	Highway (mi/gal)
Unweighted method	26.13	40.02
Weighted method ($a = b = 1$)	25.75	39.47
"Optimal" weighted method ($a = 39.41$, $b = 25.79$)	25.79	39.41

squares for solving a system of linear equations in general and indicated how it was relevant to determining gas mileage but left it to the students to apply the method to obtain explicit formulas. Then, for homework, I gave the students data that I had collected and had them actually determine my gas mileage. The data set used, together with calculated results, are given in Tables 1 and 2 above. (The weighted method is explained below.) The problem of determining gas mileage is an excellent application to cover in a linear algebra class because it is of genuine interest (I really wanted to know the answer!) and because it reinforces many important ideas in linear algebra, including the column space, the null space, linear independence, orthogonality, and the question of when a linear system has a solution.

A better method The method just discussed is quite satisfactory in practice, but it is not optimal. A similar but better method, described below, makes use of inner products other than the standard dot product. While the improved method cannot be included in a course in which the only inner product considered is the dot product, the method would be especially suitable for a course that treats general inner products, since it shows that they arise naturally and are genuinely useful.

To see why the method considered so far is not completely satisfactory consider the special case in which each time between filling the tank the car is driven only in the city or only on the highway. In this case we know how much gas was used in the city and how much on the highway. Hence we can calculate the average number of miles driven per gallon of gas for each type of driving, and clearly these averages are the "correct" values for the gas mileages. That is, we should have

$$y_1 = \frac{\sum c_j}{\sum\limits_{c_j \neq 0} g_j} \quad \text{and} \quad y_2 = \frac{\sum h_j}{\sum\limits_{h_j \neq 0} g_j}. \tag{4}$$

However, since for each j either $c_j = 0$ or $h_j = 0$, our solution (3) gives

$$y_1 = \frac{\sum c_j^2}{\sum c_j g_j} \quad \text{and} \quad y_2 = \frac{\sum h_j^2}{\sum h_j g_j}.$$

Thus, in this special case, the method of least squares gives a *weighted* average with an *inappropriate weighting*.

We rectify this by using *weighted* least squares. Recall that we want to "solve" the linear system (1). Let us call the left hand side of the j^{th} equation the amount of gas predicted to have been used during the j^{th} interval. (Of course the amount of gas *actually* used during the j^{th} interval is g_j.) It is natural to require that the total amount of gas predicted to have been used equal the total amount of gas actually used. (If we were computing just *one* gas mileage we would take the average which is determined by exactly this kind of requirement.) That is, we require that $\Sigma(c_j x_1 + h_j x_2) = \Sigma g_j$. Setting $r_j = c_j x_1 + h_j x_2 - g_j$, this becomes $\Sigma r_j = 0$. If we let $r = (r_1, \ldots, r_m)$, then, in fact, $r = Ax - g$. The method of least squares described above minimizes Σr_j^2 by requiring that r be orthogonal to the columns of A:

$$r \cdot c = 0 \quad \text{and} \quad r \cdot h = 0.$$

The condition $\Sigma r_j = 0$ that we now wish to impose says that r is orthogonal to the vector $(1, \ldots, 1)$. So now we will choose a new inner product $\langle , \rangle$ on $\mathbb{R}^m$ with the property that

$$\text{if } \langle r, c \rangle = 0 \text{ and } \langle r, h \rangle = 0, \quad \text{then } r \cdot (1, \ldots, 1) = 0. \tag{5}$$

Recall that every positive definite symmetric $m \times m$ matrix Q induces an inner product on $\mathbb{R}^m$ by the equation $\langle v, w \rangle = v^t Q w$, and that every inner product arises in this way. Clearly, we should choose an inner product whose corresponding matrix Q is as simple as possible, and this suggests trying to make Q diagonal. The reader can easily check that the inner product induced by a diagonal matrix Q will satisfy (5) provided Q has the form

$$Q = \begin{pmatrix} \dfrac{1}{ac_1 + bh_1} & & 0 \\ & \ddots & \\ 0 & & \dfrac{1}{ac_m + bh_m} \end{pmatrix}$$

where a and b are arbitrary positive numbers. (We require a and b to be positive in order to insure that Q is positive definite.)

Now we carry out essentially the same procedure as before, except that we use the inner product induced by Q instead of the dot product. Our solution will be the vector $\bar{x}$ that minimizes the length of $A\bar{x} - g$ with respect to our new inner product. As before, the length of $A\bar{x} - g$ is minimized when $A\bar{x} - g$ is perpendicular to the column space of A. It is easy to see that $A\bar{x} - g$ is perpendicular to the column space of A with respect to the inner product induced by Q if and only if it satisfies the equation $A^tQ(A\bar{x}-g)=0$, or, equivalently, $A^tQA\bar{x}=A^tQg$. If the columns of A are independent, then A^tQA is invertible, and we arrive at our solution:

$$\bar{x}=\left(A^tQA\right)^{-1}A^tQg.$$

Now we can write the solution out in terms of our inner product:

$$\bar{x}=\begin{pmatrix}\langle c,c\rangle & \langle c,h\rangle\\ \langle c,h\rangle & \langle h,h\rangle\end{pmatrix}^{-1}\begin{pmatrix}\langle c,g\rangle\\ \langle h,g\rangle\end{pmatrix}=\begin{pmatrix}\dfrac{\langle h,h\rangle\langle c,g\rangle-\langle c,h\rangle\langle h,g\rangle}{\langle c,c\rangle\langle h,h\rangle-\langle c,h\rangle^2}\\ \dfrac{\langle c,c\rangle\langle h,g\rangle-\langle c,h\rangle\langle c,g\rangle}{\langle c,c\rangle\langle h,h\rangle-\langle c,h\rangle^2}\end{pmatrix}.$$

Note that this is exactly the same solution as before except that in place of the dot product we have the inner product $\langle,\rangle$ given by the formula

$$\langle v,w\rangle=\sum\frac{v_jw_j}{ac_j+bh_j}.$$

Note also that in the special case in which each time between filling the tank the car is driven only in the city or only on the highway, the above solution reduces to (4), as desired.

What values should be used for a and b? The optimal values depend on how the variations in gas mileage in the city and on the highway compare. If we assume that the two variations are the same, then one can show that we should take $a=x_1$ and $b=x_2$, but of course we can't actually do that since we are trying to *find* x_1 and x_2. We could use an iterative procedure, but there is really no need to do so because the result is not very sensitive to the values used. It is quite satisfactory simply to take $a=b=1$. (Note that only the *ratio* a/b matters.) For instance, as shown in Tables 1 and 2, the results obtained from the data there with $a=x_1$ and $b=x_2$ differ by less than 0.1 mi/gal from the results obtained with $a=b=1$.

It is clear that, at least in theory, the weighted method is superior to the unweighted method. In practice, the two methods yield similar results, although the values obtained using the unweighted method tend to be slightly too high. This is illustrated by the results in Table 1. Using the gas mileages computed by the unweighted method, the total amount of gas predicted to have been used is 480 gallons, whereas the total amount actually used is 487 gallons. Naturally this discrepancy does not occur with the weighted method. It is easy to understand why the results of the unweighted method tend to be too high: the unweighted method in effect gives too much weight to those intervals in which many miles are driven, and these tend to be the intervals with higher gas mileage since higher gas mileage allows more miles to be driven before refueling.

Acknowledgment. This paper was first presented to the MAA on January 12, 1994, under the title *Gas mileage: an application of linear algebra.* The author was partially supported by an NSF postdoctoral fellowship.

REFERENCES

1. Bill Jacob, *Linear Functions and Matrix Theory*, Springer-Verlag, New York, NY, 1995.
2. Gilbert Strang, *Linear Algebra and its Applications*, 3rd edition, Harcourt Brace Jovanovich, Orlando, FL, 1988.

The Poisson Variation of Montmort's Matching Problem

DON RAWLINGS
California Polytechnic State University
San Luis Obispo, CA 93407

The matching problem In 1708, Pierre Rémond de Montmort [**6**] proposed and solved the following problem:

> **Matching problem** From the top of a shuffled deck of n cards having face values $1, 2, \ldots, n$, cards are drawn one at a time. A match occurs if the face value on a card coincides with the order in which it is drawn. For instance, if the face values of a five-card deck appear in the order 2 5 3 4 1, then there are two matches (3 and 4). What is the probability that no match occurs?

Known as *le problème des rencontres*, Montmort's problem has become a classic. Treated in almost every modern text on combinatorics, it has been recast in a variety of amusing guises including the random switching of n hats by a possibly tipsy cloakroom attendant, the random placement of n letters in n envelopes by a secretary apparently having a bad day, and even in terms of a wager on the ordering in which the horses finish in a given race (for instance, see [**3**], [**4**], and [**8**]).

As is well known (e.g., see [**2**, p. 107]), the principle of inclusion-exclusion may be used to show that the probability $M_{n,k}$ of k matches occurring with an n-card deck is

$$M_{n,k} = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}. \tag{1}$$

For $k = 0$, of course, (1) gives the solution to Montmort's problem. The celebrated fact that the probability of no match occurring is asymptotic to e^{-1}, that is,

$$\lim_{n \to \infty} M_{n,0} = e^{-1} \tag{2}$$

also follows from (1). Incidentally, the distribution of matches $\{M_{n,k}\}_{0 \le k \le n}$ has applications in psychology and in testing psychic powers (e.g., see [**1**], [**2**, p. 108], or [**9**]).

The Poisson variation: marking the cards The following twist, henceforth referred to as the *Poisson variation*, is added to Montmort's problem: Suppose that as each card is drawn it is marked (by underlining its face value) with probability $(1 - t)$ where t is a real number between 0 and 1. As underlined integers are evidently distinct from ordinary ones, they cannot be involved in a match. So, if from a five-card deck the sequence of marked and unmarked face values $\underline{1}\,2\,5\,\underline{4}\,3$ appear, then the only match involves 2.

Perhaps odd at first glance, this twist gives rise to some aesthetically pleasing mathematics. To be precise, it leads to the following natural extensions of (1) and (2) from $t = 1$ to the entire interval $[0, 1]$.

THEOREM 1. *The probability* $P_{n,k}$ *of* k *matches occurring in the Poisson variation of the matching problem with an* n*-card deck is given by*

$$P_{n,k} = \frac{t^k}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j t^j}{j!}.$$

Moreover, $\{P_{n,k}\}_{0 \le k \le n}$ *is asymptotically Poisson with parameter* t:

$$\lim_{n \to \infty} P_{n,k} = \frac{t^k}{k!} e^{-t}.$$

So the asymptotic probability of no matches occurring is now e^{-t}!

Proof. To formalize matters, let Ω_n denote the set generated by first randomly permuting the integers 1 through n and then underlining each integer with probability $(1-t)$ where $0 \le t \le 1$. The probability of $w = \underline{2}\,4\,\underline{1}\,\underline{3} \in \Omega_4$ being generated is therefore $t(1-t)^3/4!$. For $t = 1$, note that Ω_n coincides with the set of permutations S_n of $\{1, 2, \dots, n\}$.

A *fixed point* of $w = w_1 w_2 \dots w_n \in \Omega_n$ is an integer j satisfying $w_j = j$. For example, $w = \underline{1}\,2\,5\,\underline{4}\,3 \in \Omega_5$ has one fixed point ($j = 2$). Furthermore, an element in Ω_n having no fixed points is said to be a *derangement*.

Obviously, a sequence of marked and unmarked face values appearing in the Poisson variation of the matching problem may be naturally identified with an element of Ω_n. Moreover, under this identification, matches correspond to fixed points and the event of no matches evidently coincides with a derangement.

Let $F_{n,k}$ be the number of permutations in S_n having k fixed points. Making the usual appeal that such a permutation may be constructed by selecting k integers to fix and then "deranging" the remaining $(n-k)$ integers gives the well known identity $F_{n,k} = \binom{n}{k} F_{n-k,0}$.

As $M_{n,k} = F_{n,k}/n!$, it follows that

$$M_{n,k} = \frac{F_{n-k,0}}{(n-k)!k!}. \tag{3}$$

Since $\{M_{n,k}\}_{0 \le k \le n}$ is a probability distribution, we also have

$$\sum_{k=0}^{n} M_{n,k} = 1. \tag{4}$$

An element in Ω_n having k fixed points may be generated by first selecting a permutation $\sigma \in S_n$ with l fixed points where $k \le l \le n$ and then underlining $(l-k)$ of σ's fixed points. From the properties of elementary probability, (3), and the binomial theorem, it follows that

$$P_{n,k} = \sum_{l=k}^{n} M_{n,l} \binom{l}{k} t^k (1-t)^{l-k} = \sum_{l=k}^{n} \frac{F_{n-l,0}}{l!(n-l)!} \binom{l}{k} t^k \sum_{j=0}^{l-k} (-1)^j \binom{l-k}{j} t^j$$

$$= \frac{t^k}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j t^j}{j!} \sum_{l=j+k}^{n} \frac{F_{n-l,0}}{(n-l)!(l-k-j)!}. \tag{5}$$

To complete the proof, just note that an index shift and (4) imply that the rightmost sum at the end of (5) is

$$\sum_{\nu=0}^{n-k-j} \frac{F_{n-k-j-\nu\,,0}}{(n-k-j-\nu)!\nu!} = \sum_{\nu=0}^{n-k-j} M_{n-k-j\,,\,\nu} = 1.$$

Concluding remarks Montmort's interest in the matching problem was motivated by the more colorful and complicated problem of determining the banker's advantage in a gambling game known as Treize. A detailed account of Treize and of the work of Montmort and Nicholas Bernoulli on the matching problem is given in **[5]**. Montmort's problem has been extended in a variety of other ways. In this regard, the articles **[1]**, **[3]**, **[7]**, and **[8]** are highly recommended. The question of whether marking cards provides nice extensions of the results in **[1]**, **[3]**, and **[8]** is left open.

Finally, Theorem 1 may be deduced from a result in **[7]** that involves Bernoulli trials, partitions, and q-combinatorics. The preceding proof is far less encumbered.

REFERENCES

1. D. E. Barton, The matching distributions: Poisson limiting forms and derived methods of approximation, *J. Royal Stat. Soc.* 20 (1958), 73–92.
2. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, John Wiley and Sons, New York, NY, 1990.

→ S. Fisk, The secretary's packet problem, this MAGAZINE 61 (1988), 103–105.

4. R. P. Grimaldi, *Discrete and Combinatorial Mathematics: An Applied Introduction*, Addison-Wesley, Reading, MA, 1990.
5. A. Hald, *A History of Probability and Statistics and Their Applications before* 1750, John Wiley and Sons, New York, NY, 1989.
6. P. R. de Montmort, *Essai d'Analyse sur les Jeux de Hazard*, Paris, 1708. Reprinted by Chelsea Publishing, New York, NY, 1980.
7. D. P. Rawlings, A probabilistic approach to some of Euler's number theoretic identities, *Trans. Amer. Math. Soc.* 350 (1998), 2939–2951.
8. G. R. Sanchez, Swapping hats: A generalization of Montmort's problem, this MAGAZINE 71 (1998), 53–57.
9. P. E. Vernon, The matching method applied to investigations of personality, *Psych. Bull.* 33 (1936), 149–177.

Chaos and Love Affairs

P.-M. BINDER
Departamento de Física
Universidad de Los Andes
Apartado Aéreo 4976
Bogotá, Colombia

In [**1**] and [**2**], Steven Strogatz explores the possible dynamics of a two-dimensional system of linear differential equations by interpreting the variables R and J as Romeo's and Juliet's love for each other, respectively. The dynamics is governed by the equations $\dot{R} = aR + bJ$ and $\dot{J} = cR + dJ$, where the constant coefficients a, b, c, d are real numbers, which may be positive or negative. Each combination of signs correspond to a possible human behavior; for example, a "cautious lover" has $a < 0$, $b > 0$. Different values of these coefficients give rise to different dynamics, which can be studied by means of fixed points and linear stability analysis. For these, one can find amusing (or tragic) interpretations, such as love-hate cycles, mutual indifference (fixed point at the origin), and love feasts or war (trajectories that go off to infinity).

In this note I examine the following question: How complicated does a model of love affairs need to be in order to exhibit chaos? To do this, I will summarize the necessary conditions for continuous time models with increasing number of variables. A 1-variable model needs time-delay (external periodic driving is not enough); a 2-variable model needs external periodic driving; and a 3-variable model just needs one nonlinear term (no external forcing is needed). The R–J space interpretations of these conditions suggest a rather simple mnemonic device for students to remember them. In more detail, the necessary conditions for chaos in continuous time systems are as follows:

1. A one-dimensional nonlinear time-delayed equation of the form $\dot{x} = f(x - \tau)$ can be chaotic. This requires defining $f(t)$ in the interval $-\tau < t < 0$, which in principle makes the system infinite-dimensional. In practice one observes chaotic attractors with an effective dimensionality that increases with τ; see the discussion in [**3**, Ch. 8], and references therein. The best known example of this situation is the Mackey–Glass equation for the regeneration of blood cells in patients with leukemia.
2. For a two-dimensional autonomous system, the Poincaré–Bendixson theorem implies that only closed orbits can occur. These are limited to fixed points, limit cycles, and cycle graphs. Therefore, there can be no chaos: see [**2**, pp. 203–210]. However, a system of two nonlinear differential equations with external forcing can exhibit chaos. There is a nice study of the forced nonlinear pendulum in [**4**, Ch. 3], which includes several plots of strange attractors. A forced linear pendulum can exhibit resonance, but not chaos.
3. A system of three or more coupled differential equations with just one nonlinear term (for example, a product of two variables), can have bundles of trajectories that flow, stretch, and fold at the same time; this can result in bounded, deterministic nonperiodic flow with sensitive dependence to initial conditions, in a word, chaos. The simplest example, called the Rössler attractor, is discussed in [**2**, Ch. 12], along with more complicated and better-known examples such as the Lorenz system.

While in no way do I want to imply a rigorous connection between the above statements and the R–J psychological space or its equivalent in other dimensions, they suggest the following mnemonic device to remember when chaos can occur:

To have chaos you need a single person obsessed with the past, a couple affected by the phases of the moon, or a ménage à trois.

It would be a nice exercise to incorporate into this sentence the conditions necessary for chaos in discrete-time dissipative systems, and for Hamiltonian chaos in general; however, I have been unable to find proper interpretations for these in R–J space.

REFERENCES

1. S. H. Strogatz, Love affairs and differential equations, this MAGAZINE 61 (1988), 35.
2. S. H. Strogatz, *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry and Engineering*, Addison-Wesley, Reading, MA, 1996.
3. A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, Springer-Verlag, New York, NY, 1992.
4. G. L. Baker and J. P. Gollub, *Chaotic Dynamics: An Introduction*, Cambridge Univ. Pr., Cambridge, UK, 1990.

Edge-Length of Tetrahedra

HANS SAMELSON
Stanford University
Stanford, CA 94305

For a tetrahedron Δ in Euclidean 3-space, with vertices P, Q, R, S, we define the edge-length $E(\Delta)$ as the sum of the lengths of the six edges $PQ, PR, \dots$. Let two tetrahedra Δ and Δ' with vertices A, B, C, D and A', B', C', D' be given, with Δ' contained in Δ. We shall show that the relation

$$E(\Delta') < \tfrac{4}{3} \cdot E(\Delta) \tag{1}$$

holds, and that the constant $4/3$ is sharp in the sense that there are pairs Δ, Δ' with $E(\Delta')$ as close to $4/3 \cdot E(\Delta)$ as we want. In particular $E(\Delta')$ can be larger than $E(\Delta)$!

We begin by considering degenerate tetrahedra δ with vertices a, b, c, d (not necessarily different from each other) lying in that order on some straight line in our space. The concept edge-length applies to these in the obvious way. The segment ab is covered three times, by ab, ac, and ad; similarly for cd. But bc is covered four times, by ad, ac, bd, and bc. Thus (with $|ab|$ denoting the length of ab) we have the inequality

$$3|ad| \le E(\delta) \le 4|ad|. \tag{2}$$

The extremes can occur. For equality on the left we must have $b = c$; for that on the right we must have $a = b$ and $c = d$.

Next take any segment AB in space, and consider its orthogonal projection $(AB)_l$ on any directed line l through the origin. The set of these lines has a natural identification with the unit-sphere around the origin (represent each line by the point of intersection of its forward half with the sphere). Thus we have on it a measure, namely the usual surface area on the sphere (given by $dS = \sin\theta\, d\theta\, d\phi$ in spherical coordinates); this measure is invariant under rotations of space around the origin.

With fixed A, B and variable l the length $|(AB)_l|$ gives a function on the sphere. We form the average of $|(AB)_l|$, i.e., the integral of this function over the sphere, divided by the total measure 4π. Clearly this average depends only on the length $|AB|$ of the segment AB and not on its position (the integrand is invariant under translations of AB, and the integral is invariant under rotations around the origin).

Just as clearly, it is a multiple $c\cdot|AB|$ of $|AB|$ for some non-zero c. (By computing $\int\int|\cos\theta|\sin\phi\, d\theta d\phi$ one finds $c = \frac{1}{2}$.)

We are now ready to prove our assertion (Equation (1)) about $E(\Delta)$ and $E(\Delta')$. Take any directed line l through the origin, and project the two tetrahedra orthogonally onto the line, producing degenerate tetrahedra δ_l and δ'_l with vertices a, b, c, d and a', b', c', d'. The segment $a'd'$ is part of the segment ad, and so $|ad|$ is at least as large as $|a'\,d'|$. By Equation (2) we have the inequalities $E(\delta'_l) \leq 4|a'\,d'|$ and $3|ad| \leq E(\delta_l)$. From the remark after Equation (2) we see easily that equality in either relation holds only for a set of lines of measure 0 (consisting of a finite number of great circles). We conclude that

$$3E(\delta'_l) < 4E(\delta_l)$$

except for a set of lines l of measure 0.

Averaging over all lines l this turns into the inequality $3c\cdot E(\Delta') < 4c\cdot E(\Delta)$, and so we have

$$3E(\Delta') < 4E(\Delta),$$

which was our claim in Equation (1). ∎

We come to the promised examples. They are immediate if we allow degenerate tetrahedra, by Equation (2): Take $a = a' = b'$ and $d = d' = c'$ and take $b = c$, preferably near the middle of the segment ad. Then $E(\delta) = 3|ad|$ and $E(\delta') = 4|ad|$; thus $E(\delta') = 4/3\cdot E(\delta)$. To get non-degenerate tetrahedra, take $A = a$, $D = d$, and shift b and c slightly to new positions B and C such that A, B, C, D form a non-degenerate tetrahedron Δ. Next move b' and c' slightly to B' and C' such that the tetrahedron Δ' so obtained is non-degenerate and contained in Δ. Then $E(\Delta)$ will be slightly more than $3|ad|$, and $E(\Delta')$ will be at worst slightly less than $4|ad|$. One could even construct Δ' in the interior of Δ.

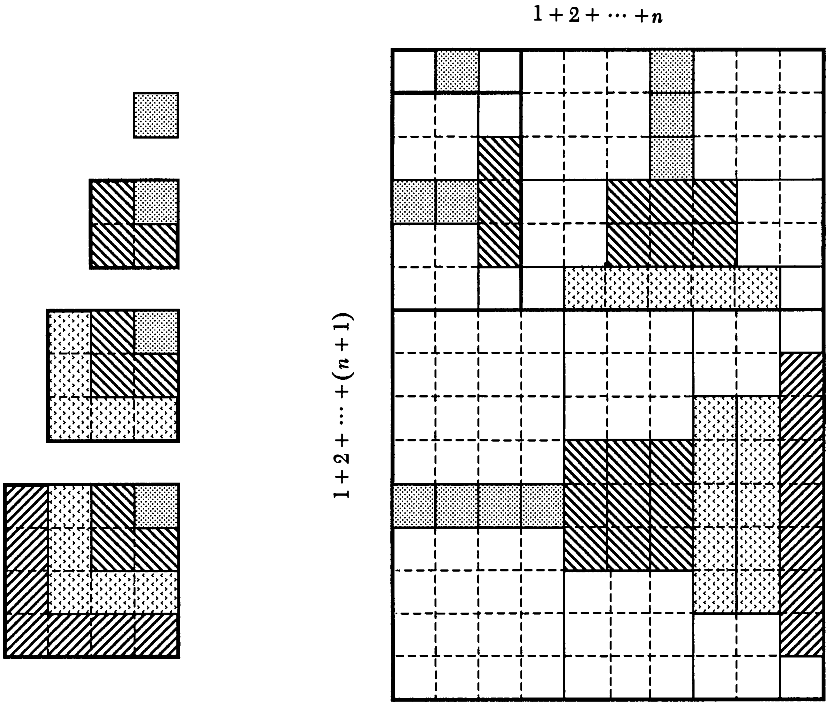
The examples are long thin tetrahedra. The construction shows in fact that any Δ with $E(\Delta)$ less than four times the length of the longest edge (the diameter of Δ) contains a Δ' with $E(\Delta) < E(\Delta')$ in its interior. This is different at the other extreme, the regular tetrahedron, whose edge-length equals six times the diameter; any inscribed Δ', $\neq \Delta$, has some edge shorter than the diameter of Δ. I don't know what happens in between, but suspect that down to edge-length four all inscribed tetrahedra have shorter edge-length.

The result and argument generalize to the edge-length of n-simplices in Euclidean n-space; the factor $4/3$ gets replaced by $(n+1)^2/4n$ if n is odd and by $(n+2)/4$ if n is even.

Acknowledgments. I heard about this problem from Paul Halmos. I thank the referee for informing me that it was proposed by "Proof" in *Crux Mathematicorum* in December 1996, and was circulated as Macalester Problem of the Week number 834. A Macalester student, David Castro, showed one could get as close as desired to 4/3, and Dan Velleman conjectured that 4/3 is an upper bound.

Proof Without Words: Sums of Sums of Squares

$$\sum_{k=1}^{n} \sum_{i=1}^{k} i^2 = \frac{1}{3}\binom{n+1}{2}\binom{n+2}{2}$$



$$3(1^2) + 3(1^2 + 2^2) + 3(1^2 + 2^2 + 3^2) + \cdots + 3(1^2 + 2^2 + \cdots + n^2)$$
$$= \binom{n+1}{2}\binom{n+2}{2}$$

—C. G. Wastun
Washington State University
Pullman, WA 99164

PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

ZE-LI DOU, KEN RICHARDSON, and SUSAN G. STAPLES, *Assistant Editors*
Texas Christian University

Proposals

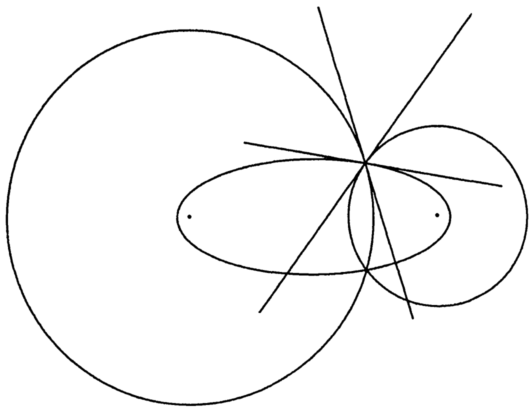
To be considered for publication, solutions should be received by November 1, 2000.

1599. *Proposed by Ice B. Risteski, Skopje, Macedonia.*

Given $\alpha > \beta > 0$ and $f(x) = x^{\alpha}(1-x)^{\beta}$. If $0 < a < b < 1$ and $f(a) = f(b)$, show that $f'(a) < -f'(b)$.

1600. *Proposed by Juan-Bosco Romero Márquez, Universidad de Valladolid, Valladolid, Spain.*

Let $\mathscr{C}$ be either an ellipse or a hyperbola. For P a point on $\mathscr{C}$, prove that the tangent line to $\mathscr{C}$ at P bisects one of the angles formed by the tangents to the circles passing through P with centers at the foci of $\mathscr{C}$.



We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution.

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

Solutions and new proposals should be mailed to Elgin Johnston, Problems Editor, Department of Mathematics, Iowa State University, Ames, IA 50011, or mailed electronically (ideally as a LATEX file) to `johnston@math.iastate.edu`. *Readers who use e-mail should also provide an e-mail address.*

1601. *Proposed by John Clough, State University of New York at Buffalo, Buffalo, New York; Jack Douthett, TVI Community College, Albuquerque, New Mexico; and Roger Entringer, University of New Mexico, Albuquerque, New Mexico.*

Let a and b be positive integers. Place a white points on a circle so that they form the vertices of a regular a-gon. Place b black points on the same circle so that they form the vertices of a regular b-gon and so that white and black points are distinct. Beginning with a black point whose clockwise distance *from* the nearest white point is a minimum and proceeding clockwise, label the points with the integers 0 through $a+b-1$. Prove that the black points have labels $\lfloor k(a+b)/b \rfloor$, $k = 0, 1, \ldots, b-1$, and the white points have labels $\lceil k(a+b)/a \rceil - 1$, $k = 1, 2, \ldots, a$.

1602. *Proposed by Michael Golomb, Professor Emeritus, Purdue University, West Lafayette, Indiana.*

Suppose S is a bounded set of points in $\mathbb{R}^n$, $n \geq 2$, such that every n-simplex whose vertices are points of S has volume at most V. Prove that there exists an n-prism of volume at most $2n(n-1)^{n-1}V$ that contains S.

(An n-prism is an n-dimensional solid bounded by two parallel hyperplanes and a finite number of hyperplanes, each of which contains a line parallel to a fixed line intersecting these two hyperplanes.)

1577. *Proposed by Philip Korman, University of Cincinnati, Cincinnati, Ohio.*

Consider the differential equation $x''(t) + a(t)x^3(t) = 0$ on $0 \leq t < \infty$, where $a(t)$ is continuously differentiable and $a(t) \geq \kappa > 0$.

(a) If $a'(t)$ has only finitely many changes of sign, prove that any solution $x(t)$ is bounded.
(b) If one does not assume that $a'(t)$ has only finitely many sign changes, is $x(t)$ necessarily bounded?

(This problem was originally published with a nontrivial typographical error in the differential equation. It has been fixed in the above.)

Quickies

Answers to the Quickies are on page 245.

Q901. *Proposed by Jorge-Nuno Silva, University of Lisbon, Lisbon, Portugal.*

Begin with a finite pile of balls each labeled with a positive integer, with duplicates permitted. A move consists of removing a ball and possibly replacing it with any finite number of balls labeled with smaller positive numbers. Martin Gardner (*Scientific American* and *The Last Recreations*) publicized the following result of Raymond M. Smullyan (*Annals of the New York Academy of Sciences* 321 (1979), pp. 86–90): The pile will always disappear after a finite number of moves.

View this as an alternate move game between two players, won by the player removing the last ball. Describe those piles for which the player with the first move has a winning strategy.

Q902. *Proposed by K. R. S. Sastry, Bangalore, India.*

Let non-obtuse $\triangle ABC$ have sides of integral length satisfying $BC < AC < AB$. Let D denote the foot of the perpendicular from B to AC. Prove that $AD - CD = 4$ if and only if $AB - BC = 2$.

Solutions

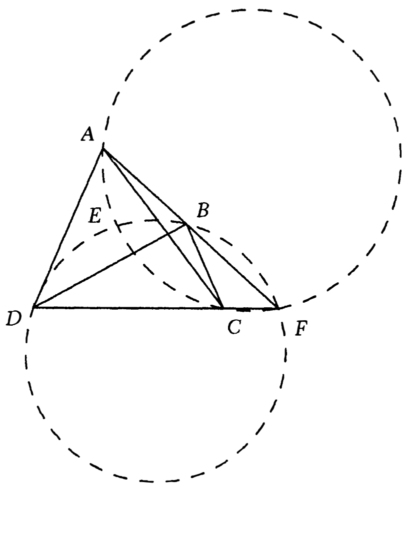
Similar Triangles in a Convex Quadrilateral — June 1999

1574. *Proposed by Larry Hoehn, Austin Peay University, Clarksville, Tennessee.*

Let $ABCD$ be a convex quadrilateral. Prove or disprove: There exists a point E in the plane of $ABCD$ such that $\triangle ABE \sim \triangle CDE$.

I. Solution by Marty Getz and Dixon Jones, University of Alaska Fairbanks, Fairbanks, Alaska.

There always exists such a point E. If AB is parallel to CD, then E is the intersection of the diagonals AC and BD. Otherwise, let AB and CD intersect at F. Let E be the second intersection of the circumcircle of $\triangle ACF$ and the circumcircle of $\triangle BDF$. Because quadrilateral $AECF$ is cyclic, $\angle EAF$ and $\angle ECF$ are supplementary. Thus, $\angle EAB = \angle ECD$. Similarly, $\angle EDC = \angle EBA$. Thus, $\triangle ABE \sim \triangle CDE$.



II. Solution by Gerald D. Brown, Mission Viejo, California.

Represent A, B, C, D, and E as complex numbers. Then $\triangle ABE \sim \triangle CDE$ if and only if $(E-A)/(B-A) = (E-C)/(D-C)$ if and only if

$$E = \frac{AD - BC}{A - B - C + D}.$$

If the point E does not exist, then the denominator $A - B - C + D = 0$. However, this would require that $ABDC$ be a parallelogram, which is impossible because $ABCD$ is given convex.

Also solved by Herb Bailey, Jordi Dou (Spain), Peter Hohler (Switzerland), Hans Kappus (Switzerland), Victor Y. Kutsenok, Neela Lakshmanan, Harry Sedinger, Paul J. Zwier, and the proposer.

Ratio of Areas of Two Triangles Sharing a Side **June 1999**

1575. *Proposed by Peter Y. Woo, Biola University, La Mirada, California.*

Given a convex quadrilateral $ABCD$, find the ratio of the area of $\triangle ABD$ to that of $\triangle BCD$ in terms of $\angle BAC$, $\angle BCA$, $\angle DAC$, and $\angle DCA$.

Solution by Waihung Robert Pun, student, Shippensburg University, Shippensburg, Pennsylvania.

Let $\alpha_1 = \angle BAC$, $\alpha_2 = \angle DAC$, $\gamma_1 = \angle BCA$, and $\gamma_2 = \angle DCA$. By the law of sines, we have

$$\frac{AB}{BC} = \frac{\sin\gamma_1}{\sin\alpha_1} \quad \text{and} \quad \frac{AD}{CD} = \frac{\sin\gamma_2}{\sin\alpha_2}.$$

Then the ratio of the area of $\triangle ABD$ to that of $\triangle BCD$ is

$$\begin{aligned}\frac{\frac{1}{2}AB\cdot AD\cdot\sin(\alpha_1+\alpha_2)}{\frac{1}{2}BC\cdot CD\cdot\sin(\gamma_1+\gamma_2)} &= \frac{\sin\gamma_1\cdot\sin\gamma_2\cdot\sin(\alpha_1+\alpha_2)}{\sin\alpha_1\cdot\sin\alpha_2\cdot\sin(\gamma_1+\gamma_2)}\\ &= \frac{\sin\gamma_1\cdot\sin\gamma_2\cdot(\sin\alpha_1\cdot\cos\alpha_2+\cos\alpha_1\cdot\sin\alpha_2)}{\sin\alpha_1\cdot\sin\alpha_2\cdot(\sin\gamma_1\cdot\cos\gamma_2+\cos\gamma_1\cdot\sin\gamma_2)}\\ &= \frac{\cot\alpha_2+\cot\alpha_1}{\cot\gamma_2+\cot\gamma_1}.\end{aligned}$$

Also solved by Herb Bailey, Gerald D. Brown, Russell Euler and Jawad Sadek, Peter Hohler (Switzerland), Geoffrey A. Kandall, Hans Kappus (Switzerland), Victor Y. Kutsenok, Neela Lakshmanan, Mark Littlefield, Can A. Minh (graduate student), Stephen Noltie, Volkhard Schindler (Germany), Heinz-Jürgen Seiffert (Germany), Chin W. Yang, Paul J. Zwier, and the proposer.

Bisecting a Convex Polygon **June 1999**

1576. *Proposed by Mircea Radu, Bielefeld University, Bielefeld, Germany.*

For $\mathscr{P}$ a convex polygon, prove that the following are equivalent.

(1) The straight lines that divide $\mathscr{P}$ into two polygons of equal perimeter have a common point.
(2) The straight lines that divide $\mathscr{P}$ into two polygons of equal area have a common point.
(3) $\mathscr{P}$ is centrally symmetric.

Solution by Jim Delany, California Polytechnic State University, San Luis Obispo, California.

It is clear that (3) implies (1) and (2). We will show (2) $\Rightarrow$ (3) and (1) $\Rightarrow$ (3). In each case we will work in polar coordinates, taking the common point of intersection of the bisectors to be the origin. We suppose that the equation of the polygon is $r = f(\theta)$, where f is positive and has period 2π.

(2) ⇒ (3). If A is the area enclosed by the polygon, then the formula for area in polar coordinates implies

$$\int_{\theta}^{\theta+\pi} \frac{1}{2}[f(t)]^2\,dt = \frac{A}{2}$$

for all θ. Differentiating with respect to θ yields

$$\frac{1}{2}[f(\theta+\pi)]^2 - \frac{1}{2}[f(\theta)]^2 = 0,$$

so $f(\theta+\pi) = f(\theta)$ for all θ. It is clear that the preceding argument applies not only to polygons but to any polar curve $r = f(\theta)$, where f is positive, continuous, and has period 2π.

(1) ⇒ (3). Consider the function

$$\Delta(\theta) = f(\theta+\pi) - f(\theta).$$

It is continuous and $\Delta(\pi) = -\Delta(0)$, so there is an angle $\alpha \in [0,\pi]$ for which $\Delta(\alpha) = 0$ and $f(\alpha+\pi) = f(\alpha)$.

Let O denote the origin, A denote the point with polar coordinates $(f(\alpha),\alpha)$, and B denote the first polygonal vertex $(f(\beta),\beta)$ for which $\beta > \alpha$. Let A' and B' be the points opposite A and B, namely $(f(\alpha+\pi), \alpha+\pi)$, and $(f(\beta+\pi), \beta+\pi)$. We want to show that the triangles $\triangle OAB$ and $\triangle OA'B'$ are congruent. Condition (1) implies that the lengths AB and $A'B'$ are equal. Let B'' be the point on line BO such that $\triangle OAB \cong \triangle OA'B''$. Let M be the midpoint of AB. Clearly line MO intersects segment $A'B''$ in its midpoint M''. Furthermore, condition (1) also implies that line MO intersects segment $A'B'$ in its midpoint M'. If $B' \neq B''$, then $M'M''$ is parallel to the line through B' and B'', yet intersects this line at O. Therefore, $B' = B''$ and $\triangle OAB \cong \triangle OA'B'$. We conclude $f(\gamma+\pi) = f(\gamma)$ for $\alpha \le \gamma \le \beta$. Continuing through each of the vertices in this manner yields (3).

We remark that when $r = f(\theta)$ with f continuously differentiable, an argument similar to that for (2) ⇒ (3) above shows that (1) ⇒ (3). The example $r = f(\theta)$ where $f(\theta) = \cos\theta$ for $|\theta| \le \pi/3$ and f has period $2\pi/3$ shows that (1) may hold while (3) fails for $r = f(\theta)$ with f piecewise continuously differentiable.

Also solved by Jiro Fukuta (Professor Emeritus, Japan), L. R. King, Victor Y. Kutsenok, Neela Lakshmanan, Victor Pambuccian, Michael Reid, and the proposer.

1577. This problem appeared with a critical typographical error. (The incorrect superscript in galleys appeared correct to the section editor's poor eyes.) The corrected version is among this month's new proposals.

A Binomial Symmetry **June 1999**

1578. *Proposed by Chu Wenchang, Paris, France, and Alberto Marini, IAMI-CNR-Milano, Italy.*

For nonnegative integers m, n, and p and complex numbers a, b, and z, define

$$S_{m,n,p}(a,b,z) := \sum_{k=0}^{m} (-1)^k \binom{m}{k} \frac{z}{z+k} \cdot \frac{(a+bk)^p}{\binom{z+k+n}{n}}.$$

For $p \le \max\{m,n\}$, show that $S_{m,n,p}(a,b,z) = S_{n,m,p}(a-bz, -b, z)$.

Solution by Paul J. Zwier, Calvin College, Grand Rapids, Michigan.

We prove the claim for $p \le m+n$. Set

$$P_{m,n,p}(a,b,z) := \binom{m+n}{n}\binom{z+m+n}{m+n} S_{m,n,p}(a,b,z).$$

Then

$$\begin{aligned} P_{m,n,p}(a,b,z) &= \sum_{k=0}^{m} (-1)^k (a+bk)^p \binom{z+k-1}{k}\binom{z+m+n}{m-k} \\ &= \sum_{k=0}^{m} (a+bk)^p \binom{-z}{k}\binom{z+m+n}{m-k} \end{aligned}$$

is a polynomial in z. Furthermore, $S_{m,n,p}(a,b,z) = S_{n,m,p}(a-bz,-b,z)$ if and only if $P_{m,n,p}(a,b,z) = P_{n,m,p}(a-bz,-b,z)$.

We first use induction to show that $P_{m,n,p}(a,b,z)$ and $P_{n,m,p}(a-bz,-b,z)$ have degree at most p in z. For $z = -m-n,\dots,-1,0$, $P_{m,n,0}(a,b,z)$ is the number of subsets with m elements of a set with $m+n$ elements, i.e., $\binom{m+n}{m}$. Because $P_{m,n,p}(a,b,z)$ has degree at most m, it follows that $P_{m,n,0}(a,b,z)$ is the unique polynomial of degree at most $m+n$ with $P_{m,n,0}(a,b,z) = \binom{m+n}{m}$ for $z=-m-n,\dots,-1,0$, hence $P_{m,n,0}(a,b,z) \equiv \binom{m+n}{m}$. To apply induction, first note that $P_{0,n,p}(a,b,z) = a^p$. For $m \ge 1$

$$\begin{aligned} P_{m,n,p+1}(a,b,z) &= aP_{m,n,p}(a,b,z) + b\sum_{k=0}^{m} k(a+bk)^p \binom{-z}{k}\binom{z+m+n}{m-k} \\ &= aP_{m,n,p}(a,b,z) - bz\sum_{k=1}^{m} ((a+b) \\ &\qquad + b(k-1))^p \binom{-z-1}{k-1}\binom{(z+1)+(m-1)+n}{(m-1)-(k-1)} \\ &= aP_{m,n,p}(a,b,z) \\ &\qquad - bz\sum_{k=0}^{m-1} ((a+b)+bk)^p \binom{-z-1}{k}\binom{(z+1)+(m-1)+n}{(m-1)-k} \\ &= aP_{m,n,p}(a,b,z) - bzP_{m-1,n,p}(a+b,b,z+1). \end{aligned}$$

Similarly, $P_{0,m,p}(a-bz,-b,z) = (a-bz)^p$ and, for $n \ge 1$,

$$\begin{aligned} P_{n,m,p+1}(a-bz,-b,z) &= (a-bz)P_{n,m,p}(a-bz,-b,z) \\ &\qquad + bzP_{n-1,m,p}(a-b-bz,-b,z+1). \end{aligned}$$

For $z = -m-n, \dots, -1, 0$,

$$\begin{aligned} P_{n,m,p}(a-bz) &= \sum_{k=0}^{n} (a-b(z+k))^p \binom{-z}{k}\binom{z+m+n}{n-k} \\ &= \sum_{k=0}^{-z} (a-b(z+k))^p \binom{-z}{k}\binom{z+m+n}{n-k} \\ &= \sum_{j=0}^{-z} (a+bj)^p \binom{-z}{-z-j}\binom{z+m+n}{z+n+j} \\ &= \sum_{j=0}^{-z} (a+bj)^p \binom{-z}{j}\binom{z+m+n}{m-j} \\ &= \sum_{j=0}^{m} (a+bj)^p \binom{-z}{j}\binom{z+m+n}{m-j} = P_{m,n,p}(a,b,z). \end{aligned}$$

Thus, $P_{n,m,p}(a-bz, -b, z) = P_{m,n,p}(a,b,z)$ for $m+n+1$ values of z. If $p \le m+n$, this implies $P_{n,m,p}(a-bz, -b, z) \equiv P_{m,n,p}(a,b,z)$.

Comments. Heinz-Jürgen Seiffert proved the original claim by induction on m. Hjalmar Rosengren reports that the above generalization is a special case of a transformation formula discussed by George Gasper in "Summation formulas for basic hypergeometric series," *SIAM J. Math. Anal.* 12 (1981), 196–200.

Also solved by Hjalmar Rosengren (Sweden), Heinz-Jürgen Seiffert (Germany), and the proposers.

Answers

Solutions to the Quickies on page 241.

A901. The first player has a winning strategy exactly when there is an odd number of balls with some label. The player should remove a ball with the largest label having an odd number of balls and add balls with smaller labels so that there are an even number of balls in the pile with each label. A player facing such a position must leave an odd number of balls for at least one of the labels and, in particular, cannot win with one move.

A902. Let $a = BC$, $b = AC$, $c = AB$, and $d = AD - CD$. Then

$$BD^2 = c^2 - \left(\frac{b+d}{2}\right)^2 = a^2 - \left(\frac{b-d}{2}\right)^2,$$

or $c^2 - a^2 = bd$.

If $c - a = 2$, then $c = b+1$ and $a = b-1$. Thus, $4b = bd$, hence $d = 4$.

Conversely, if $d = 4$, then $(c+a)(c-a) = 4b$. From $c + a > b$, it follows that $c - a$ is 1, 2, or 3. If $c - a = 1$, then $c + a = 4b > 2a + b$, so that $c > a + b$, a contradiction. If $c - a = 3$, then $c + a = 4a/3$, hence $c = (4a+9)/6$, which cannot be an integer. Therefore, $c - a = 2$.

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of mathematics literature. Readers are invited to suggest items for review to the editors.

Schechter, Bruce, In the life of pure reason, prizes have their place, *New York Times* (25 April 2000) (National Edition) D5, (City Edition) F5. Ahuja, Anjana, A million-dollar maths question, *London Times* (16 March 2000), `http://www.the-times.co.uk/news/pages/tim/2000/03/16/timfeafea02004.html` . Doxiadis, Apostolos K., *Uncle Petros and Goldbach's Conjecture*, Bloomsbury Books. ISBN 1–58234067–6

Well, after the conquest of Fermat's Last Theorem, what's left to do that can be explained easily to the public? Goldbach's Conjecture, of course: that every even integer greater than two is the sum of two primes. Moreover, a correct proof by 15 March 2002 will win $1 million, courtesy of publishers Faber & Faber and Bloomsbury Books. They are promoting their book *Uncle Petros and Goldbach's Conjecture*, about an imaginary Greek colleague of Hardy, Littlewood, and Ramanujan who withdraws into isolation to try to prove the conjecture. The publishers paid a premium "in the five figures" to insure against having to pay out the prize. It would be more realistic to allow the contest to run for 10 years; but the premium would be greater, and by then the book would be out of print.

Mackenzie, Dana, Fermat's Last Theorem's first cousin, *Science* 287 (4 February 2000) 792–793. Laumon, G., La correspondance de Langlands sur les corps de fonctions [d'aprs Laurent Lafforgue]. G. Laumon. 62 pages. Séminaire Bourbaki, 52ème année, 1999-2000, n° 873 (text of a Bourbaki lecture given 12 March 2000), `http://arXiv.org/PS_cache/math/pdf/0003/0003131.pdf` .

Another part of the *Langlands program*, a collection of assertions about deep connections between Galois representations and automorphic forms, may have been proved. The program, of which Fermat's Last Theorem is a small part, was proposed by Robert Langlands (Princeton University) in 1967 in a letter to André Weil. The proof of the Langlands conjecture for algebraic function fields (finite extensions of the field $\mathbb{C}(z)$ of rational functions in an indeterminate) has been announced by Laurent Lafforgue (Université de Paris–Sud) but not yet submitted for publication. The conjecture for local fields was proven in 1998, but still open is the main case of algebraic number fields (finite extensions of $\mathbb{Q}$).

Zome. Kits of assorted sizes, ranging from 72 pieces ($9.95) to 1406 pieces ($249.95), with support materials: *Lesson Plans 1.0*, $19.95 (individual lesson plans downloadable free); *Zome Primer* by Steve Baer, $4.50; and *Zometool Manual* by David Booth, $4.50. Zometool, Inc., 1526 South Pearl St., Denver, CO 80210; `www.zomesystem.com` .

Zome kits consist of precision-molded plastic hubs and struts from which to build geometric structures: polyhedra, domes, DNA strands, crystals, lattices, buckyballs. The struts, which come in three lengths, are color- and shape-coded for easy assembly. Models can be dipped in soap solution to examine soap films. Kits and support materials are available on free loan to teachers for risk-free evaluation. (Note: Not yet mentioned at the Web site are the new angled green struts, which greatly enlarge the universe of buildable models.)

Quinn, Jennifer, Eric Libicki, and Arthur Benjamin, *The Number Years*, http://www.jquinn.oxy.edu/numberyears/ .

Now mathematicians and their students have their own trivia game show format, *The Number Years*, suitable for play in large groups. (Confession: I was an enthusiastic competitor and finalist—but didn't win—the contest at the mathematics meetings in Washington DC in January.) The Web site includes questions from previous contests and instructions on how to play directly from the Web. Sample Round 1 multiple-choice questions: Who first proved L'Hôpital's Rule? Which telephone area code is prime? Which city has not recognized the importance of Euclid's contribution to mathematics by naming a street after him? In Reverse Polish Notation "1+2*3" is written as ... ? According to Webster's dictionary, which of the following mathematical terms also means "tartar"? Which of the following mathematicians was *not* born on today's date? Round 2 questions are short-answer; Round 3 contains mathematical word puzzles.

Beck, Anatole, Michael N. Bleicher, and Donald W. Crowe, *Excursions into Mathematics*, 2nd ("Millennium") ed., A K Peters, 2000; xxv + 499 pp, $34 (P). ISBN 1-56881-115-2.

This is a mathematically beautiful and appealingly designed book, saved from obscurity by a reprint of the 1969 edition (with 7 pp of corrections and updates but without the color frontispiece showing 7-coloring of a torus). "[D]esigned to acquaint the general reader with some of the flavor of mathematics," it contains 6 independent "excursions," on polyhedra, perfect numbers, area, noneuclidean geometries, games, and number representations. None is a rehash of "standard" topics, and some results were first published here. Each excursion is designed to take up to 40 classroom hours, and the authors in part planned the book for high-school seniors. "The conceptual level ... is intended to be about the same as the calculus, and the student should be prepared for some hard thinking. ... Each of [the first] three chapters will introduce the student to the principle of mathematical induction. Each of the last three chapters presupposes that the student has mastered this principle." Surveying the book, which is chockablock with imaginative proofs and conjectures, those of us in the profession through the intervening 30 years are confronted with a discouraging truth, one that we already know in our souls too well: *The target audience has almost disappeared* (in the U.S., at least). In the 1960s, many high-school students mastered induction, which they saw used again in calculus; today, many sophomore and junior mathematics and computer science majors struggle in a discrete mathematics course with the topic but can't quite think hard enough to "get" it. Even in the 1970s, though, this book was not widely used in liberal-arts mathematics courses, much less by high schools. For the publisher of the first edition, it was a loss-leader "prestige" book (a representative told me at one point that they had given away 9,000 "examination" copies). The conceptual level of the book is above calculus as it is now taught. Moreover, none of the topics in the book is motivated by or directed to applications (except for matrix games). Paralleling the cultural decline of ideas in favor of pragmatism, of political ideals in favor of economic values, the campaign for relevancy in education succeeded admirably: Students value little else and have scant desire for hard thinking divorced from immediate application. Nevertheless—*what a magnificent living fossil of a book!* So, where's today's audience? You yourself will find the book entertaining, enlightening, and delightful, and so will curious mathematics students at all levels who admire abstract thought and are eager to indulge in it.

Knuth, Donald E., *Digital Typography*, Center for the Study of Language and Information (CSLI), 1999; xv + 685 pp (P). ISBN 1-57586-010-4.

"I guess I must have ink in my veins." This collection of more than 30 articles by Don Knuth chronicles his "decades-long quest to produce beautiful books with the help of computers."

NEWS AND LETTERS

Twenty-Eighth Annual USA Mathematical Olympiad – Problems and Solutions

1. Some checkers placed on an $n \times n$ checkerboard satisfy the following conditions:

(a) every square that does not contain a checker shares a side with one that does;

(b) given any pair of squares that contain checkers, there is a sequence of squares containing checkers, starting and ending with the given squares, such that every two consecutive squares of the sequence share a side.

Prove that at least $(n^2 - 2)/3$ checkers have been placed on the board.

Solution. It suffices to show that if m checkers are placed so as to satisfy condition (b), then the number of squares they either cover or are adjacent to is at most $3m + 2$. But this is easily seen by induction: it is obvious for $m = 1$, and if m checkers are so placed, some checker can be removed so that the remaining checkers still satisfy (b); they cover at most $3m - 1$ squares, and the new checker allows us to count at most 3 new squares (since the square it occupies was already counted, and one of its neighbors is occupied).

2. Let $ABCD$ be a cyclic quadrilateral. Prove that

$$|AB - CD| + |AD - BC| \geq 2|AC - BD|.$$

First Solution (by Po-Ru Loh). Let E be the intersection of AC and BD. Then the triangles ABE and DCE are similar, so if we let $x = AE, y = BE, z = AB$, then there exists k such that $kx = DE, ky = CE, kz = CD$. Now

$$|AB - CD| = |k - 1|z$$

and

$$|AC - BD| = |(kx + y) - (ky + x)| = |k - 1| \cdot |x - y|.$$

Since $|x - y| \leq z$ by the triangle inequality, we conclude $|AB - CD| \geq |AC - BD|$, and similarly $|BC - DA| \geq |AC - BD|$. These two inequalities imply the desired result.

Second Solution. Let $2\alpha, 2\beta, 2\gamma, 2\delta$ be the measures of the arcs subtended by AB, BC, CD, DA, respectively, and take the radius of the circumcircle of $ABCD$ to be 1. Assume without loss of generality that $\beta \leq \delta$. Then $\alpha + \beta + \gamma + \delta = \pi$, and (by the Extended Law of Sines)

$$|AB - CD| = 2|\sin\alpha - \sin\gamma| = \left|\sin\frac{\alpha - \gamma}{2}\cos\frac{\alpha + \gamma}{2}\right|$$

and

$$|AC - BD| = 2|\sin(\alpha+\beta) - \sin(\beta+\gamma)| = \left|\sin\frac{\alpha-\gamma}{2}\cos\left(\frac{\alpha+\gamma}{2}+\beta\right)\right|.$$

Since $0 \leq (\alpha+\gamma)/2 \leq (\alpha+\gamma)/2+\beta \leq \pi/2$ (by the assumption $\beta \leq \delta$) and the cosine function is nonnegative and decreasing on $[0,\pi/2]$, we conclude that $|AB - CD| \geq |AC - BD|$, and similarly $|AD - BC| \geq |AC - BD|$.

3. Let $p > 2$ be a prime and let a, b, c, d be integers not divisible by p, such that

$$\{ra/p\} + \{rb/p\} + \{rc/p\} + \{rd/p\} = 2$$

for any integer r not divisible by p. Prove that at least two of the numbers $a+b$, $a+c$, $a+d$, $b+c$, $b+d$, $c+d$ are divisible by p. (Note: $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of x.)

Solution. For convenience, we write $[x]$ for the unique integer in $\{0,\ldots,p-1\}$ congruent to x modulo p. In this notation, the given condition can be written

$$[ra] + [rb] + [rc] + [rd] = 2p \quad \text{for all } r \text{ not divisible by } p. \tag{1}$$

The conditions of the problem are preserved by replacing a, b, c, d with sa, sb, sc, sd for any integer s relatively prime to p. If we choose s so that $sa \equiv 1 \pmod p$, then replace a, b, c, d with $[sa],[sb],[sc],[sd]$ respectively, we end up in the case $a = 1$ and $b, c, d \in \{1,\ldots,p-1\}$. Applying (1) with $r = 1$, we see moreover that $a+b+c+d = 2p$.

Now observe that

$$[(r+1)x] - [rx] = \begin{cases} [x] & [rx] < p - [x] \\ -p + [x] & [rx] \geq p - [x]. \end{cases}$$

Comparing (1) with $r = s$ and $r = s+1$ and using the observation, we see that for $r = 1,\ldots,p-2$, two of the quantities

$$p - a - [ra], p - b - [rb], p - c - [rc], p - d - [rd]$$

are positive and two are negative. We say that a pair (r,x) is *positive* if $[rx] < p - [x]$ and *negative* otherwise; then for each r, $(r,1)$ is positive, so exactly one of $(r,b),(r,c),(r,d)$ is also positive.

Lemma 1 *If $r_1, r_2, x \in \{1,\ldots,p-1\}$ have the property that (r_1,x) and (r_2,x) are positive but (r,x) is negative for $r_1 < r < r_2$, then*

$$r_2 - r_1 = \left\lfloor \frac{p}{x} \right\rfloor \quad \textit{or} \quad r_2 - r_1 = \left\lfloor \frac{p}{x} \right\rfloor + 1.$$

Proof: Note that (r,x) is negative if and only if $\{rx+1,\ldots,(r+1)x\}$ contains a multiple of p. In particular, between r_1x and r_2x, there lies exactly one multiple of p. Since $[r_1x]$ and $[r_2x]$ are distinct elements of $\{p-[x],\ldots,p-1\}$, we have

$$p - x + 1 < r_2x - r_1x < p + x - 1,$$

from which the lemma follows. □

Recall that exactly one of $(1,b)$, $(1,c)$, $(1,d)$ is positive; we may as well assume $(1,b)$ is positive, which is to say $b < p/2$ and $c,d > p/2$. Put $s_1 = \lfloor p/b \rfloor$, so that s_1 is the smallest positive integer such that (s_1,b) is negative. Then exactly one of (s_1,c) and (s_1,d) is positive, say the former. Since s_1 is also the smallest positive integer such that (s_1,c) is negative, or equivalently such that $(s_1,p-c)$ is positive, we have $s_1 = \lfloor p/(p-c) \rfloor$. The lemma states that consecutive values of s for which (s,b) is negative differ by either s_1 or s_1+1. It also states (when applied with $x = p - c$) that consecutive values of s for which (s,c) is positive differ by either s_1 or s_1+1.

From these observations we will show that (s,d) is always negative. Indeed, if this were not the case, there would exist a smallest positive integer s such that (s,d) is positive; then (s,b) and (s,c) are both negative. If s' is the last integer before s such that (s',b) is negative (which exists because (s_1,d) is negative, so s' cannot equal s_1), then

$$s - s' = s_1 \quad \text{or} \quad s - s' = s_1 + 1.$$

Likewise, if t were the next integer after s' such that (t,c) were positive, then

$$t - s' = s_1 \quad \text{or} \quad t - s' = s_1 + 1.$$

From these we deduce that $|t-s| \le 1$. However, if $t \ne s$, then (t,b) must also be negative, and any two values of x for which (x,b) is negative differ by at least $s_1 \ge 2$. Thus $t = s$, contradicting the choice of s as a minimal counterexample, and proving the claim.

r	1		s_1	s_1+1		s'	$s'+1$		$s-1$	s	t
(r,b)	+		−	+		−	+		+	−	−
(r,c)	−	…	+	−	…	+	−	…	−	−	+
(r,d)	−		−	−		−	−		−	+	−

Therefore (s,d) is always negative. On the other hand, if $d \ne p-1$, the unique $s \in \{1,\dots,p-1\}$ such that $[ds] = 1$ is not equal to $p-1$, and (s,d) is positive, contradiction. Thus $d = p-1$ and $a+d$ and $b+c$ are divisible by p, as desired.

4. Let $a_1, a_2, \dots, a_n$ $(n > 3)$ be real numbers such that

$$a_1 + a_2 + \cdots + a_n \ge n \qquad \text{and} \qquad a_1^2 + a_2^2 + \cdots + a_n^2 \ge n^2.$$

Prove that $\max(a_1, a_2, \dots, a_n) \ge 2$.

Solution. Let $b_i = 2 - a_i$, and let $S = \sum b_i$ and $T = \sum b_i^2$. Then the given conditions are that

$$(2 - a_1) + \cdots + (2 - a_n) \ge n$$

and

$$(4 - 4b_1 + b_1^2) + \cdots + (4 - 4b_n + b_n^2) \ge n^2,$$

which is to say $S \le n$ and $T \ge n^2 - 4n + 4S$.

From these inequalities, we obtain

$$T \geq n^2 - 4n + 4S \geq (n-4)S + 4S = nS.$$

On the other hand, if $b_i > 0$ for $i = 1, \ldots, n$, then certainly $b_i < \sum b_i = S \leq n$, and so

$$T = b_1^2 + \cdots + b_n^2 < nb_1 + \cdots + nb_n = nS.$$

Thus we cannot have $b_i > 0$ for $i = 1, \ldots, n$, so $b_i \leq 0$ for some i, and $a_i \geq 2$ for that i, proving the claim.

5. The Y2K Game is played on a 1×2000 grid as follows. Two players in turn write either an S or an O in an empty square. The first player who produces three consecutive boxes that spell SOS wins. If all boxes are filled without producing SOS then the game is a draw. Prove that the second player has a winning strategy.

Call a partially filled board *stable* if there is no SOS and no single move can produce an SOS; otherwise call it *unstable*. For a stable board call an empty square *bad* if either an S or an O played in that square produces an unstable board. Thus a player will lose if the only empty squares available to him are bad, but otherwise he can at least be guaranteed another turn with a correct play.

Claim: A square is bad if and only if it is in a block of 4 consecutive squares of the form S – – S.

Proof: If a square is bad, then an O played there must give an unstable board. Thus the bad square must have an S on one side and an empty square on the other side. An S played there must also give an unstable board, so there must be another S on the other side of the empty square. □

From the claim it follows that there are always an even number of bad squares. Thus the second player has the following winning strategy:

(a) If the board is unstable at any time, play the winning move, otherwise continue as below.

(b) On the first move, play an S at least four squares away from either end and from the first player's first move. (The board is long enough that this is possible.)

(c) On the second move, play an S three squares away from the second player's first move, so that the squares in between are empty. (Regardless of the first player's second move, this must be possible on at least one side.) This produces two bad squares; whoever plays in one of them first will lose. Thus the game will not be a draw.

(d) On any subsequent move, play in a square which is not bad. Such a square will always exist because if the board is stable, there will be an odd number of empty squares and an even number of bad squares.

Since there exist bad squares after the second player's second move, the game cannot end in a draw, and since the second player can always leave the board stable, the first player cannot win. Therefore eventually the second player will win.

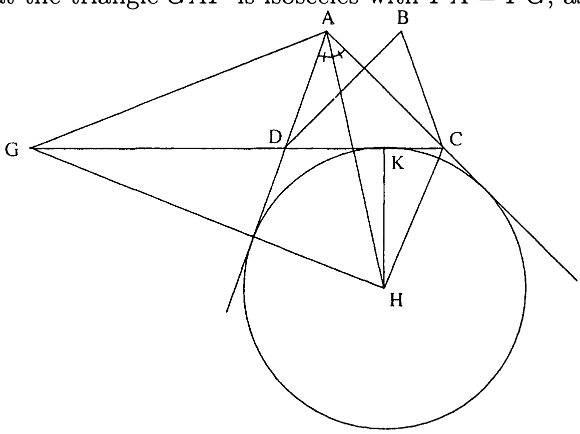
6. Let $ABCD$ be an isosceles trapezoid with $AB \parallel CD$. The inscribed circle ω of triangle BCD meets CD at E. Let F be a point on the (internal) angle bisector of $\angle DAC$ such that $EF \perp CD$. Let the circumscribed circle of triangle ACF meet line CD at C and G. Prove that the triangle AFG is isosceles.

Solution. We will show that $FA = FG$. Let H be the center of the escribed circle (also known as the excircle) of triangle ACD opposite vertex A. Then H lies on the angle bisector AF. Let K be the point where this escribed circle touches CD. By a standard computation using equal tangents, we see that $CK = (AD+CD-AC)/2$. By a similar computation in triangle BCD, we see that $CE = (BC+CD-BD)/2 = CK$. Therefore $E = K$ and $F = H$.

Since F is now known to be an excenter, we have that FC is the external angle bisector of $\angle DCA = \angle GCA$. Therefore

$$\angle GAF = \angle GCF = \frac{\pi}{2} - \frac{1}{2}\angle GCA = \frac{\pi}{2} - \frac{1}{2}\angle GFA.$$

We conclude that the triangle GAF is isosceles with $FA = FG$, as desired.



Fortieth Annual International Mathematical Olympiad – Problems

1. Find all finite sets S of at least three points in the plane such that for all distinct points A, B in S, the perpendicular bisector of AB is an axis of symmetry for S.

2. Let $n \geq 2$ be a fixed integer.

(a) Find the least constant C such that for all nonnegative real numbers $x_1, \ldots, x_n$,

$$\sum_{1 \leq i < j \leq n} x_i x_j (x_i^2 + x_j^2) \leq C \left(\sum_{i=1}^{n} x_i \right)^4 .$$

(b) Determine when equality occurs for this value of C.

3. We are given an $n \times n$ square board, with n even. Two distinct squares of the board are said to be adjacent if they share a common side. (A square is *not* adjacent to itself.) Find the minimum number of squares that can be marked so that every square (marked or not) is adjacent to at least one marked square.

4. Find all pairs (n, p) of positive integers such that

- p is prime;
- $n \leq 2p$;
- $(p-1)^n + 1$ is divisible by n^{p-1}.

5. The circles Γ_1 and Γ_2 lie inside circle Γ, and are tangent to it at M and N, respectively. It is given that Γ_1 passes through the center of Γ_2. The common chord of Γ_1 and Γ_2, when extended, meets Γ at A and B. The lines MA and MB meet Γ_1 again at C and D. Prove that the line CD is tangent to Γ_2.

6. Determine all functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(x - f(y)) = f(f(y)) + xf(y) + f(x) - 1.$$

for all $x, y \in \mathbb{R}$.

Notes

The top eight students on the 1999 USAMO were (in alphabetical order):

Reid W. Barton	Arlington, MA
Gabriel D. Carroll	Oakland, CA
Lawrence O. Detlor	New York, NY
Stephen E. Haas	Sunnyvale, CA
Po-Shen Loh	Madison, WI
Alexander B. Schwartz	Bryn Mawr, PA
Paul A. Valiant	Belmont, MA
Melanie E. Wood	Indianapolis, IN

Alexander (Sasha) Schwartz was the winner of the Samuel Greitzer–Murray Klamkin award, given to the top scorer on the USAMO. Newly introduced this year was the Clay Mathematics Institute (CMI) award, to be presented (at the discretion of the USAMO graders) for a solution of outstanding elegance, and carrying a $1000 cash prize. The CMI award was presented to Po-Ru Loh (Madison, WI; brother of Po-Shen Loh) for his solution to USAMO Problem 2, included herein.

Members of the USA team at the 1999 IMO (Bucharest, Romania) were Reid Barton, Gabriel Carroll, Lawrence Detlor, Po-Shen Loh, Paul Valiant, and Melanie Wood. Titu Andreescu (Director of the American Mathematics Competitions) and Kiran Kedlaya (Massachusetts Institute of Technology) served as team leader and deputy leader, respectively. The team was also accompanied by Walter Mientka (University of Nebraska, Lincoln), who served as secretary to the IMO Advisory Board.

At the 1999 IMO, gold medals were awarded to students scoring between 28 and 39 points (39 was the highest score obtained on this extremely difficult exam), silver medals to students scoring between 19 and 27 points, and bronze medals to students scoring between 12 and 18 points. Barton and Valiant received gold medals, Carroll, Loh and Wood received silver medals, and Detlor received a bronze medal.

In terms of total score (out of a maximum of 252), the highest ranking of the 81 participating teams were as follows:

China	182	Republic of Korea	164
Russia	182	Iran	159
Vietnam	177	Taiwan	153
Romania	173	United States	150
Bulgaria	170	Hungary	147
Belarus	167	Ukraine	136

The 2000 IMO is scheduled to be held in Seoul and Taejon, Republic of Korea. Special note: the 2001 IMO is scheduled to be held in Washington, DC and Fairfax, VA, USA. For more information about the 2001 IMO, send email to `imo2001@amc.unl.edu`.

The 1999 USAMO was prepared by Titu Andreescu, Zuming Feng, Kiran Kedlaya, Alexander Soifer and Zvezdelina Stankova-Frenkel. The training program to prepare the USA team for the IMO, the Mathematical Olympiad Summer Program (MOSP), was held at the University of Nebraska, Lincoln. Titu Andreescu, Zuming Feng, Kiran Kedlaya, and Zvezdelina Stankova-Frenkel served as instructors, assisted by Andrei Gnepp and Daniel Stronger.

The booklet *Mathematical Olympiads 1999* presents additional solutions to problems on the 28th USAMO and solutions to the 40th IMO. Such a booklet has been published every year since 1976. Copies are $5.00 for each year 1976-1997. They are available from:

Titu Andreescu, American Mathematics Competitions, 1740 Vine St., Lincoln, NE 68588-0858.

For further information about the sequence of examinations leading to the USAMO, MOSP and IMO, contact the Director of American Mathematics Competitions, Titu Andreescu, at the above address, or visit the AMC web site at `http://www.unl.edu/amc`.

This report was prepared by Titu Andreescu and Kiran Kedlaya.

THE MATHEMATICAL ASSOCIATION OF AMERICA

New from William Dunham, award-winning author of Journey through Genius: The Great Theorems of Mathematics, and The Mathematical Universe....

Euler

The Master of Us All

William Dunham

Series: Dolciani Mathematical Expositions

Without question, Leonhard Euler (1707-1783) ranks among history's greatest mathematicians. Across six decades of unmatched productivity, and despite a visual impairment that grew ever worse, he charted the course of mathematics throughout the eighteenth century and beyond. His reputation is captured in Laplace's famous admonition, "Read Euler, read Euler. He is the master of us all."

Written for the mathematically literate reader, this book provides a glimpse of Euler in action. Following an introductory biographical sketch are chapters describing his contributions to eight different topics—number theory, logarithms, infinite series, analytic number theory, complex variables, algebra, geometry, and combinatorics. Each chapter begins with a prologue to establish the historical context and then proceeds to a detailed consideration of one or more Eulerian theorems on the subject at hand. Each chapter concludes with an epilogue surveying subsequent developments or addressing related questions that remain unanswered to this day. At the end of the book is a brief outline of Euler's collected works, the monumental *Opera Omnia*, whose publication has consumed virtually all of the twentieth century.

In all, the book contains three dozen proofs from this remarkable individual. Yet this is merely the tip of the scholarly iceberg, for Euler produced over 30,000 pages of pure and applied mathematics during his lifetime. *Euler: The Master of Us All* samples the work of a mathematician whose influence, industry, and ingenuity are of the very highest order.

Catalog Code: DOL-22/JR
192 pp., Paperbound, ISBN- 0-88385-328-0
List: $29.95 MAA Member: $23.95

Phone in Your Order Now!* ☎ *1-800-331-1622

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers

Name ______________________

Address ______________________

City ____________ State ____ Zip ______

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	DOL-22/JR	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________


Attend The
Largest Annual
Summer Meeting
of Mathematicians!
REGISTER
NOW AT
www.maa.org
Invited Addresses
THE MATHEMATICS OF "TOY STORY 2"
Edwin Catmull, Pixar Animation Studios
MATHEMATICS DEPARTMENTS IN THE 21st CENTURY: ROLE, RELEVANCE, AND RESPONSIBILITY
William Kirwan, Ohio State University
HOW MATHEMATICIANS STUDY CHAOS?
Yakov Sinai, Princeton University
Lecture 1: Chaos and Disorder
Lecture 2: Chaos as Dynamical Disorder
Lecture 3: Chaos and Hyperbolicity
FINITE QUANTUM CHAOS
Audrey Terras, University of California, San Diego
SINGULARITY THEORY AND GRAVITATIONAL LENSING
Arlie Petters, Duke University
THE MATHEMATICS OF COMPUTERS
John Ewing, American Mathematical Society
THE "3x + 1" PROBLEM
Jeffrey Lagarias, AT&T Laboratory Research
PROFOUND UNDERSTANDING OF FUNDAMENTAL MATHEMATICS: WHAT AN ELEMENTARY MATHEMATICS TEACHER SHOULD KNOW IN ORDER TO TEACH IT
Li Ping Ma, Carnegie Foundation
THE ALTERNATIVE LIFE OF E.T. BELL
Constance Reid, San Francisco, CA
CALIFORNIA
MATHFEST
2000
Los Angeles
AUGUST 3-5
2000 • UCLA
MATHEMATICAL ASSOCIATION OF AMERICA
FOR MORE INFORMATION, PLEASE CALL
1-800-331-1622
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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