Vol. 74, No. 1, February 2001

MATHEMATICS MAGAZINE

$$1 = \sqrt[3]{2+\sqrt{5}} + \sqrt[3]{2-\sqrt{5}} =$$

$$\sqrt[6]{9+4\sqrt{5}} - \sqrt[6]{9-4\sqrt{5}} = \sqrt[9]{38+17\sqrt{5}} + \sqrt[9]{38-17\sqrt{5}}$$

$$2 = \sqrt[3]{7+5\sqrt{2}} + \sqrt[3]{7-5\sqrt{2}} =$$

$$\sqrt[5]{41+29\sqrt{2}} + \sqrt[5]{41-29\sqrt{2}} = \sqrt[6]{99+70\sqrt{5}} + \sqrt[6]{99-70\sqrt{5}}$$

. . . and other radicals.

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. 74, pp. 75–76, 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.

Submit new manuscripts to Frank A. Farris, Editor, Mathematics Magazine, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053-0373. Manuscripts should be laser printed, with wide line spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should mail three copies and keep one copy. In addition, authors should supply the full five-symbol 2000 Mathematics Subject Classification number, as described in *Mathematical Reviews*. Copies of figures should be supplied on separate sheets, both with and without lettering added.

Cover image by Jason Challas, who lectures on Computer Art (and other radicals) at Santa Clara University. Thanks to Al and Grace Sowa. Photograph of Martin Luther King, Jr. by Marion S. Trikosko, 1964 March 26. Library of Congress, Prints and Photographs Division. Reproduction Number: LC-U9-11696-frame #9A

AUTHORS

G. Don Chakerian received his Ph.D. from the University of California, Berkeley in 1960. He taught for three years at the California Institute of Technology before joining the faculty of the University of California, Davis, where he is now Professor Emeritus. His research deals with convex geometry, integral geometry, and geometric inequalities. The present article is related to the author's lifelong interest in astronomy and an ongoing quest to understand what his high school physics teacher was talking about.

B. Carter Edwards, known as Bagel to his friends, received his B.A. from Reed College in 1998. He has acted in various theatrical productions. Bagel wrote his senior thesis on origami under the supervision of **Jerry Shurman**, who has been at Reed since 1989 after earning a B.A. from Stanford in 1982 and a Ph.D. from Princeton in 1988.

Thomas J. Osler is professor of mathematics at Rowan University. He received his Ph.D. from the Courant Institute at New York University in 1970 and is the author of 32 mathematical papers. This paper originated while teaching Cardan's method of solving cubics—certain nested radicals appeared in the solutions that could not be simplified using elementary algebra. In addition to teaching university mathematics for the past 39 years, Tom has a passion for long distance running, in which he has competed for the past 46 consecutive years. Included in his over 1600 races are wins in three national championships in the late 1960s at distances from 25 kilometers to 50 miles. He is the author of two books about running.

Vol. 74, No. 1, February 2001

MATHEMATICS MAGAZINE

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 *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 Dave Riska (*driska@maa.org*), 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

Central Force Laws, Hodographs, and Polar Reciprocals

DON CHAKERIAN
University of California, Davis
Davis, CA 95616-8633

> It is not difficult, when one knows the calculus, and to write the differential equations and to solve them, to show that it's an ellipse. I believe in the lectures here—or at least in the book—[you] calculated the orbit by numerical methods and saw that it looked like an ellipse. That's not exactly the same thing as *proving* that it is exactly an ellipse. The Mathematics Department ordinarily is left the job of proving that it's an ellipse, so that they have something to do over there with their differential equations. [Laughter]
>
> —Richard Feynman, from *Feynman's Lost Lecture* [**4**]

1. Introduction

Sir Isaac Newton's *Philosophiae Naturalis Principia Mathematica* (or simply, the *Principia* [**14**]) ushered in the era of modern science. But contemporary students often find Newton's geometric style of exposition fiercely demanding, particularly where it leans heavily on recondite facts concerning conic sections, universally ignored in the present-day school curriculum. Sherman Stein [**17**] has done us a service in illuminating the basic geometric method underlying Newton's analysis of Keplerian orbits, making this more accessible to the modern reader. Many years before Stein's article appeared, the brilliant physicist Richard Feynman, preparing an undergraduate lecture on Newton's treatment of elliptical orbits and confounded by the esoteric properties of conics employed by Newton, was led at one point to substitute a variant of his own invention involving the *hodograph*, or velocity diagram, of an orbit. (See Goodstein and Goodstein [**4**] for a reconstruction of Feynman's lecture. The reviews by Griffiths [**5**] and Weinstock [**20**] give an excellent overview of this intriguing lecture and also point out some shortcomings of the reconstruction.)

The hodograph of an orbit is obtained by translating each velocity vector so the tail is at the origin; the locus of the heads of the resulting vectors is the hodograph. Feynman's simplified approach hinged on the fact that the hodograph of an orbit satisfying Kepler's laws must be circular. Andrew Lenard [**9**] gives an elegant geometrical derivation of the inverse square law of gravitation from Kepler's laws of planetary motion based on the circular hodograph. A similar, and very concise, derivation appears in Maxwell's lovely little book [**10**], where Sir William Rowan Hamilton is credited as the originator of the hodograph technique. In Hamilton [**8**] one finds several applications of the hodograph method, including a proof that planetary orbits are conic sections if and only if the central force law is inverse square. But Hamilton's exposition is difficult to follow without an intimate acquaintance with the algebraic machinery of quaternions. Milnor [**11**] gives a lucid account of the geometry of Kepler orbits and a noneuclidean metric that can be imposed on the set of velocity vectors of orbits such that the hodographs are the geodesics.

The main purpose of this article is to call attention to a useful fact that appears not to have been explicitly exploited in these matters, namely, that the hodograph of an orbit associated with any central force law is geometrically similar to the polar reciprocal of the orbit itself; see (3.4) below. (Guggenheimer [**7**] develops a geometric approach to the oscillatory properties of certain linear second order differential equations, including Hill's equation, based on the similarity of hodographs and polar reciprocals.) This observation enables one to adapt the arguments used in [**9**] and [**10**] for the inverse square law, virtually unchanged, to derive the force laws associated with orbits other than Keplerian ellipses. Another advantage is the light cast on "converse" results. By this we mean the following:

Kepler's three laws of planetary motion, based on extensive empirical data, are

(I) The orbits followed by the planets are ellipses with the sun at one focus.
(II) As a planet moves along its orbit, the radius drawn from the sun to the planet sweeps out equal areas in equal times.
(III) If the major axis of an elliptical orbit has length $2a$ and the period of revolution is T, then the quantity a^3/T^2 is the same for all the planets in our solar system.

There is no question that Newton gave an impeccable derivation of his universal law of gravitation from Kepler's laws. In other words, the *Principia* contains complete and satisfactory arguments that if the orbits of planets satisfy (I), (II), and (III), then the acceleration of the planets can be attributed to an inverse square central force whose origin is at the sun. But there is a continuing controversy as to whether the *Principia* provides adequate justification of the converse, namely, that the orbit of a planet under the influence of an inverse square central force must necessarily be a conic section. Weinstock [**18**],[**19**] gives the argument against Newton. The pros and cons are presented in a set of articles in the May 1994 issue of *The College Mathematics Journal*, together with numerous related references. The article by Nauenberg [**13**] in that issue lends support to the position that Newton's constructions are sufficient to justify the converse. A point of contention is whether Newton was really aware of the required uniqueness theorem (equivalent to the uniqueness of the solution of a second order differential equation given an initial position and initial velocity vector). Arnol'd [**1**, p. 33] gives the benefit of the doubt to Newton's basic understanding of the required uniqueness theorem, much as a professional basketball referee might do for Michael Jordan on a close call in a playoff game:

> Of course, one could raise the objection that Newton did not know this theorem. In fact, he did not state it in the form that we have just used. But he certainly knew it in essence, as well as many other applications of the theory of perturbations—the mathematical analysis of Newton is to a considerable extent a well-developed theory of perturbations.

In Section 6 we indicate how the hodograph method leads readily to the converse, once one knows that a Keplerian orbit is a conic section with the sun at one focus if and only if the polar reciprocal is a circle. The expression for the force law given below by (4.6), involving the curvature of the polar reciprocal, is the main tool used for this and other applications.

Finally, in order not to disappoint Richard Feynman, we briefly discuss differential equations in Section 12. Another benefit attendant to viewing the hodograph in terms of the polar reciprocal is the insight this brings to the standard analytic treatment of the orbits. In particular one sees why the substitution $u = 1/r$, where $r = r(\theta)$ is the polar equation of the orbit, brings about a simplification of the corresponding differential equation. As we shall see, this substitution transfers the analysis from the radial function of the orbit to the support function of its polar reciprocal, in terms of which the corresponding force law has a very simple form.

2. Polar reciprocals

Fix a point S as origin in a Euclidean plane (S will later serve as the sun) and consider a smooth curve C in that plane (C will later serve as a planetary orbit). To each point P on C we associate a point P^* in the plane as follows: let ℓ be the tangent line to C at P and suppose $p > 0$ is the distance from S to ℓ. Then P^* is defined to be that point at distance $1/p$ from S such that the ray $\overrightarrow{SP^*}$ is perpendicular to ℓ and intersects ℓ. In case ℓ passes through S, we associate with P a "point at infinity" in a direction perpendicular to ℓ. Then C^* is defined to be the locus of points P^* as P varies over C (see FIGURE 1 and FIGURE 5).

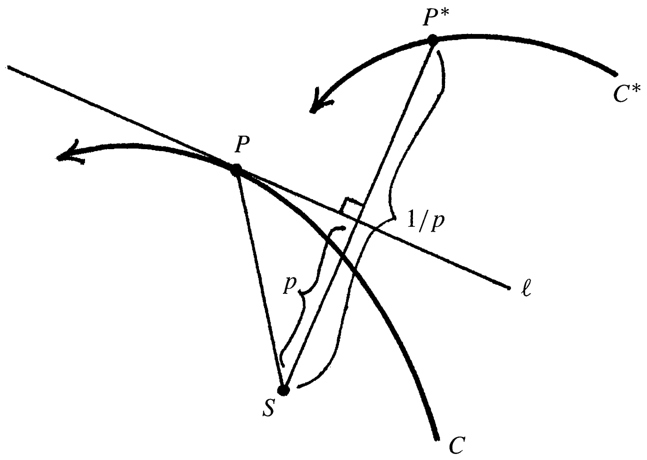


Figure 1 The definition of C^*.

The curve C^* is called the *polar reciprocal* of C, or more precisely, the polar reciprocal of C with respect to the unit circle centered at S. A readable exposition of properties of polar reciprocals from the viewpoint of projective geometry is given in Cremona [**2**, Chap. XXII]. One also finds a useful treatment in Guggenheimer [**7**], where C^* is called the *projective polar* of C. A key property is

$$C^* \textit{ is a conic section if and only if } C \textit{ is a conic section.} \tag{2.1}$$

Another central property is duality, namely,

$$(C^*)^* = C. \tag{2.2}$$

Thus, in FIGURE 2, with $r = SP$, $r^* = SP^*$, $p =$ the distance from S to ℓ, and $p^* =$ the distance from S to ℓ^* (where ℓ^* is the tangent line to C^* at P^*), we have

$$\ell \perp \overleftrightarrow{SP^*} \text{ and } r^* = 1/p \text{ by definition, and } \ell^* \perp \overleftrightarrow{SP} \text{ and } r = 1/p^* \text{ by duality.} \tag{2.3}$$

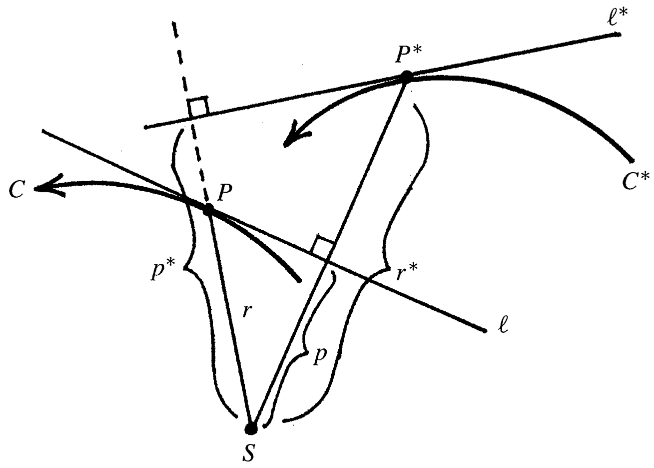


Figure 2 The dual relationship between C and C^*. ($p^* = 1/r$ follows by similar right triangles from $p = 1/r^*$.)

Those familiar with the theory of convex sets will recognize that if C is the boundary of a convex region $\bar{C}$ having S as an interior point, then C^* is the boundary of the usual polar body, or dual, of $\bar{C}$ (see [**3**] or [**16**] for the standard properties). The formulas (2.3) are familiar from this perspective. Letting θ be the polar angle of C (i.e., the angle between a fixed horizontal ray and $\overrightarrow{SP}$), as in FIGURE 3, and θ^* the polar angle of C^*, we see that $r = r(\theta)$ is the radial function of C and $r^* = r^*(\theta^*)$ the radial function of C^*. Then θ is the support angle of C^* (i.e., the angle between the fixed ray and the perpendicular to the tangent line at P^*) and θ^* is the support angle of C. The function $p = p(\theta^*)$ is the support function of C, while $p^* = p^*(\theta)$ is the support function of C^*.

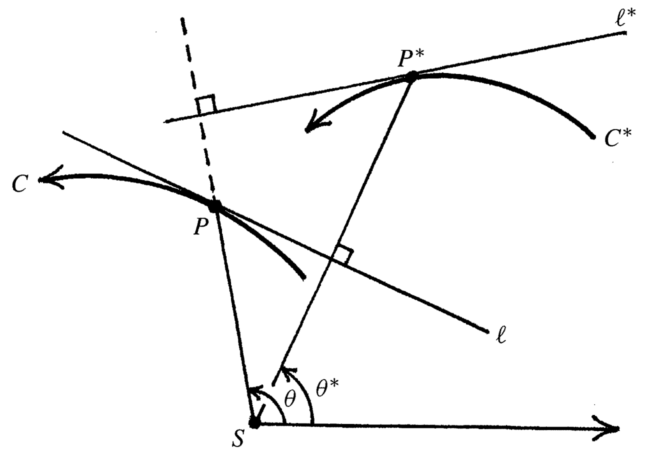


Figure 3 The polar and support angles of C and C^*.

The curvature of a plane curve is the rate of change of the angle of the tangent with respect to arclength, or equivalently, the rate of change of the support angle. Thus if ds and ds^* are the arclength elements of C and C^* respectively, the curvatures are given by $d\theta^*/ds$ and $d\theta/ds^*$ respectively. Let $\rho = \rho(\theta^*)$ and $\rho^* = \rho^*(\theta)$ be the radii of curvature (reciprocals of curvature) of C and C^* respectively. Then we have

$$ds = \rho d\theta^* \text{ and } ds^* = \rho^* d\theta. \tag{2.4}$$

We alert the reader that (2.4) naturally requires that ρ and ρ^* be well defined, which means that the corresponding curvatures should not vanish. In our later examples this will be true but for one exceptional case, which will require extra comment.

3. Hodographs of Keplerian orbits

In the following we shall view a curve C as the orbit of a planet P governed by some force law. Note that the definition of the polar reciprocal C^* does not depend on how C is parametrized, but is a purely geometric construct. On the other hand, when we view C as the orbit of a planet P, we have in mind a particular parametrization for which we interpret the parameter t physically as time. If P moves along C according to Kepler's second law (II), then we shall refer to C as a *Keplerian orbit*. In that case, if dA is the sectorial element of area bounded by an infinitesimal arc of length ds along C and by the radii drawn from S to the endpoints of the arc, we have that dA/dt is a positive constant. Consistent with traditional notation, we denote this constant by $h/2$. (In fact, h is the angular momentum of the planet P.) One could also say that a Keplerian orbit is a curve C parametrized by a constant multiple of the sectorial area. Now $dA = (1/2)p\,ds$ (see FIGURE 4), so if $v = ds/dt$ is the speed of P along C, we have

$$\frac{h}{2} = \frac{dA}{dt} = \frac{1}{2}p\frac{ds}{dt} = \frac{1}{2}pv = \frac{1}{2}\frac{v}{r^*}, \tag{3.1}$$

where we have used $p = 1/r^*$ from (2.3).

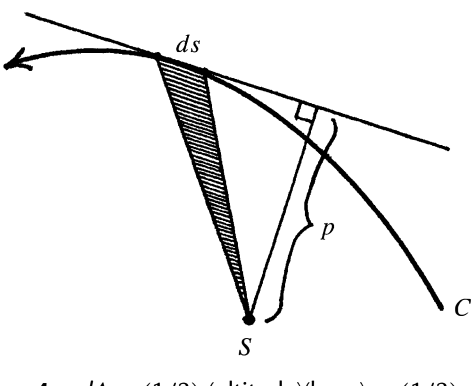


Figure 4 $dA = (1/2)$ (altitude)(base) $= (1/2)\,pds$.

From (3.1) we see that v is a constant multiple of r^*, that is,

$$v = hr^* = h(SP^*). \tag{3.2}$$

This result is in fact Proposition I, Cor. I, of the *Principia*. Let $\vec{v}$ be the velocity vector at P, so $|\vec{v}| = v = h(SP^*)$. Since the vector $\overrightarrow{SP^*}$ is perpendicular to ℓ (FIGURE 1) and $\vec{v}$ is parallel to ℓ, we see that $\vec{v}$ is obtained from $h(\overrightarrow{SP^*})$ by a counterclockwise rotation through 90°. We therefore have

$$\vec{v} = hI(\overrightarrow{SP^*}), \tag{3.3}$$

where I denotes 90° counterclockwise rotation. (If we think of C as embedded in the complex plane with S as origin, we can identify I with the complex imaginary unit i and interpret the operation of I in (3.3) as multiplication by i.)

The locus of the heads of the velocity vectors along C (with their tails all placed at S) is the hodograph, or velocity diagram, of the orbit C. From (3.3) we see that the hodograph of C is similar to C^*. In fact,

the hodograph of a Keplerian orbit C is obtained from C^* by rotating counterclockwise through 90° *and rescaling by the factor h.* (3.4)

4. The force law

Deducing the force law causing a planet P to follow a given orbit C in accordance with Kepler's second law (II), with S playing the role of the sun, requires an examination of the acceleration vector $\overrightarrow{accel} = d\vec{v}/dt$. From (3.3), using the obvious fact that differentiation commutes with the rotation I, we have

$$\overrightarrow{accel} = \frac{d\vec{v}}{dt} = hI\frac{d}{dt}(\overrightarrow{SP^*}). \tag{4.1}$$

Note that $d(\overrightarrow{SP^*})/dt$ is the rate of change of the position vector of C^* and is therefore parallel to ℓ^*. Since by duality ℓ^* is perpendicular to $\overrightarrow{SP}$, a 90° rotation turns it parallel to $\overrightarrow{SP}$. Consequently (4.1) implies that $\overrightarrow{accel}$ is parallel to $\overrightarrow{SP}$ and points toward S. By Newton's second law of motion, $\overrightarrow{accel} = \vec{F}/m$, where m is the mass of the planet P and $\vec{F}$ is the force acting on P. We thence conclude that

the force $\vec{F}$ acting on P is at all times directed toward S. (4.2)

Having established (4.2), our next objective is to find the magnitude $|\vec{F}|$ of the force $\vec{F}$. By Newton's second law this is tantamount to evaluating the magnitude $|\overrightarrow{accel}|$ of the acceleration vector of P. To this end, from (4.1) we find

$$|\overrightarrow{accel}| = h\left|\frac{d}{dt}(\overrightarrow{SP^*})\right| = h\frac{ds^*}{dt}, \tag{4.3}$$

since $|d(\overrightarrow{SP^*})| = ds^* =$ the arclength element of C^*.

From (2.4) we have $ds^* = \rho^* d\theta$, so (4.3) gives

$$|\overrightarrow{accel}| = h\rho^*\frac{d\theta}{dt}. \tag{4.4}$$

But the sectorial element of area depicted in FIGURE 4 also has the polar coordinate form $dA = (1/2)r^2\, d\theta$. That is,

$$\frac{1}{2}p\, ds = dA = \frac{1}{2}r^2\, d\theta. \tag{4.5}$$

Therefore,

$$r^2\frac{d\theta}{dt} = 2\frac{dA}{dt} = h,$$

so

$$\frac{d\theta}{dt} = \frac{h}{r^2}.$$

Consequently, from (4.4),

$$|\overrightarrow{accel}| = h^2 \frac{\rho^*}{r^2}. \tag{4.6}$$

In other words, the magnitude of the force acting on a planet P following a Keplerian orbit C is inversely proportional to the square of the distance from S to P and directly proportional to the radius of curvature of the polar reciprocal C^* at P^*. (One finds the equivalent result, framed in terms of the hodograph, in Hamilton [**8**, p. 302, eq. XXVI].)

5. When the orbit is a conic section

Imagine a planet P following a Keplerian orbit C that is a conic section. From (2.1) we know C^* is also a conic section. But of more importance to us is the fact that

if a conic section C has S at one focus, then C^* is in fact a circle. (5.1)

See Cremona [**2**, p. 266] for a proof. A direct proof is very easy using the polar coordinate equation for a conic section. FIGURE 5 shows an ellipse and its polar reciprocal.

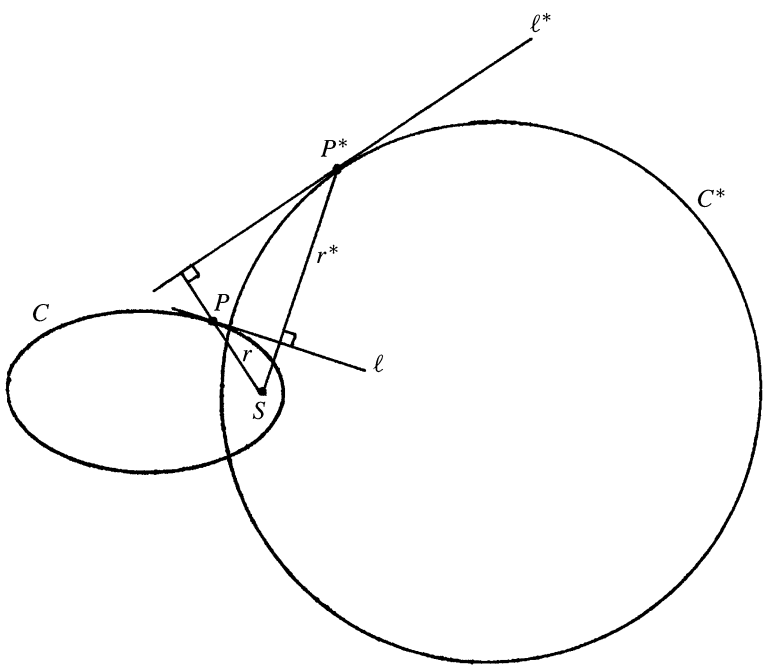


Figure 5 The polar reciprocal of an ellipse C with respect to a circle centered at a focus S is a circle C^*.

Thus, for a Keplerian conic section orbit with S at one focus, C^* is a circle, so $\rho^* \equiv$ *constant*. Therefore (4.6) shows that $|\overrightarrow{accel}|$ is proportional to $1/r^2$. From this and (4.2), we have Propositions XI, XII, and XIII in the *Principia*, rephrased:

If a planet P follows a conic section orbit with S at one focus and satisfies Kepler's second law (II), then the force acting on P is always directed toward S and has magnitude inversely proportional to the square of the distance from P to S. (5.2)

One obtains quantitative results more precise than (5.2) by finding the radius of C^*. For example, assuming the orbit C is an ellipse with one focus at S and having major and minor axes of lengths $2a$ and $2b$ respectively, one can show that C^* is a circle of radius a/b^2, so $\rho^* = a/b^2$. Then (4.6) gives

$$|\overrightarrow{accel}| = \frac{ah^2}{b^2}\frac{1}{r^2}. \tag{5.3}$$

Let T be the period of revolution, the time it takes to travel once around the ellipse. Since $dA/dt = h/2$, we have $\pi ab = (h/2)T$, or

$$h = \frac{2\pi ab}{T}. \tag{5.4}$$

Substituting this in (5.3) then gives

$$|\overrightarrow{accel}| = \frac{4\pi^2 a^3}{T^2}\frac{1}{r^2}. \tag{5.5}$$

Thus a planet of mass m satisfying Kepler's laws (I) and (II) must be acted on by a force $\vec{F}$ that is directed toward the sun and has magnitude

$$|\vec{F}| = \frac{4\pi^2 a^3}{T^2}\frac{m}{r^2}. \tag{5.6}$$

Under the further assumption that Kepler's third law (III) is satisfied, we are led to conclude from this that the magnitude of the force acting on each planet in our solar system is directly proportional to the mass of the planet and inversely proportional to the square of its distance from the sun, *where the proportionality constant $4\pi^2 a^3/T^2$ is independent of the particular planet*. Newton's experience with apples, together with Galileo's fabled Pisa experiments, indicated that the gravitational force of attraction is proportional to the product of the masses of the bodies involved, so in fact this proportionality constant must be of the form

$$\frac{4\pi^2 a^3}{T^2} = GM, \tag{5.7}$$

where M is the mass of the sun S, and G is a universal constant, the universal gravitational constant. Thus the magnitude of the force in Newton's law of universal gravitation takes the form

$$|\vec{F}| = \frac{GmM}{r^2} \tag{5.8}$$

6. The converse

Consider now the converse of (5.2). We ask whether a planet P under the influence of an inverse square force directed toward S must necessarily follow an orbit C that is a conic section with one focus at S. Observe that the argument we gave using polar reciprocals in Section 4, leading to (4.2), showed only that a Keplerian orbit is necessarily the result of a central force, while now we need to know that a central force implies a Keplerian orbit. The famous argument by Newton in the *Principia* [**14**, Propositions I and II] gives the result simply and elegantly in both directions, namely, an orbit satisfies Kepler's second law (II) and lies in a plane if and only if it is acted on by a force directed toward the fixed point S. Once we know this, the converse of (5.2) is easily established, for in this case (4.6) applies, and from $|\overrightarrow{accel}| = k/r^2$ we can conclude that $\rho^* \equiv constant$. Therefore C^*, being a plane curve of constant nonzero curvature, must be a circle (or arc of a circle). But then $C = (C^*)^*$ is the polar reciprocal of a circle and is consequently a conic section with S at one focus.

The cases where C is an ellipse, parabola, or hyperbola correspond respectively to whether S is inside, on, or outside the circle C^*. This is a consequence of the refinement of (5.1) given in Cremona [**2**, p. 266] but is also easy to establish using polar coordinates.

One anomalous case is not taken into account here. Equation (4.6) was derived under the assumption that the curvature of C^* is never zero, so that ρ^* would be well defined. Another possible orbit C above, of interest only to astronauts whose onboard computers have gone awry, lies along a straight line through S.

7. The force in terms of the geometry of the orbit

While (4.6) has a pleasing simplicity, it is often convenient to have at hand an expression for the force involving only quantities directly related to C. This is made possible by the relation

$$\rho\rho^* = \frac{r^3}{p^3}, \tag{7.1}$$

derived as follows. From (4.5), substituting $ds = \rho d\theta^*$, we have $p\rho d\theta^* = r^2\, d\theta$, or

$$\rho = \frac{r^2}{p}\frac{d\theta}{d\theta^*}. \tag{7.2}$$

The corresponding relation for C^* is

$$\rho^* = \frac{r^{*2}}{p^*}\frac{d\theta^*}{d\theta}. \tag{7.3}$$

Multiplication of (7.2) and (7.3) eliminates the pesky $d\theta^*/d\theta$ and we have

$$\rho\rho^* = \frac{r^2 r^{*2}}{pp^*}. \tag{7.4}$$

From this we obtain (7.1) by using $pr^* = p^*r = 1$.

Applying (7.1), we now substitute for ρ^* in (4.6) to obtain

$$|\overrightarrow{accel}| = h^2\frac{\rho^*}{r^2} = h^2\frac{r}{\rho p^3}, \tag{7.5}$$

the desired expression for the acceleration in terms of quantities directly related to C.

As an aside, if γ is the angle formed by $\overrightarrow{SP}$ and $\overrightarrow{SP^*}$ (in FIGURE 3 we have $\gamma = \theta - \theta^*$), then $\sec\gamma = r/p$. Thus (7.1) gives

$$\rho^* = (\sec^3\gamma)/\rho.$$

Substituting this in (4.6) gives an expression for the force law as presented, for example, in Needham [**12**, eq. (7)].

8. Circular orbits

In Proposition VII of the *Principia*, Newton treated the case of circular orbits where S is not necessarily the center of the circle. We use (7.5) to deal with this case, assuming C is a circle of radius a and S is any point inside or on C. Let c be the distance from S to the center of C and define b by $b^2 = a^2 - c^2$.

Recalling that $r = SP$, and that p is the distance from S to the tangent line of C at P, with two applications of the Pythagorean theorem in FIGURE 6, we have $(p-a)^2 + (r^2 - p^2) = c^2$. Thus $2ap = r^2 + a^2 - c^2 = r^2 + b^2$, giving finally

$$p = \frac{r^2 + b^2}{2a}. \tag{8.1}$$

Substituting this expression for p in (7.5), and using $\rho \equiv a$, gives

$$|\overrightarrow{accel}| = 8h^2a^2\frac{r}{(r^2+b^2)^3}. \tag{8.2}$$

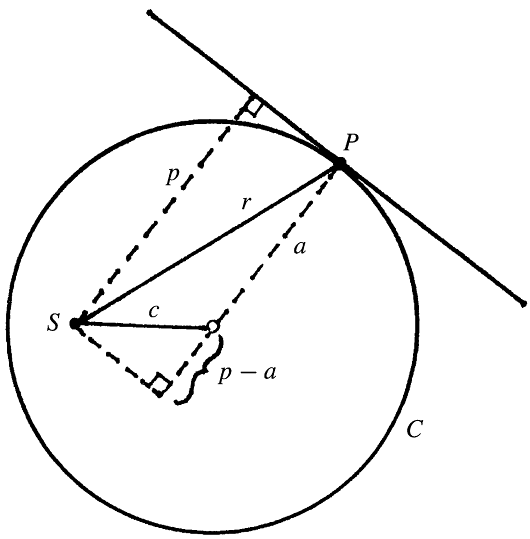


Figure 6 $c^2 = a^2 + r^2 - 2ap$.

The case where S lies on the circle is particularly interesting. Here $c = a$, so $b = 0$, and

$$|\overrightarrow{accel}| = \frac{8a^2h^2}{r^5}. \tag{8.3}$$

Therefore, if a central force causes a planet P to follow a circular orbit passing through the center of force S, then the magnitude of the force must be inversely proportional to the fifth power of the distance from P to S. (This is Corollary I to Proposition VII in the *Principia*.)

9. Elliptical orbits with the sun at the center

Proposition X of the *Principia* deals with an elliptical orbit C where the center of force S is at the center of the ellipse rather than at a focus. In order to apply (7.5) we use the fact that the support function p of an ellipse with semiaxes a and b, when the origin S is at the center of the ellipse, has the form

$$p(\theta^*) = (a^2\cos^2\theta^* + b^2\sin^2\theta^*)^{1/2}.$$

See [**6**, p. 21]. As mentioned in Section 12, the radius of curvature can be expressed in terms of the support function by $\rho = p + p''$. Then a straightforward, if somewhat tedious, calculation gives

$$\rho = \frac{a^2b^2}{p^3}. \tag{9.1}$$

Another proof is found in Salmon [**15**, pp. 169 and 228]. Then (7.5) immediately gives

$$|\overrightarrow{accel}| = \frac{h^2}{a^2b^2}r, \tag{9.2}$$

so the magnitude of the force acting on P must be directly proportional to the distance from S to P. Introducing the period of revolution T, and the fact that $\pi ab = (h/2)T$ (from $dA/dt = h/2$), we have

$$|\overrightarrow{accel}| = \frac{4\pi^2}{T^2}r. \tag{9.3}$$

10. Transmutation of the force law

Suppose P traverses an orbit C under the influence of a central force directed toward S. Given another fixed point $\tilde{S}$, what force directed toward $\tilde{S}$ would cause P to move along the same orbit with the same period of revolution? Newton answered this in Corollary III to Proposition VII of the *Principia*. In Needham [**12**] one finds a nice exposition of this and related matters in terms of complex analysis. We give here the equivalent result obtained via equation (7.5).

Let r and p be the radial and support functions of C relative to the origin S, and $\tilde{r}$ and $\tilde{p}$ the corresponding functions relative to $\tilde{S}$. Let F denote the magnitude of the force directed toward S and $\tilde{F}$ that of the force directed toward $\tilde{S}$. We now proceed under Newton's assumption that the "periodic times" are the same in both cases, which is equivalent to the assumption that the rates dA/dt and $d\tilde{A}/dt$ at which the sectorial areas are swept out are the same (although the planet traverses the orbit at different velocities when under the influence of the different forces). From (7.5) we have

$$F = mh^2 \frac{r}{\rho p^3} \text{ and } \tilde{F} = m\tilde{h}^2 \frac{\tilde{r}}{\tilde{\rho} \tilde{p}^3}, \tag{10.1}$$

for any given point P on C. But $\rho = \tilde{\rho}$ at P, and the assumption about periodic times implies $h = \tilde{h}$, so we have

$$\tilde{F} = \frac{(\tilde{r}/\tilde{p}^3)}{(r/p^3)} F. \tag{10.2}$$

One example among those given by Newton is that of an ellipse C with S as center and $\tilde{S}$ at one focus. In FIGURE 7, let A be the other focus and B the reflection of A across the tangent line ℓ at P.

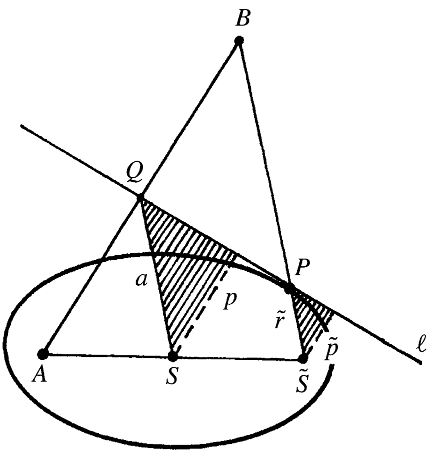


Figure 7 Showing that $p/\tilde{p} = a/\tilde{r}$.

The reflection property of the ellipse implies that $\tilde{S}B$ passes through P and has length $2a$ equal to the major axis of the ellipse, since ℓ is the perpendicular bisector of segment AB. It follows that SQ is parallel to $\tilde{S}B$ and has length a. Hence the shaded right triangles are similar, implying that

$$\frac{p}{\tilde{p}} = \frac{a}{\tilde{r}}. \tag{10.3}$$

Using (10.3) in (10.2) gives

$$\tilde{F} = \frac{a^3}{r\tilde{r}^2} F. \tag{10.4}$$

Thus, given a knowledge of the force law (9.3) for the magnitude of the force centered at S, we obtain from (10.4) the inverse square law,

$$\tilde{F} = \frac{a^3}{r\tilde{r}^2} \frac{4\pi^2 r}{T^2} = \frac{4\pi^2 a^3}{T^2} \frac{1}{\tilde{r}^2} = \frac{4\pi^2 a^3}{\tilde{T}^2} \frac{1}{\tilde{r}^2},$$

for the force centered at the focus $\tilde{S}$. Similarly, one deduces (9.3) from (5.5) via (10.4).

As an exercise, the reader may apply (10.2) to relate the force laws corresponding to a circular orbit, where S is chosen at the center of the circle and $\tilde{S}$ any point on the circle itself.

11. The connection with Newton's approach

In order to find the force law associated with a given Keplerian orbit C, Newton considered the planet moving along C from P to Q over a short period of time Δt, and drew the line through Q parallel to $\overrightarrow{SP}$ and intersecting ℓ at R, where ℓ is the tangent line to C at P (see FIGURE 8).

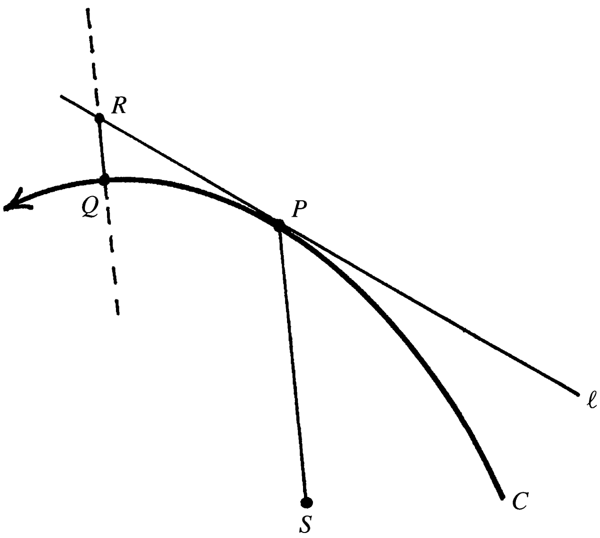


Figure 8 The planet moves from P to Q in time Δt.

The expression he used to calculate the magnitude of the force acting on P is proportional to

$$\lim_{Q\to P} \frac{RQ}{(\Delta A)^2}, \tag{11.1}$$

where ΔA is the area of the sector bounded by the arc of C from P to Q and by the radii drawn from S to P and Q. (Note that RQ is the magnitude of the displacement of P from the rectilinear path it would have followed had the force been absent. Since $\Delta A = (h/2)\Delta t$, we see that the expression (11.1) is proportional to the magnitude of the acceleration toward S, although we won't assume this in the following.)

To connect (11.1) with (4.6), we begin by giving a geometric argument showing that for any curve C, not necessarily a Keplerian orbit,

$$RQ \sim \frac{\Delta s \Delta s^*}{r^*}, \tag{11.2}$$

where Δs is the arclength along C from P to Q, Δs^* is the arclength along C^* from P^* to Q^*, and of course $r^* = SP^*$. Here $f \sim g$ means that f is asymptotic to g as $\Delta t \to 0$, that is, $\lim_{\Delta t\to 0}(f/g) = 1$.

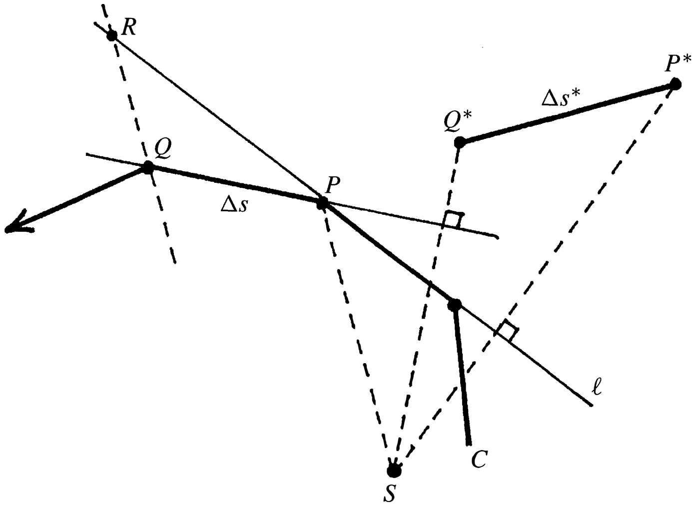


Figure 9 Polygonal approximations of C and C^*.

To establish (11.2) we proceed in the spirit of Newton, replacing C and C^* by polygons as in FIGURE 9.

In FIGURE 9, ℓ plays the role of the tangent line to C at P and $\overleftrightarrow{PQ}$ the role of the tangent line at Q. The point R is the intersection of ℓ with the line through Q parallel to $\overleftrightarrow{SP}$. By the polar duality of C^* and C, the line $\overleftrightarrow{P^*Q^*}$ is perpendicular to $\overleftrightarrow{PS}$ and therefore also perpendicular to $\overleftrightarrow{RQ}$. But $\overleftrightarrow{Q^*S}$ is perpendicular to $\overleftrightarrow{PQ}$, and $\overleftrightarrow{SP^*}$ is perpendicular to $\overleftrightarrow{PR}$. Therefore triangle P^*Q^*S is similar to triangle RQP, so

$$\frac{RQ}{QP} = \frac{P^*Q^*}{Q^*S}, \tag{11.3}$$

or

$$\frac{RQ}{\Delta s} = \frac{\Delta s^*}{Q^*S}. \tag{11.4}$$

Since $Q^*S \to r^*$ as $Q \to P$, we obtain from (11.4) the desired relation (11.2).

Now we note that $\Delta s^* \sim \rho^*\Delta\theta$ and $r^* = 1/p$, so (11.2) implies

$$RQ \sim (p\Delta s)(\rho^*\Delta\theta) = (p\Delta s)(r^2\Delta\theta)\frac{\rho^*}{r^2}. \tag{11.5}$$

But, from (4.5), $\Delta A \sim (1/2)(p\Delta s)$ and also $\Delta A \sim (1/2)(r^2\Delta\theta)$, so (11.5) yields

$$\frac{RQ}{(\Delta A)^2} \sim 4\frac{\rho^*}{r^2}. \tag{11.6}$$

Assuming now that C is a Keplerian orbit, so $\Delta t = 2\Delta A/h$, we obtain from (11.6)

$$\frac{RQ}{(\Delta t)^2} \sim h^2\frac{\rho^*}{r^2}. \tag{11.7}$$

Thus (11.2) shows that the expression for force given by (11.1) is equivalent to (4.6). (The reader may note that (11.7) could also be obtained from (11.2) using the results in Section 4, starting with (4.3).)

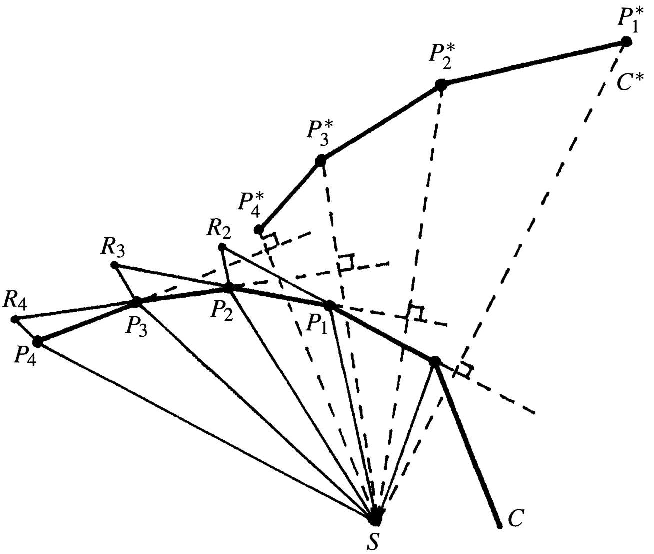


Figure 10 *A polygon C and its polar reciprocal C*.*

In FIGURE 10 we find a fuller picture of the relationship between a polygon $C = P_1P_2P_3\ldots$ and its polar reciprocal $C^* = P_1^*P_2^*P_3^*\ldots$.

In the figure, point R_n is on $\overleftrightarrow{P_{n-2}P_{n-1}}$ and $\overleftrightarrow{R_nP_n}$ is parallel to $\overleftrightarrow{SP}_{n-1}$. The argument leading to (11.2) shows that, for each $n = 1, 2, \ldots$, triangle $P_nP_{n+1}R_{n+1}$ is similar to triangle $SP_{n+1}^*P_n^*$. If in addition $P_nR_{n+1} = P_{n-1}P_n$ for $n = 2, 3, \ldots$, then C corresponds to a Keplerian orbit, and sliding each triangle $P_nP_{n+1}R_{n+1}$ so that P_n ends up at S produces the hodograph as a similar copy of C^*.

12. Doing something with our differential equations

A useful analytic expression for the radius of curvature of C in terms of its support function is

$$\rho(\theta^*) = p(\theta^*) + p''(\theta^*). \tag{12.1}$$

The proof of this relationship is traditionally presented in the context of convex curves (see [**3**, p. 26], [**6**, p. 22], or [**16**, p. 110]) but is easily carried over to more general plane curves, although in these cases some care is required in the definition of the angle function θ^*.

We want to apply (12.1) in the case of the polar reciprocal C^*, where we have

$$\rho^*(\theta) = p^*(\theta) + p^{*\prime\prime}(\theta). \tag{12.2}$$

To simplify notation we let $u = p^*$, so $\rho^* = u + u''$. If the force law is such that $|\overrightarrow{accel}| = f(r)$, then (4.6) gives

$$u'' + u = \frac{r^2}{h^2} f(r). \tag{12.3}$$

Since $r = 1/p^* = 1/u$, this is of the form

$$u'' + u = \frac{1}{h^2u^2} f(1/u) = g(u). \tag{12.4}$$

This last equation is often encountered in a treatment of planetary motion, obtained by first deriving the differential equation for the motion in terms of polar coordinates, and then substituting $r = 1/u$ to reach a simplified form. We see that this substitution can be viewed as essentially taking us from the analysis of the radial function of the orbit C to the support function of its polar reciprocal C^*. In other words, solutions of (12.4) determine the support function of C^*, and this in turn determines the radial function of C.

Acknowledgment. I am indebted to Sherman Stein for several suggestions contributing to the improvement of this paper. I am also deeply grateful to the two referees for their careful reading of the manuscript and for their numerous comments that also significantly improved the exposition.

REFERENCES

1. V. I. Arnol'd, *Huygens and Barrow, Newton and Hooke*, Birkhäuser Verlag, Berlin, Germany, 1990.
2. L. Cremona, *Elements of Projective Geometry.* Reprint, Dover, New York, NY, 1960.
3. R. J. Gardner, *Geometric Tomography*, Cambridge University Press, Cambridge, UK, 1995.
4. D. L. Goodstein and J. R. Goodstein, *Feynman's Lost Lecture*, W. W. Norton and Company, New York, NY, 1996.
5. G. W. Griffiths, Review of [**4**], *Math. Intelligencer* **20** (3) (1998), 68–70.
6. H. Groemer, *Geometric Applications of Fourier Series and Spherical Harmonics*, Cambridge University Press, Cambridge, UK, 1996.
7. H. Guggenheimer, *Applicable Geometry*, Robert E. Kreiger Publishing Company, New York, NY, 1977.
8. W. R. Hamilton, *Elements of Quaternions*, Volume II. Reprint, Chelsea, New York, NY, 1969.
9. A. Lenard, Kepler orbits *More Geometrico*, *College Math. J.* **25** (1994), 90–98.
10. J. C. Maxwell, *Matter and Motion*, 1877. Reprint, Dover, New York, NY.
11. J. Milnor, On the geometry of the Kepler problem, *Amer. Math. Monthly* **90** (1983), 353–365.
12. T. Needham, Newton and the transmutation of force, *Amer. Math. Monthly* **100** (1993), 119–137.
13. M. Nauenberg, Newton's *Principia* and inverse-square orbits, *College Math. J.* **25** (1994), 212–221.
14. I. Newton, *Principia*, Motte's translation revised by Cajori, University of California Press, Berkeley, CA, 1934.
15. G. Salmon, *A Treatise on Conic Sections*, Longmans, Green, and Co., London, UK, 1900.
16. R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*, Cambridge University Press, Cambridge, UK, 1993.
17. S. Stein, Exactly how did Newton deal with his planets? *Math. Intelligencer* **18**, (2) (1996), 6–11.
18. R. Weinstock, Dismantling a centuries-old myth: Newton's *Principia* and inverse-square orbits, *American Journal of Physics* **50** (1982), 610–617.
19. R. Weinstock, Isaac Newton: Credit where credit won't do, *College Math. J.* **25** (1994), 179–192.
20. R. Weinstock, Review of [**4**], *Math. Intelligencer* **21** (3) (1999), 71–73.

---

40 years ago in the MAGAZINE (Vol. 34, No. 3, Jan.–Feb., 1961):

Announcement: Commencing with Volume 35, the MATHEMATICS MAGAZINE will become an official publication of The Mathematical Association of America. It will be owned and supported by the Association. Friends and readers of the MAGAZINE will be pleased to know that such a strong and prominent professional organization saw fit to guarantee the future of the MAGAZINE as a journal devoted to collegiate mathematics.

Folding Quartic Roots

B. CARTER EDWARDS
JERRY SHURMAN
Reed College
Portland, OR 97202

1. Introduction

The geometric tools of antiquity—straightedge and compass—solve linear and quadratic equations but fail at cubics, to the persistent consternation of angle trisectors. Recent articles show how other geometric systems—origami [**6**] or Mira* [**4**]—go beyond straightedge and compass to solve cubic equations as well; the key is to construct a common tangent to two parabolas. Solving cubics and even equations of higher degree by parabolas or other conics is an old idea ([**7, 1, 3**], and, more recently, [**5**]), but origami and Mira methods share the novel feature of constructing common tangents to parabolas from the parabolas' foci and directrices without needing the parabolas themselves. Thus these methods really use only points and lines.

Since algebra reduces the quartic equation to equations of lower degree, we know in principle how to solve the quartic by tangents to parabolas as well, but translating the algebraic reduction into an origami- or Mira-based procedure is geometrically unmotivated and unappealingly tortuous. Instead, this note shows how to solve certain quartics (to be detailed in a moment) by the common tangents to a parabola and a circle. Granting field operations and two square roots, the actual construction is a geometric hybrid, requiring one compass operation (i.e., a circle) and one origami-type fold. Our tools operate in the Euclidean plane, so of course we are considering real quartics and are constructing real roots.

The path from common tangents to polynomial roots is as follows: Every conic has a so-called dual conic whose points represent the tangents to the original, so a common tangent to two conics leads naturally to a common point of their duals. Having this common point amounts to solving simultaneous quadratic equations; choosing the conics judiciously and substituting one equation into the other shows that in fact we have a solution to a cubic or quartic equation.

Specifically, we will solve a reduced quartic (one with no x^3 term),

$$x^4 + bx^2 + 2cx + d = 0,$$

obtained from the general quartic by translating the variable. We have to impose the conditions

$$c^2 - bd < 0, \quad d < 0;$$

i.e., the quadratic part $bx^2 + 2cx + d$ is negative for all x. While these constraints are unnecessarily restrictive for a real root, they certainly guarantee one and they arise naturally from our construction.

This work stems from the first author's undergraduate thesis under the direction of the second author.

*Mira is a registered trademark of the Mira-Math Company, Willowdale, Ontario, Canada. It refers to a transparent and reflective piece of plastic that allows constructions based on being able to see both a figure and its reflection.

2. Origami folds

In [6], R. Geretschläger lays out a set of axioms for origami-generated geometry. He observes that folding a given point F onto any point Q of a line $\mathcal{D}$ constructs a tangent to the parabola $\mathcal{P}$ having F and $\mathcal{D}$ for its focus and directrix. See FIGURE 1.

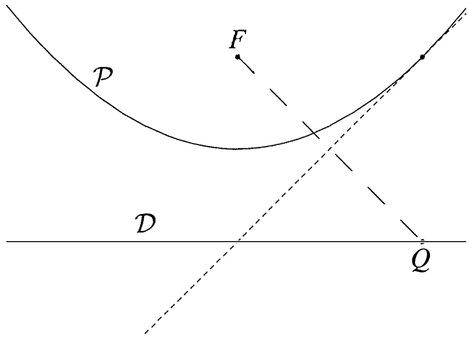


Figure 1 Folding a tangent to a parabola.

Suppose that two parabolas $\mathcal{P}_1$, $\mathcal{P}_2$, with foci F_1, F_2 and directrices $\mathcal{D}_1$, $\mathcal{D}_2$, have common tangents. A continuum of foldings takes F_1 to points of $\mathcal{D}_1$, constructing all tangents to $\mathcal{P}_1$. In particular, sliding F_1 along $\mathcal{D}_1$ until F_2 also lies on $\mathcal{D}_2$ constructs the common tangents to $\mathcal{P}_1$ and $\mathcal{P}_2$. Thus we have the axiom that the common tangents to two parabolas, each specified by its focus and directrix, are constructible by origami when they exist. Of course, they needn't exist at all—consider the case when one parabola lies entirely within the convex hull of the other.

In a similar vein, this paper uses the common tangents to a parabola and a circle. Suppose that a parabola $\mathcal{P}$, with focus F and directrix $\mathcal{D}$, and a circle $\mathcal{C}$, with center O and radius r, have a common tangent. Use a compass to plot the circle $2\mathcal{C}$ with center O and radius $2r$. Then sliding F along $\mathcal{D}$ until O lies on $2\mathcal{C}$ gives all folds tangent to $\mathcal{P}$ and $\mathcal{C}$.

Introducing a compass construction into origami geometry is allowable since origami subsumes ruler and compass [**4, 6**]. Using the compass may be aesthetically unappealing, but it will let us give a palatable folding procedure for solving the quartic equation.

3. Results from projective space

Working with conics is easy in projective space, where the computations are handled nicely by matrix algebra. (See [**8**] for a lovely introduction to projective geometry.) Very briefly, homogenizing polynomial equations with an extra variable symmetrizes calculations and corresponds to completing space out at infinity, giving problems the right number of solutions. Specifically, the quadratic equation of a conic in the plane homogenizes to projective form

$$\mathcal{C} : a_{11}x^2 + 2a_{12}xy + 2a_{13}xz + a_{22}y^2 + 2a_{23}yz + a_{33}z^2 = 0\,,$$

or, using matrices,

$$\mathcal{C} : v^t M v = 0 \qquad \text{where} \qquad v = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad \text{and} \qquad M = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}.$$

Note that the symmetric matrix M is defined only up to nonzero scalar multiple. The conic $\mathcal{C}$ is nondegenerate (not the union of two lines) when $\det M \neq 0$ [**8**, Section 1.7], and these are the only conics we consider.

The dual conic of $\mathcal{C}$ is described by the inverse matrix:

$$\widehat{\mathcal{C}} : v^t M^{-1} v = 0.$$

Multiplying by M gives a bijection from $\mathcal{C}$ to $\widehat{\mathcal{C}}$, for if $q = Mp$ with $p \in \mathcal{C}$ then substituting $p = M^{-1}q$ into $p^t M p = 0$ gives $q^t M^{-1} q = 0$.

We are interested in the dual $\widehat{\mathcal{C}}$ because its points describe the tangents to the original conic $\mathcal{C}$. To see this, compute that the partial derivative of $v^t M v$ with respect to the ith variable is $2v^t m_i$ where m_i is the ith column of M; thus the gradient of $\mathcal{C}$ at a point p is $\nabla_p(v^t M v) = 2p^t M$ (as a row vector), and the tangent line to $\mathcal{C}$ at p, being orthogonal to the gradient, is $\mathcal{L} : p^t M v = 0$. Again letting $q = Mp \in \widehat{\mathcal{C}}$, we see that the tangent is then

$$\mathcal{L}_q : q^t \cdot v = 0.$$

Thus the tangent to $\mathcal{C}$ at p is described by the point $q \in \widehat{\mathcal{C}}$, as desired. Decompressing the notation, we see that the tangent line's equation $Ax + By + Cz = 0$ comes directly from the coordinates of $q = [A \quad B \quad C]^t$.

It follows that if $\mathcal{C}_1$ and $\mathcal{C}_2$ are conics then finding their common tangents is equivalent to finding the points where their duals meet. By Bezout's theorem (see, e.g., [**2**]), since the duals have quadratic equations there are four such points in the complex projective plane, counting multiplicity. Since $\mathcal{C}_1$ and $\mathcal{C}_2$ have real matrices, so do their duals and these points occur in complex conjugate pairs, meaning that an even number of them are real. In sum, $\mathcal{C}_1$ and $\mathcal{C}_2$ have zero, two, or four real common tangents, counting multiplicity.

4. Solving the cubic

Geretschläger solves the cubic equation

$$x^3 + bx^2 + cx + d = 0$$

using (essentially—his variables are $-1/2$ times ours) the parabolas

$$\mathcal{P}_1 : (y + c)^2 = -4d(x - b) \quad \text{and} \quad \mathcal{P}_2 : x^2 = -4y.$$

We reproduce his computation but work projectively to foreshadow our methods for the quartic.

The parabolas' matrices are

$$M_1 = \begin{bmatrix} 0 & 0 & 2d \\ 0 & 1 & c \\ 2d & c & c^2 - 4bd \end{bmatrix} \quad \text{and} \quad M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix}$$

with inverses

$$M_1^{-1} = \frac{1}{d} \begin{bmatrix} b & -c/2 & 1/2 \\ -c/2 & d & 0 \\ 1/2 & 0 & 0 \end{bmatrix} \quad \text{and} \quad M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix}.$$

Inspecting the inverses shows that the duals meet at points $[A \quad B \quad C]^t$ where

$$bA^2 - cAB + AC + dB^2 = 0, \qquad -BC = A^2 .$$

For nonvertical affine common tangents of the original pair of parabolas, we normalize to $B = -1$, $z = 1$, de-homogenizing the tangent $Ax + By + Cz = 0$ to $y = Ax + C$. This gives $C = A^2$ in the second relation, and the first becomes

$$A^3 + bA^2 + cA + d = 0 .$$

Thus the slopes of common tangents to two parabolas solve the general cubic equation. FIGURE 2 shows this method applied to the cubic equation $x^3 - 2x^2 - x + 2 = 0$, with roots 2, 1, and -1.

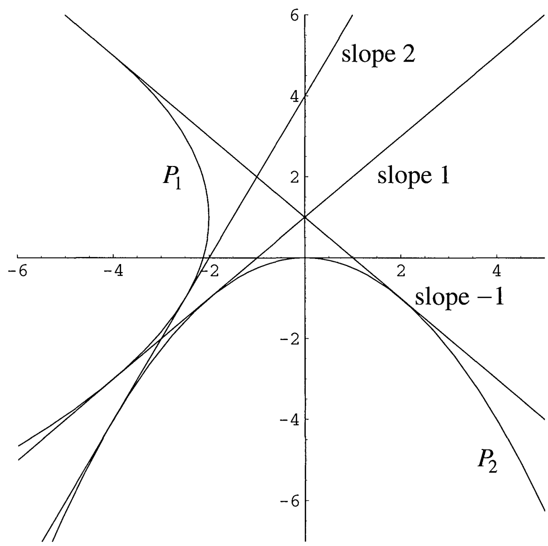


Figure 2 Solving the cubic by slopes of common tangents.

5. Solving the quartic

No construction of common tangents to two parabolas can solve the quartic, even in reduced form

$$x^4 + bx^2 + 2cx + d = 0,$$

because any two parabolas share a projective tangent out at infinity, not leaving enough affine tangents for the four possible roots.

We *can* solve the reduced quartic (assuming its quadratic part $bx^2 + 2cx + d$ has nonzero discriminant $bd - c^2$) by replacing the first parabola in the cubic method with a nonparabolic conic. Set

$$M_1 = \begin{bmatrix} d & c & 0 \\ c & b & 0 \\ 0 & 0 & bd - c^2 \end{bmatrix} \quad \left(\text{so } M_1^{-1} = \frac{1}{bd - c^2} \begin{bmatrix} b & -c & 0 \\ -c & d & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$$

and keep M_2 from the cubic case. Then, by the same algebraic method, the slopes A of the common tangents to the conic

$$\mathcal{C} : dx^2 + 2cxy + by^2 + bd - c^2 = 0$$

and the parabola $\mathcal{P}_2 : x^2 = -4y$ solve the reduced quartic equation. That is,

$$A^4 + bA^2 + 2cA + d = 0, \qquad \text{assuming} \quad bd - c^2 \neq 0.$$

Unfortunately, geometric methods for constructing the common tangent in question aren't immediately clear (at least to the authors), since the conic $\mathcal{C}$ is not a parabola.

As a second approach—less elegant algebraically but more realizable geometrically—we consider the common tangents to a circle at the origin, constructible by compass once we specify its radius, and to a variable parabola.

Thus, let b, c, d be given with $c^2 - bd < 0$ and $d < 0$. Set

$$e = \pm\frac{\sqrt{bd - c^2}}{d} \text{ (either value)}, \qquad r = |e|\sqrt{-d},$$

so that $-de^2/r^2 = 1$ and $de^2 = b - c^2/d$. Consider the dualized conics $\widehat{\mathcal{C}}$ and $\widehat{\mathcal{P}}$ with matrices

$$M_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/r^2 \end{bmatrix} \quad \text{and} \quad M_2^{-1} = \begin{bmatrix} 0 & 0 & de/2 \\ 0 & -d & c/2 \\ de/2 & c/2 & 0 \end{bmatrix}.$$

Computing the matrices M_1 and M_2 shows that the original conics $\mathcal{C}$ and $\mathcal{P}$ are respectively a circle of radius r and a parabola in affine space:

$$\mathcal{C} : x^2 + y^2 = r^2 \quad \text{and} \quad \mathcal{P} : c^2x^2 - 2cdexy + d^2e^2y^2 - 4d^2ex = 0.$$

The duals $\widehat{\mathcal{C}}$ and $\widehat{\mathcal{P}}$ intersect at points $[A \quad B \quad C]^t$ where

$$A^2 + B^2 = (1/r^2)C^2, \qquad deAC = dB^2 - cBC.$$

Again this describes a common tangent line $Ax + By + Cz = 0$ to the original circle and parabola, and setting $B = -1$, $z = 1$ gives the affine form $y = Ax + C$ with y-intercept C,

$$A^2 + 1 = (1/r^2)C^2, \qquad deAC = cC + d.$$

Multiply the first equation by de^2C^2 and square the second to get

$$de^2A^2C^2 + de^2C^2 = (de^2/r^2)C^4, \qquad de^2A^2C^2 = (c^2/d)C^2 + 2cC + d.$$

Substitute the second equation into the first to obtain

$$(-de^2/r^2)C^4 + (de^2 + c^2/d)C^2 + 2cC + d = 0,$$

and recall that $-de^2/r^2 = 1$ and $de^2 = b - c^2/d$, so that

$$C^4 + bC^2 + 2cC + d = 0.$$

Thus the y-intercept solves the reduced quartic when its quadratic part is negative.

This method is really no different algebraically from that used to solve the cubic equation, despite finding roots as intercepts rather than slopes. Projectively the intercepts and slopes are both coefficients; a slope in one affine piece of the projective plane is an intercept in another.

6. The geometric procedure

Translating the preceding algebra into a geometric construction requires locating the focus and directrix of the parabola $\mathcal{P}$. Its equation,

$$\mathcal{P} : c^2x^2 - 2cdexy + d^2e^2y^2 - 4d^2ex = 0,$$

becomes, after a change of variables,

$$\mathcal{P} : z^2 = 4d^3e^2w$$

where

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} c/(bd) & de & c^2e/b^2 \\ -e/b & c & -(bc + cde^2)/b^2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z \\ w \\ 1 \end{bmatrix}.$$

In (z, w)-coordinates, the parabola $\mathcal{P}$ has focus $F = (0, d^3e^2)$ and directrix $\mathcal{D} : w = -d^3e^2$. The (z, w)-origin has (x, y)-coordinates

$$O' = \left(\frac{c^2e}{b^2}, -\frac{bc + cde^2}{b^2} \right),$$

and the focus and directrix in (x, y)-coordinates are

$$F' = O' + d^3e^2(de, c) \quad \text{and} \quad \mathcal{D}' : (x, y) = O' - d^3e^2(de, c) + t(c, -de), \quad t \in \mathbf{R}.$$

With the calculations complete, we have the procedure for solving the reduced quartics whose coefficients satisfy the conditions $c^2 - bd < 0$, $d < 0$. Given b, c, d, compute e and r and use a compass to draw the circle $2\mathcal{C}$ at the origin of radius $2r$. Plot the focus F' and the directrix $\mathcal{D}'$. Fold the paper in a fashion that takes the origin onto the circle and the focus onto the directrix. All such folds are common tangents to $\mathcal{C}$ and $\mathcal{P}$, so their y-intercepts are roots.

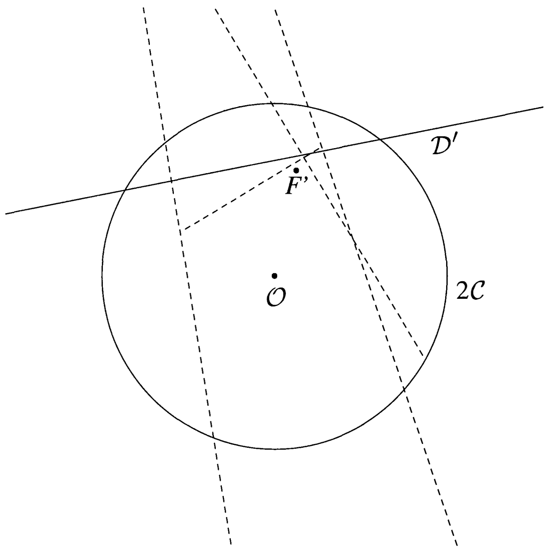


Figure 3 The four origami folds taking $\mathcal{O}$ to $2\mathcal{C}$ and F' to $\mathcal{D}'$.

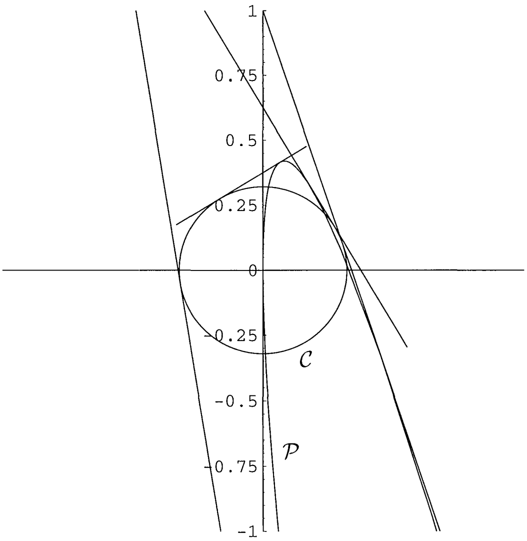


Figure 4 Solving the quartic by y-intercepts of common tangents.

7. An example

Consider the quartic polynomial

$$q(x) = x^4 - \frac{177}{64}x^2 + \frac{143}{64}x - \frac{15}{32} = \left(x - \frac{3}{8}\right)\left(x - \frac{5}{8}\right)(x-1)(x+2).$$

The origin $\mathcal{O}$, the circle $2\mathcal{C}$, the focus F', and the directrix $\mathcal{D}'$ are plotted in FIGURE 3, along with the four folds taking the origin to the circle and the focus to the directrix.

In FIGURE 4, these folds are seen as the common tangents of the circle $\mathcal{C}$ and the parabola $\mathcal{P}$. They visibly intersect the y-axis at three of the roots, 3/8, 5/8, 1; the fourth root −2 is out of the plot range.

REFERENCES

1. J.L. Berggren, *Episodes in the Mathematics of Medieval Islam*, Springer-Verlag, New York, NY, 1986, pp. 120–124.
2. Egbert Brieskorn and Horst Knorrer, *Plane Algebraic Curves*, Birkhäuser, Boston, MA, 1986.
3. Rene Descartes, *The Geometry of Rene Descartes*, Book III, translated from the French and Latin by David Eugene Smith and Marcia L. Latham, Dover, New York, NY, 1954.
4. John W. Emert, Kay I. Meeks, Roger B. Nelson, "Reflections on a Mira," *Amer. Math. Monthly* **101** (1994), 544–549.
5. William M. Faucette, "A Geometric Interpretation of the Solution of the General Quartic Polynomial," *Amer. Math. Monthly* 103 (1996), 51–57.
6. Robert Geretschläger, "Euclidean Constructions and the Geometry of Origami," this MAGAZINE **68** (1995), 357–371.
7. Sir Thomas Heath, *A History of Greek Mathematics, Vol. I*, Dover, New York, NY, 1981, pp. 241–244, 251–255.
8. Pierre Samuel, *Projective Geometry*, Springer-Verlag, New York, NY, 1988.

Cardan Polynomials and the Reduction of Radicals

THOMAS J. OSLER
Rowan University
Glassboro, NJ 08028

1. Introduction

Expressions of the form $\sqrt[n]{a+\sqrt{b}}$ occur frequently in the literature. To begin with, Cardan's solution of the cubic equation

$$x^3 - 3cx = 2a \tag{1}$$

is given by the sum of two radicals

$$x = \sqrt[3]{a+\sqrt{a^2-c^3}} + \sqrt[3]{a-\sqrt{a^2-c^3}}. \tag{2}$$

The author was teaching a course in the history of mathematics and designing simple problems of this type for solution by his students when he encountered an awkward algebraic result. With $c = -1$ and $a = 2$, equation 1 becomes

$$x^3 + 3x = 4$$

which has the solution $x = 1$. However, the solution given by (2) is

$$x = \sqrt[3]{2+\sqrt{5}} + \sqrt[3]{2-\sqrt{5}}. \tag{3}$$

A quick check with a calculator shows that this expression is really 1, but how can one algebraically manipulate it into that result? All attempts to do so led the author back to the original cubic from which it started. (The reader might try now to reduce (3) to the value 1.) That such a combination of square and cube roots should produce a rational number, let alone an integer, is startling. Later we will see that the individual radicals are related to the golden section and its reciprocal:

$$\sqrt[3]{2\pm\sqrt{5}} = \frac{1\pm\sqrt{5}}{2}. \tag{4}$$

This last relation is especially interesting because it shows that these cube roots are constructible with straightedge and compass. In general, we can construct any number defined by an expression composed of a finite number of additions, subtractions, multiplications, divisions and square roots of rational numbers. It is known that the number $\sqrt[3]{2}$ is not constructible. Yet (4) shows that $\sqrt[3]{2+\sqrt{5}}$, which looks more complicated than $\sqrt[3]{2}$, is constructible because it can be expressed by the golden section. This is indeed an unexpected result. It is surprises of this type that make it a joy to be a mathematician.

In general, expressions like $\sqrt[n]{a\pm\sqrt{b}}$ with rational a and b are not constructible and their sum

$$x = \sqrt[n]{a+\sqrt{b}} + \sqrt[n]{a-\sqrt{b}}$$

is not a rational number. However for special rational values of a and b the individual radicals are constructible and their sum is rational. We will investigate how to determine these values.

Following are several interesting radicals with simple values:

$$\sqrt[3]{2 \pm \sqrt{5}} = \frac{1 \pm \sqrt{5}}{2}, \quad \sqrt[6]{9 \pm 4\sqrt{5}} = \frac{1 \pm \sqrt{5}}{2}, \quad \sqrt[9]{38 \pm 17\sqrt{5}} = \frac{1 \pm \sqrt{5}}{2},$$

$$\sqrt[3]{7 \pm 5\sqrt{2}} = 1 \pm \sqrt{2}, \quad \sqrt[5]{41 \pm 29\sqrt{2}} = 1 \pm \sqrt{2}, \quad \sqrt[6]{99 \pm 70\sqrt{2}} = 1 \pm \sqrt{2},$$

$$\sqrt[3]{-5 \pm i\sqrt{2}} = 1 \pm i\sqrt{2}, \quad \sqrt[5]{1 \pm 11i\sqrt{2}} = 1 \pm i\sqrt{2}, \quad \sqrt[9]{17 \pm 56i\sqrt{2}} = 1 \pm i\sqrt{2}.$$

In each case, the sum of pairs of conjugate radicals is an integer because the radicals cancel. For example $\sqrt[5]{41 + 29\sqrt{2}} + \sqrt[5]{41 - 29\sqrt{2}} = 2$. (Many radical expressions, some like those mentioned here, appear in the work of Ramanujan. Our method for simplifying them has been referred to in [**2**] and [**3**].)

Our main tools for examining a given radical expression for possible simplification will be what we will call the *Cardan polynomials* which we denote by $C_n(c, x)$. (We use the name "Cardan polynomials" because the expression for their roots closely resembles Cardan's solution of the cubic equation.) The Cardan polynomials are related to the familiar Chebyshev polynomials by

$$C_n(c, x) = 2c^{n/2} T_n\left(\frac{x}{2\sqrt{c}}\right).$$

Cardan polynomials seem to have first appeared in a note by J. E. Woko [**12**] in a different context. The use of Cardan polynomials in simplifying radicals may be new. For other approaches to the simplification of radicals see [**6**], [**8**] and [**9**]. (In [**2**], Chebyshev polynomials are used to reduce a radical expression.)

A note on radical values: A radical expression of the form $\sqrt[n]{a + \sqrt{b}}$ has, in general, n complex values. When such radicals appear in an equation, it may be that the equation is true for only one of the n possible values. When two such radicals appear in a single equation, then appropriate pairs of values must be selected. In specific cases, the meaning is usually clear in context. For instance, in (3), we take the real cube roots.

2. Cardan polynomials

Let us start with the expression

$$x = \sqrt[n]{a + \sqrt{b}} + \sqrt[n]{a - \sqrt{b}} \tag{5}$$

We will define c as

$$c = \sqrt[n]{a + \sqrt{b}}\,\sqrt[n]{a - \sqrt{b}} = \sqrt[n]{a^2 - b}, \tag{6}$$

so $c^n = a^2 - b$.

If we raise both sides of (5) to the power n, and make appropriate substitutions, we will generate a polynomial equation in x, of degree n, with integer coefficients. We write the equation in the form $C_n(c, x) = 2a$; this defines the *Cardan polynomial of* *degree n*. In the case $n = 3$, for instance, cubing (5) gives

$$
\begin{aligned}
x^3 &= \left(\sqrt[3]{a+\sqrt{b}} + \sqrt[3]{a-\sqrt{b}}\right)^3 \\
&= \left(\sqrt[3]{a+\sqrt{b}}\right)^3 + 3\left(\sqrt[3]{a+\sqrt{b}}\right)^2\left(\sqrt[3]{a-\sqrt{b}}\right) \\
&\quad + 3\left(\sqrt[3]{a+\sqrt{b}}\right)\left(\sqrt[3]{a-\sqrt{b}}\right)^2 + \left(\sqrt[3]{a-\sqrt{b}}\right)^3 \\
&= \left(\sqrt[3]{a+\sqrt{b}}\right)^3 + 3\sqrt[3]{a+\sqrt{b}}\sqrt[3]{a-\sqrt{b}}\left(\sqrt[3]{a+\sqrt{b}} + \sqrt[3]{a-\sqrt{b}}\right) \\
&\quad + \left(\sqrt[3]{a-\sqrt{b}}\right)^3 .
\end{aligned}
$$

Using (5) and (6) to simplify the middle summand we get

$$x^3 = a + \sqrt{b} + 3cx + a - \sqrt{b} = 3cx + 2a.$$

We now have $x^3 - 3cx = 2a$. If we write $C_3(c, x) = x^3 - 3cx$, then $C_3(c, x) = 2a$. The reader can continue in this way and derive algebraically the Cardan polynomials $C_n(c, x)$ of higher degree, and find that

$$C_n(c, x) = 2a. \tag{7}$$

Another approach to our Cardan polynomials is simpler, and reveals their connection to the Chebyshev polynomials. Writing

$$\sqrt[n]{a \pm \sqrt{b}} = r \exp(\pm i\theta) \tag{8}$$

(where r and θ need not be real) gives

$$a \pm \sqrt{b} = r^n \exp(\pm in\theta). \tag{9}$$

Using (8) to rewrite (5) we get

$$x = \sqrt[n]{a+\sqrt{b}} + \sqrt[n]{a-\sqrt{b}} = re^{i\theta} + re^{-i\theta} = 2r\cos\theta.$$

From (9) we have $r^n e^{in\theta} + r^n e^{-in\theta} = 2a$, so

$$\cos n\theta = \frac{a}{r^n}. \tag{10}$$

The Chebyshev polynomials $T_n(\cos\theta)$ are defined by $T_n(\cos\theta) = \cos n\theta$. Now (10) gives

$$T_n(x/2r) = \frac{a}{r^n} \tag{11}$$

Comparing (11) with (7) we have

$$C_n(c, x) = 2r^n T_n(x/2r). \tag{12}$$

To determine r, use (6) and (8) to get

$$c = \sqrt[n]{a+\sqrt{b}}\sqrt[n]{a-\sqrt{b}} = re^{i\theta}re^{-i\theta} = r^2.$$

Now (12) becomes

$$C_n(c,x) = 2c^{n/2}T_n\left(\frac{x}{2\sqrt{c}}\right). \tag{13}$$

We see that the exact nature of the n^{th} Cardan polynomial can be determined from the n^{th} Chebyshev polynomial.

3. Constructing a table of Cardan polynomials

Starting with a table of the familiar Chebyshev polynomials $T_n(x)$ and using (13) we can tabulate the $C_n(c,x)$:

TABLE 1: Cardan polynomials

$C_1(c,x)$	$=$	x
$C_2(c,x)$	$=$	x^2-2c
$C_3(c,x)$	$=$	x^3-3cx
$C_4(c,x)$	$=$	$x^4-4cx^2+2c^2$
$C_5(c,x)$	$=$	$x^5-5cx^3+5c^2x$
$C_6(c,x)$	$=$	$x^6-6cx^4+9c^2x^2-2c^3$
$C_7(c,x)$	$=$	$x^7-7cx^5+14c^2x^3-7c^3x$
$C_8(c,x)$	$=$	$x^8-8cx^6+20c^2x^4-16c^3x^2+2c^4$
$C_9(c,x)$	$=$	$x^9-9cx^7+27c^2x^5-30c^3x^3+9c^4x$
$C_{10}(c,x)$	$=$	$x^{10}-10cx^8+35c^2x^6-50c^3x^4+25c^4x^2-2c^5$
$C_{11}(c,x)$	$=$	$x^{11}-11cx^9+44c^2x^7-77c^3x^5+55c^4x^3-11c^5x$

If we denote the coefficients by $d(n,k)$, where

$$C_n(c,x) = x^n - d(n,1)cx^{n-2} + d(n,2)c^2x^{n-4} - d(n,3)c^3x^{n-6} + \cdots,$$

the table above suggests that

$$d(n,k) = d(n-1,k) + d(n-2,k-1). \tag{14}$$

This means that any coefficient in the above table is the sum of the one directly above it and the one two rows above and to the left. Using this simple recursion relation, we can extend the table indefinitely. This is simpler than transforming the Chebyshev polynomials with (13).

Relation (14) follows from $C_n(c,x) = xC_{n-1}(c,x) - cC_{n-2}(c,x)$, which in turn follows from the known relation [**10**] $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$.

4. The quadratic connection

Another important relation is

$$\sqrt[n]{a \pm \sqrt{b}} = \frac{x \pm \sqrt{x^2 - 4c}}{2}, \tag{15}$$

where, as before, $x = \sqrt[n]{a + \sqrt{b}} + \sqrt[n]{a - \sqrt{b}}$ and $c = \sqrt[n]{a + \sqrt{b}}\sqrt[n]{a - \sqrt{b}} = \sqrt[n]{a^2 - b}$. To derive (15), we start with

$$y = \sqrt[n]{a + \sqrt{b}}. \tag{16}$$

Multiply (16) by $\sqrt[n]{a - \sqrt{b}}$ and use (6) to get

$$y\sqrt[n]{a - \sqrt{b}} = c. \tag{17}$$

It follows from (16) and (17) that $x = \sqrt[n]{a + \sqrt{b}} + \sqrt[n]{a - \sqrt{b}} = y + c/y$, which implies the quadratic equation $y^2 - xy + c = 0$, with the solutions

$$y = \frac{x \pm \sqrt{x^2 - 4c}}{2} \tag{18}$$

Comparing (16) and (18) we see that (15) is true.

5. A technique for resolving radicals

To decide whether a radical $\sqrt[n]{a \pm \sqrt{b}}$ can be reduced to some nice form, we first solve

$$b = a^2 - c^n \tag{19}$$

for c. If c is an integer or rational number, there is hope that the radical can be simplified, however, even when this is not the case do not give up. Next, examine the equation

$$C_n(c, x) = 2a \tag{20}$$

using the table of Cardan polynomials, and search for an integer or rational root x, where

$$\sqrt[n]{a + \sqrt{b}} + \sqrt[n]{a - \sqrt{b}} = x. \tag{21}$$

If such a root is found, then our radical can be written as

$$\sqrt[n]{a \pm \sqrt{b}} = \frac{x \pm \sqrt{x^2 - 4c}}{2}. \tag{22}$$

If the above procedure fails to simplify the radical then try the same steps with n replaced by any divisor d of n, where $n = d\,e$. In this case, we examine $\sqrt[e]{\sqrt[d]{a \pm \sqrt{b}}}$

for simplification. Start with the largest divisor $d < n$ and continue trying successively smaller divisors until all divisors are exhausted or a nice simplified expression emerges. We illustrate with several examples.

Example 1: A calculator shows that $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}} = 1.000000000$ out to nine zeroes. Show that the expression is exactly 1.

Solution: Comparing our expression with (21) we see that we need $n = 3$, $a = 2$, and $b = 5$. Using (19) we get $c = -1$. From the table given in Section 3 and (20) we get $C_3(c, x) = x^3 - 3cx = 2a$, which reduces to $x^3 + 3x = 4$, of which $x = 1$ is an exact root. Notice also that (22) yields $\sqrt[3]{2 \pm \sqrt{5}} = \frac{1\pm\sqrt{5}}{2}$, from which it follows that $\sqrt[3]{2+\sqrt{5}}+\sqrt[3]{2-\sqrt{5}} = 1$. (This problem is also considered in [**11**].)

Example 2: Show that $\sqrt[7]{\frac{29}{2}+\frac{1}{2}\sqrt{845}}+\sqrt[7]{\frac{29}{2}-\frac{1}{2}\sqrt{845}} = 1$.

Solution: In this case $n = 7$, $a = 29/2$, $b = 845/4$, and $c = -1$. Relation (20) becomes now $C_7(-1, x) = 29$. From the table in section 3 we get $x^7 + 7x^5 + 14x^3 + 7x = 29$. Since $x = 1$ is a root, we are finished. We see, also, that the quadratic relation (22) implies

$$\sqrt[7]{\frac{29}{2} \pm \frac{1}{2}\sqrt{845}} = \frac{1 \pm \sqrt{5}}{2}.$$

Example 3: (See also [**5**] and [**7**].) For what positive numbers b is $\sqrt[3]{2+\sqrt{b}} + \sqrt[3]{2-\sqrt{b}}$ an integer?

Solution: The desired values of b are 5 and 100/27. To see why, let $a = 2$ and $n = 3$ in (19) to get $c^3 = 4 - b$. Also (20) becomes $x^3 - 3cx - 4 = 0$, and solving for c we get

$$c = \frac{x^3 - 4}{3x}. \tag{23}$$

We can now look at the problem another way: Let x be an integer, and use (23) to find c. Now $b = 4 - c^3$; that this last expression must be positive limits the number of possible solutions. If we try $x = 1$, (23) gives us $c = -1$, so $b = 4 - c^3 = 5$. If we try $x = 2$, then $c = 2/3$ and $b = 100/27$. All other values of x yield values of c that give $b < 0$.

Example 4: The dedication for the paper [**1**] reads: "In memory of Ramanujan on the

$$\left(32\left(\frac{146410001}{48400}\right)^3 - 6\left(\frac{146410001}{48400}\right)\right.$$
$$\left.+\sqrt{\left(32\left(\frac{146410001}{48400}\right)^3 - 6\left(\frac{146410001}{48400}\right)\right)^2 - 1}\right)^{1/6} \text{th}$$

anniversary of his birth." What is this number?

Solution: We see at once that with $a = 32(\frac{146410001}{48400})^3 - 6(\frac{146410001}{48400})$, and $b = a^2 - 1$, the complicated number above can be written as $y = (a + \sqrt{b})^{1/6}$. Since $b = a^2 - c^6$, we see that $c = 1$ or -1. Also, $146410001 = 110^4 + 1$ and $48400 = 4(110)^2$, so $2a = (\frac{110^4+1}{110^2})^3 - 3(\frac{110^4+1}{110^2})$. We recognize this last expression as $2a = C_3(1, x)$, with

$x = \frac{110^4+1}{110^2}$. Hence, using (22) we get $\sqrt[3]{a+\sqrt{b}} = \frac{x+\sqrt{x^2-4c}}{2} = 110^2$, and our original number is $y = 110$.

Notice that the original problem involved a sixth root and therefore we would expect to look at the sixth Cardan polynomial. The sixth Cardan polynomial would have failed to help. When this occurs, we should use the fact that $\sqrt[mn]{z} = \sqrt[m]{\sqrt[n]{z}}$ to see if Cardan polynomials of lower order can reduce the radical.

Example 5: (This example is motivated by [**12**], where the author approaches the Cardan polynomials from an entirely different direction.) Let α and β be the roots of the quadratic equation $y^2 - (\alpha+\beta)y + \alpha\beta = 0$. Show that $\alpha^n + \beta^n = C_n(\alpha\beta, \alpha+\beta)$. (This is an expansion of the symmetric function of the roots $\alpha^n + \beta^n$, in powers of the coefficients $\alpha\beta$ and $\alpha+\beta$ using Cardan polynomials.)

Solution: Consider the quadratic connection (18) which is derived from $y^2 - xy + c = 0$. If α and β are the two roots of this quadratic in y, then $c = \alpha\beta$ and $x = \alpha + \beta$. From (15) we see that $\alpha = \sqrt[n]{a+\sqrt{b}}$ and $\beta = \sqrt[n]{a-\sqrt{b}}$, so $\alpha^n + \beta^n = 2a$. The result now follows immediately from the fact that $C_n(c, x) = 2a$.

Acknowledgment. The author wishes to thank Richard Askey, Bruce Berndt, Heng Huat Chan, Theodore Rivlin, and Marcus Wright for their help and encouragement. The author also thanks the referee for several important references and other improvements.

REFERENCES

1. B. C. Berndt, H. H. Chan, L.-C. Zhang, Ramanujan's association with radicals in India, *Amer. Math. Monthly* **104** (1997), 905–911.
2. B. C. Berndt, H. H. Chan, and L.-C. Zhang, Radicals and units in Ramanujan's work, *Acta Arith.* **87** (1998), 145–158.
3. B. C. Berndt, Y.-S. Choi, and S.-Y. Kang, Problems submitted by Ramanujan to the Journal of the Indian Mathematical Society. (In *Continued Fractions: From Analytic Number Theory to Constructive Approximation*, B. C. Berndt and F. Gesztesy, eds., Amer. Math. Soc., Providence, RI, pp. 15–56.)
4. G. Cardano, *Ars Magna or the Rules of Algebra*, Dover, New York, NY, 1993.
5. G. H. Hardy, *A Course in Pure Mathematics*, Cambridge Univ. Press, New York, NY, 1952.
6. M. Honsbeek, Denesting certain nested radicals of depth two, to appear in *Journal de Théorie de Nombres de Bordeaux*.
7. D. C. Kay, Problem 1529, this MAGAZINE **70** (1997), 298.
8. S. Landau, How to tangle with a nested radical, *Math. Intelligencer*, **16** (1994), 49–55.
9. S. Landau, Simplification of nested radicals, *SIAM Journal of Computing* **21** (1992), 85–110.
10. T. J. Rivlin, *Chebyshev Polynomials*, John Wiley & Sons, New York, NY, 1990.
11. A. Sofo, Generalization of a radical identity, *Math. Gazette* **83** (1999), 274–276.
12. J. E. Woko, A Pascal-like triangle for $\alpha^n + \beta^n$, *Math. Gazette* **81** (1997), 75–79.

The Editor owes a debt of gratitude to computer consultants Eric V. Level of 47 Software and Scott A. Dick of Sun Vista Technologies for their expert help in designing a database to track articles, authors, and referees for the MAGAZINE.

NOTES

Avoiding Your Spouse at a Bridge Party

BARBARA H. MARGOLIUS
Cleveland State University
Cleveland, OH 44115
margolius@math.csuohio.edu

Introduction In his note in the February, 2000 issue of this MAGAZINE, "Dinner, Dancing and Tennis, Anyone?" James Brawner [**1**] poses the question (rephrased here in terms of bridge instead of dancing):

> *The Bridge Couples Problem*: Suppose n married couples ($2n$ people) are invited to a bridge party. Bridge partners are chosen at random, without regard to gender. What is the probability that no one will be paired with his or her spouse?

I was intrigued by the convergence behavior of this probability. Brawner conjectures, correctly, that this probability converges to $e^{-\frac{1}{2}}$, and he also notes that the convergence is very slow. In fact, to achieve ten decimal places of accuracy in our estimate for $e^{-\frac{1}{2}}$ based on our bridge problem probabilities, we would have to invite more than 2 billion couples to our party.

The problem, in a first reading, may sound like Montmort's matching problem:

> *The Hat Matching Problem*: n men, attending a banquet, check their hats. When each man leaves he takes a hat at random. What is the probability that no man gets his own hat?

(This problem was first proposed by the mathematician Pierre Remond de Montmort [**6**], [**7**]. A more complete discussion of it appears in this MAGAZINE in an article by Gabriela Sanchis [**8**] and in the excellent history of probability by Anders Hald [**5**]. It is also discussed in Brawner's article [**1**].) The well-known solution to this problem is $h_n = \sum_{j=0}^{n} \frac{(-1)^j}{j!}$ which converges rapidly to e^{-1}. The difference between this classical problem and the bridge couples problem is that in the bridge couples problem, it makes sense for one man to be paired with another, but in the hat matching problem, no man is going to select another man in place of his hat.

In this note, I derive a nonrecursive formula for the bridge couples probabilities, and provide a simple approximation formula for computing these probabilities as well as a procedure for getting the exact probabilities to an arbitrary number of decimal places. We use the inclusion-exclusion principle from probability and properties of alternating series to find the nonrecursive formula and establish its limit as n, the number of couples, tends to infinity.

Let b_n be the probability that of n couples, no one is assigned his or her spouse as a bridge partner.

A second order recursive formula for the bridge couples problem We compute a recursive formula for the bridge couples probabilities and use it to show that for $n > 2$, the b_n are strictly increasing.

From the recursive formulas that Brawner derived or by reasoning directly, as below, one can show that the b_n satisfy the recursive formula:

$$b_n = \frac{2n-2}{2n-1}\left[\frac{1}{2n-3}b_{n-2} + b_{n-1}\right]. \tag{1}$$

We know that the first person assigned a bridge partner was assigned someone other than his or her spouse with probability $\frac{2n-2}{2n-1}$. Say Priscilla Smith was assigned to George Washington. If no couples are assigned their spouses as bridge partners, then either Martha Washington is assigned to John Smith or she is not. Martha Washington is assigned John Smith with probability $1/(2n-3)$. The probability that none of the other $n-2$ couples are assigned one another as bridge partners is b_{n-2}, so the joint probability that Martha is assigned to John and the other couples are not assigned their spouses is $\frac{1}{2n-3}b_{n-2}$. The other possibility is that Martha is not assigned to John and none of the other couples are assigned one another. Since we can think of Martha Washington and John Smith as a couple who are not assigned to each other, this probability is b_{n-1}. Hence, the recursive formula (1).

It is easy to see that $b_0 = 1$ and $b_1 = 0$. Consider the case where $n = 2$. Three pairings are possible: George Washington could be paired with Martha Washington, John Smith or Priscilla Smith, and the other two would have each other as partners. One of these three pairings results in spouses being with their mates and two do not, so $b_2 = 2/3$.

Claim. *For $n > 2$, the b_n are strictly increasing in n.*

Proof. Intuitively, it is clear that this should be so: as the number of couples increases, there are more possible alternative partners for each bridge player other than the spouse.

We proceed by showing that $b_n - b_{n-1}$ is strictly positive. We may rewrite (1) as

$$b_n - \frac{2n-2}{2n-1}b_{n-1} = \frac{2n-2}{2n-1}\left[\frac{1}{2n-3}b_{n-2}\right]. \tag{2}$$

We will use this relation to make substitutions twice in what follows.

Subtracting b_{n-1} from both sides of (1) gives

$$\begin{aligned} b_n - b_{n-1} &= \frac{2n-2}{2n-1}\left[\frac{1}{2n-3}b_{n-2} + b_{n-1}\right] - b_{n-1} \\ &= \frac{1}{2n-1}\left[-\left(b_{n-1} - \frac{2n-4}{2n-3}b_{n-2}\right) + \frac{2}{2n-3}b_{n-2}\right] \end{aligned}$$

We may substitute for the quantity in parentheses using (2) to obtain

$$\begin{aligned} b_n - b_{n-1} &= \frac{1}{2n-1}\left[-\left(\frac{2n-4}{2n-3}\left[\frac{1}{2n-5}b_{n-3}\right]\right) + \frac{2}{2n-3}b_{n-2}\right] \\ &= \frac{2n-4}{(2n-1)(2n-3)(2n-6)}\left[\left(b_{n-2} - \frac{2n-6}{2n-5}b_{n-3}\right) + \frac{n-4}{n-2}b_{n-2}\right]. \end{aligned}$$

Substituting a second time for the quantity in parentheses using (2) we have

$$b_n - b_{n-1} = \frac{1}{(2n-1)(2n-3)}\left[\frac{2n-4}{(2n-5)(2n-7)}b_{n-4} + \frac{n-4}{n-3}b_{n-2}\right]. \tag{3}$$

We can show by evaluating the recursive formula that b_2 through b_4 are strictly positive, so the right-hand side of (3) is positive for all $n > 4$. By direct computation, $b_2 = 2/3$, $b_3 = 8/15$, $b_4 = 4/7$. So the b_n for $n > 2$ are strictly increasing in n. This completes the proof. ■

The $b_n \leq 1$ (because they are probabilities), so since they are strictly increasing, $\lim_{n\to\infty} b_n = b_\infty$ exists. To obtain this limit, we require a nonrecursive formula for the b_n.

A nonrecursive formula for the bridge couples probabilities Begin by counting the number of ways that $2n$ people can be paired into bridge partners. Suppose that we assign people to partners by having each person report to a seat along a long dining table where we have placed n seats on each side of the table and labelled the seats "1L" and "1R" through "nL" and "nR" according to whether the seat is on the left or the right of the table. Guests on the left are paired with their opposite number on the right side of the table. There are $(2n)!$ such assignments. Of course, we are concerned with the number of possible distinct pairings of $2n$ people into n pairs, so we are not concerned with who is on the right or who on the left, nor are we concerned with the order in which the couples are seated. To correct for double counting from these two factors we divide by $2^n n!$. The number of ways to assign $2n$ people bridge partners is

$$\frac{(2n)!}{2^n n!}.$$

Now suppose that we are interested in the probability that Priscilla Smith has been paired with her husband John. Let r_1 be the probability that a particular couple has been assigned as bridge partners. There are

$$\frac{(2(n-1))!}{2^{n-1}(n-1)!}$$

ways that this could have happened since the other $(n-1)$ couples may be assigned any available bridge partners, so the probability is

$$r_1 = \left(\frac{(2(n-1))!}{2^{n-1}(n-1)!}\right) \bigg/ \left(\frac{(2n)!}{2^n n!}\right).$$

Similarly, the probability that any particular k couples have been assigned their spouses as bridge partners is

$$r_k = \left(\frac{(2(n-k))!}{2^{n-k}(n-k)!}\right) \bigg/ \left(\frac{(2n)!}{2^n n!}\right).$$

Since there are n married couples, it may seem that the probability of no couples being assigned one another as bridge partners is simply $1 - nr_n$. This is not the case because the events are not disjoint. There are some assignments of bridge partners in which both the Smiths and the Washingtons are paired with their spouses. To correct for this double counting, we add back in the probability that two couples have been paired with their spouses. For a particular two couples, this probability is r_2, but there are $\binom{n}{2}$ ways in which two couples may be chosen from a group of n couples so we multiply this probability by $\binom{n}{2}$. Now we have $1 - nr_1 + \binom{n}{2}r_2$. Again, we have the problem that the event that the Smiths and Washingtons have been assigned their spouses as bridge partners does not preclude the possibility that the Clintons may also have been assigned as bridge partners. If we continue to make the necessary corrections, we have that the probability that no one has his or her spouse for a bridge partner is

$$\begin{aligned} b_n &= 1 - \binom{n}{1} r_1 + \binom{n}{2} r_2 - \cdots + (-1)^n \binom{n}{n} r_n \\ &= \sum_{j=0}^{n} (-1)^j \binom{n}{j} r_j. \end{aligned} \tag{4}$$

This is an example of the inclusion-exclusion principle.

Note that more generally, we have

$$P\{\text{exactly } k \text{ couples are paired}\} = \sum_{j=k}^{n} (-1)^{j-k} \binom{n}{j} r_j$$

is the probability that exactly k couples are paired with their spouses.

Substituting for r_j in (4), we have

$$b_n = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left(\frac{(2(n-j))!}{2^{n-j}(n-j)!} \right) \Big/ \left(\frac{(2n)!}{2^n n!} \right).$$

By rearranging terms and cancelling common factors, we can obtain

$$b_n = \sum_{j=0}^{n} \frac{\left(-\frac{1}{2}\right)^j}{j!} \prod_{m=0}^{j-1} \left(1 + \frac{1}{(2(n-m)-1)} \right). \tag{5}$$

(See the appendix for a detailed derivation of (5)).

From this formulation of b_n, it is clear that if $a_{n,j}$ represents the jth summand,

$$\lim_{n\to\infty} a_{n,j} = \lim_{n\to\infty} \frac{\left(-\frac{1}{2}\right)^j}{j!} \prod_{m=0}^{j-1} \left(1 + \frac{1}{(2(n-m)-1)} \right) = \frac{\left(-\frac{1}{2}\right)^j}{j!}.$$

This leads us to the suspect that the following theorem may be true:

Theorem. *The limit of the bridge couple probabilities is*

$$\lim_{n\to\infty} b_n = e^{-\frac{1}{2}}.$$

Proof. We will show that for every $\epsilon > 0$ there is an n such that $|b_n - e^{-\frac{1}{2}}| < \epsilon$. Choose k so that $1/k! < \epsilon/3$. For $n > k$, we may rewrite b_n as two finite sums:

$$\begin{aligned} b_n = &\sum_{j=0}^{k} \frac{\left(-\frac{1}{2}\right)^j}{j!} \prod_{m=0}^{j-1} \left(1 + \frac{1}{(2(n-m)-1)} \right) \\ &+ \sum_{j=k+1}^{n} \frac{\left(-\frac{1}{2}\right)^j}{j!} \prod_{m=0}^{j-1} \left(1 + \frac{1}{(2(n-m)-1)} \right). \end{aligned}$$

Note that

$$\prod_{m=0}^{j-1} \left(1 + \frac{1}{(2(n-m)-1)} \right) < 2^j,$$

so that

$$\left|\sum_{j=k+1}^{n}\frac{\left(-\frac{1}{2}\right)^j}{j!}\prod_{m=0}^{j-1}\left(1+\frac{1}{(2(n-m)-1)}\right)\right| < \left|\sum_{j=k+1}^{n}\frac{(-1)^j}{j!}\right|.$$

Consider the difference,

$$\begin{aligned}
&|b_n - e^{-\frac{1}{2}}| \\
&= \left|\sum_{j=0}^{k}\frac{\left(-\frac{1}{2}\right)^j}{j!}\prod_{m=0}^{j-1}\left(1+\frac{1}{(2(n-m)-1)}\right)\right. \\
&\quad \left. + \sum_{j=k+1}^{n}\frac{\left(-\frac{1}{2}\right)^j}{j!}\prod_{m=0}^{j-1}\left(1+\frac{1}{(2(n-m)-1)}\right) - \sum_{j=0}^{\infty}\frac{\left(-\frac{1}{2}\right)^j}{j!}\right| \\
&\le \left|\sum_{j=0}^{k}\frac{\left(-\frac{1}{2}\right)^j}{j!}\left[\prod_{m=0}^{j-1}\left(1+\frac{1}{(2(n-m)-1)}\right)-1\right]\right| \\
&\quad + \left|\sum_{j=k+1}^{n}\frac{\left(-\frac{1}{2}\right)^j}{j!}\left[\prod_{m=0}^{j-1}\left(1+\frac{1}{(2(n-m)-1)}\right)-1\right]\right| + \left|\sum_{j=n+1}^{\infty}\frac{\left(-\frac{1}{2}\right)^j}{j!}\right|. \qquad (6)
\end{aligned}$$

We bound the first term in absolute values on the right-hand side of (6):

$$\begin{aligned}
&\left|\sum_{j=0}^{k}\frac{\left(-\frac{1}{2}\right)^j}{j!}\left[\prod_{m=0}^{j-1}\left(1+\frac{1}{(2(n-m)-1)}\right)-1\right]\right| \\
&< \left|\sum_{j=0}^{k}\frac{\left(-\frac{1}{2}\right)^j}{j!}\right|\left[\left(1+\frac{1}{(2(n-k)-1)}\right)^k-1\right].
\end{aligned}$$

Since k is fixed, we may choose n large enough so that

$$\left[\left(1+\frac{1}{(2(n-k)-1)}\right)^k-1\right] < \epsilon/3,$$

so the first term is less than $\epsilon/3$. (Why?) The terms in the last line of (6) are an alternating sum and an alternating series with terms of decreasing magnitude; hence they are each bounded by the absolute value of their first terms, $1/(k+1)!$ and $1/(2(n+1)!)$, respectively. These quantities in turn are each bounded by $\epsilon/3$. The difference $|b_n - e^{-\frac{1}{2}}|$ is therefore bounded by ϵ, and since ϵ was arbitrary we have the required result.

From the formula for $a_{n,j}$, we can make another observation: if n is large relative to j, then $a_{n,j} \approx \frac{(-\frac{1}{2})^j}{j!}(1+\frac{1}{(2n)})^j$. If n and j are of approximately the same magnitude and n is large enough, then $\frac{(-\frac{1}{2})^j}{j!}$ is small and the term $a_{n,j}$ contributes little to the sum b_n. What this means is that a good approximation to b_n is $e^{-\frac{1}{2}-\frac{1}{4n}}$. (In fact, for $n > 4$, this approximation serves as an upper bound for the b_n and the difference is less than $1/(24n^2)$. See the second exercise at the end of this note.) This is illustrated in FIGURE 1, which shows the sixth through twentieth values of b_n as small circles. The curved line is $e^{-\frac{1}{2}-\frac{1}{4n}}$, and the horizontal line is the constant $e^{-\frac{1}{2}}$. The circles are below the curve $e^{-\frac{1}{2}-\frac{1}{4n}}$, but this is difficult to see as the approximation improves as n increases.

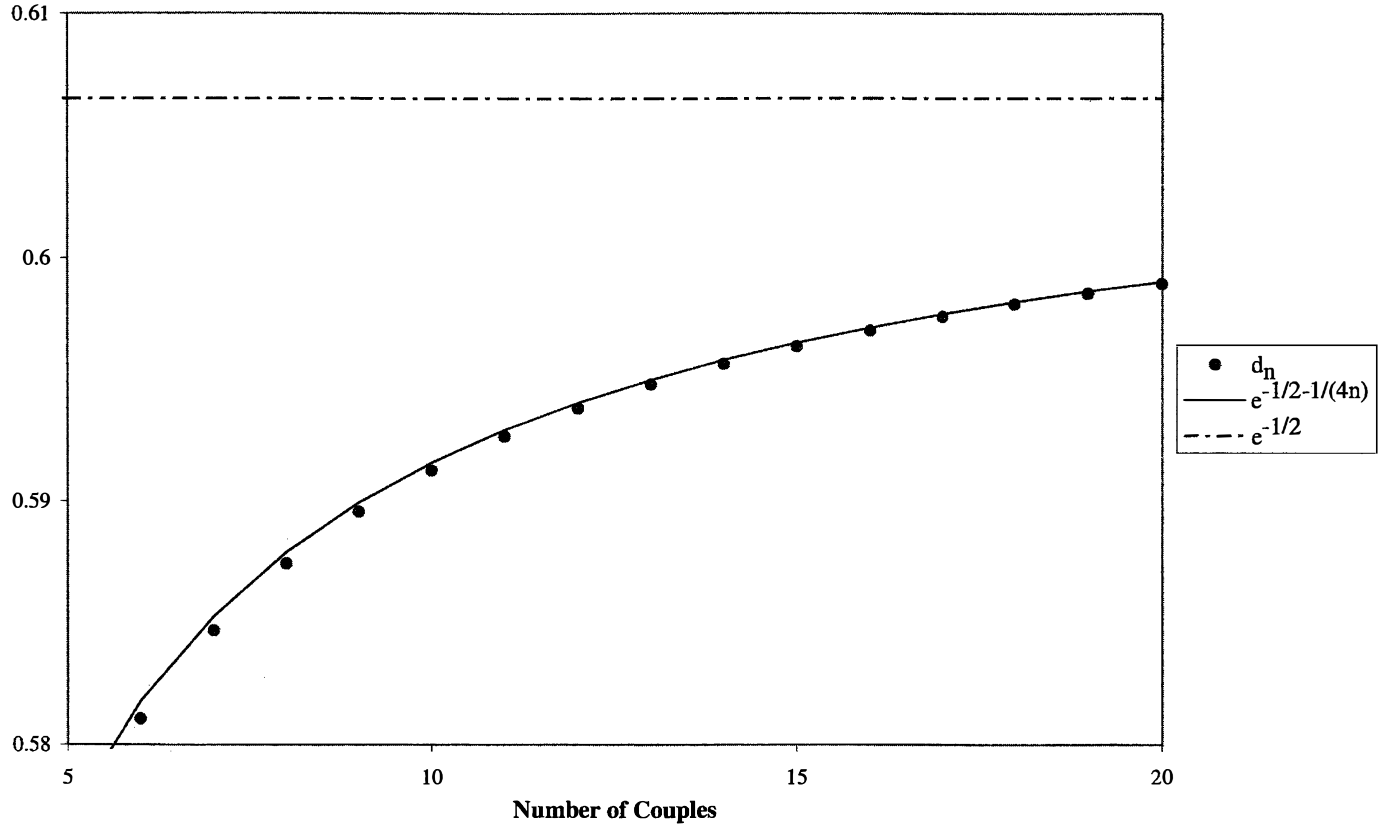


Figure 1 Exact bridge couple probabilities and exponential estimates.

Notice that b_n is an alternating sum with decreasing terms. (This is a consequence of our having formulated the expression for b_n using the inclusion-exclusion principle.) Indeed, if $a_{n,j}$ is the jth summand, then

$$a_{n,j+1} = -\frac{1}{(j+1)}\left(\frac{n-j}{2(n-j)-1}\right)a_{n,j}$$

and $|\frac{1}{(j+1)}(\frac{n-j}{2(n-j)-1})| < 1$. Let $S_{n,k}$ be the kth partial sum of b_n ($k \leq n$) so

$$S_{n,k} = \sum_{j=0}^{k} \frac{(-\frac{1}{2})^j}{j!} \prod_{m=0}^{j-1}\left(1 + \frac{1}{(2(n-m)-1)}\right), \tag{7}$$

then

$$S_{n,k-1} < b_n < S_{n,k} \tag{8}$$

for k even. The graph below illustrates this. The horizontal line is $e^{-\frac{1}{2}}$. The top curved lined is $S_{n,4}$. The next line is actually two curves that are hard to distinguish on the graph: $S_{n,5}$, and $S_{n,6}$. The bottom curve is $S_{n,3}$. $S_{n,k}$ for $k > 6$ if shown on the graph would appear on the graph between $S_{n,5}$ and $S_{n,6}$. At this scale, they would be indistinguishable from the graphs of b_n and $e^{-\frac{1}{2}-\frac{1}{4n}}$ which would also be between $S_{n,5}$, and $S_{n,6}$. Although convergence of the b_n to $e^{-\frac{1}{2}}$ is slow, convergence of the partial sums to b_n is fast.

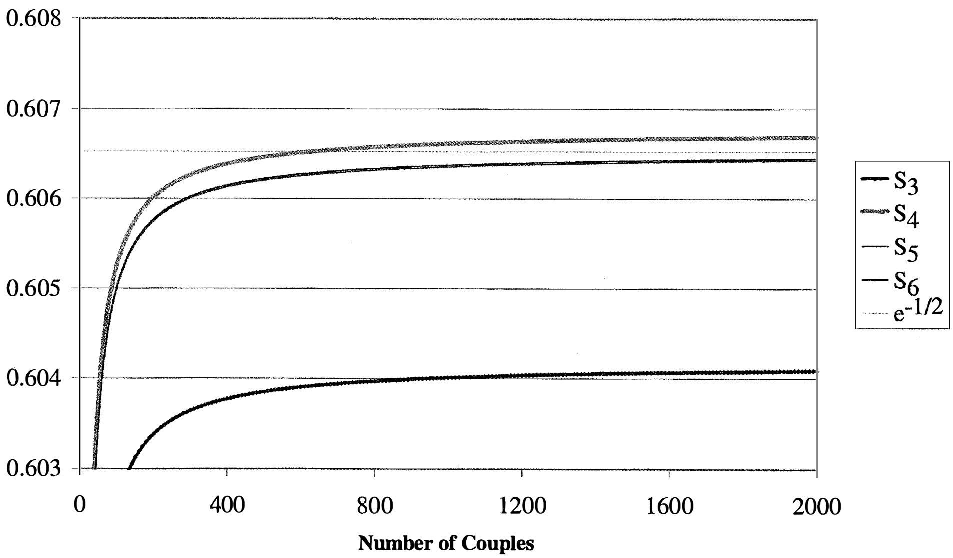


Figure 2 Partial sums of bridge couple probabilities

In Table 1 below, the first column is n, the number of couples invited to the bridge party; the second column shows values of b_n for each n. The third column gives the number k of the first $S_{n,k}$ that bounds b_n above, but is less than $e^{-\frac{1}{2}}$, and the fourth column gives the value of that $S_{n,k}$. With a partial sum of just a few terms you can see that b_n is less than $e^{-\frac{1}{2}}$. The fifth column is $e^{-\frac{1}{2}-\frac{1}{4n}}$ which closely approximates b_n.

TABLE 1: The Bridge Couples Problem Probabilities, b_n

n	b_n	k	An upper bound on b_n $S_{n,k}$	An approximation of b_n $e^{-\frac{1}{2}-\frac{1}{4n}}$
10	0.5912368196	4	0.5915694213	0.5915553644
100	0.6050124905	4	0.6052587849	0.6050162269
1,000	0.6063790081	6	0.6063804716	0.6063790460
10,000	0.6065154963	6	0.6065169551	0.6065154966
100,000	0.6065291434	6	0.6065306018	0.6065291434
1,000,000	0.6065305081	8	0.6065305132	0.6065305081
2,420,954,467	0.6065306597	10	0.6065306597	0.6065306597

The probabilities b_n given in the second column are computed to ten decimal places of accuracy by taking as b_n the first $S_{n,k}$ that agrees with $S_{n,k+1}$ to 10 decimal places. Since the partial sums alternate as upper and lower bounds for b_n, it is not necessary to sum all $n+1$ terms $a_0, a_1, \ldots, a_n$ to obtain this accuracy. The probabilities for $b_{1,000}$, $b_{10,000}$ and $b_{100,000}$ differ in the least significant digits from those computed by Brawner. For example, he obtained $b_n = 0.6065154929$ for $n = 10{,}000$. This is probably due to an accumulation of rounding errors from computing the b_n using a recursion formula requiring computation of 10,000 (or more) probabilities rather than computing an eleven term sum ($S_{n,10}$) as was required to compute most of these probabilities to 10 decimal places of accuracy.

For comparison to the numbers in the table, the limit, $b_\infty = e^{-\frac{1}{2}} = 0.6065306597$ to ten decimal places of accuracy. ■

We leave the following as exercises for the interested reader:

1. Suppose that we have a dinner party for n families and each family has m members. Guests are assigned to tables randomly. Each of the n tables seats m guests. Find a nonrecursive formula for the probability that every table has members from more than one family seated at it. Show that for $m > 2$, this probability converges rapidly to 1. (For $m = 2$, this is the bridge couples problem.)
2. Obtain a bound for the difference $|e^{-\frac{1}{2}-\frac{1}{4n}} - b_n|$. *Hint:* Write the terms in the difference as a finite sum from 0 to n and the tail of the exponential series for $e^{-\frac{1}{2}-\frac{1}{4n}}$ from $n+1$ to infinity. Show that for n large enough that the finite sum is an alternating sum with terms of decreasing absolute magnitude. Use the partial sums of the alternating sum and a bound on the tail of the exponential series to obtain the required bound.

Acknowledgment. Thanks are due to two anonymous referees for their helpful suggestions.

Appendix Simplification of b_n to obtain (5):

$$b_n = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left(\frac{(2(n-j))!}{2^{n-j}(n-j)!} \right) \Big/ \left(\frac{(2n)!}{2^n n!} \right)$$

$$= \sum_{j=0}^{n} (-2)^j \binom{n}{j} \left(\frac{(2(n-j))!}{(n-j)!} \right) \Big/ \left(\frac{(2n)!}{n!} \right)$$

$$= \sum_{j=0}^{n} \frac{(-2)^j}{j!} \left(\frac{n!}{(n-j)!} \right)^2 \left(\frac{(2(n-j))!}{(2n)!} \right)$$

$$= \sum_{j=0}^{n} \frac{(-1)^j}{j!} \frac{n(n-1)\cdots(n-j+1)}{(2(n-1)+1)\cdots(2(n-j)+1)}$$

$$= \sum_{j=0}^{n} \frac{(-1)^j}{j!} \prod_{m=0}^{j-1} \frac{n-m}{2(n-m)-1}$$

$$= \sum_{j=0}^{n} \frac{\left(-\frac{1}{2}\right)^j}{j!} \prod_{m=0}^{j-1} \left(1 + \frac{1}{(2(n-m)-1)}\right).$$

REFERENCES

1. J. N. Brawner, Dinner, Dancing and Tennis, Anyone?, this MAGAZINE **73** (2000), 29–35.
2. R. A. Brualdi, *Introductory Combinatorics*, North-Holland, New York, NY, 1992.
3. R. J. Clarke, Derangements and Bell numbers, this MAGAZINE **66** (1993), 299–303.
4. W. Feller, *An introduction to probability theory and its applications*, second edition, John Wiley and Sons, New York, NY, 1971.
5. A. Hald, *A History of Probability and Statistics and Their Applications Before 1750*, John Wiley and Sons, New York, NY, 1990.
6. P. R. de Montmort, *Essay d'Analyse sur les Jeux de Hazard*, Quillau, Paris, published anonymously, 1708.
7. P. R. de Montmort, *Essay d'Analyse sur les Jeux de Hazard*, Seconde Edition, Revûe et augmentée de plusiers Lettre, Quillau, Paris, 1713. Reprinted 1714. Published anonymously. Reprinted by Chelsea, New York, NY, 1980.
8. G. R. Sanchis, Swapping hats: a generalization of Montmort's problem, this MAGAZINE **71** (1998), 53–57.

The Cwatset of a Graph

NANCY-ELIZABETH BUSH
PAUL A. ISIHARA
Wheaton College
Wheaton, IL 60187

Cwatsets were introduced just a few years ago in this journal with applications in statistics and coding theory [**8**]. Since then, cwatset theory has become a hot topic for original undergraduate research [**2**],[**5**],[**6**],[**7**]. In fact, D. K. Biss was awarded the 1998 Morgan Prize for excellence in undergraduate research partly for his contribution to this growing field [**1**]. Here we will introduce the idea that every graph Γ has an associated cwatset, denoted cwat(Γ), and describe a few basic facts about the cwatset of a graph. We suggest how further investigation of cwat(Γ) could benefit both cwatset theory and graph theory by i) exploiting a new graph-based method for generating cwatsets and ii) exploring new isomorphic invariants that might be useful in the design of a cwatset-based test for graph isomorphism.

To begin, let us review the basic definitions needed from each theory:

- A *cwatset* C is a special subset of the abelian group $\mathbb{Z}_2^n$ that is "**c**losed **w**ith **a** **t**wist", meaning that, given any $c \in C$, there is a permutation σ of the set $\{1, 2, \ldots, n\}$ such that the coset $C + c$ is equal to C^σ. (Following [**8**], we use the term *coset* even though cwatsets need not be subgroups of $\mathbb{Z}_2^n$.) The permutation σ appearing as a superscript of C indicates how the n entries in each element of C are twisted or re-arranged. For instance, take the cwatset $C = \{000, 011, 101\}$ in $\mathbb{Z}_2^3$ and any one of its elements, let's say $c = 011$. The coset $C + 011$ is equal to $C^{(2,3)}$ since on one hand

$$C + 011 = \{000, 011, 101\} + 011 = \{011, 000, 110\},$$

and, on the other hand, $C^{(2,3)}$ is computed by interchanging the 2nd and 3rd entries of each element in C:

$$C^{(2,3)} = \{000, 011, 101\}^{(2,3)} = \{000, 011, 110\}.$$

In a similar way, one can show that the coset $C + 101$ is equal to $C^{(1,3)}$ and the coset $C + 000$ is equal to $C^{()}$, where $()$ denotes the identity permutation. For the remainder of this Note, n will always indicate an integer greater than 1.

- A *graph* Γ consists of a vertex set V and an edge set E. The type of graph we will consider, sometimes called a *simple graph*, is one in which each edge $e \in E$ is specified by a subset of V that has exactly two elements. In what follows, we will always assume that the vertex set is given by $V = \{1, 2, \dots n\}$. For example, taking $n = 3$, a graph that we shall call Γ_C having edge set $\{\{1, 2\}, \{1, 3\}\}$ is shown in FIGURE 1. In a moment we will link this graph with the cwatset $C = \{000, 011, 101\}$ of $\mathbb{Z}_2^3$.

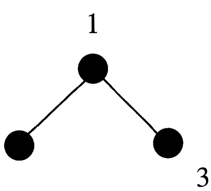


Figure 1 Graph Γ_C with vertex set {1,2,3} and edge set {{1,2},{1,3}}.

- *Isomorphic graphs* are graphs with essentially the same structure. Two graphs are isomorphic if the vertices of one graph may be re-labeled in such a way that the resulting graph is identical to the other graph (i.e. has the same vertex set and edge set). Isomorphism is used to define an equivalence relation on the set of all graphs. Given a graph Γ, the equivalence class $[\Gamma]$ denotes the set of all graphs that are isomorphic to Γ. FIGURE 2 shows the three graphs in the equivalence class $[\Gamma_C]$.

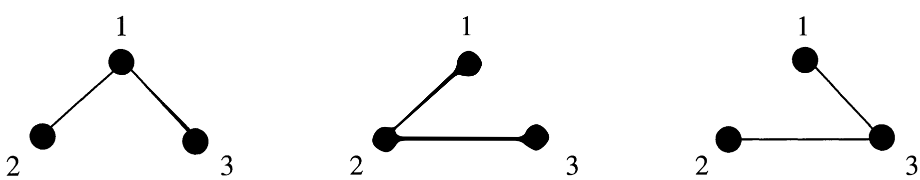


Figure 2 The three isomorphic graphs in $[\Gamma_C]$.

Our connection between graph theory and cwatset theory starts with the observation that each graph Γ with vertex set $\{1, 2, \dots, n\}$ may be represented as an element

$b_1 b_2 \ldots b_{n(n-1)/2} \in \mathbb{Z}^{n(n-1)/2}$. For example, Γ_C with $n = 3$ may be represented by the element $110 \in \mathbb{Z}_2^3$. To see why, first note that the edge set of a graph Γ with vertex set $\{1, 2, 3\}$ must be a subset of the set $\{e_1, e_2, e_3\}$ where the edges e_i are defined "lexicographically", meaning that $e_1 = \{1, 2\}$, $e_2 = \{1, 3\}$, and $e_3 = \{2, 3\}$. The presence or absence of an edge e_i in the edge set of Γ is indicated respectively by the value 1 or 0 assigned to b_i in the representation $b_1 b_2 b_3$ of Γ. That Γ_C is represented by $110 \in \mathbb{Z}_2^3$ indicates that its edge set contains edges $\{1, 2\}$ ($b_1 = 1$) and $\{1, 3\}$ ($b_2 = 1$), but not edge $\{2, 3\}$ ($b_3 = 0$).

In FIGURE 3, we have placed the three graphs in the equivalence class $[\Gamma_C]$ at the corners of a triangular arrangement, labeling them with their representations in $\mathbb{Z}_2^3$. In addition, the labeled double arrows indicate pairwise differences between these representations. We use this figure to illustrate the following facts that hold for every simple graph Γ with vertex set $\{1, 2, \ldots, n\}$:

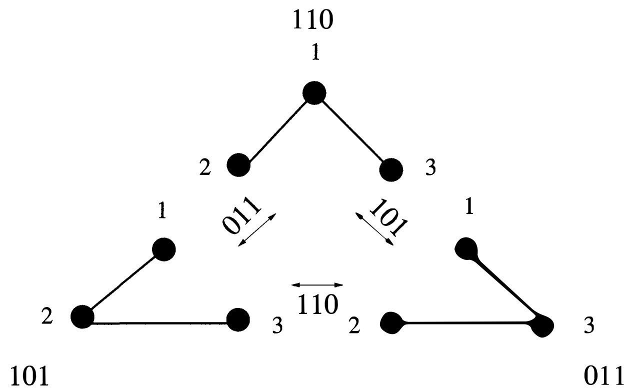


Figure 3 Isomorphic graphs in $[\Gamma_C]$ and their differences in $\mathbb{Z}_2^3$.

Fact 1 *Let cwat*(Γ) *denote the set of all differences between the representations in* $\mathbb{Z}_2^{n(n-1)/2}$ *of* Γ *and its isomorphically equivalent graphs. Then cwat*(Γ) *is a cwatset called the cwatset of* Γ. For example, cwat(Γ_C) = {000,011,101} is the cwatset C of $\mathbb{Z}_2^3$ that we studied earlier.

Fact 2 *The cwatsets of isomorphic graphs are twists of each other.* For example, the graphs Γ_C (represented by 110) and Γ'_C (represented by 101) are isomorphic. One can check that $(\text{cwat}(\Gamma_C))^{(2,3)} = \text{cwat}(\Gamma'_C)$.

These facts can be proved by observing that i) if γ and γ' respectively represent in $\mathbb{Z}_2^{n(n-1)/2}$ two isomorphic graphs Γ and Γ', and if $\gamma + c = \gamma'$, then since $\text{cwat}(\Gamma) + \gamma$ is equal to $\text{cwat}(\Gamma') + \gamma'$ (each set consists of the representations of all the graphs in $[\Gamma]$, or, equivalently, of $[\Gamma']$), it follows that $\text{cwat}(\Gamma) + c = \text{cwat}(\Gamma')$; ii) a permutation τ that re-labels the vertices of Γ to produce an isomorphic graph Γ', also induces a twist σ_τ such that $(\text{cwat}(\Gamma))^{\sigma_\tau} = \text{cwat}(\Gamma')$.

In addition to these facts, an interesting conjecture ([**3**]), which we call the "**0/1** conjecture" is the following:

0/1 Conjecture *The sum of the representations of the graphs in an equivalence class is always either* $\mathbf{0} = 00\ldots0$ *or* $\mathbf{1} = 11\ldots1$. For example, the sum of the representations of the graphs in $[\Gamma_C]$ is given by $110 + 101 + 011 = 000$.

In FIGURE 4, we have drawn a similar diagram for a graph Γ_G with edge set $\{\{1,2\},\{1,3\},\{1,4\}\}$ and representation 111000 in $\mathbb{Z}_2^6$.

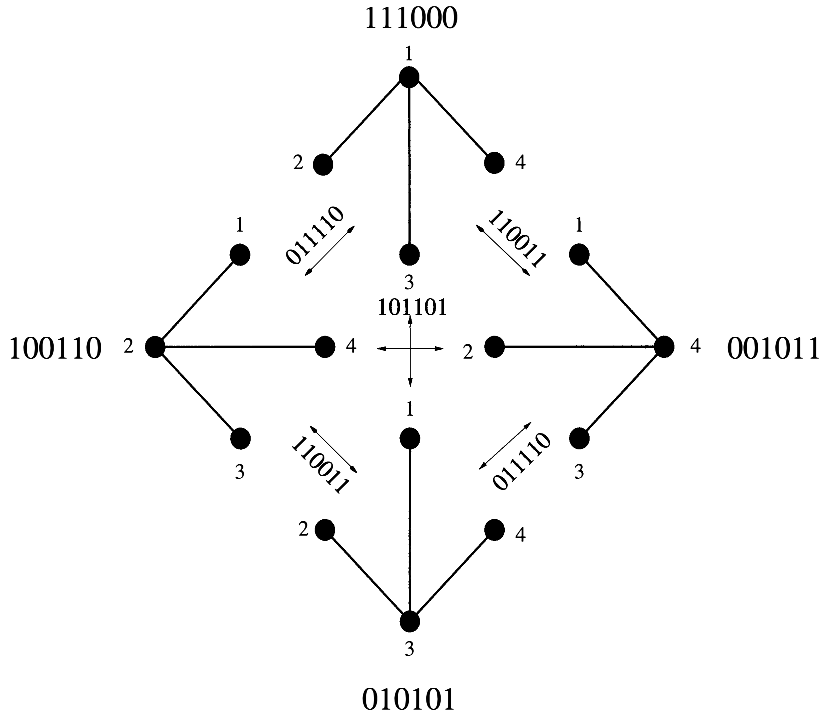


Figure 4 Isomorphic graphs in $[\Gamma_G]$ and their differences in $\mathbb{Z}_2^6$.

In this case cwat(Γ_G), is not only a cwatset but also a subgroup $G = \{000000, 011110, 101101, 110011\}$ of $\mathbb{Z}_2^6$. (Check that G is closed under addition.) Observe further that the cwatsets of the graphs in $[\Gamma_G]$ are all the same. This is not coincidental, as indicated by

Fact 3 *Cwat*(Γ) *is a group if and only if cwat*$(\Gamma) =$ *cwat*(Γ') *for each* $\Gamma' \in [\Gamma]$.

This fact is also illustrated by cwat(Γ_C), which is not a group, and a simple check, using FIGURE 3, that the cwatsets of the graphs in $[\Gamma_C]$ are indeed different. (Try proving Fact 3 by noting that cwat(Γ) is a subgroup if and only if cwat$(\Gamma) + c =$ cwat(Γ) for each $c \in$ cwat(Γ).)

Let γ be the representation in $\mathbb{Z}_2^{n(n-1)/2}$ of a graph Γ. Then the complement of Γ (denoted $\bar{\Gamma}$) is the graph with representation $\mathbf{1} - \gamma$. For example, since Γ_G is represented by 111000, $\bar{\Gamma}_G$ is represented by $111111 - 111000 = 000111$. In FIGURE 5 we have drawn the complementary graphs corresponding to the graphs in FIGURE 4.

Note that while the graphs positioned at the corners of FIGURE 4 have been replaced by their complements in FIGURE 5, the differences between graphs are the same in both figures. Consequently, cwat(Γ_G) must equal cwat$(\overline{\Gamma}_G)$. The observation that the difference between the representations in $\mathbb{Z}_2^{n(n-1)/2}$ of two graphs is the same as the difference between the representations of their complements gives

Fact 4 *The cwatset of a graph always equals the cwatset of its complement* (i.e. cwat$(\Gamma) =$ cwat$(\bar{\Gamma})$).

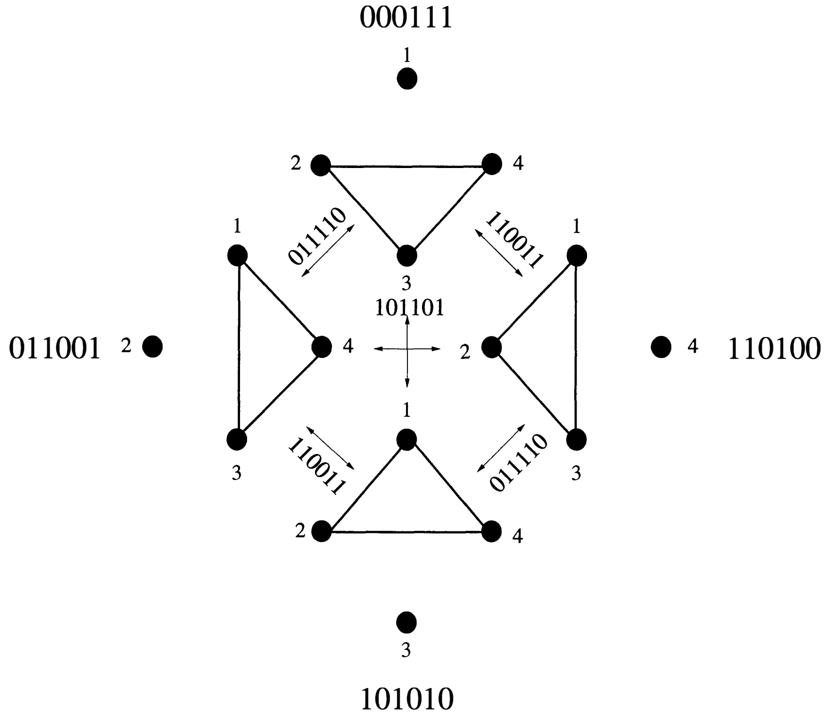


Figure 5 Isomorphic graphs in $[\bar{\Gamma}_G]$ and their differences in $\mathbb{Z}_2^6$.

This brings us to some ideas about how further investigation of cwat(Γ) might benefit both cwatset theory and graph theory:

i) *Explore the use of graphs (and perhaps other structures such as digraphs, multigraphs and directed multigraphs) to generate cwatsets.* Cwatsets constructible as cwat(Γ) for some graph Γ are infinite in number, though not all cwatsets are cwatsets of graphs. This graph-based method of construction appears typically to generate cwatsets that are not subgroups. This is fortunate, since there are very few techniques for constructing non-group cwatsets [7]. For a given n, how many of the cwatsets that are not groups are cwatsets of graphs (or other structures)?

As for the special case when the cwatset of a graph is a group, what can be said about the groups that so occur? How may we characterize the equivalence classes of graphs whose cwatsets are groups?

The standard construction of cwat(Γ) (as we illustrated in FIGURES 3 and 4 when computing cwat(Γ_C) and cwat(Γ_G) respectively) requires explicit knowledge of all the isomorphic graphs in $[\Gamma]$. Let us use the term *special method* to denote a method that does not require a "brute force" enumeration of the entire equivalence class of a graph as is the case in the standard construction. Can the cwatset of a graph be computed by a special method? To this end, a special type of cwatset called a *cyclic cwatset* may have an important role since such cwatsets are constructible by an efficient recursive algorithm:

Generation of Cyclic Cwatsets: Given $\sigma \in S_n$ and $b \in \mathbb{Z}_2^n$, the subset of $\mathbb{Z}_2^n$ defined by $\{b^{\sigma^k} + \cdots + b^{\sigma} + b \mid k \geq 0\}$ is called the *cyclic cwatset* generated by σ and b. Elements in such a cwatset may be generated recursively:

- $b_0 = b$;
- $b_{k+1} = b_k^\sigma + b$.

For example, cwat(Γ_C) is cyclic, being generated, for instance, by $\sigma = (1, 3, 2)$ and $b = 011$:

- $b_0 = b = 011$.
- $b_1 = b_0^\sigma + b = (011)^{(1,3,2)} + 011 = 110 + 011 = 101$.
- $b_2 = b_1^\sigma + b = (101)^{(1,3,2)} + 011 = 011 + 011 = 000$.
- $b_3 = b_0, b_4 = b_1$, etc..

Cwat(Γ_G) is also cyclic, being generated, for instance, by $\sigma = (1, 3)(4, 6)$ and $b = 011110$. For which Γ is cwat(Γ) cyclic? Can the cyclicity of cwat(Γ) be determined by a special method? If cwat(Γ) is cyclic, can its generators be determined by a special method? Investigation of properties that can be determined by special methods is a key to the following:

ii) Explore the design of a cwatset-based algorithm to test whether two graphs are isomorphic.

Referring back to Fact 3, we know that if cwat(Γ) is equal to a group G, then cwat(Γ') is also equal to G for all $\Gamma' \in [\Gamma]$. In this case we say that the property "cwat(Γ) is a group" is an *isomorphic invariant.* An invariant property, if true about Γ, must also be true about every graph isomorphic to Γ. Hence if cwat(Γ) is a group, and cwat(Γ') is not a group, Γ and Γ' cannot be isomorphic. Note, however, that unless there is a special method to determine whether the cwatset of a graph is a group, this particular isomorphic invariant is not really useful as a test for isomorphism.

Ideally, to design an effective isomorphism test, we need a *special invariant property*, meaning a property P for which, given any graph Γ:

- A special method can be used to determine whether Γ satisfies P.
- P holds for Γ if and only if it holds for every graph in $[\Gamma]$.
- If Γ satisfies P, then cwat(Γ) can be computed by a special method.

Note that answers to the questions we asked earlier about cyclic cwatsets would settle whether "cwat(Γ) is cyclic" is a special invariant property. The discovery of a special invariant property P would lead to the following algorithm to test whether two graphs Γ and Γ' represented respectively by $\gamma, \gamma' \in \mathbb{Z}_2^{n(n-1)/2}$ are isomorphic:

Step 1: Use a special method to determine whether P holds for both Γ and Γ'.
Step 2: a) If neither graph satisfies P, exit (the test fails).

b) If only one of the two graphs satisfies P, exit (the graphs are not isomorphic).
Step 3: Compute cwat(Γ) by special method and then check whether the difference $c = \gamma' - \gamma$ is in cwat(Γ). If so, Γ and Γ' are isomorphic; if not, Γ and Γ' are not isomorphic.

Should cwatset research lead to the discovery of a special invariant property P, light would be shed on a famous unsolved problem in graph theory, the so-called "graph-isomorphism problem" [**4**], that involves finding an efficient algorithm to determine whether two graphs are isomorphic.

Acknowledgment. We would like to thank Dr. Gary Sherman for his inspiration and the referees for their helpful suggestions. Nancy-Elizabeth Bush did this work while an undergraduate student at Wheaton College.

REFERENCES

1. D. K. Biss, On the Symmetry Groups of Hypergraphs of Perfect Cwatsets, *Ars Combinatoria*, to appear.
2. C. M. Girod, M. Lepinski, J. R. Mileti and J. R. Paulhus, Cwatset Isomorphism and its Consequences, *Rose-Hulman Technical Report Series*, MS TR 00-01 (2000).
3. L. Granlund, D. Shaw, and R. Yates, Private Communication (April 2000).
4. J. Gross and J. Yellen, *Graph Theory and Its Applications*, CRC Press, Boca Raton, FL, 1999.
5. J. Kerr, Hypergraph Representations and Orders of Cwatsets, *Rose Hulman Technical Report Series*, MS TR 93-02 (1993).
6. M. Lepinski, Automorphic Subsets of the n-Dimensional Cube are Translations of Cwatsets, *Rose-Hulman Technical Report Series*, MS TR 99-01 (1999).
7. R. Mohr, CWATSETS: Weights, Cardinalities and Generalizations, *Rose-Hulman Technical Report Series*, MS TR 96-03 (1996).
8. G. J. Sherman and M. Wattenberg, *Introducing ... Cwatsets!*, this MAGAZINE **67** (1994), pp. 109–117.

Period-3 Orbits via Sylvester's Theorem and Resultants

JACQUELINE BURM
Project SEED
Detroit, MI 48201

PAUL FISHBACK
Grand Valley State University
Allendale, MI 49401

Introduction One of the most familiar functions in nonlinear dynamics is the logistic function

$$f_r(x) = rx(1-x),$$

where r and x are real numbers. The sequence of iterates of this function, f_r, $f_r(f_r)$, $f_r(f_r(f_r))$, ... leads to a dynamical system illustrating such important concepts as bifurcations, period-doubling, and chaos.

Several recent articles ([**1**],[**5**],[**7**]) have investigated real, period-3 orbits of this system. Specifically, they consider the third iterate $f_r^3 = f_r(f_r(f_r))$ and prove that $r_0 = 1+\sqrt{8}$ is the smallest positive value of r for which the equation $f_r^3(x) = x$ has a solution x_0 that is not already a fixed point of f_r. The sequence x_0, $x_1 = f_r(x_0)$, $x_2 = f_r(x_1)$ is called a period-3 orbit of f_r. Sarkovskii's theorem [**3**] states that when a period-3 orbit exists for a continuous function on the real line, so do orbits of all other periods as well.

The proofs given in [**1**], [**5**], and [**7**] vary substantially. The authors of [**7**] use a change of variables and "careful algebraic maneuvers" to prove their result. The proof given in [**1**] is simpler than that in [**7**] but "does not readily lend itself to generalization." The approach of [**5**] makes clever use of the discrete Fourier transform. The purpose of this note is partly to provide another derivation of the fact that $r_0 = 1+\sqrt{8}$ using variable elimination and Sylvester's theorem. More importantly the method used generalizes to many other single variable and multivariable polynomial mappings and illustrates the value of using resultants in the study of bifurcation structure.

The birth of period-3 Let r_0 denote the smallest positive value r for which $f_r(x) = rx(1-x)$ possesses a period-3 orbit. When $r = r_0$ the graph of f_r^3 intersects the line $y = x$ at five values of x. Two of these values, $x = 0$ and $x = \frac{r_0 - 1}{r_0}$, correspond to fixed points of f_r. The remaining three values of x constitute a period-3 orbit. At each orbit value the graph of f_r^3 is tangent to $y = x$. (See FIGURE 1.)

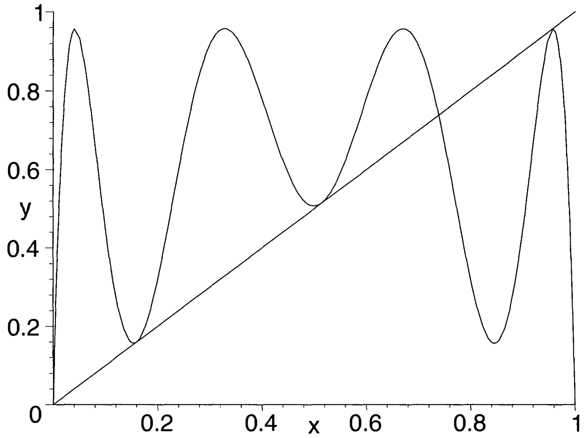


Figure 1 The birth of the period-3 orbit.

If we denote

$$\begin{aligned} G_r(x) &= f_r^{(3)}(x) - x \\ &= -r^7x^8 + 4r^7x^7 - (6r^7 + 2r^6)x^6 + (4r^7 + 6r^6)x^5 - (r^7 + 6r^6 + r^5 + r^4)x^4 \\ &\quad + (2r^6 + 2r^5 + 2r^4)x^3 - (r^5 + r^4 + r^3)x^2 + (r^3 - 1)x, \end{aligned}$$

then for the special choice $r = r_0$, G_r is tangent to the x-axis. In other words, when $r = r_0$, the polynomials G_r and $\frac{d}{dx}G_r$ have a common root, which implies that the *resultant* of G_r and $\frac{d}{dx}G_r$ must vanish.

The resultant and Sylvester's theorem Let $P(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0$ be a polynomial of degree m, whose roots are denoted by $t_1, t_2, \ldots, t_m$, and let $Q(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_0$ be a second polynomial in x of degree n. Then the resultant of P and Q is defined by

$$R(P, Q) = a_m^n Q(t_1) Q(t_2) \cdots Q(t_m).$$

Clearly $R(P, Q)$ vanishes if and only if P and Q have a common root.

Various techniques exist for computing resultants, but the earliest known method using only the coefficients of P and Q was established by J. J. Sylvester in 1840:

THEOREM 1. *Let P and Q be polynomials as given above. Define S to be the $(m+n)$ by $(m+n)$ "Sylvester" matrix given by*

$$
\begin{array}{l}
n \text{ rows} \left\{ \vphantom{\begin{array}{c}a\\a\\a\\a\\a\end{array}} \right. \\
m \text{ rows} \left\{ \vphantom{\begin{array}{c}a\\a\\a\\a\end{array}} \right.
\end{array}
\left[\begin{array}{cccccccccccccc}
a_m & a_{m-1} & a_{m-2} & \cdot & \cdot & \cdot & a_0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\
0 & a_m & a_{m-1} & a_{m-2} & \cdot & & & a_0 & 0 & \cdot & \cdot & \cdot & 0 \\
0 & 0 & a_m & a_{m-1} & a_{m-2} & \cdot & \cdot & \cdot & a_0 & 0 & \cdot & \cdot & 0 \\
\vdots & & & & & \vdots & & & & & & & \vdots \\
0 & 0 & 0 & & & a_m & & & & & \cdot & a_1 & a_0 \\
b_n & b_{n-1} & \cdot & \cdot & \cdot & \cdot & \cdot & b_0 & 0 & \cdot & \cdot & \cdot & 0 \\
0 & b_n & b_{n-1} & \cdot & \cdot & \cdot & \cdot & \cdot & b_0 & 0 & \cdot & \cdot & 0 \\
\vdots & & & & & & & & & & & & \vdots \\
0 & \cdot & \cdot & \cdot & 0 & b_n & b_{n-1} & \cdot & \cdot & \cdot & \cdot & b_1 & b_0
\end{array}\right].
$$

Then

$$R(P, Q) = \det(S).$$

The proof is elementary, requiring only basic properties of polynomials and determinants (see e.g., [**2**]).

Using Sylvester's theorem and a computer algebra system, one may verify that

$$R\left(G_r, \frac{d}{dx}G_r\right) = -r^{49}(r-1)^2(r^2-2r-7)^3(r^2+r+1)^4(r^2-5r+7)^4.$$

The only real, positive root of this polynomial that does not correspond to a fixed point of f_r occurs at $r_0 = 1+\sqrt{8} \approx 3.8284$.

The value $r_0 = 1+\sqrt{8}$ corresponds to the onset of two real, period-3 orbits for f_r. For r slightly greater than r_0, one of the corresponding period-3 orbits is *stable*, or *attracting*. This means that if x_0 is a value in this orbit, then $|(f_r^3)'(x_0)| < 1$, and for x sufficiently close to x_0, $f^{3n}(x) \longrightarrow x_0$ as $n \longrightarrow \infty$.

Eventually, as r increases beyond a certain value, this corresponding orbit becomes unstable or repelling. When this change or *bifurcation* occurs, the slope of $f_r^{(3)}$ at each point in this orbit equals -1. The corresponding value of r is determined using

$$R\left(f_r^3(x) - x, \frac{d}{dx}f_r^3(x) + 1\right) = -r^{49}(r-3)(r-1)(r^2-3r+3)(r^2-r+1)$$
$$(r^6 - 6r^5 + 4r^4 + 24r^3 - 14r^2 - 36r - 81)^3.$$

The only real, positive root of this polynomial that does not correspond to a fixed point or to a period-2 orbit of f_r is the single real, positive root r_1 of

$$r^6 - 6r^5 + 4r^4 + 24r^3 - 14r^2 - 36r - 81 = 0.$$

This equation was also derived in [**5**], where it was shown that

$$r_1 = 1 + \left(\frac{11}{3} + \left(\frac{1915}{54} + \frac{5\sqrt{201}}{2}\right)^{\frac{1}{3}} + \left(\frac{1915}{54} - \frac{5\sqrt{201}}{2}\right)^{\frac{1}{3}}\right)^{\frac{1}{2}} \approx 3.8415.$$

Hence f_r possesses stable period-3 orbits when $r_0 < r < r_1$.

Higher dimensions The resultants in these two examples were calculated using the classical method of Sylvester. Over the years, systems of polynomial equations have arisen in a variety of applications including robotics, computer graphics, chemical engineering, and architecture. *Multivariate resultants* are a tool for determining the solvability of such systems.

The multivariate resultant corresponding to a system of $n+1$ polynomials in n variables is an expression in the coefficients of the polynomials that must equal zero when the system has a real solution. A variety of techniques now exist for computing multivariate resultants.

That the multivariate resultant is useful for investigating higher-dimensional mappings may be seen by investigating the *Hénon mapping*

$$H_{a,b}(x, y) = (1 + y - ax^2, bx),$$

where $-1 \le b \le 1$. Consider the problem of determining the region in the (a, b)-plane corresponding to real, stable, period-3 orbits. Analytic solutions to this problem have been derived elsewhere, (see e.g., [**6**]), but without the aid of resultants.

The periodicity condition requires that the system of two polynomial equations arising from $H^3_{a,b}(x, y) = (x, y)$ has a solution x and y that does not correspond to a fixed point. To generate a third polynomial equation, which is needed for the elimination of x and y, we use the fact that at each point (x, y) in a stable period-3 orbit, the eigenvalues of the Jacobian of $H^3_{a,b}(x, y)$ in x and y, denoted by J, must both have modulus less than one. By the Schur-Cohn stability criterion, this condition is satisfied if and only if $-1 \pm \text{trace}(J) < \det(J) < 1$, [**4**].

Thus the boundary of the desired region is determined in three steps. Each step involves using a resultant to eliminate x and y from a system of three polynomials consisting of $H^3_{a,b}(x, y) = (x, y)$ and one of the following:

i. $\text{trace}(J) = \det(J) + 1$; ii. $\text{trace}(J) = -\det(J) - 1$; iii. $\det(J) = 1$.

The three resulting relations in a and b form what are called the "divergence," the "flip," and the "flutter" boundary components, respectively. They correspond to cases in which at least one eigenvalue of J is equal to 1, at least one eigenvalue is equal to -1, and both eigenvalues are complex and of modulus 1.

A *Maple* implementation of a multivariate resultant algorithm developed by Saxena in [**8**] (available for download at `http://www2.gvsu.edu/~fishbacp/henon/article.html`), yields the following three respective boundary components:

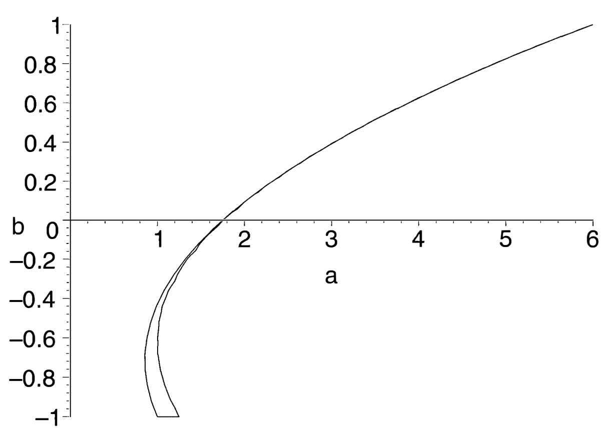


Figure 2 Period-3 stability region for the Hénon mapping.

i. $4a - 7b^2 - 7 - 10b = 0$;

ii. $-81 + 72a - 128a^2 + 64a^3 - 54b - 216ba - 32ba^2 - 18b^2 - 252ab^2 - 128b^2a^2 + 90b^3 - 216b^3a - 18b^4 + 72b^4a - 54b^5 - 81b^6 = 0$;

iii. $b = -1$.

The boundary obtained from these curves is plotted in FIGURE 2 in the (a, b)-plane. (See also [**6**].) Points interior to this boundary correspond to real, stable, period-3 orbits of the Hénon mapping.

Acknowledgment. Jacqueline Burm did this work while a student at Grand Valley State University. Both she and the faculty co-author were supported by the Grand Valley State University Summer Undergraduate Research Program.

REFERENCES

1. J. Bechhoefer, The birth of period 3, revisited, this MAGAZINE **69** (1996), 115–118.
2. W. Conkwright, *The Theory of Equations*, Ginn and Company, New York, NY, 1957.
3. R. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd edition, Addison Wesley, Reading, MA, 1989.
4. S. Elaydi, *An Introduction to Difference Equations*, 2nd edition, Springer-Verlag, New York, NY, 1999.
5. W. Gordon, Period three trajectories of the logistic map, this MAGAZINE **69** (1996), 118–120.
6. D. Hitzl and F. Zele, An exploration of the Hénon quadratic map, *Physica D* **14** (1985), 305–326.
7. P. Saha and S. Strogatz, The birth of period three, this MAGAZINE **68** (1995), 42–47.
8. Saxena, Tushar, Efficient Variable Elimination Using Resultants, doctoral dissertation, SUNY-Albany, Albany, NY, 1998.

Writing Numbers in Base 3, the Hard Way

GARY E. MICHALEK
La Salle University
Philadelphia, PA 19141-1199

Introduction An important feature of our system of writing numbers is the significance attached to the position of the numerals. (The Romans had it almost, but not quite, right.) When we write 563, for example, we really mean

$$563 = 5(10^2) + 6(10^1) + 3(10^0),$$

where each numeral is the coefficient of the power of 10 corresponding to its position. This is the base 10 system of expressing numbers. Similarly, in base 2, nonnegative integers are expressed using the numbers 0 and 1 as coefficients of powers of 2. For example, 27 is written as 11011, since

$$27 = 1(2^4) + 1(2^3) + 0(2^2) + 1(2^1) + 1(2^0).$$

Negative integers can be expressed in a similar fashion, if we allow the use of -1 as a coefficient. That is, every integer can be expressed in base 2 using the set $\{-1, 0, 1\}$ as a *coefficient system.* Do we *need* to use the set $\{-1, 0, 1\}$, or might another set of numbers do as well? This question was considered in a more complicated context by the Hungarian number theorist Imre Kátai, and most of the following results are implicit in his work [**1**]. The question is interesting in its own right, and the answers provide nice illustrations of modular arithmetic and suggest some interesting exercises in computer programming.

We consider base 3, which provides richer examples than base 2. Let A denote a set of integers. If an integer m can be written in the form

$$m = c_n(3^n) + c_{n-1}(3^{n-1}) + \cdots + c_1(3^1) + c_0(3^0)$$

for some nonnegative integer n, with each coefficient c_i in A, we say that m is *expressible in* A. The set A is a *coefficient system* if every integer is expressible in A.

There are some obvious restrictions on the sets that can constitute coefficient systems. For instance, if m is expressed in the form above, m and c_0 must be congruent modulo 3. Therefore, for every integer to be expressible, A must contain at least one representative from each congruence class. For the moment, consider sets of exactly three elements, say $A = \{a_0, a_1, a_2\}$, where $a_i \equiv i \pmod 3$. A coefficient system of minimal size (three elements, for base 3) is called a *number system*. For simplicity, in this note we will assume a number system has $a_0 = 0$. The following observations shorten the list of possible base three number systems:

(i) a_1 and a_2 must have opposite signs so that both positive and negative numbers are expressible;

(ii) If a_1 and a_2 have a common divisor k, then every number expressible in A is divisible by k. Hence $\gcd(a_1, a_2) = 1$ for a number system.

The remaining candidates for number systems are sets of the form $\{0, a_1, a_2\}$, where a_1 and a_2 are relatively prime, of opposite signs, and congruent to 1 and 2 modulo 3, respectively. To gain further insight into which triples form number systems, we consider how to find the expression for a number (if such an expression exists).

Expressing a number Given $A = \{a_0 = 0, a_1, a_2\}$, define a function $F_A : \mathbb{Z} \to \mathbb{Z}$ by $F_A(m) = (m - a_i)/3$, where a_i is the element of A congruent to m modulo 3. Consider, for instance, the expression for 10 in $A = \{0, -11, 2\}$:

$$10 = \mathbf{2}(3^3) + \mathbf{0}(3^2) + \mathbf{-11}(3^1) + \mathbf{-11}$$

Repeated application of F_A gives

$$F_A(10) = \frac{10 - \mathbf{-11}}{3} = 7; \qquad F_A{}^2(10) = \frac{7 - \mathbf{-11}}{3} = 6$$

$$F_A{}^3(10) = \frac{6 - \mathbf{0}}{3} = 2; \qquad F_A{}^4(10) = \frac{2 - \mathbf{2}}{3} = 0.$$

At each stage, F_A subtracts the appropriate member of A and divides by 3. These appropriate numbers are, of course, precisely the coefficients in the expression for 10. If m is expressible in A, then clearly $F_A{}^n(m) = 0$ for some positive integer n. Conversely, if $F_A{}^n(m) = 0$, we can find the expression for m by keeping track of which element of A is subtracted at each stage. In short, m is expressible in A if and only if $F_A{}^n(m) = 0$ for some positive integer n.

Number system or not? How can we check whether *all* numbers are expressible, as we must do to decide whether A is a number system? The key is a fortuitous property of the function F_A, which is easily established by checking the inequalities involved. As F_A is applied repeatedly, the results decrease in absolute value, and eventually the result is in the interval $I_A = [-M, M]$, where M is the integer part of $\max\{|a_1|/2, |a_2|/2\}$. With $A = \{0, -11, 2\}$, where $I_A = [-5, 5]$, we have

$$F_A(100) = 37, \quad F_A(37) = 16, \quad F_A(16) = 9, \quad F_A(9) = 3.$$

Furthermore, $F_A(I_A) \subset I_A$, so one can think of the interval as a black hole. Once inside the interval, only two things can happen as you continue to apply F_A: Either 0 is eventually reached (in which case the number you began with is expressible) or numbers begin to repeat (in which case 0 will never be reached and the number is not expressible). In the case above, one can continue to apply F_A to see that 100 is *not* expressible. These repeating nonzero numbers (those m for which $F^k(m) = m$ for some positive k) are called *periodic* numbers. The existence of periodic numbers makes clear that A cannot be a number system.

If periodic numbers exist, they must lie in the "black hole interval" for F_A. We therefore have a way to check (in a finite amount of time!) whether A is a number system: Display the action of F_A on I_A in a directed graph, where an integer m is connected by an arrow to $F_A(m)$. The directed graph for $A = \{0, -11, 2\}$ is shown in FIGURE 1. The presence of the nonzero periodic numbers 1, 4, 5, and -1 demonstrates that this A is not a number system.

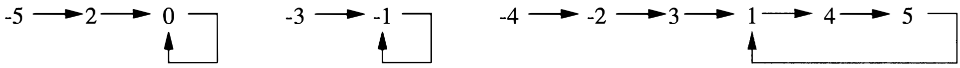


Figure 1 Directed graph for $\{0, -11, 2\}$

For $A = \{0, -11, 2\}$, only -5 and 2 are expressible, as well as any numbers outside I_A that connect to -5 and 2. Contrast this with the directed graph of an actual number system, $A = \{0, -11, 5\}$, shown in FIGURE 2. Here, all integers in $I_A = [-5, 5]$, and therefore all integers, connect to 0.

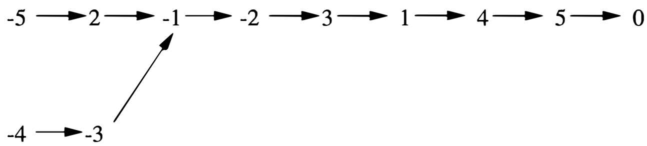


Figure 2 Directed graph for $\{0, -11, 5\}$

Note that there are no number systems in base 2: If $a \in A$, then $-a$ will not be expressible, since

$$F_A(-a) = \frac{-a - a}{2} = -a.$$

New criteria and a theorem Our analysis reveals some additional requirements for a base three number system:

(iii) A number system cannot contain an even nonzero number, for if a is even and $a \in A$, then $F_A(-a/2) = -a/2$, so $-a/2$ will not be expressible in A.

(iv) A number system cannot contain the number 13. If $13 \in A$, we have ${F_A}^3(-5) = -5$, as the reader may verify. (Notice, in this computation the other nonzero number in A doesn't come into play. This phenomenon is similar to the even number problem. Even numbers and 13 carry their "own loops" with them and so cannot belong to a number system. Can you find other numbers that have their "own loops"?)

(v) The set A will not be a number system if both nonzero numbers in A lie outside I_A. If, say, $A = \{0, 13, -19\}$ with $I_A = [-9, 9]$, then the only expressible number in I_A is 0: The only nonzero numbers m with $F_A(m) = 0$ are 13 and -19.

However, the nonzero numbers in I_A cannot connect to 13 and -19, and so will not connect to 0 in the directed graph.

Must a set that meets all the requirements (i)–(v) be a number system? The answer is no. Consider $A = \{0, 7, -25\}$, for which 4 is a periodic number. In general, no quick inspection of the members of A will determine whether A is a number system, and the directed graph must be drawn to determine whether periodic numbers exist.

It may also be interesting to examine a set that fails to meet the listed requirements in an extreme way. Consider $A = \{0, 7, -7\}$, where both nonzero numbers lie outside I_A, and are certainly not relatively prime. The directed graph here has an interesting shape. It is a good exercise in modular arithmetic to show that repeated application of F_A connects an integer m to an element of I_A that is congruent to m modulo 7.

One positive result of observations (i)–(v) relates to the earlier definition of a coefficient system. Here the set A is not necessarily minimal in size, and, unlike the case for number systems, the expression for a number need not be unique. Consider two arbitrarily chosen *positive* integers a and b, with $a < b$, where one of the numbers is congruent to 1 and the other is congruent to 2 modulo 3. We know $A = \{0, a, b\}$ is not a coefficient system, but we can expand the set in a natural way to make it so. If we consider instead the set $A - A$, consisting of all differences between elements of A, the obstacles presented by remarks (i) and (v) are removed. Since $A - A$ contains $-a$ and $-b$, we have the necessary mixture of signs. Furthermore, some number must lie in I_A, since either $b - a$ or a is less than $b/2$. In fact, we have the following nice theorem, conjectured by Kátai after extensive testing on a computer. (It is an interesting exercise to write a computer program to check whether a given set is a coefficient system.) The proof of the theorem, rather long but using only simple algebraic techniques, appears in [**2**].

Theorem. *Let $A = \{0, a, b\}$, where a and b are congruent to* 1 *and* 2 *modulo* 3. *Then $A - A = \{0, a, b, -a, -b, b - a, a - b\}$ is a coefficient system if and only if* $\gcd(a, b) = 1$.

For example, $A = \{0, 7, -25\}$ is not a coefficient system, but the expanded set $A - A = \{0, 7, -25, -7, 25, 32, -32\}$ is, since $\gcd(7, -25) = 1$. A theorem recently proved by the author [**3**] generalizes this result to other bases:

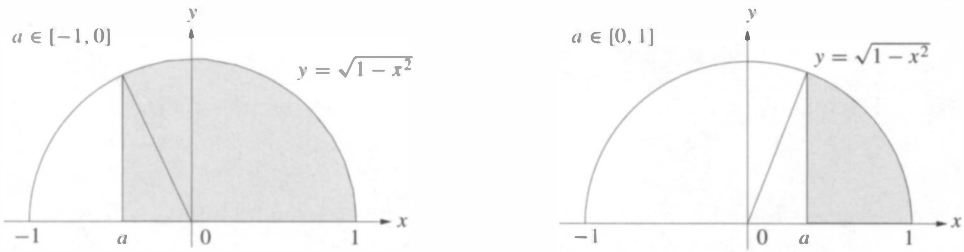
Theorem. Let $A = \{a_0 = 0, a_1, a_2, \dots, a_{n-1}\}$ be a set of representatives of the congruence classes modulo n, with $n > 1$. Then $A - A$ is a coefficient system in base n if and only if $\gcd(a_1, a_2, \dots, a_{n-1}) = 1$.

REFERENCES

1. Imre Kátai, *Generalized Number Systems and Fractal Geometry*, Monograph, Janus Pannonius University, Pécs, Hungary, 1995.
2. Gary Michalek, Base three just touching covering systems, *Publicationes Mathematicae Debrecen*, Tomus **51**/3-4 (1997), 241–263.
3. Gary Michalek, Base n just touching covering systems, *Publicationes Mathematicae Debrecen*, to appear.

Proof Without Words: Look Ma, No Substitution!

$$\int_a^1 \sqrt{1-x^2}dx = \frac{\cos^{-1} a}{2} - \frac{a\sqrt{1-a^2}}{2}, \; a \in [-1, 1].$$



$$\int_a^1 \sqrt{1-x^2}dx = \frac{\cos^{-1} a}{2} + \frac{(-a)\sqrt{1-a^2}}{2} \qquad \int_a^1 \sqrt{1-x^2}dx = \frac{\cos^{-1} a}{2} - \frac{a\sqrt{1-a^2}}{2}$$

—MARC CHAMBERLAND
GRINNELL COLLEGE
GRINNELL, IA 50112

Permutations and Coin-Tossing Sequences

DAVID CALLAN
Department of Statistics
University of Wisconsin
Madison, WI 53706-1693

"The answer can be restated in quite a striking form, like this

THEOREM 3.7.1. *Let a positive integer n be fixed. The probabilities of the following two events are equal:*

(a) a permutation is chosen at random from among those of n letters, and it has an even number of cycles, all of whose lengths are odd

(b) a coin is tossed n times and exactly n/2 heads occur."

This quote is from Herbert Wilf's delightful book *generatingfunctionology* [**1**, p. 75]. It occurs in the chapter on the exponential formula, a powerful technique for counting labelled structures formed from "connected" components. Such structures include various types of graphs, permutations (formed from cycles), and partitions (a union of blocks). Applied to permutations on n letters comprising an even number of cycles all of whose lengths are odd, the exponential formula shows that there are $\binom{n}{n/2}\frac{n!}{2^n}$ of them. Thus the probability in (a) is $\binom{n}{n/2}\frac{1}{2^n}$; this is obviously also the probability in (b), and the quoted theorem follows. The purpose of the present note is to offer a combinatorial explanation of this "striking" result.

Both probabilities are 0 if n is odd; so assume n is even. Now let $\mathcal{A}_n$ denote the permutations in (a) and let $\mathcal{B}_n$ denote the coin-tossing sequences in (b). Thus $\mathcal{A}_n$ is the set of permutations on $[n] = \{1, 2, \ldots, n\}$ all of whose cycles are of odd length (their number has to be even since n is even). We can take $\mathcal{B}_n$ to be the set of 0-1 sequences comprising m 1s and m 0s (where $m = n/2$). Let $\mathcal{S}_n$ denote the set of all $n!$ permutations on $[n]$ and $\mathcal{T}_n$ the set of all 2^n 0-1 sequences of length n. Then the Theorem asserts

$$\frac{|\mathcal{A}_n|}{|\mathcal{S}_n|} = \frac{|\mathcal{B}_n|}{|\mathcal{T}_n|} \tag{1}$$

We will "explain" this coincidence of probabilities by constructing a bijection

$$\mathcal{A}_n \times \mathcal{T}_n \longrightarrow \mathcal{B}_n \times \mathcal{S}_n \tag{2}$$

First, we give a bijection from $\mathcal{A}_n$—the permutations on $[n]$ with odd-length cycles—to $\mathcal{C}_n$, the permutations on $[n]$ with even-length cycles. To do so, say a permutation is in *standard cycle form* when its cycles are arranged so that the largest element in each cycle occurs first, and these first elements are in increasing order. Thus $(5, 3, 1)(8, 2, 6)(9)(12, 10, 4, 11, 7)$ is in standard cycle form. Given $\pi \in \mathcal{A}_n$ in standard form, move the last element of the cycles in the 1st, 3rd, 5th, ... positions to the end, respectively, of the cycles in 2nd, 4th, 6th, ... positions. Thus $(5, 3, 1)(8, 2, 6)(9)(12, 10, 4, 11, 7) \to (5, 3)(8, 2, 6, 1)(12, 10, 4, 11, 7, 9)$. This is the desired bijection. Note the resulting permutation is in $\mathcal{C}_n$ and is again in standard form. To reverse the mapping, delete the last element of the last cycle. Place it at the end of the preceding cycle provided this gives a legitimate cycle (largest element first); otherwise create a new 1-cycle consisting of this element alone. Proceed similarly so that each of the original even-length cycles either has its last element deleted or acquires a new last element.

Second, we give a bijection between $\mathcal{C}_n$—the permutations on $[n]$ with even-length cycles—and $\mathcal{D}_n \times \mathcal{D}_n$ where $\mathcal{D}_n$ is the permutations on $[n]$ with all cycles of length 2 (transpositions). An element $\pi \in \mathcal{D}_n$ can be represented as a graph on the vertices $[n]$ in which each vertex has degree 1 (two vertices are joined precisely when they occur in the same transposition). Thus $(\pi_1, \pi_2) \in \mathcal{D}_n \times \mathcal{D}_n$ is a pair of such graphs as illustrated in FIGURE 1 (with $n = 6$; solid edges for the first graph, dotted edges for the second).

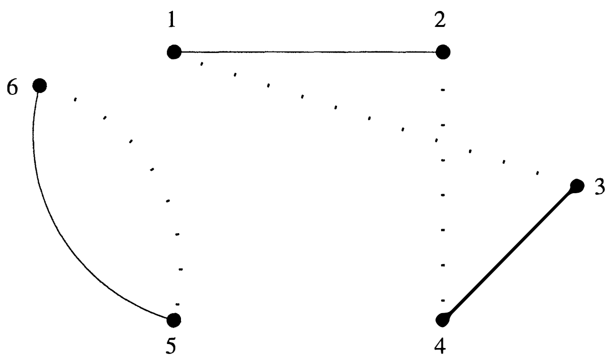


Figure 1

The union of the edge sets is a collection of (unoriented) cycles of even length ≥ 2 with alternating solid and dotted edges. This yields even-length cycles on $[n]$ just as we want except that we must orient the cycles of length ≥ 4 in one of two possible ways. But there are two possible patterns for the solid and dotted edges in such a cycle, so all is well. For definiteness, orient each cycle in the direction of, say, the solid edge emanating from its smallest vertex.

These bijections show that the left side of (2) is $\approx \mathcal{D}_n \times \mathcal{D}_n \times \mathcal{T}_n$ and hence $\approx \mathcal{D}_{2m} \times \mathcal{D}_{2m} \times \mathcal{T}_m \times \mathcal{T}_m$ (recall $n = 2m$). Turning to the right side of (2), we observe that there is a bijection $\mathcal{S}_{2m} \longrightarrow \mathcal{B}_{2m} \times \mathcal{S}_m \times \mathcal{S}_m$. Given $\pi \in \mathcal{S}_{2m}$, the *locations* of $1, 2, \dots, m$ in π give an element of $\mathcal{B}_{2m}$, the *order* of $1, 2, \dots, m$ in π gives an element of $\mathcal{S}_m$, and the order of $m+1, m+2, \dots, 2m$ gives another. For example, $\left(\begin{smallmatrix}1&2&3&4&5&6\\4&6&3&5&1&2\end{smallmatrix}\right)$ yields $(0\,0\,1\,0\,1\,1) \times \left(\begin{smallmatrix}1&2&3\\3&1&2\end{smallmatrix}\right) \times \left(\begin{smallmatrix}1&2&3\\1&3&2\end{smallmatrix}\right) \in \mathcal{B}_{2m} \times \mathcal{S}_m \times \mathcal{S}_m$ where in the last permutation $1\,3\,2$ is the rank ordering of $4\,6\,5$. Hence the right side of (2) is $\approx \mathcal{B}_{2m} \times \mathcal{B}_{2m} \times \mathcal{S}_m \times \mathcal{S}_m$.

Thus we can identify a "square root" of each side of (2) and it now suffices to exhibit a bijection

$$\mathcal{D}_{2m}\left(\begin{smallmatrix}\text{products of}\\ \text{disjoint 2-cycles}\end{smallmatrix}\right) \times \mathcal{T}_m\left(\begin{smallmatrix}\text{unrestricted}\\ \text{0-1 sequences}\end{smallmatrix}\right) \longrightarrow \mathcal{B}_{2m}\left(\begin{smallmatrix}\text{sequences of}\\ m\text{ 0s, }m\text{ 1s}\end{smallmatrix}\right) \times \mathcal{S}_m\left(\begin{smallmatrix}\text{unrestricted}\\ \text{permutations}\end{smallmatrix}\right) \quad (3)$$

This is quite easy: given $(\pi, \epsilon) \in \mathcal{D}_{2m} \times \mathcal{T}_m$, start with π in standard cycle form. Reverse the transpositions located in those positions where ϵ has a 1, and then arrange the transpositions in the order of their first elements. For example, with $m = 4$, $\pi = (3, 2)(5, 1)(6, 4)(8, 7)$ (in standard cycle form), and $\epsilon = (1, 0, 1, 1)$, $(\pi, \epsilon) \to (2, 3)(5, 1)(4, 6)(7, 8) \to (2, 3)(4, 6)(5, 1)(7, 8)$. The first elements of the final product of transpositions form an m-element subset of $[2m]$ determining an element of $\mathcal{B}_{2m}$, while the rank ordering of the second elements is a permutation in $\mathcal{S}_m$. The example yields $\{2, 4, 5, 7\} \to (0, 1, 0, 1, 1, 0, 1, 0) \in \mathcal{B}_8$ and $(3, 6, 1, 8) \to (2, 3, 1, 4) \in \mathcal{S}_4$. The original pair (π, ϵ) can be uniquely retrieved, and the bijection (3) is established.

REFERENCES

1. Herbert S. Wilf, *generatingfunctionology*, Academic Press, New York, NY, 1990.

An Easy Solution to *Mini Lights Out*

JENNIE MISSIGMAN
RICHARD WEIDA
Lycoming College
Williamsport, PA 17701

In this MAGAZINE [**1**], Anderson and Feil demonstrated how to use linear algebra to solve the game *Lights Out*, which consists of a 5×5 array of lighted buttons; each light may be on or off. Pushing any button changes the on/off state of that light as well as the states of all its vertical and horizontal neighbors. Given a particular configuration of lights which are turned on, the object of the game is to turn out all the lights. While the computation of the solution in [**1**] is relatively straightforward, it certainly cannot be accomplished by hand in a reasonable amount of time. Analysis of a somewhat similar 3×3 game, *Merlin's Magic Square*, can be found in [**2, 3**].

Tiger Electronics has recently released a new version of the game, called *Mini Lights Out*. This consists of a 4×4 array of lighted buttons, but this time, unlike the original 5×5 version, "on" a torus. That is, the uppermost and lowermost rows are considered neighbors and likewise the leftmost and rightmost columns are considered neighbors.

Applying the techniques and notation of [**1**], we obtain not only the results predicted by Anderson and Feil, but also a complete, easy-to-compute, winning strategy for the 4×4 "mini" game.

We define the *neighborhood* of a button to be the set of buttons affected by pushing that button. That is, the neighborhood contains the button itself and its vertical and horizontal neighbors. In the original 5×5 *Lights Out* game, a neighborhood could consist of either three, four, or five buttons depending on whether the button was at a corner, on an edge, or in the interior, respectively. For the 4×4 *Mini Lights Out* game the neighborhood will always consist of exactly five buttons.

Since pushing a button twice is the same as not pushing the button at all, we can concentrate on solutions that "use" each button at most once. Also, the order the buttons are pushed does not matter. Thus we can represent any strategy by a 16×1 column vector x where each component is 1 if that button is to be pushed and 0 otherwise. In particular, the button in row i and column j is represented by the component $4(i-1)+j$.

In a similar fashion we can represent any configuration of the game by a 16×1 column vector b where each component is 1 if that button is lit and 0 otherwise.

Furthermore, any move of the game (i.e. pushing the ith button) can be represented by a 16×1 column vector v_i consisting of 1s for each button in the neighborhood and 0s elsewhere. Consider the 16×16 matrix $A = [v_1 \mid v_2 \mid \cdots \mid v_{16}]$. This matrix is most easily given by

$$A = \begin{bmatrix} B & I & O & I \\ I & B & I & O \\ O & I & B & I \\ I & O & I & B \end{bmatrix},$$

where

$$B = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \; I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{ and } O = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

One can check that the result of applying a strategy x to an originally unlit board is simply the vector Ax. Also, due to the binary nature of the set-up, the strategy for turning off a given set of lights is identical to the strategy for turning those same lights on from an unlit board. Therefore, given an initial configuration b, our goal is to find a strategy vector x so that $Ax \equiv b \pmod 2$. This is generally done using Gauss-Jordan elimination, mod 2. However, we note that for our matrices, $B^2 \equiv I_4 \pmod 2$ and thus, $A^2 \equiv I_{16} \pmod 2$, where I_n is the $n \times n$ identity matrix. Hence, $x \equiv Ix \equiv A^2x \equiv Ab \pmod 2$ is the unique solution. Furthermore, this guarantees that every initial configuration has a winning strategy.

We note that because of the specific size of this puzzle, the neighborhoods of two distinct buttons will always intersect in either 0 or 2 buttons. Call two buttons *disjoint* if their neighborhoods are disjoint.

Define the *count* of a button to be the number of buttons in its neighborhood that are currently turned on. Furthermore, let the *parity* of a button be the parity of its count. That is, the parity of a button is 1 if an odd number of lights in its neighborhood are currently turned on, and 0 if an even number of lights are currently turned on.

Suppose a button has a current count of n. Pushing that button will change its count to $5-n$, and thus change its parity. Pushing a disjoint button will not change the original button's count or its parity. Pushing a non-disjoint button will make the original button's count either $n+2$, $n-2$, or n, depending on whether the buttons common to both neighborhoods were, respectively, originally both off, both on, or one on and one off. Thus the original button's parity will not be changed by pushing a non-disjoint button. This leads to the following result.

THEOREM 1. *The parity of a button is changed if and only if that button is pushed.*

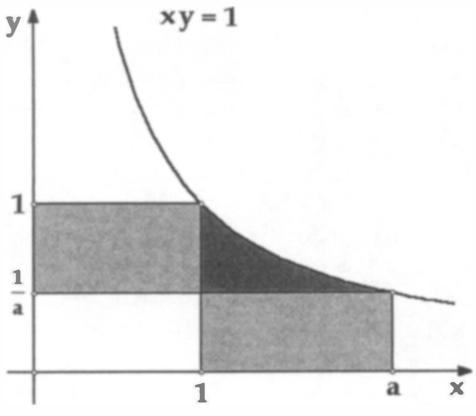
Combining Theorem 1 and the fact that every game is winnable, we get an easily implemented method to solve the *Mini Lights Out* game. Clearly, at the end of the game each button must have parity 0. By Theorem 1, this parity can only be changed by pushing that button. Thus we merely push those buttons whose parity is one.

Acknowledgment. We wish to thank the referees for their comments, which helped clarify our explanation. Jennie Missignman is an undergraduate student at Lycoming College.

REFERENCES

1. Marlow Anderson and Todd Feil, Turning lights out with linear algebra, this MAGAZINE **71** (1998), 300–303.
2. Donald H. Pelletier, Merlin's Magic Square, *Amer. Math. Monthly* **94** (1987), 143–150.
3. Daniel L. Stock, Merlin's Magic Square revisited, *Amer. Math. Monthly* **96** (1989), 608–610.

Proof Without Words: Logarithm of a Number and Its Reciprocal



$$\int_{1/a}^{1} \frac{1}{y}\,dy = \int_{1}^{a} \frac{1}{x}\,dx, \qquad a > 0$$

$$-\ln\left(\frac{1}{a}\right) = \ln a$$

—VINCENT FERLINI
KEENE STATE COLLEGE
KEENE, NH 03435

The Classification of Groups of Order $2p$

JOSEPH A. GALLIAN
University of Minnesota Duluth
Duluth, MN 55812

In [**3**] Sinefakopoulos uses only normality and a counting lemma to prove that every group of order p^2 where p is a prime is abelian. From this, the classification of groups of order p^2 is immediate: If $|G| = p^2$, then $G \approx \mathbb{Z}_{p^2}$ or $G \approx \mathbb{Z}_p \oplus \mathbb{Z}_p$. In most books this result is given as a corollary of the class equation and occurs late in the text (see [**1**, p. 404] and [**2**, p. 479]). Unfortunately, many undergraduate courses in abstract algebra never get to the latter chapters of the book. This is definitely true at schools that offer only one semester of abstract algebra. Another particularly nice theorem that falls into this category is the classification of groups of order $2p$ where p is an odd prime. This result is typically given as an application of Cauchy's theorem ("If a prime p divides $|G|$, then G has an element of order p") or is not stated at all. In this note we give a proof of this theorem as an application of the most fundamental of all facts about finite groups, Lagrange's Theorem.

THEOREM 2. *Let G be a group of order $2p$ where p is a prime greater than 2. Then G is isomorphic to the cycle group $\mathbb{Z}_{2p}$ or the dihedral group D_p.*

Proof. Under the assumption that G does not have an element of order $2p$ we will show that $G \approx D_p$. We begin by first showing that G must have an element of order p. By our assumption and Lagrange's Theorem, any nonidentity element of G must have order 2 or p. To prove that G has an element of order p it suffices to show that not every nonidentity element of G has order 2. If this were the case, then for all a and b in the group $ab = (ab)^{-1} = b^{-1}a^{-1} = ba$ so that G would be abelian. Then for any non-identity elements $a, b \in G$ with $a \neq b$, the set $\{e, a, b, ab\}$ would be closed and therefore a subgroup of G of order 4. Since this contradicts Lagrange's Theorem, we conclude that G must have an element of order p; call it a.

Now let b be any element not in $\langle a \rangle$, the cyclic subgroup generated by a. Then $b\langle a \rangle \neq \langle a \rangle$ and $G = \langle a \rangle \cup b\langle a \rangle$ since a group of order $2p$ has only two distinct cosets with p elements. We next claim that $|b| = 2$. To see this, observe that since $\langle a \rangle$ and $b\langle a \rangle$ are the only two distinct cosets of $\langle a \rangle$ in G we must have $b^2\langle a \rangle = \langle a \rangle$ or $b^2\langle a \rangle = b\langle a \rangle$. We may rule out $b^2\langle a \rangle = b\langle a \rangle$ for then b^2 can be written in the form ba^i so that $b = a^i \in \langle a \rangle$, a contradiction. On the other hand, $b^2\langle a \rangle = \langle a \rangle$ implies that $b^2 \in \langle a \rangle$ and therefore $|b^2| = p$ or $|b^2| = 1$. But $|b^2| = p$ implies that $|b| = 2p$, which we have ruled out, or $|b| = p$. However, if both $|b^2| = p$ and $|b| = p$, then $\langle b^2 \rangle = \langle b \rangle$ and therefore $b \in \langle b^2 \rangle \subseteq \langle a \rangle$, which is not true. Thus, any element of G not in $\langle a \rangle$ has order 2.

Next consider ab. Since $ab \notin \langle a \rangle$ our argument above shows that $|ab| = 2$. Then $ab = (ab)^{-1} = b^{-1}a^{-1} = ba^{-1}$. Moreover, because every element of G has the form a^i or ba^i the relation $ab = ba^{-1}$ completely determines the multiplication table for G. (For example, $a^3(ba^4) = a^2(ab)a^4 = a^2(ba^{-1})a^4 = a(ab)a^3 = a(ba^{-1})a^3 = (ab)a^2 = (ba^{-1})a^2 = ba$.) And since the multiplication table for all non-cyclic groups of order $2p$ is uniquely determined, all non-cyclic groups of order $2p$ must be isomorphic to each other. But of course, D_p, the dihedral group of order $2p$, is one such group. ∎

As an immediate corollary we have that S_3, the symmetric group of degree 3, is isomorphic to D_3.

Acknowledgment. I wish to thank the referees for their comments.

REFERENCES

1. Joseph A. Gallian, *Contemporary Abstract Algebra, 4th ed.*, Houghton Mifflin, Boston, MA 1998.
2. W. Keith Nicholson, *Introduction to Abstract Algebra*, PWS-Kent, Boston, MA 1993.
3. Achilleas Sinefakopoulos, On groups of order p^2, this MAGAZINE **70** (1997), 212–213.

A Characterization of Infinite Cyclic Groups

CHARLES LANSKI
University of Southern California
Los Angeles, CA 90089-1113

Introduction We prove an interesting and nontrivial characterization of infinite cyclic groups, using only basic notions about groups: orders of elements, cyclic groups, cosets, commutators, and quotients.

THEOREM 1. *An infinite group is cyclic when each of its nonidentity subgroups has finite index.*

Why should we expect the number of cosets of subgroups to be crucial in determining the structure of infinite groups? An observation about cosets of subgroups of cyclic groups will motivate the theorem, but first we recall the definition of a cyclic group. For $\mathbb{Z} = (\mathbb{Z}, +)$ the group of integers under addition, G any group, and $g \in G$, let $\langle g \rangle = \{g^k \in G \mid k \in \mathbb{Z}\}$ denote the cyclic subgroup of G generated by g. A group G is called *cyclic* when $G = \langle g \rangle$ for some $g \in G$.

An important property of a cyclic group $G = \langle g \rangle$ is that each of its subgroups is cyclic and has the form $\langle g^m \rangle$ ([**2**, p. 59] or [**3**, p. 75]). The subgroups of $(\mathbb{Z}, +)$ are the $n\mathbb{Z} = \{nk \in \mathbb{Z} \mid k \in \mathbb{Z}\} = \langle n \rangle$ (the group operation is addition!) for all nonnegative integers n. When $n \geq 1$ the cosets of $n\mathbb{Z}$ in $\mathbb{Z}$, giving the integers modulo n, are $n\mathbb{Z}$, $n\mathbb{Z} + 1$, $n\mathbb{Z} + 2, \ldots,$ and $n\mathbb{Z} + (n-1)$, so $n\mathbb{Z}$ has exactly n distinct cosets in $\mathbb{Z}$. This simple but key observation is a special case of a result about cosets of subgroups in an arbitrary cyclic group which we state below.

A test to show groups are not cyclic When the collection $\{Hg \mid g \in G\}$ of all right cosets of a subgroup H in a group G is finite, we let $[G : H]$ denote the number of distinct cosets and call $[G : H]$ the *index* of H in G.

PROPOSITION. *Every nonidentity subgroup of a cyclic group has finite index.*

Proof. Let the cyclic group be $G = \langle g \rangle$ and let a nonidentity subgroup be $H = \langle g^m \rangle$ for some $m \geq 1$. We claim that $H = He, Hg, \ldots, Hg^{m-1}$ are all of the right cosets of H in G. Any right coset of H in G is Hg^t for some $t \in \mathbb{Z}$. Use the division algorithm in the integers to write $t = qm + r$ with $0 \leq r < m$. Since $(g^m)^z \in H$ for all $z \in \mathbb{Z}$, it follows that $Hg^t = H(g^m)^q g^r = Hg^r$. ■

The fact that every nonidentity subgroup of a cyclic group has finite index is not very surprising, especially for finite groups! However, we can use the Proposition to show that certain groups are not cyclic. Our examples and some later arguments require this important fact: for any group G and subgroup H, $Hx = Hy$ exactly when $xy^{-1} \in H$ ([**2**, p. 81] or [**3**, p. 133]). An easy but useful application is that $Hx = H = He$ exactly when $x \in H$.

Example 1. $G = \mathbb{Z} \oplus \mathbb{Z}$ is not cyclic.

Consider the subgroup $H = \mathbb{Z} \oplus \langle 0 \rangle$ and the collection $\{H + (0, z) \mid z \in \mathbb{Z}\}$ of right cosets. If two of these cosets were equal, say $H + (0, z) = H + (0, z')$, for some $z, z' \in \mathbb{Z}$ then we would have $(0, z) - (0, z') = (0, z - z') \in H$ forcing $z = z'$. Therefore different elements of $\mathbb{Z}$ give rise to different cosets $H + (0, z)$ of H in G. Since there are infinitely many cosets and $H \neq \langle e_G \rangle$, we must conclude from the Proposition that G is not cyclic.

Example 2. $(\mathbb{Q}, +)$ is not cyclic.

Consider the cosets $\{\mathbb{Z} + 1/p \mid p \text{ is a prime}\}$ of the subgroup $\mathbb{Z}$ of $\mathbb{Q}$. If $\mathbb{Z} + 1/p = \mathbb{Z} + 1/q$ for different primes p and q, then we must have $1/p - 1/q = (q - p)/pq \in \mathbb{Z}$. Thus $q - p = pqz$ for some $z \in \mathbb{Z}$. This is impossible, since then each prime would divide the other. Hence there are infinitely many cosets of $\mathbb{Z}$ in $(\mathbb{Q}, +)$, at least one for each prime, so as in Example 1 $(\mathbb{Q}, +)$ cannot be cyclic by the Proposition.

Exercise. Show that the group $(\mathbb{Q}^+, \cdot)$ of positive rationals under multiplication is not cyclic.

Exercise. Show that the subgroup $G = \{(a, b, c) \in \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z} \mid a + b + c = 0\}$ of $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$ is not cyclic.

About the theorem and its proof The converse of the Proposition for infinite groups is THEOREM 1. Of course the converse of a true statement need not be true, so why should we expect THEOREM 1 to be true? A counterexample does not seem easily found, and while this is not exactly convincing evidence for the truth of the assertion, it does make it more reasonable. In fact, Y. Fedorov [**1**] proved THEOREM 1 in 1951, but his result is not widely known. Although the statement of THEOREM 1 is quite elementary, its proof is not obvious, at least by elementary methods. The utility of the result itself is questionable since it is hard to see how we would know that each subgroup of a given infinite group has finite index. However, the result is appealing because of its simplicity and the definitive nature of the characterization it gives. Aside from this, it provides a nice example of how basic notions about groups, which are covered in a first abstract algebra course, can be combined to prove a nonstandard and pretty result about infinite groups.

Fedorov's theorem is not commonly found in textbooks, but a proof appears in W. R. Scott's Group Theory [**5**, p. 446], using rather sophisticated results, including the fundamental theorem of abelian groups, the transfer map, and results on FC-groups. Our goal is to present a proof using only basic facts and ideas about groups.

We will recall the relevant definitions as they are needed, but start with a description of our approach. There are two main steps. First we show that THEOREM 1 holds when G is abelian, and then observe that G must be "almost abelian": its center has finite index. Using this, the second step proves and applies an interesting result of I. Schur, which shows that the commutator subgroup of G must be finite. This will result in a contradiction when G is not abelian. Let us begin the argument.

Throughout the proof, we will assume that G is an infinite group and if H is any nonidentity subgroup of G then $[G : H]$ is finite. We write $H \leq G$ when H is a subgroup of G, and when the order of $g \in G$ is finite we denote it by $o(g)$.

The proof for abelian groups Our first lemma concerns orders of elements and intersections of cyclic subgroups in G.

LEMMA 1.

(i) If $H \leq G$, then either $H = \langle e \rangle$ or H is infinite.

(ii) If $g \in G \setminus \{e\}$, then g has infinite order.

(iii) If $x, y \in G \setminus \{e\}$, then $\langle x \rangle \cap \langle y \rangle = \langle x^a \rangle = \langle y^b \rangle$ for some $a, b \in \mathbb{Z} \setminus \{0\}$.

Proof. For (i), if $H \neq \langle e \rangle$ then by our basic assumption $[G : H]$ is finite. Thus G is the union of finitely many right cosets of H, say $G = Hg_1 \cup \cdots \cup Hg_k$. If H is finite then each of these Hg_i has the same number of elements as H, as in the proof of Lagrange's theorem (see [**2**], [**3**], or [**4**]), and this forces G to be finite. Hence, H must be infinite. Now (i) implies (ii) since $H = \langle g \rangle \leq G$, and $\langle g \rangle$ is finite when $o(g)$ is finite ([**2**, p. 56] or [**3**, p. 72]). In (iii), the set $\{\langle x \rangle y^i \mid i \in \mathbb{Z}\}$ of cosets of $\langle x \rangle$ in G must be finite, so $\langle x \rangle y^i = \langle x \rangle y^j$ for some $i > j$. This forces $y^{i-j} = y^i (y^j)^{-1} \in \langle x \rangle$ and shows that $y^{i-j} \in \langle x \rangle \cap \langle y \rangle$. Now $y^{i-j} \neq e$ since by (ii) $y \in G \backslash \langle e \rangle$ has infinite order. Therefore $\langle x \rangle \cap \langle y \rangle$ is a nonidentity subgroup of both $\langle x \rangle$ and $\langle y \rangle$, so $\langle x^a \rangle = \langle x \rangle \cap \langle y \rangle = \langle y^b \rangle$ with both $a, b \in \mathbb{Z} \setminus \{0\}$ ([**2**, p. 59] or [**3**, p. 75]). ■

We need another lemma about indices of subgroups in order to prove THEOREM 1 for abelian groups. The lemma is a consequence of the multiplicative property of indices which is easy to prove for finite groups by using Lagrange's theorem ([**2**, p. 91], [**3**, p. 143] or [**4**, p. 41]).

LEMMA 2. *In any group A, let $H \leq K \leq A$ with $[A : H]$ finite. If $H \neq K$ then $[A : K] < [A : H]$.*

Proof. Let $Ha_1, \ldots, Ha_m$ be the distinct right cosets of H in A, so $A = Ha_1 \cup \cdots \cup Ha_m$. We may assume that $a_1 = e$. Now $Ha_j \subseteq Ka_j$ for each j implies that $A = Ka_1 \cup \cdots \cup Ka_m$. It follows that for any $g \in G$, $g \in Ka_j$ for some j, and so $Kg = Ka_j$. Thus $\{Ka_i\}$ contains all the right cosets of K in G. We want to show that $Ka_i = Ka_j$ for $i \neq j$. There is some $k \in K \backslash H$, and since $k \in A$, $k \in Ha_s$ for some $s > 1$. Therefore $a_s \in Hk \subseteq K$, so $Ka_s = K = Ke = Ka_1$ and $[A : K] < m = [A : H]$. ■

LEMMA 3. *If G as above is an abelian group then G is cyclic.*

Proof. Let G be an abelian group. If $h \in G \backslash \langle e \rangle$ then $[G : \langle h \rangle]$ is finite, so there must be $x \in G$ for which $[G : \langle x \rangle]$ is minimal. If $\langle x \rangle = G$ the proof is finished, so assume that $y \in G \backslash \langle x \rangle$. From Lemma 1 we have that $x^s = y^k$ for some $s, k \in \mathbb{Z} \setminus \{0\}$. Set $\gcd(s, k) = d$ and suppose that $d > 1$. Write $k = da$, $s = db$, and use $xy = yx$ to get $(y^a x^{-b})^d = y^{ad} x^{-bd} = y^k x^{-s} = e$. Any nonidentity element of G has infinite order by Lemma 1, forcing $y^a = x^b$. However, $gcd(a, b) = 1$, so we may assume from the start that $d = 1$, and therefore that $1 = uk + vs$ for some $u, v \in \mathbb{Z}$ (see [**2**], [**3**], or [**4**]). Now set $g = y^v x^u$. Using again the fact that $xy = yx$, we compute that $g^k = y^{vk} x^{uk} = (y^k)^v x^{uk} = x^{sv} x^{uk} = x^{vs+uk} = x$. Similarly $g^s = y^{vs} x^{us} = y^{vs} (x^s)^u = y^{vs} y^{ku} = y^{vs+uk} = y$. These computations show that $x, y \in \langle g \rangle$; thus $\langle x \rangle \leq \langle g \rangle$ but $\langle x \rangle \neq \langle g \rangle$. By Lemma 2 $[G : \langle g \rangle] < [G : \langle x \rangle]$, contradicting the minimality of $[G : \langle x \rangle]$. Therefore, $y \in G \backslash \langle x \rangle$ is impossible, so $G = \langle x \rangle$ is cyclic. ■

To prepare for the final step in the proof of THEOREM 1 we need to see that G is not far from being abelian, in the sense that its center has finite index. Recall that the *center* of G is $Z(G) = \{x \in G \mid xg = gx \text{ for all } g \in G\}$. It is straightforward to prove that $\langle e\rangle \le Z(G) \le G$.

LEMMA 4. *$Z(G) \neq \langle e\rangle$, so $[G : Z(G)]$ is finite.*

Proof. Let $g \in G\backslash\langle e\rangle$. Our basic assumption shows that $G = \langle g\rangle x_1 \cup \cdots \cup \langle g\rangle x_k$ for some $x_1 = e, x_2, \ldots, x_k \in G$. By Lemma 1, for each $1 \le i \le k$ there is some positive power $m(i)$ of g with $g^{m(i)} \in \langle x_i\rangle$. Set $m = m(1)\ldots m(k)$ and write $m = m(i)c(i)$. Then $g^m = (g^{m(i)})^{c(i)} \in \langle x_i\rangle$ for all x_i, which forces $g^m x_i = x_i g^m$. But G is the union of the various $\langle g\rangle x_j$, so for any $h \in G$, $h \in \langle g\rangle x_i$ for some i. Thus $h = g^t x_i$ for some $t \in \mathbb{Z}$ and now $g^m h = hg^m$, so we must have $g^m \in Z(G)$. In view of Lemma 1, $g^m \neq e$. Therefore $Z(G) \neq \langle e\rangle$ and it follows that $[G : Z(G)]$ is finite. ■

A theorem of Schur Suppose we could show that Lemma 4 forces G to have a finite nonidentity subgroup when G is not abelian. Since G cannot have a finite nonidentity subgroup by Lemma 1, it must be abelian, so applying Lemma 3 would finish the proof. The most direct way to use Lemma 4 to produce a finite subgroup of G is to cite a result of I. Schur [**5**, p. 443] involving the *commutator* G' of G, i.e., the subgroup of G generated by all $x^{-1}y^{-1}xy$ for $x, y \in G$. Thus G' is the collection of all finite products of the $x^{-1}y^{-1}xy$ and their inverses. Standard facts about G' are that it is normal in G and that the quotient group G/G' is abelian ([**2**, p. 171] or [**4**, p. 65]).

Schur's result states that G' is finite when $[G : Z(G)]$ is finite. This pretty theorem is fairly well known, but it is not found in the standard introductory algebra texts, or even in many group theory texts. Its proof in [**5**] uses results on FC-groups. We provide an elementary and standard computational proof.

THEOREM 2 (SCHUR). *In any group A, if $[A : Z(A)] = k$ is finite then the commutator A' is finite.*

Proof. The proof is in 3 steps.

Step 1. There are only finitely many simple commutators $g^{-1}h^{-1}gh$ with $g, h \in A$. Let the distinct cosets of $Z(A)$ in A be $\{Z(A)a_1, \ldots, Z(A)a_k\}$. Then any $g \in A$ may be written $g = za_j$ for some $z \in Z(A)$ and some a_j. It follows from this and the definition of $Z(A)$ that, for $g, h \in A$ and some i and j, the simple commutator $g^{-1}h^{-1}gh = z_1^{-1}a_i^{-1}z_2^{-1}a_j^{-1}z_1a_iz_2a_j = a_i^{-1}a_j^{-1}a_ia_j$, and so there are only finitely many such.

Observe that $y^{-1}a_i^{-1}a_j^{-1}a_ia_jy = (y^{-1}a_iy)^{-1}(y^{-1}a_jy)^{-1}(y^{-1}a_iy)(y^{-1}a_jy)$, so every conjugate of a simple commutator is a simple commutator. Also note that, since $Z(A)$ a normal subgroup of A of index k, the quotient $A/Z(A)$ is a group of order k, so $g^k \in Z(A)$ for each $g \in A$ ([**2**, p. 91], [**3**, p. 136], or [**4**, p. 43]).

Step 2. For $g, h \in A$ and $n \ge 1$, $(g^{-1}h^{-1}gh)^n = ((hg)^{-1})^n(gh)^nc_1c_2\ldots c_{n-1}$, where each c_j is a simple commutator. If $n = 1$ then $g^{-1}h^{-1}gh = (hg)^{-1}gh$ as required. Assume the claim holds for the exponent $n \ge 1$. Then

$$\begin{aligned}(g^{-1}h^{-1}gh)^{n+1} &= (g^{-1}h^{-1}gh)^n(g^{-1}h^{-1}gh)\\ &= ((hg)^{-1})^n(gh)^nc_1c_2\cdots c_{n-1}(hg)^{-1}(gh)\\ &= ((hg)^{-1})^{n+1}(gh)^{n+1}(gh)^{-n-1}(hg)(gh)^nc_1\cdots c_{n-1}(hg)^{-1}(gh)\\ &= ((hg)^{-1})^{n+1}(gh)^{n+1}((gh)^{-1}((gh)^{-n}(hg)(gh)^n(hg)^{-1})(gh))\cdot\\ &\quad y^{-1}c_1yy^{-1}c_2y\cdots y^{-1}c_{n-1}y,\end{aligned}$$

where $y = (hg)^{-1}(gh)$. By our observation just above, each $y^{-1}c_jy = d_{j+1}$ is a simple commutator, $(gh)^{-n}(hg)(gh)^n(hg)^{-1}$ is a simple commutator by definition, and its conjugate by gh, namely $d_1 = (gh)^{-1}((gh)^{-n}(hg)(gh)^n(hg)^{-1})(gh)$, is a simple commutator. Therefore we can write $(g^{-1}h^{-1}gh)^{n+1} = ((hg)^{-1})^{n+1}(gh)^{n+1}d_1d_2\dots d_n$ with each d_i a simple commutator, and the claim holds by induction.

Using Step 1, assume that there are exactly m simple commutators.

Step 3. Any product of simple commutators is equal to a product of at most $m(k-1)$ simple commutators. Consider any product of $s \geq m(k-1)+1$ simple commutators. Since there are only m different simple commutators, one of them, say c, must appear at least k times in the product. Write the product as $g_1cg_2c\cdots g_kcg_{k+1}$ where each g_j is a product of simple commutators or is the identity of A, and the number of simple commutators appearing in all the g_j is $s-k$, including multiplicity. Now write

$$g_1cg_2c\cdots g_kcg_{k+1} = c^k(c^{-k}g_1c^k)(c^{-(k-1)}g_2c^{k-1})\cdots(c^{-2}g_{k-1}c^2)(c^{-1}g_kc)g_{k+1},$$

and observe that if g_j is a product of t simple commutators, then so is $c^{-i}g_jc^i$, using $y^{-1}uv\cdots wy = y^{-1}uyy^{-1}vy\cdots y^{-1}wy$ and our observation that a conjugate of a simple commutator is a simple commutator. Thus, we have $g_1cg_2c\cdots g_kcg_{k+1} = c^kh_1h_2\cdots h_{s-k}$ with each h_j a simple commutator. If $c = b^{-1}a^{-1}ba$, use Step 2 to rewrite

$$g_1cg_2c\cdots g_kcg_{k+1} = ((ab)^{-1})^k(ba)^kd_1\cdots d_{k-1}h_1\cdots h_{s-k},$$

with all d_j and h_j simple commutators. Since $(ba)^k \in Z(A)$, as we saw above, it follows that $(ba)^k = b^{-1}(ba)^kb = (ab)^k$, so $((ab)^{-1})^k(ba)^k = e$. Consequently, $g_1cg_2c\cdots g_kcg_{k+1} = d_1\cdots d_{k-1}h_1\cdots h_{s-k}$ is a product of $s-1$ simple commutators. This argument may be repeated as long as we have more than $m(k-1)$ simple commutators, and so produces a product of at most $m(k-1)$ simple commutators which is equal to the product of the s simple commutators to start with.

Since $(g^{-1}h^{-1}gh)^{-1} = h^{-1}g^{-1}hg$, the inverse of a simple commutator is again a simple commutator. Therefore, the elements in A' are just all finite products of simple commutators. Using Step 3, there are only finitely many such products since all the products of at most $m(k-1)$ of the m simple commutators will give all possible finite products of simple commutators, completing the proof. ∎

The proof of Theorem 1 We put the preceding pieces together to complete the proof of THEOREM 1. Using Lemma 4 we see that $[G : Z(G)] = k$ is finite, and now Schur's theorem yields that G' is finite. But the only finite subgroup of G is $\langle e\rangle$ by Lemma 1, so $G' = \langle e\rangle$. Thus, for all $g, h \in G$, $g^{-1}h^{-1}gh = e$, or equivalently $gh = hg$, and G is abelian. We conclude that G is cyclic from Lemma 3.

REFERENCES

1. Y. Fedorov, On infinite groups of which all nontrivial subgroups have a finite index, *Uspekhi Mat. Nauk.* 6 (1951), 187–189.
2. D. S. Dummit and R.M. Foote, *Abstract Algebra,* 2nd edition, Prentice-Hall, Upper Saddle River, NJ, 1999.
3. J. Gallian, *Contemporary Abstract Algebra*, 4th edition, Houghton Mifflin, New York, NY, 1998.
4. I. N. Herstein, *Topics in Algebra*, 2nd edition, Xerox College Publishing, Toronto, Ontario, Canada, 1975.
5. W. R. Scott, *Group Theory*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1964.

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

Assistant Editors: RAZVAN GELCA, Texas Tech University; ROBERT GREGORAC, Iowa State University; GERALD HEUER, Concordia College; PAUL ZEITZ, The University of San Francisco

Proposals

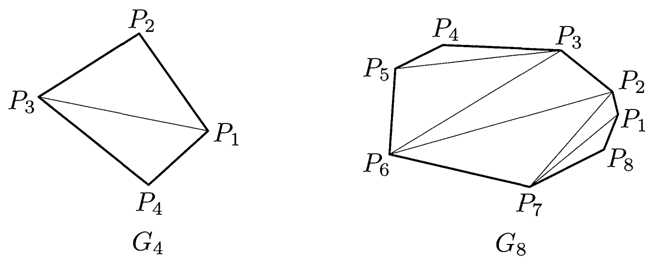
To be considered for publication, solutions should be received by July 1, 2001.

1613. *Proposed by Emeric Deutsch, Polytechnic University, Brooklyn, NY.*

Given a convex n-gon $P_1P_2\cdots P_n$, let G_n be the graph obtained as a result of tracing the zig-zag polygonal path

$$P_1P_{n-1}P_2P_{n-2}P_3P_{n-3}\cdots P_m,$$

where $m=(n-1)/2$ if n is odd and $m=(n+2)/2$ if n is even. (In the figure below we show G_4 and G_8.) For $n\geq 3$, let a_n be the number of sets of non-adjacent edges from G_n. Find a recurrence relation satisfied by the a_n's. (As an example, $a_4=8$ because in addition to the empty set, G_4 has 5 sets consisting of one edge each and 2 sets consisting of two non-adjacent edges.)



1614. *Proposed by Achilleas Sinefakopoulos, student, University of Athens, Athens, Greece.*

Determine the minimum values of each of $x+y-xy$ and $x+y+xy$, where x and y are positive real numbers such that $(x+y-xy)(x+y+xy)=xy$.

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.

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 ehjohnst@iastate.edu. All communications should include the readers name, full address, and an e-mail address and/or FAX number.

1615. *Proposed by M. N. Deshpande, Nagpur, India.*

An urn contains 3 white balls and $2n-3$ black balls, with $n \geq 2$. Balls are drawn from the urn two at a time, at random and without replacement. At each stage, the two balls drawn are inspected. If they are of the same color, they are set aside and two more balls are drawn. This process continues until a drawn pair consists of two balls of different color, after which the process stops. Let E_n denote the expected number of balls drawn before the process is stopped. Prove that $(2n-1)E_n$ is a perfect square.

1616. *Proposed by Erwin Just and Norman Schaumberger, Bronx Community College, NY.*

Let $k \geq 2$ be a positive integer, and let S_k be the set of all numbers of the form $\sum_{j=1}^{k} \frac{a_j}{a_j+b_j}$, where $a_j,\ b_j > 0$ for $1 \leq j \leq k$ and $\sum_{j=1}^k a_j = \sum_{j=1}^k b_j$. Determine the greatest lower bound and the least upper bound of S_k.

1617. *Proposed by Zhang Yun, First Middle School of Jin Chang City, Gan Su Province, China.*

Let $A_1A_2A_3A_4$ be a cyclic quadrilateral that also has an inscribed circle. Let B_1, B_2, B_3, and B_4, respectively, be the points on sides A_1A_2, A_2A_3, A_3A_4, and A_4A_1 at which the inscribed circle is tangent to the quadrilateral. Prove that

$$\left(\frac{A_1A_2}{B_1B_2}\right)^2 + \left(\frac{A_2A_3}{B_2B_3}\right)^2 + \left(\frac{A_3A_4}{B_3B_4}\right)^2 + \left(\frac{A_4A_1}{B_4B_1}\right)^2 \geq 8.$$

Quickies

Answers to the Quickies are on page 71.

Q907. *Proposed by Martin Gardner, Hendersonville, NC.*

Each of the letters A, D, E, I, N, O, P, R, S is written on a card. The nine cards are placed face up on a table so each card is visible. Beside the cards is the list of words: ASP, AID, SIN, RIO, ARE, PIE, END, POD. Two players play a game in which they take turns selecting a card without replacement. The first player to get the letters needed to form a word from the list is the winner. If both players play optimally, what is the outcome of the game?

Q908. *Proposed by Norman Schaumberger, Bronx Community College, NY.*

Given that $a_n > a_{n-1} > \cdots > a_0 = 0$, show that

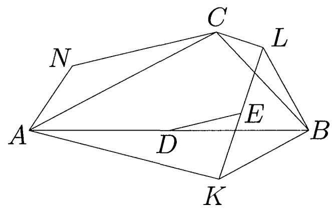
$$\frac{1^2}{a_1} + \frac{2^2}{a_2} + \cdots + \frac{n^2}{a_n} \leq \frac{n}{a_1 - a_0} + \frac{n-1}{a_2 - a_1} + \cdots + \frac{1}{a_n - a_{n-1}}.$$

Solutions

Similar Triangles Built on a Triangle **February 2000**

1589. *Proposed by Ayoub B. Ayoub, Pennsylvania State University, Abington, PA.*

On the sides of $\triangle ABC$, three similar triangles, AKB, BLC, and CNA, are drawn outward. If AB and KL are bisected by D and E respectively, prove that that DE is parallel to NC and determine DE/NC.



Solution by Hans Kappus, Rodersdorf, Switzerland.

Without loss of generality, place $\triangle ABC$ in the complex plane so that $A = 0$, $B = 1$, and $\mathrm{Re}(C) > 0$. If $\angle BAK = \phi$ and $AK = k$, $(k > 0)$ then $K = ke^{-i\phi}$ and $L = 1 + \lambda(C-1)e^{-i\phi}$ for some $\lambda \in \mathbb{R}$. Now $LB : BC = KA : AB$, so that $\lambda = k$, and hence $L = 1 + k(C-1)e^{-i\phi}$. Similarly, $N = C - kCe^{-i\phi}$.

Because $D = 1/2$ and $E = (K+L)/2 = (1 + kCe^{-i\phi})/2$, we have

$$E - D = \frac{1}{2}kCe^{-i\phi} = \frac{1}{2}(C - N).$$

This proves that DE is parallel to NC, and $DE/NC = 1/2$.

Comments: A few readers observed that the result remains true if the three triangles are drawn inward or if they are degenerate. The solution presented here can be modified slightly to apply to these additional cases.

Also solved by Michel Bataille (France), J. C. Binz (Switzerland), Gerald D. Brown, Geoffrey A. Kandall, Neela Lakshmanan, Nick Lord (England), Stephen Noltie, Duong Nguyen, Richard E. Pfiefer, Volkhard Schindler (Germany), Raul A. Simon (Chile), Achilleas Sinefakopoulos (Greece), Raymond E. Spaulding, R. S. Tiberio, Hermann Vollath (Germany), Robert L. Young, Tom Zerger, Li Zhou, and the proposer.

Bounding Sine by Power Functions **February 2000**

1590. *Proposed by Constantin P. Niculescu, University of Craiova, Craiova, Romania.*

For given a, $0 < a \le \pi/2$, determine the minimum value of $\alpha \ge 0$ and the maximum value of $\beta \ge 0$ for which

$$\left(\frac{x}{a}\right)^{\alpha} \le \frac{\sin x}{\sin a} \le \left(\frac{x}{a}\right)^{\beta}$$

holds for $0 < x \le a$.

(This generalizes the well known inequality due to Jordan, which asserts that $2x/\pi \le \sin x \le 1$ on $[0, \pi/2]$.)

Solution by Kee-Wai Lau, Hong Kong, China.

We show that the minimum value of α is 1 and the maximum value of β is $a \cot a$. For $0 < x < a \le \pi/2$, the desired equality is equivalent to

$$\beta \le \frac{\ln \sin x - \ln \sin a}{\ln x - \ln a} \le \alpha.$$

By taking the limit as $x \to 0^+$ and $x \to a^-$ we obtain, respectively,

$$\alpha \geq 1 \qquad \text{and} \qquad \beta \leq a \cot a. \tag{1}$$

Let $f(x) = (\sin x)/x$ so that

$$f'(x) = \frac{\sin x}{x^2}(x \cot x - 1) < 0$$

for $0 < x \leq a$. It follows that on this interval $f(x) \geq f(a)$, and so

$$\frac{\sin x}{\sin a} \geq \frac{x}{a}.$$

Thus the minimum value of α does not exceed 1. Hence by (1) the minimum value of α is 1.

Now let $g(x) = (\sin x)/x^k$ and $h(x) = x \cot x - k$, where $k = a \cot a$. For $0 < x \leq a$ we have

$$g'(x) = x^{-k-1} h(x) \sin x \qquad \text{and} \qquad h'(x) = \frac{\sin x \cos x - x}{\sin^2 x} < 0.$$

Hence $h(x) \geq h(a) = 0$, so $g'(x) \geq 0$, and $g(x) \leq g(a)$. Thus,

$$\frac{\sin x}{\sin a} \leq \left(\frac{x}{a}\right)^k,$$

showing that the maximum value of β is at least k. Combining with (1) we conclude that the maximum value of β is $k = a \cot a$.

Also solved by Michel Bataille (France), Paul Bracken (Quebec), Mordechai Falkowityz (Ontario), Marty Getz and Dixon Jones, Nick Lord (England), Duong Nguyen, Stephen Noltie, Heinz-Jurgen Seifert, Volkhard Schindler (Germany), Akalu Tefera, Li Zhou, and the proposer.

3 × 3 Matrices of Triangular Triples **February 2000**

1591. *Proposed by Western Maryland College Problems Group, Westminster, MD.*

We call a 3-tuple (a, b, c) of positive integers a *triangular triple* if

$$T_a + T_b = T_c,$$

where $T_n = n(n+1)/2$ is the nth triangular number. Given an integer k, prove that there are infinitely many triples of distinct triangular triples, (a_i, b_i, c_i) for $i = 1, 2, 3$, such that

$$\det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = k^3.$$

Solution by Mark Kidwell, U.S. Naval Academy, Annapolis, MD.

We prove the stronger statement that any integer can be achieved in infinitely many ways as the value of such a determinant.

We use the two identities

$$T_{3m+2} + T_{4m+2} = T_{5m+3}, \qquad m = 0, 1, 2, 3, \ldots$$
$$T_{3n} + T_{4n+1} = T_{5n+1}, \qquad n = 1, 2, 3, \ldots.$$

These identities can be verified by elementary algebra.

Next observe that for positive integers m and n,

$$\det\begin{pmatrix} 2 & 2 & 3 \\ 3m+2 & 4m+2 & 5m+3 \\ 3n & 4n+1 & 5n+1 \end{pmatrix} = m.$$

Letting $n = 1, 2, 3, \ldots$ we obtain m in infinitely many ways. Interchange the first two rows of this matrix to achieve a determinant with value $-m \le -1$. To achieve a determinant of 0 in infinitely many ways, take three distinct triples that satisfy the first identity. For example,

$$\det\begin{pmatrix} 2 & 2 & 3 \\ 5 & 6 & 8 \\ 3m+2 & 4m+2 & 5m+3 \end{pmatrix} = 0$$

for $m = 2, 3, 4, \ldots$.

Also solved by J. C. Binz (Switzerland), Stan Byrd and Ronald L. Smith, Con Amore Problem Group (Denmark), Jim Delany, Marty Getz and Dixon Jones, Carl Libis, Allen J. Mauney, Stephen Noltie, Volkhard Schindler (Germany), Paul J. Zweir, and the proposer. One partial solution was submitted.

Cotangents in a Cyclic Pentagon **February 2000**

1592. *Proposed by Ho-joo Lee, student, Kwangwoon University, Seoul, South Korea.*

Let $ABCDE$ be a cyclic pentagon. Prove that

$$\cot\angle ABC + \cot\angle ACB = \cot\angle AED + \cot\angle ADE$$

if and only if

$$\cot\angle ABD + \cot\angle ADB = \cot\angle AEC + \cot\angle ACE.$$

Solution by Li Zhou, Polk Community College, Winter Haven, Florida.

Because $ABCDE$ is cyclic,

$$\angle ABC = \pi - \angle AEC \qquad \text{and} \qquad \angle AED = \pi - \angle ABD.$$

In addition,

$$\angle ACB = \angle ADB \qquad \text{and} \qquad \angle ADE = \angle ACE.$$

Because $\cot(\pi - \theta) = -\cot\theta$, the following statements are equivalent:

$$\cot\angle ABC + \cot\angle ACB = \cot\angle AED + \cot\angle ADE$$
$$\cot(\pi - \angle AEC) + \cot\angle ADB = \cot(\pi - \angle ABD) + \cot\angle ACE$$
$$-\cot\angle AEC + \cot\angle ADB = -\cot\angle ABD + \cot\angle ACE$$
$$\cot\angle ABD + \cot\angle ADB = \cot\angle AEC + \cot\angle ACE.$$

Also solved by Neela Lakshmanan, Volkhard Schindler (Germany), Raul A. Simon (Chile), and the proposer.

Converging Compositions? **February 2000**

1593. *Proposed by Jon Florin, Chur, Switzerland.*

Let $(f_n)_{n\ge1}$ be a sequence of continuous, monotonically increasing functions on the interval [0, 1] such that $f_n(0) = 0$ and $f_n(1) = 1$. Furthermore, assume

$$\sum_{n=1}^{\infty} \max|f_n(x) - x| < \infty.$$

(In particular, $(f_n)_{n\geq 1}$ converges to the identity on [0, 1].)

(a) Must $(f_n \circ f_{n-1} \circ \cdots \circ f_2 \circ f_1)_{n\geq 1}$ converge to a continuous function?

(b) Must $(f_1 \circ f_2 \circ \cdots \circ f_{n-1} \circ f_n)_{n\geq 1}$ converge to a continuous function?

Solution by the proposer.

(a) Let $\phi_n = f_n \circ f_{n-1} \circ \cdots \circ f_2 \circ f_1$ and let I be the identity on [0, 1]. The sequence (ϕ_n) converges to a continuous function. We have

$$\phi_{n+1} - \phi_n = (f_{n+1} - I) \circ f_n \circ \cdots \circ f_1,$$

and therefore

$$\max_{[0,1]} |\phi_{n+1} - \phi_n| = \max_{[0,1]} |f_{n+1} - I|.$$

By the Weierstrass M-test $\sum_{n=1}^{\infty} (\phi_{n+1}(x) - \phi_n(x))$ converges uniformly on [0, 1]. It follows that

$$\lim_{N\to\infty} \phi_N = \phi_1 + \lim_{N\to\infty} \sum_{n=1}^{N-1} (\phi_{n+1} - \phi_n)$$

is continuous on [0, 1].

(b) Here the situation is different as the following example shows. Let f_i be linear on $[0,\ 2^{-(i+1)}]$ and linear on $[2^{-(i+1)},\ 1]$ with $f_i\left(2^{-(i+1)}\right) = 2^{-i}$. With these functions we have

$$\max_{[0,1]} |f_i(x) - I(x)| = (f_i - I)\,(2^{-(i+1)}) = 2^{-(i+1)},$$

which is finitely summable. Let $\phi_n = f_1 \circ \cdots \circ f_n$. Because $f_k(x) \geq x$, it follows that $\phi_{n+1}(x) \geq \phi_n(x)$. Thus $\lim_{n\to\infty} \phi_n(x)$ exists for each x, and for positive integer k,

$$\lim_{n\to\infty} \phi_n(2^{-k-1}) \geq \phi_k(2^{-k-1}) = \frac{1}{2}.$$

Because $\phi_n(0) = 0$, it follows that (ϕ_n) does not converge to a continuous function.

Also solved by Daniele Donini (Italy), Nick Lord (England), and Arlo W. Schulre (Guam).

Answers

Solutions to the Quickies from page 67.

A907. The nine letters can be arranged in a square with the eight words in the rows, columns and diagonals:

$$\begin{matrix} A & R & E \\ S & I & N \\ P & O & D \end{matrix}$$

This shows that the game is isomorphic to tic-tac-toe. Hence, if both players play optimally the game will end in a draw.

A908. Using

$$(x_1 + x_2 + \cdots + x_r)\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_r}\right) \geq r^2, \qquad x_j > 0,$$

we have

$$\begin{aligned} &a_k\left(\frac{1}{a_1-a_0}+\frac{1}{a_2-a_1}+\cdots+\frac{1}{a_k-a_{k-1}}\right)\\ &=[(a_1-a_0)+(a_2-a_1)+\cdots\\ &\quad+(a_k-a_{k-1})]\left(\frac{1}{a_1-a_0}+\frac{1}{a_2-a_1}+\cdots+\frac{1}{a_k-a_{k-1}}\right)\\ &\geq k^2. \end{aligned}$$

It follows that

$$\begin{aligned} &\frac{n}{a_1-a_0}+\frac{n-1}{a_2-a_1}+\cdots+\frac{1}{a_n-a_{n-1}}\\ &\qquad=\sum_{k=1}^{n}\left(\frac{1}{a_1-a_0}+\frac{1}{a_2-a_1}+\cdots+\frac{1}{a_k-a_{k-1}}\right)\geq\sum_{k=1}^{n}\frac{k^2}{a_k}. \end{aligned}$$

25 years ago in the MAGAZINE (Vol. 49, No. 1, January, 1976):

OUR NEW LOOK

Regular readers of *Mathematics Magazine* will no doubt notice the changed format of this issue and may wonder as to the reasons for them. Some of the reasons were aesthetic, some economic. We adopted the larger page size of the *Monthly* for both reasons: in this format we can get more words per page (thereby publishing the same amount of mathematics in fewer pages), and at the same time reduce the crowded appearance of the former *Magazine* page. The cover illustration, together with the redesigned interior layout was motivated by our attempt to make the appearance as well as the content of the *Magazine* as attractive as possible.

(The above was typed as camera-ready copy, probably using the Letter Gothic font of an IBM Selectric.)

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.

Guterman, Lila, Are mathematicians past their prime at 35?, *Chronicle of Higher Education* (1 December 2000), `http://chronicle.com/free/v47/i14/14a01801.htm` . Stephens, Mitchell, Hitting the wall, *Feed* (5 October 2000) `http://www.feedmag.com/essay/es400.shtml` .

Are mathematicians "washed up" by age 40? And is that so because our "working memory" declines? Guterman offers typical "anecdote" journalism, with lots of two-sentence interviews with young mathematicians. Stephens notes, though, that "anecdotes about brilliance in twenty-somethings and impressions of [mental] stiffness in thirty-, forty- or fifty-somethings do not, however, prove the existence of that chronological wall." Both articles cite a "historiometric" survey of 2,000 well-known scientists by Dean Keith Simonton, which attempts to determine at what age scientists make "best contributions" and "last major contribution." Here is the advice of Howard Gardner, professor of cognition and education at Harvard, about "avoiding the sort of obligations that tend to come with age": "Don't get married and don't become head of the department or give lots of post-Nobel Prize speeches." Of course, physiology may be involved; as Alan Lightman, an astrophysicist notes, "When I was in my twenties, I could work on a problem for forty-eight hours straight without sleep. When we get into the forties and fifties, most of us don't have that kind of stamina." (48 hrs??? Some of us never did.)

Herz-Fischler, Roger, *The Shape of the Great Pyramid*, Wilfrid Laurier University Press, 2000; xii + 294 pp, $29.95 (P). ISBN 0–88920–324–5.

Roger Herz-Fischler is a careful scholar whose earlier work *A Mathematical History of Division in Extreme and Mean Ratio* (1989) debunked most claims about the use of Fibonacci numbers and the golden ratio in art and architecture. His new book tries to answer "What was the *geometrical* basis, if any, that was used to determine the shape of the Great Pyramid?" He cites measurements of the pyramid and presents a comparative summary of 11 theories concerning its shape, including the well-known theory that its height is $4/\pi$ times the length of a side, a theory whose sociology he investigates in detail. After expositing each of the theories and its history, Herz-Fischler sets out philosophical considerations, particularly criteria that a theory should satisfy. Based on those, he reaches tentative conclusions; as in the case of a mystery thriller, I should not reveal those here (except perhaps to suggest that you not bet on the pi-theory).

Math in the Media: Highlights of math news from science literature and the current media, `http://www.ams.org/new-in-math/` .

Each month the AMS posts a paragraph each about half a dozen mathematical stories that have appeared in the media. A smaller number of specific articles are summarized in *Math Digest*, `http://www.ams.org/new-in-math/mathdigest/index.html` .

Glastonbury, Marion, Natural wonders: Responding to autistic lives, *Changing English: Studies in Reading and Culture* 7 (1) (March 2000) 75–88. Osborne, Lawrence, The little professor syndrome, *New York Times Magazine* (18 June 2000) 54–59, `http://www.nytimes.com/library/magazine/home/20000618mag-asperger.html` . Seymour, Liz, Condition bears gifts, frustrations, *Washington Post* (7 August 2000) B1; `http://washingtonpost.com/wp-adv/archives/` under "Asperger".

Glastonbury contends that the eccentricities of Paul Erdős and Alan Turing were clues to their having autistic personalities. They certainly were not fully autistic, but perhaps they suffered from Asperger's syndrome, a neurological disorder characterized by difficulties in social interaction, narrow interests, and obsessive behavior. Osborne interviews children diagnosed with Asperger's and speculates about their future as adults. Seymour gives anecdotes of several individuals with Asperger's, including one who earned an undergraduate degree in mathematics at Princeton but whose affliction resulted in his leaving a graduate program in mathematics.

Berlinski, David, *The Advent of the Algorithm: The Idea That Rules the World*, Harcourt, 2000; xix + 345 pp, $28. ISBN 0–15–100338–6.

Mathematicians and computer scientists probably think that works about their subject are immune to the whims of literary style. But not so. Disregard the hyperbole of this book's dust jacket, which proclaims this "the book of Genesis for the computer revolution," and turn to the publication data page. There—but not in the preface or in the introduction—you find a curious "Note to the Reader": "This is a work of scholarship. The author has woven stories, involving imagined people and incidents into the text, the better to enable the reader to enjoy the technical discussions. Or to endure them. All of these inventions of the author begin and end with [a special symbol]." We forgive the mispunctuation as we try to imagine the connection between the first sentence and the second. Are such inventions the author's definition of scholarship? After all, the book contains no references, much less citations to them. The text is as overwritten as the hyperbole on the dust jacket, and the "inventions" are distractions full of irrelevant details. But maybe the general public will be captivated and learn a bit about the fundamental concept underlying computer science.

Nahin, Paul J., *Duelling Idiots and Other Probability Puzzlers*, Princeton University Press, 2000; xviii + 271 pp, $24.95. ISBN 0–691–00979–1.

"This is a book for people who really like probability puzzles"—and who have taken a course in mathematical probability and can write simple computer programs in their favorite language. It goes beyond the simple conventional puzzles to more unusual ones (e.g., the Monty Hall car-and-goats problem is not here) and to ones that require a computer to find the answers (e.g., by simulation). The book includes the context and statements of 21 problems (80 pp), solutions and elaborations (94 pp), a section on random number generators (20 pp), and a collection of MATLAB programs (65 pp).

Nielsen, Michael A., and Isaac L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000; xiii + 676 pp, (P).

This book is the first comprehensive work on quantum computation. Part I and the Appendices include 200 pp of background from computer science, mathematics, and physics, all at a fast pace, suitable for a graduate student. Part II, quantum computation, treats quantum circuits, the quantum Fourier transform, quantum search algorithms, and the physical realization of quantum computers. Part III, quantum information, covers quantum noise, distance measures, quantum error-correction, and quantum information theory. We know a great deal about how quantum computers should work before one has even been built!

NEWS AND LETTERS

Guidelines for Authors

What do *you* like to read? What kind of writing can grab the interest of an undergraduate mathematics major? How can MATHEMATICS MAGAZINE serve to remind us all why we chose to study mathematics in the first place? If you keep these questions firmly in mind, you will be well on the way to meeting our editorial guidelines.

General information MATHEMATICS MAGAZINE is an expository journal of undergraduate mathematics. In this section, we amplify our meaning of these words.

Articles submitted to the MAGAZINE should be written in a clear and lively *expository* style. The MAGAZINE is not a research journal; papers in a terse "theorem-proof" style are unsuitable for publication. The best contributions provide a context for the mathematics they deliver, with examples, applications, illustrations, and historical background. We especially welcome papers with historical content, and ones that draw connections among various branches of the mathematical sciences, or connect mathematics to other disciplines.

Every article should contain interesting *mathematics*. Thus, for instance, articles on mathematical pedagogy alone, or articles that consist mainly of computer programs, would be unsuitable.

The MAGAZINE is an *undergraduate* journal in the broad sense that its intended audience is teachers of collegiate mathematics and their students. One goal of the MAGAZINE is to provide stimulating supplements for undergraduate mathematics courses, especially at the upper undergraduate level. Another goal is to inform and refresh the teachers of these courses by revealing new connections or giving a new perspective on history. We also encourage articles that arise from undergraduate research or pose questions to inspire it. In writing for the MAGAZINE, make your work attractive and accessible to non-specialists, including well-prepared undergraduates.

Writing and revising MATHEMATICS MAGAZINE is responsible first to its readers and then to its authors. A manuscript's publishability therefore depends as much on the quality of exposition as the mathematical significance. Our general advice is simple: Say something new in an appealing way, or say something old in a refreshing, new way. But say it clearly and directly, assuming a minimum of background. Our searchable database of past pieces from the MAGAZINE and the *College Mathematics Journal* (see `http://www.maa.org/mathmag.html`), can help you check the novelty of your idea.

Make your writing vigorous, expressive, and informal, using the active voice. Give plenty of examples and minimize computation. Help the reader understand your motivation and share your insights. Illustrate your ideas with visually appealing graphics, including figures, tables, drawings, and photographs.

First impressions are vital. Choose a short, descriptive, and attractive title; feel free to make it funny, if that would draw the reader in. Be sure that the opening sentences provide a welcoming introduction to the entire paper. Readers should know why they ought to invest time reading your work.

Our referees are asked to give detailed suggestions on style, as well as check for mathematical accuracy. In practice, almost every paper requires a careful revision by the author, followed by further editing in our office. To shorten this process, be sure to read your own work carefully, possibly after putting it away for a cooling-off period.

Provide a generous list of references to invite readers—including students—to pursue ideas further. Bibliographies may contain suggested reading along with sources actually referenced. In all cases, cite sources that are currently and readily available.

Since 1976, the Carl B. Allendoerfer Prize has been awarded annually to recognize expository excellence in the MAGAZINE. In addition to these models of style, many useful references are available. Some are listed at the end of these guidelines.

Style and format We assume that our authors are at least sometime-readers of the MAGAZINE, with some knowledge of its traditions. If so, they know that most papers are published either as Articles or as Notes. Articles have a broader scope than Notes and usually run longer than 2000 words. Notes are typically shorter and more narrowly focused. Articles should be divided into a few sections, each with a carefully chosen title. Notes, being shorter, usually need less formal sectioning. Footnotes and subsectioning are almost never used in the MAGAZINE.

In addition to expository pieces, we accept a limited number of Math Bites, poems, cartoons, Proofs Without Words, and other miscellanea.

List references either alphabetically or in the order cited in the text, adhering closely to the MAGAZINE's style for capitalization, use of italics, etc.

We recommend using simple, unadorned LaTeX in the preparation of your manuscript. Whatever technology you use, your manuscript should be generously spaced, with the title, author, and author's address at the top of the first page. Templates with further stylistic details are posted at our website in a variety of formats. Number the pages, but number only those equations that you refer to in the text. Whether LaTeX is used or not, we hope to receive some electronic version of every article accepted.

For initial submission, graphical material may be interspersed with text. Each figure should be numbered, and referenced by number in the text. Authors themselves are responsible for providing images of suitable quality. If a piece is to appear in the MAGAZINE, separate copies of all illustrations must be supplied, both with and without added lettering. We hope authors will be able to provide electronic versions of all figures, preferably in a PostScript format.

Submitting manuscripts Please submit three copies and provide your full contact information, including an email address, if available. In cases of multiple authorship, specify one corresponding author. Direct all correspondence to

Frank A. Farris, Editor, MATHEMATICS MAGAZINE, Santa Clara University,
500 El Camino Real, Santa Clara, CA 95053-0373

Electronic submission is possible in limited circumstances; ultimately, we need two hard copies to send to referees and one to retain. To request that we do this printing for you, please inquire at `mathmag@scu.edu` or 408-554-4122.

SUGGESTED READING

1. R.P. Boas, Can we make mathematics intelligible? *Amer. Math. Monthly* **88** (1981), 727–731.
2. Paul Halmos, How to write mathematics, *Enseign. Math.* **16** (1970), 123–152. Reprinted in Halmos, *Selecta, expository writings*, Vol. 2, Springer, New York, 1983, 157–186.
3. Andrew Hwang, Writing in the age of LaTeX *AMS Notices* **42** (1995), 878–882.
4. D.E. Knuth, T. Larrabee, and P.M. Roberts, *Mathematical Writing*, MAA Notes #14, 1989.
5. Steven G. Krantz, *A Primer of Mathematical Writing*, American Mathematical Society, 1997.
6. N. David Mermin, *Boojums All the Way Through*, Cambridge Univ. Pr., Cambridge, UK, 1990.

The 61st Annual William Lowell Putnam Examination

Editor's Note: The *American Mathematical Monthly* will print additional Putnam solutions later in the year.

A1 Let A be a positive real number. What are the possible values of $\sum_{j=0}^{\infty} x_j^2$, given that $x_0, x_1, x_2, \ldots$ are positive numbers for which $\sum_{j=0}^{\infty} x_j = A$?

A2 Prove that there exist infinitely many integers n such that n, $n+1$, and $n+2$ are each the sum of two squares of integers.
[*Example:* $0 = 0^2 + 0^2$, $1 = 0^2 + 1^2$, and $2 = 1^2 + 1^2$.]

A3 The octagon $P_1P_2P_3P_4P_5P_6P_7P_8$ is inscribed in a circle, with the vertices around the circumference in the given order. Given that the polygon $P_1P_3P_5P_7$ is a square of area 5 and the polygon $P_2P_4P_6P_8$ is a rectangle of area 4, find the maximum possible area of the octagon.

A4 Show that the improper integral

$$\lim_{B\to\infty} \int_0^B \sin(x)\sin(x^2)\,dx$$

converges.

A5 Three distinct points with integer coordinates lie in the plane on a circle of radius $r > 0$. Show that two of these points are separated by a distance of at least $r^{1/3}$.

A6 Let $f(x)$ be a polynomial with integer coefficients. Define a sequence $a_0, a_1, \ldots$ of integers such that $a_0 = 0$ and $a_{n+1} = f(a_n)$ for all $n \geq 0$. Prove that if there exists a positive integer m for which $a_m = 0$ then either $a_1 = 0$ or $a_2 = 0$.

B1 Let a_j, b_j, and c_j be integers for $1 \leq j \leq N$. Assume, for each j, that at least one of a_j, b_j, c_j is odd. Show that there exist integers r, s, t such that $ra_j + sb_j + tc_j$ is odd for at least $4N/7$ values of j, $1 \leq j \leq N$.

B2 Prove that the expression

$$\frac{\gcd(m,n)}{n}\binom{n}{m}$$

is an integer for all pairs of integers $n \geq m \geq 1$. [Here $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ and $\gcd(m,n)$ is the greatest common divisor of m and n.]

B3 Let $f(t) = \sum_{j=1}^{N} a_n \sin(2\pi jt)$, where each a_j is real and $a_N \neq 0$. Let N_k denote the number of zeros (including multiplicities) of $\dfrac{d^k f}{dt^k}$ for t in $[0, 1)$ [1]. Prove that

$$N_0 \leq N_1 \leq N_2 \leq \cdots \qquad \text{and} \lim_{k\to\infty} N_k = 2N.$$

B4 Let $f(x)$ be a continuous function such that $f(2x^2 - 1) = 2xf(x)$ for all x. Show that $f(x) = 0$ for $-1 \leq x \leq 1$.

[1]This restriction on t was unfortunately omitted from the original statement.

B5 Let S_0 be a finite set of positive integers. We define finite sets $S_1, S_2, \ldots$ of positive integers as follows:

Integer a is in S_{n+1} if and only if exactly one of $a-1$ or a is in S_n.

Show that there exist infinitely many integers N for which $S_N = S_0 \cup \{N + a : a \in S_0\}$.

B6 Let B be a set of more than $2^{n+1}/n$ distinct points with coordinates of the form $(\pm 1, \pm 1, \ldots, \pm 1)$ in n-dimensional space, with $n \geq 3$. Show that there are three distinct points in B which are the vertices of an equilateral triangle.

Solution to A1 Answer: Every positive number less than A^2 is a possible value of this sum.

To see this, first note that $0 < \sum_{j=0}^{\infty} x_j^2 < \left(\sum_{j=0}^{\infty} x_j\right)^2$ since every x_j is positive. We show by example that every positive number less than A^2 is a possible value of the sum. Let B be such a number and take $x_j = ar^j$, where a and r, $0 < r < 1$, are yet to be determined. We want to choose a and r so that

$$\sum_{j=0}^{\infty} ar^j = A \qquad \text{and} \qquad \sum_{j=0}^{\infty} a^2 r^{2j} = B,$$

or equivalently,

$$\frac{a}{1-r} = A \qquad \text{and} \qquad \frac{a^2}{1-r^2} = B.$$

Eliminating a from these two equations yields $r = (A^2 - B)/(A^2 + B)$. Note that $0 < r < 1$ provided that $0 < B < A^2$, so the proof is complete.

Solution to A2 For any positive integer k, let $n = (2^{2k-1})^2 + (2^k)^2$. Then

$$n + 1 = 0^2 + (2^{4k-2} + 2^{2k} + 1) = 0^2 + (2^{2k-1} + 1)^2$$

and $n + 2 = 1^2 + (2^{2k-1} + 1)^2$. Two other infinite families meeting the criterion arise from

$$n = 4k^4 + 4k^2 \text{ and } n = (2k(k+1))^2.$$

Solution to A3 The maximum area is $3\sqrt{5}$, as follows:

Let O be the center of the circle. Since the square has side $\sqrt{5}$, the radius r of the circle is $\sqrt{5/2}$. Since the rectangle has area 4 and diagonal $\sqrt{10}$ (= diameter), its side lengths are $\sqrt{2}$ and $2\sqrt{2}$ (by solving the system $ab = 4$, $a^2 + b^2 = 10$). Without loss of generality, $\overline{P_2P_4} = \sqrt{2}$. Then $\cos(\angle P_2OP_4) = 3/5$ (e.g., using the Law of Cosines in triangle P_2OP_4).

The area of the octagon is the sum of the areas of triangles $OP_1P_2, \ldots OP_7P_8$, OP_8P_1. Recalling that the area of $\triangle OP_iP_j$ is $(r^2/2)\sin(\angle P_iOP_j)$, we must maximize

$$\begin{aligned} K &= \frac{5}{4}\big(\sin(\angle P_1OP_2) + \cdots + \sin(\angle P_7OP_8) + \sin(\angle P_8OP_1)\big) \\ &= \frac{5}{2}\big(\sin(\angle P_1OP_2) + \sin(\angle P_2OP_3) + \sin(\angle P_3OP_4) + \sin(\angle P_4OP_5)\big) \end{aligned}$$

since P_5, P_6, P_7, P_8 are, respectively, diametrically opposite P_1, P_2, P_3, P_4.

Set $\alpha = \angle P_1OP_2$ and $\beta = \angle P_2OP_4$. Then $\cos\beta = 3/5$ and so $\sin\beta = 4/5$. We have

$$\angle P_2OP_3 = \angle P_1OP_3 - \alpha = \pi/2 - \alpha,$$
$$\angle P_3OP_4 = \angle P_2OP_4 - \angle P_2OP_3 = \alpha + \beta - \pi/2,$$
$$\angle P_4OP_5 = \angle P_3OP_5 - P_3OP_4 = \pi/2 - (\alpha + \beta - \pi/2) = \pi - (\alpha + \beta)$$

So

$$\begin{aligned} K &= \frac{5}{2}\big(\sin\alpha + \sin(\pi/2 - \alpha) + \sin(\alpha + \beta - \pi/2) + \sin(\pi - \alpha - \beta)\big) \\ &= \frac{5}{2}\big(\sin\alpha + \cos\alpha - \cos(\alpha+\beta) + \sin(\alpha+\beta)\big) \\ &= \frac{5}{2}\big(\sin\alpha(1 + \sin\beta + \cos\beta) + \cos\alpha(1 - \cos\beta + \sin\beta)\big) \\ &= 6\sin\alpha + 3\cos\alpha. \end{aligned}$$

Therefore, the maximum value of K is $\sqrt{6^2 + 3^2} = 3\sqrt{5}$.

Alternate Solution to A3 The maximum area is $3\sqrt{5}$.

Let O be the center of the circle. Since the square has side $\sqrt{5}$, the radius r of the circle is $\sqrt{5/2}$. Since the rectangle has area 4 and diagonal $\sqrt{10}$, its side lengths are $\sqrt{2}$ and $2\sqrt{2}$. Without loss of generality, $\overline{P_2P_4} = \sqrt{2}$ and $\overline{P_8P_2} = 2\sqrt{2}$.

The area of the octagon is the area of the rectangle $P_2P_4P_6P_8$ plus twice the areas of triangles $P_8P_1P_2$ and $P_2P_3P_4$. The rectangle can be rotated relative to the square so that these triangles each have maximal area (when they are isosceles—i.e., when $\overline{P_8P_1} = \overline{P_1P_2}$ and $\overline{P_2P_3} = \overline{P_3P_4}$, because this gives maximal altitudes to their respective bases P_8P_2 and P_2P_4).

Thus, the maximum area of the octagon is

$$4 + \sqrt{2}\left(\sqrt{5/2} - \sqrt{5/2 - 1/2}\right) + 2\sqrt{2}\left(\sqrt{5/2} - \sqrt{5/2 - 2}\right) = 3\sqrt{5}.$$

Solution to A4 For any B,

$$\begin{aligned} &\int_0^B \sin(x)\sin(x^2)dx \\ &= \frac{1}{2}\int_0^B \cos(x^2 - x)dx - \frac{1}{2}\int_0^B \cos(x^2 + x)dx \\ &= \frac{1}{2}\int_0^B \cos(x^2 - x)dx - \frac{1}{2}\int_1^{B+1} \cos(t^2 - t)dt \text{ (by substituting } t = x + 1) \\ &= \frac{1}{2}\int_0^1 \cos(x^2 - x)dx - \frac{1}{2}\int_B^{B+1} \cos(x^2 - x)dx. \end{aligned}$$

To complete the proof we must show that

$$\lim_{B\to\infty}\int_B^{B+1} \cos(x^2 - x)dx = 0.$$

Toward this end, let $\{y\}$ denote the largest $y_0 \le y$ such that $y_0^2 - y_0$ is an odd multiple of $\pi/2$. Since $\lim_{y\to\infty}(y - \{y\}) = 0$, it suffices to show that

$$\lim_{B\to\infty}\int_{\{B\}}^{\{B+1\}} \cos(x^2 - x)dx = 0.$$

This follows from the observation that the integral is a sum

$$\pm\sum_{n=n_0}^{n_1}\int_{x^2-x=\frac{(2n-1)\pi}{2}}^{x^2-x=\frac{(2n+1)\pi}{2}} \cos(x^2 - x)dx$$

with summands of alternating sign and decreasing absolute value, and with first summand $\to 0$ as $B \to \infty$. To see this, substituting $t = x^2 - x - \pi n$, we have

$$\int_{x^2-x=\frac{(2n-1)\pi}{2}}^{x^2-x=\frac{(2n+1)\pi}{2}} \cos(x^2 - x)dx = (-1)^n\int_{-\pi/2}^{\pi/2}\frac{\cos t}{\sqrt{1+4n+4t}}dt.$$

Solution to A5 We may suppose that one of the points on the circle has coordinates $P = (0, 0)$, and the other two have coordinates $Q = (a, b)$ and $R = (c, d)$, for integers a, b, c, and d, where $ad - bc \neq 0$. Let (h, k) be the coordinates of the center of the circle. Then (h, k) is the intersection of the three perpendicular bisectors

$$y - \frac{b}{2} = -\frac{a}{b}\left(x - \frac{a}{2}\right)$$
$$y - \frac{d}{2} = -\frac{c}{d}\left(x - \frac{c}{2}\right)$$
$$y - \frac{b+d}{2} = -\frac{c-a}{b-d}\left(x - \frac{a+c}{2}\right)$$

From the first two equations:

$$h = \frac{d(a^2+b^2) - b(c^2+d^2)}{2(ad-bc)}, \text{ and}$$
$$k = \frac{-c(a^2+b^2) + a(c^2+d^2)}{2(ad-bc)}.$$

It follows that

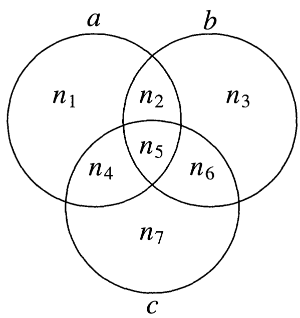
$$r^2 = h^2 + k^2 = \frac{(a^2+b^2)(c^2+d^2)((a-c)^2+(b-d)^2)}{4(ad-bc)^2}$$
$$= \frac{(\overline{PQ})^2(\overline{PR})^2(\overline{QR})^2}{4(ad-bc)^2}$$

Let $L = \max\{\overline{PR}, \overline{PQ}, \overline{QR}\}$. Then

$$r^2 \le \frac{L^6}{4(ad-bc)^2} \le \frac{L^6}{4}$$
$$r^{\frac{1}{3}} \le \frac{L}{2^{\frac{1}{3}}} \le L$$

Solution to A6 If y is an integer, then $(f(x) - f(y))/(x - y)$ is a polynomial with integer coefficients, so for any two distinct integers a and b, $b - a$ divides $f(b) - f(a)$. In particular, $a_{i+1} - a_i$ divides $f(a_{i+1}) - f(a_i) = a_{i+2} - a_{i+1}$, for each $i = 0, 1, 2, \ldots, m-1$. Since $a_{m+1} - a_m = a_1 - a_0$, $a_{i+1} - a_i$ must all be equal in absolute value. Either that absolute value is zero, in which case $a_1 = 0$, or d, where d is nonzero. However, in the latter case, if they all agree in sign, $0 = a_m - a_0 = \sum_{i=0}^{m-1}(a_{i+1} - a_i) = md$, giving a contradiction. Therefore, there must exist j such that $a_{j+2} - a_{j+1} = -(a_{j+1} - a_j)$, so that $a_{j+2} = a_j$. Applying f to each side $m - j$ times gives $a_2 = a_0 = 0$.

Solution to B1 The three circles in the Venn diagram, labeled a, b, c, represent the triples (a, b, c) whose corresponding coordinates are odd, and the labels within the regions denote the number of j for which (a_j, b_j, c_j) has the corresponding parity pattern (for example n_1 is the number of triples (a_j, b_j, c_j) with a_j odd and b_j, c_j even).



With this notation we can enumerate all the cases to obtain the following table.

r	s	t	Number of j for which $ra_j + sb_j + tc_j$ is odd
even	even	odd	$n_4 + n_5 + n_6 + n_7$
even	odd	even	$n_2 + n_3 + n_5 + n_6$
even	odd	odd	$n_2 + n_3 + n_4 + n_7$
odd	even	even	$n_1 + n_2 + n_4 + n_5$
odd	even	odd	$n_1 + n_2 + n_6 + n_7$
odd	odd	even	$n_1 + n_3 + n_4 + n_6$
odd	odd	odd	$n_1 + n_3 + n_5 + n_7$

The sum of the seven numbers in the right column is $4N$, and therefore one of the numbers in this list is $\geq 4N/7$, which is what we needed to show.

Solution to B2 Let $S = \left\{x \in \mathbb{Z} : \frac{x}{n}\binom{n}{m} \text{ is an integer}\right\}$. Clearly, $n \in S$; but $m \in S$ as well since

$$\frac{m}{n}\binom{n}{m} = \frac{m}{n}\frac{n}{m}\binom{n-1}{m-1} = \binom{n-1}{m-1}.$$

It follows that linear combinations of m and n belong to S; in particular, $\gcd(m, n)$ is in S because there are integers s and t such that $\gcd(m, n) = sm + tn$.

Solution to B3 Think of $f(t)$ as a real-valued continuous, infinitely differentiable function on the circle. By Rolle's theorem there is always a zero of $f^{(k+1)}$ between any two zeros of $f^{(k)}$. This evidently includes zeros counted with multiplicity, so $N_{k+1} \geq N_k$ for all k.

The function $g_k(t) := f^{(4k+1)}(t)/(2\pi)^{4k+1} = \sum_{j=1}^{N} a_j j^{4k+1} \cos(2\pi j t)$ is real valued, and is dominated by the last term when k is sufficiently large. To make this precise, note that for $n = 0, 1, 2, \ldots, 2N$, we have

$$(-1)^n \frac{g_k\left(\frac{n}{2N}\right)}{a_N} = \sum_{j=1}^{N-1} \frac{(-1)^n j^{4k+1} a_j}{a_N} \cos\left(2\pi j \left(\frac{n}{2N}\right)\right) + N^{4k+1}$$

$$\geq N^{4k+1} - \sum_{j=1}^{N-1} \frac{|a_j|\, j^{4k+1}}{|a_N|} > 0$$

if k is sufficiently large. Thus, by the Intermediate Value Theorem, g_k has a zero in each interval $(n/2N, (n+1)/2N)$ and so $N_{4k+1} \geq 2N$. Combining this with the first part we have $\lim_{k\to\infty} N_k \geq 2N$.

Finally, suppose we write $z = e^{2i\pi t}$. With this notation, every $f^{(k)}(t)$ has the form $\sum_{n=1}^{N}(c_n z^n + d_n z^{-n})$. Multiplying through by z^N we get a polynomial of degree $2N$, so it has no more than $2N$ roots. Thus, $N_k \leq 2N$.

Solution to B4 For $x \neq 0$,

$$f(x) = \frac{f(2x^2-1)}{2x} = -\frac{f(2x^2-1)}{-2x} = -f(-x).$$

Thus,

$$f(0) = \lim_{x\to 0} f(x) = -\lim_{x\to 0} f(-x) = -f(0),$$

so that $f(0) = 0$.

For all θ,

$$f(\cos 2\theta) = f(2\cos^2\theta - 1) = 2\cos\theta\, f(\cos\theta),$$

so that

$$f(\cos\theta) = 0 \text{ iff } f(\cos 2\theta) = 0$$

(the necessity uses $f(0) = 0$). Clearly,

$$f(\cos\theta) = 0 \text{ iff } f(\cos(\theta + 2\pi)) = 0.$$

Then, since $f(\cos\pi/2) = 0$, we have for all integers m, n,

$$f\left(\cos\frac{\pi/2 + 2m\pi}{2^n}\right) = 0,$$

so that $f(x) = 0$ for a dense set of points in $[-1, 1]$. The result follows by continuity.

Note: $f(x)$ is not necessarily 0 outside of $[-1, 1]$; e.g.,

$$f(x) = \begin{cases} -\sqrt{x^2-1} & \text{if } x < -1, \\ 0 & \text{if } -1 \leq x \leq 1, \\ \sqrt{x^2-1} & \text{if } x > 1. \end{cases}$$

Solution to B5 For each $j \geq 0$, let $f_j(x) = \sum_{a \in S_j} x^a$. Note that $f_0(x)$ is a polynomial and that $f_{j+1}(x) = (1+x) f_j(x) \bmod 2$. Therefore $f_N(x) = (1+x)^N f_0(x) \bmod 2$, by induction. But $(1+x)^2 \equiv 1 + x^2 \bmod 2$, so $(1+x^2)^2 \equiv 1 + x^4 \bmod 2$, and by induction, $(1+x)^{2^k} \equiv 1 + x^{2^k} \bmod 2$ for all integers $k \geq 0$.

Now take any k for which 2^k is larger than the degree of $f_0(x)$. Then for $N = 2^k$,

$$f_N(x) \equiv (1+x)^N f_0(x) \equiv (1+x^N) f_0(x) = \sum_{x \in S_0} x^a + \sum_{x \in S_0} x^{N+a} \mod 2,$$

which implies the result.

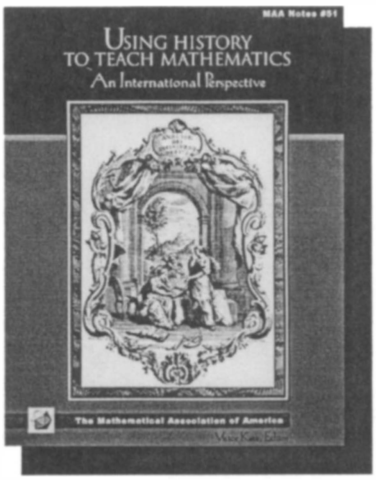
Solution to B6 Let A denote the set of 2^n points in n-dimensional space of the form $(\pm 1, \pm 1, \ldots, \pm 1)$. For each $a \in A$, let $N_a = \{b \in B : |a - b| = 2\}$. The (multiset) union of the sets N_a for $a \in A$ will contain each element of B exactly n times (because for each $b \in B$ there are exactly n points a in A for which $|a - b| = 2$). Thus, the (multiset) union of the sets N_a for $a \in A$ contains $n|B|$ elements. Because $n|B| > 2^{n+1}$ and A contains 2^n elements, the Pigeonhole Principle implies that there is an $a \in A$ such that N_a contains at least 3 elements, say b_1, b_2, b_3. These three distinct points of B are the vertices of an equilateral triangle with side length $2\sqrt{2}$.

The Mathematical Association of America

Using History to Teach Mathematics

Victor Katz, editor

Series: MAA Notes



This book is a collection of articles by international specialists in the history of mathematics and its use in teaching, based on presentations given at an international conference in 1996. Although the articles vary in technical or educational level and in the level of generality, they show how and why an understanding of the history of mathematics is necessary for informed teaching of various subjects in the mathematics curriculum, both at secondary and at university levels. Many of the articles can serve teachers directly as the basis of classroom lessons, while others will give teachers plenty to think about in designing courses or entire curricula. For example, there are articles dealing with the teaching of geometry and quadratic equations to high school students, of the notion of pi at various levels, and of linear algebra, combinatorics, and geometry to university students. But there is also an article showing how to use historical problems in various courses and one dealing with mathematical anomalies and their classroom use.

Although the primary aim of the book is the teaching of mathematics through its history, some of the articles deal more directly with topics in the history of mathematics not usually found in textbooks. These articles will give teachers valuable background. They include one on the background of Mesopotamian mathematics by one of the world's experts in this field, one on the development of mathematics in Latin America by a mathematician who has done much primary research in this little known field, and another on the development of mathematics in Portugal, a country whose mathematical accomplishments are little known. Finally, an article on the reasons for studying mathematics in Medieval Islam will give all teachers food for thought when they discuss similar questions, while a short play dealing with the work of Augustus DeMorgan will help teachers provide an outlet for their class thespians.

Catalog Code: NTE-51/JR 300 pp., Paperbound, 2000 ISBN 088385-163-6 List: $32.95 MAA Member: $25.95

Name__________ Credit Card No.__________

Address__________ Signature__________ Exp. Date___/___

City__________ Qty_____ Price $_____ Amount $_____

State_____ Zip _____ Shipping and Handling $_____

Phone __________ Catalog Code: NTE-51/JR Total $_____

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

Order Via:

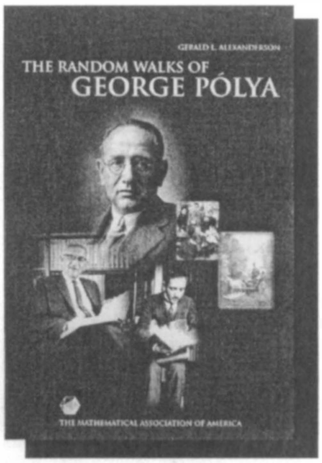
Phone: 1 (800) 331.1622
Fax: (301) 206.9789
Mail: Mathematical Association of America
PO Box 91112
Washington, DC 20090-1112
Web: www.maa.org

The Mathematical Association of America

The Random Walks of George Pólya

Gerald L. Alexanderson

Series: MAA Spectrum



In the first half of this charming book Gerald Alexanderson presents an insightful portrait of George Pólya, the great teacher and mathematician. In the second half of the book, Alexanderson assembles eight papers that describe Pólya's contribution to various fields.

George Pólya enjoyed the esteem of the mathematical community not only for his deep and influential contributions in a variety of mathematical fields, but also for his groundbreaking work in the teaching of mathematics. His standing in the latter area could rest solely on his having written one of the most widely read books in mathematics, the still-popular *How to Solve It*. In addition to his championing problem-solving, he contributed to mathematics important results in complex and real analysis, inequalitites, mathematical physics, combinatorics, probability theory, number theory, and geometry. He coined the phrases "random walk" and "central limit theorem" and gave to mathematics the Pólya Enumeration Theorem, along with many other ideas used widely today. The present work describes how such versatility came about and, along the way, tells some enlightening stories about mathematics and mathematicians.

The list of articles about Pólya's work include: *Pólya's Work in Probability*, by K.L. Chung, *Pólya's Work in Analysis*, by R.P. Boas, *Comments on Number Theory*, by D.H. Lehmer, *Pólya's Geometry*, by Doris Shattschneider, *Pólya's Enumeration Theorem*, by R.C. Read, *Pólya's Contributions in Mathematical Physics*, by M.M. Schiffer, *George Pólya and Mathematics Education*, by Alan Schoenfeld, and *Pólya's Influence-References to His Work*.

Paperbound Edition:	Casebound Edition:
Catalog Code: RWP	**Catalog Code: WAY**
320pp., Paperbound, 2000	320pp., Casebound, 2000
ISBN 0-88385-528-3	ISBN 0-88385-531-3
List Price: $29.95 Member Price: $23.95	List Price: $41.95 Member Price: $32.95

Name__________ Credit Card No.__________

Address__________ Signature__________Exp. Date___/___

City__________ Qty_____ Price $______ Amount $______

State______ Zip ______ Shipping and Handling $______

Phone __________ **Catalog Code: RWP/JR or WAY/JR** Total $______

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

Order Via:

Phone: **1 (800) 331.1622**

Fax: **(301) 206.9789**

Mail: **Mathematical Association of America**
PO Box 91112
Washington, DC 20090-1112

Web: **www.maa.org**

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

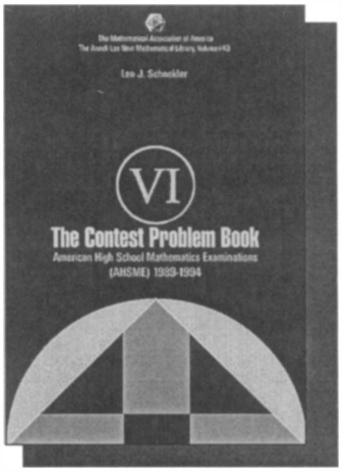
Credit Card No. ____________ Expires ___/___

Signature ____________

The Mathematical Association of America

The Contest Problem Book VI

Leo J. Schneider



American High School Mathematics Examinations (AHSME) 1989-1994

Series: Anneli Lax New Mathematical Library

The Contest Problem Book VI chronicles the high school competitions sponsored by the Mathematical Association of America. It contains 180 challenging problems from the six years of the American High School Mathematics Examinations (AHSME), 1989-1994, as well as a selection of other problems. Many problem solving techniques for problems in this book show alternative approaches that appear in print for the first time.

Some aspects of mathematical problem solving unique to competitions are discussed. Useful tools are selected from important areas of high school mathematics.

A Problems Index classifies the 180 problems in the book into subject areas: Algebra, Complex Numbers, Discrete Mathematics, Number Theory, Statistics, and Trigonometry. Outstanding problems combine elementary techniques from diverse areas of mathematics. You will find problems that really engage your students in this volume.

Catalog Code: NML-40/JR 232 pp., Paperbound, 2000 ISBN 0-88385-642-5 List: $21.95 MAA Member: $17.95

Name________ Credit Card No.________

Address________ Signature________ Exp. Date___/___

City________ Qty____ Price $____ Amount $____

State____ Zip ____ Shipping and Handling $____

Phone ________ Catalog Code: NML-40/JR Total $____

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

Order Via:

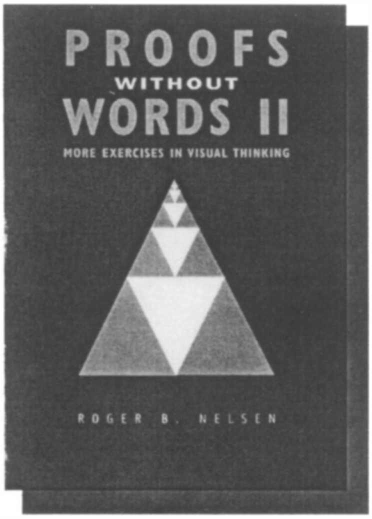
Phone: 1 (800) 331.1622
Fax: (301) 206.9789
Mail: Mathematical Association of America
PO Box 91112
Washington, DC 20090-1112
Web: www.maa.org

The Mathematical Association of America

Proofs Without Words II: *More Exercises in Visual Thinking*

Roger B. Nelsen

Series: Classroom Resource Materials



What are "proofs without words?" Many would argue that they are not really "proofs" (nor, for that matter, are many "without words," on account of equations which often accompany them). Like its predecessor *Proofs Without Words*, published by the MAA in 1993, this book is a collection of pictures or diagrams that help the reader see why a particular mathematical statement may be true, and also to see how one might begin to go about proving it true. The emphasis is on providing visual clues to the observer to stimulate mathematical thought.

Proofs without words have been around for a long time. In this volume you find modern renditions of proofs without words from ancient China, tenth century Arabia, and Renaissance Italy. While the majority of the proofs without words in this book originally appeared in journals published by the MAA, others first appeared in journals published by other organizations in the US and abroad, and on the World Wide Web.

The proofs in this collection are arranged by topic into five chapters: geometry and algebra; trigonometry, calculus and & analytic geometry; inequaltieis; integer sums, and infinite series & linear algebra. Although the proofs without words are presented primarily for the enjoyment of the reader, teachers will want to use them with students at many levels - in precalculus courses in high school, in college courses in calculus, number theory and combinatorics, and in pre-service and in-service classes for teachers.

Catalog Code: PW2/JR 142 pp., Paperbound, 2000 ISBN 088385-721-9 List: $24.00 MAA Member: $18.95

Name__________ Credit Card No.__________

Address__________ Signature__________Exp. Date___/___

City__________ Qty_____ Price $_____ Amount $_____

State_____ Zip _____ Shipping and Handling $_____

Phone __________ **Catalog Code: PW2/JR** Total $_____

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

Order Via:

Phone: **1 (800) 331.1622**
Fax: **(301) 206.9789**
Mail: **Mathematical Association of America**
PO Box 91112
Washington, DC 20090-1112
Web: **www.maa.org**

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, DC 20036