Vol. 72, No. 2 April 1999

MATHEMATICS MAGAZINE

- The Discovery of Ceres
- Marriages Made in the Heavens
- Pianos and Continued Fractions
- Anatomy of a Circle Map

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

Send new manuscripts to Paul Zorn, Editor, Department of Mathematics, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

Cover illustration. A computer graphic image based on Ritmüller's portrait of Gauss on the terrace of the observatory at Göttingen. The graphic was created by Don Teets, with help from Becky Weiss and Janet Burgoyne.

W. E. Briggs received his Ph.D. from the University of Colorado at Boulder in 1953, writing his thesis in prime number theory under Burton Jones. He was a mathematics faculty member at the University of Colorado at Boulder until his retirement in 1988. He also served as Dean of the College of Arts and Sciences from 1963 until 1980.

William L. Briggs has been at the mathematics department at the University of Colorado at Denver since 1984. He received his Ph.D. in applied mathematics from Harvard University in 1976. His current research interests are computational mathematics and mathematical problems in biology and medicine. He has just completed a textbook (Addison-Wesley Longman) for liberal arts mathematics courses.

Edward Dunne received a B.S. from Santa Clara University and a Ph.D. from Harvard University. Coming from a family of artists and musicians, his brothers once dubbed him "the white sheep of the family" for having chosen a career in mathematics.

Dan Kalman joined the mathematics faculty at Washington DC's American University in 1993, following eight years as an applied mathematician in the aerospace industry, and earlier teaching appointments in Wisconsin and South Dakota. He is an Associate Editor of *Mathematics Magazine*, author of the textbook *Elementary Mathematical Models*, and recipient of three MAA writing awards: George Polya (1994), Trevor Evans (1997), and Carl B. Allendoerfer (1998). In 1996–1997, he served as Associate Executive Director of the MAA. Kalman received his B.S. from Harvey Mudd College (1974) and his Ph.D. from the University of Wisconsin (1980).

Mark McConnell is the son of an organist/choirmaster and a minister. He plays the flute and recorder, and has sung in choruses since he was eight. He became interested in the mathematics of tuning while at Harvard (where he received his B.A.) because his college chorus kept going sharp. He discovered Madonna's "Borderline chord" one spring afternoon while sitting barefoot on the green at Brown University and studying for his Ph.D., which he received in 1987. Today he is an Associate Professor of Mathematics at Oklahoma State University, is married with two children, and sings in several choirs.

Donald Teets received his B.A. from the University of Colorado, his M.S. from Colorado State University, and his Doctor of Arts from Idaho State University. He has taught at the South Dakota School of Mines and Technology since 1988, and has been chair of the department of mathematics and computer science since 1997. The Gauss–Ceres orbit problem was brought to his attention by a historical note regarding the method of least squares in a calculus text. When he is not doing mathematics, he enjoys backpacking, cross-country skiing, and rock climbing.

Karen Whitehead received both her B.A. and Ph.D. from the University of Minnesota. She joined the faculty at the South Dakota School of Mines and Technology in 1981, serving as department head and college dean before taking on her current duties as Vice President for Academic Affairs. Her interest in the Gauss–Ceres problem was two-fold: before looking at mathematics involved, she first translated Gauss's paper and several related works from the original German. Her major avocation is music: she is a substitute church organist, and has sung in student choral groups on campus for the past fourteen years.

Vol. 72, No. 2, April 1999

MATHEMATICS MAGAZINE

EDITOR
Paul Zorn
St. Olaf College

ASSOCIATE EDITORS
Arthur Benjamin
Harvey Mudd College
Paul J. Campbell
Beloit College
Douglas Campbell
Brigham Young University
Barry Cipra
Northfield, Minnesota
Susanna Epp
DePaul University
George Gilbert
Texas Christian University
Bonnie Gold
Monmouth University
David James
Howard University
Dan Kalman
American University
Victor Katz
University of DC
David Pengelley
New Mexico State University
Harry Waldman
MAA, Washington, DC

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

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

Copyright © by the Mathematical Association of America (Incorporated), 1999, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. Permission to make copies of individual articles, in paper or electronic form, including posting on personal and class web pages, for educational and scientific use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the following copyright notice:

Copyright the Mathematical Association of America 1999. All rights reserved.

Abstracting with credit is permitted. To copy otherwise, or to republish, requires specific permission of the MAA's Director of Publication and possibly a fee.

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

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

Printed in the United States of America

ARTICLES

The Discovery of Ceres: How Gauss Became Famous

DONALD TEETS
KAREN WHITEHEAD
South Dakota School of Mines and Technology
Rapid City, SD 57701

"The Duke of Brunswick has discovered more in his country than a planet: a super-terrestrial spirit in a human body."

These words, attributed to Laplace in 1801, refer to the accomplishment of Carl Friedrich Gauss in computing the orbit of the newly discovered planetoid *Ceres Ferdinandea* from extremely limited data. Indeed, although Gauss had already achieved some fame among mathematicians, it was his work on the Ceres orbit that "made Gauss a European celebrity—this a consequence of the popular appeal which astronomy has always enjoyed..." [2]. The story of Gauss's work on this problem is a good one and is often told in biographical sketches of Gauss (e.g., [2], [3], [6]), but the mathematical details of how he solved the problem are invariably omitted from such historical works. We are left to wonder, how did he do it? *Just how did Gauss compute the orbit of Ceres?* This is the question that we shall answer in this paper!

As the reader will observe, Gauss's work offers a rare instance of solving an historically great problem in applied mathematics using only the most modest mathematical tools. It is a complicated problem, involving over 80 variables in three different coordinate systems, yet the tools that Gauss uses are largely high school algebra and trigonometry! Gauss achieves greatness in this work not through deep, abstract mathematical thinking, but rather through an incredible vision of how the various quantities in the problem are related, a vision that guides him through extraordinary computations that others would likely abandon as futile.

Thus the description of Gauss's work that follows involves much algebraic and trigonometric computation. We hope the reader can appreciate Gauss's genius by observing how difficult it is to see how the various computational steps he undertakes might reasonably lead to the final goal. We hope also to have provided enough details so that the interested reader can follow Gauss's work from start to finish.

We begin with a brief introduction to reacquaint the reader with the historical background of the Ceres orbit problem.

Historical background

The asteroid Ceres was first observed by the Italian astronomer Joseph Piazzi in Palermo on New Year's Day, 1801. Within the European scientific community at the time there had been considerable discussion of the possibility that a major planet remained to be discovered on an orbit lying between those of Mars and Jupiter. Indeed, a group of 24 astronomers including Piazzi had formed to make a systematic search for such a planet, led by Baron Xavier von Zach, director of the Seeburg observatory and editor of the astronomical journal *Monatliche Correspondenz zur Beförderung der Erd- und Himmelskunde*.

Piazzi observed Ceres until February 11, 1801, when its position in the sky became too near that of the sun for any further observation. Meanwhile, on January 24, Piazzi had sent letters reporting his discovery to his colleagues Bode in Berlin, Oriani in Milan, and Lalande in Paris. In these letters, which reached Lalande in February but took until April to reach the others, Piazzi variously referred to the new object as a comet and as a planet **[3]**, **[7]**.

Reports of Piazzi's discovery soon reached von Zach, who published in the June 1801 issue of the *Monatliche Correspondenz* a long article "on a long supposed, now probably discovered, new major planet of our solar system between Mars and Jupiter." Though Piazzi requested that the publication of his observations be delayed, they were quickly shared among the leading European astronomers of the day; thus the July issue of the *Monatliche Correspondenz* contains a preliminary orbit for Ceres computed by the astronomer Burckhardt. In the September issue, von Zach finally published Piazzi's complete observations, and in the October issue, he reported that astronomers had looked carefully during August and September for the re-emergence of Ceres, but without success **[3]**.

It is at this point that Gauss became involved in the problem. At 24 years of age, Gauss had recently completed his doctoral degree and was living in relative obscurity in Brunswick, supported by an annual stipend from the Duke of Brunswick-Wölfenbuttel. Regarding the problem of computing planetary orbits from a short sequence of observations, Gauss writes in the preface to **[5]**,

> Some ideas occurred to me in the month of September of the year 1801, ... which seemed to point to the solution of the great problem of [computing planetary orbits]... [T]hese conceptions... happily occurred at the most propitious moment for their preservation and encouragement that could have been selected. For just about this time the report of the new planet, discovered on the first day of January of that year with the telescope at Palermo, was the subject of universal conversation; and soon afterwards the observations made by that distinguished astronomer Piazzi from the above date to the eleventh of February were published. Nowhere in the annals of astronomy do we meet with so great an opportunity, and a greater one could hardly be imagined, for showing most strikingly, the value of this problem, than in this crisis and urgent necessity, when all hope of discovering in the heavens this planetary atom, among innumerable small stars after the lapse of nearly a year, rested solely upon a sufficiently approximate knowledge of its orbit to be based upon these very few observations. Could I ever have found a more seasonable opportunity to test the practical value of my conceptions, than now in employing them for the determination of the orbit of the planet Ceres, which during these forty-one days had described a geocentric arc of only three degrees, and after the lapse of a year must be looked for in a region of the heavens very remote from that in which it was last seen? This first application of the method was made in the month of October, 1801, and the first clear night, when the planet was sought for as directed by the numbers deduced from it, restored the fugitive to observation.

Gauss's earliest extant notes on Ceres were recorded in November of 1801, and it was in that month that he completed his first orbit determination. In the December 1801 issue of the *Monatliche Correspondenz*, von Zach published Gauss's predicted orbit for Ceres, writing that "Great hope for help and facilitation is accorded us by the recently shared investigation and calculation of Dr. Gauss in Brunswick" **[11]**. Although he pointed out that Gauss's orbit was significantly different from those of Burckhardt and other well known astronomers, von Zach gave arguments in its favor, concluding that "All this proves the Gaussian ellipse. What confidence it must thus awaken if astronomers recognize the precision with which it represents the collected Piazzi observations." Precision indeed, for on December 7, 1801, von Zach was able to relocate Ceres according to Gauss's predictions, and a few weeks later on New Year's Eve, the rediscovery was confirmed by Wilhelm Olbers, an amateur astronomer who later became a close friend of Gauss. And almost immediately, Gauss's reputation as a young genius was established throughout Europe.

Just how did Gauss compute the orbit of Ceres? Though his *Theoria motus corporum coelestium in sectionibus conicis solem ambientium (Theory of the motion of the heavenly bodies moving about the sun in conic sections)* of 1809 is clearly his crowning achievement in the area of planetary motion, Gauss writes in his preface to that work that "scarcely any trace of resemblance remains between the method in which the orbit of Ceres was first computed, and the form given in this work." Dunnington **[3]**, in his monumental biography of Gauss, writes that "His earliest notes on Ceres... lack clearness," and in **[8]** we find that "there is some controversy regarding precisely how he did it." Fortunately, Gauss sent a manuscript summarizing his methods in a letter to Olbers dated August 6, 1802, just seven months after the rediscovery of Ceres. The manuscript, entitled *Summarische Übersicht der zur Bestimmung der Bahnen der beiden neuen Hauptplaneten angewandten Methoden (Summary Survey of the Methods Applied in the Determination of the Orbits of Both New Planets)* was published years later in the September, 1809 issue of the *Monatliche Correspondenz*. Though the *Summarische Übersicht* had apparently already undergone certain refinements compared to the earliest methods, it is by far the most complete record of Gauss's early work on the computation of planetary orbits, and is therefore the work upon which we base our answer to the question of how Gauss computed the orbit of Ceres.

We close this portion of the narrative by recommending that the interested reader consult **[3]** for a more complete historical account of the discovery of Ceres, and **[7]** for a comparison of Gauss's earliest (unpublished) methods to those of the *Summarische Übersicht* and *Theoria Motus*.

The fundamentals of planetary orbits

To understand Gauss's work, we must first introduce the basic terminology of planetary orbits. According to *Kepler's First Law* the planet's orbit is an ellipse with the sun at one focus. As illustrated in Fig. 1, it is convenient to choose a standard rectangular coordinate system with the sun at the origin. (Typically the xy-plane is chosen to be the plane of the Earth's orbit, the so called *ecliptic* plane.) The angle i between the positive z-axis and the vector $\mathbf{n}$ normal to the planet's orbital plane is called the *inclination* of the orbit, with $0° \leq i \leq 90°$. The planet's orbital plane and the ecliptic plane intersect in the *line of nodes*, and, assuming that the direction of motion is as indicated by the arrow, the point N on this line is known as the *ascending node*. The angle Ω measured from the positive x-axis counterclockwise to the line of nodes is the *longitude of the ascending node*. Letting ω represent the angle between the line of nodes and the major axis of the planet's elliptical orbit, we define the *longitude of aphelion* $\pi = \Omega + \omega + 180°$ (the sum of angles in two

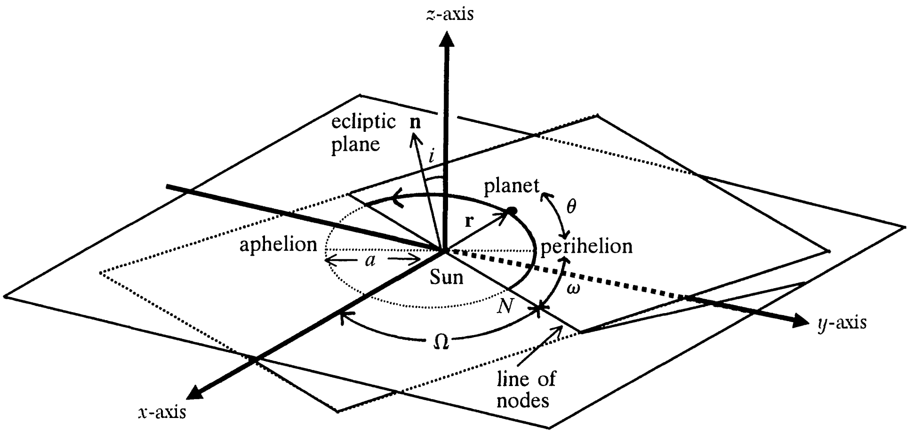


FIGURE 1
Parameters describing the planetary orbit

different planes!), which determines the orientation of the ellipse within the orbital plane. Note that *aphelion* is the point on the orbit furthest from the sun, whereas *perihelion* is the point closest to the sun. The ellipse itself is determined by a, the length of its semimajor axis, and e, its eccentricity. Finally, the position of the planet on this elliptical orbit is determined by τ_p, the time of perihelion passage. Collectively, the six quantities i, Ω, π, a, e, and τ_p are referred to as the *elements* of the orbit. For future reference, we note that the angle θ in Fig. 1 is known as the *true anomaly*, and define $v = \Omega + \omega + \theta$, again a sum of angles in two different planes. (The word *anomaly* was apparently chosen because of the discrepancy between the actual and computed values of θ in early studies of planetary motion.)

Suppose now that we know two heliocentric (sun-centered) vectors $\mathbf{r}$ and $\mathbf{r}''$ describing the planet's position at times τ and τ''. It is straightforward to compute from the normal vector $\mathbf{n} = \mathbf{r} \times \mathbf{r}''$ the inclination i, the equations of the orbital plane and the line of nodes, and Ω, the longitude of the ascending node.

Within the orbital plane, the elliptical orbit is given by the polar equation

$$r = \frac{k}{1 + e\cos\theta} = \frac{k}{1 - e\cos(v - \pi)},$$

where $k = a(1 - e^2)$. (Consistent with our previous notation, we use θ and θ'' to denote the true anomalies at times τ and τ'', respectively; similarly for v and v''.) With this notation, we note that the area of the ellipse is $\pi a^{\frac{3}{2}}\sqrt{k}$ or $\pi a^2\sqrt{1-e^2}$, letting the context distinguish between our two uses of the symbol π.

According to *Kepler's Second Law*, the vector from the sun to a planet sweeps out area at a constant rate. If α denotes the area of the elliptical sector determined by the two vectors $\mathbf{r}$ and $\mathbf{r}''$, $\Delta\tau$ denotes the elapsed time between observations, and t denotes the period of the planet's orbit, Kepler's Second Law gives us

$$\frac{\alpha}{\Delta\tau} = \frac{\pi a^{\frac{3}{2}}\sqrt{k}}{t}.$$

If we denote the period and semimajor axis for the Earth's orbit by T and A, respectively, *Kepler's Third Law* tells us that $\frac{t^2}{a^3} = \frac{T^2}{A^3}$. Using $T = 365.25$ (days) and

$A = 1$ (astronomical unit), and approximating α from $\mathbf{r}$ and $\mathbf{r}''$ with the trapezoidal rule applied to a polar area integral, k can be determined.

The value of the ratio $\dfrac{\cos\theta''}{\cos\theta} = \dfrac{\cos(\theta + \theta'' - \theta)}{\cos\theta}$ can be obtained from the ellipse equations

$$e\cos\theta = \frac{k}{\|\mathbf{r}\|} - 1 \quad \text{and} \quad e\cos\theta'' = \frac{k}{\|\mathbf{r}''\|} - 1,$$

and the angle $\theta'' - \theta$ can be found from $\mathbf{r}\cdot\mathbf{r}'' = \|\mathbf{r}\|\,\|\mathbf{r}''\|\cos(\theta'' - \theta)$. Together, these can be solved for θ using the cosine addition formula. Next, ω can be obtained from $\mathbf{r}\cdot\langle\cos\Omega, \sin\Omega, 0\rangle = \|\mathbf{r}\|\cos(\theta + \omega)$.

Once θ and ω are known, π, e and a can easily be determined from the preceding relationships.

The elements i and Ω determine the orbital plane, and π, a, and e determine the shape and orientation of the elliptical orbit within that plane. Thus the geometry of the orbit in space is completely determined, but we still do not know *where* on the orbit the planet is located at a particular time. For this we need an initial condition, namely τ_p, the time (i.e., date and hour) of perihelion passage. To determine τ_p, we introduce yet another term from astronomy: in Fig. 2, a circle of radius a is

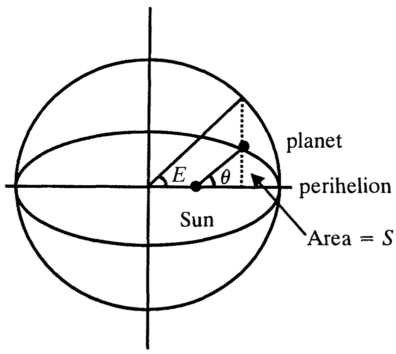


FIGURE 2
The true anomaly θ and eccentric anomaly E

circumscribed around the elliptical orbit with semimajor axis of length a, with the centers of the two coinciding. As noted above, θ is the true anomaly, whereas the angle E is known as the *eccentric anomaly*. These are related by

$$\tan\left(\frac{\theta}{2}\right) = \left(\frac{1+e}{1-e}\right)^{\frac{1}{2}}\tan\left(\frac{E}{2}\right),$$

which one derives by applying the half angle formula for tangent to angles E and θ in Fig. 2. The eccentric anomaly E can in turn be used to compute τ_p via *Kepler's equation* $E - e\sin E = \dfrac{2\pi}{t}(\tau - \tau_p)$. This latter equation is an immediate consequence of Kepler's second law: in Fig. 2, the area S of the elliptical sector is related to τ_p by

$$\frac{S}{\tau - \tau_p} = \frac{\pi a^2\sqrt{1-e^2}}{t}.$$

Substituting the computed area $S = \frac{1}{2}a^2\sqrt{1-e^2}\,(E - e\sin E)$ (found by integration), one obtains Kepler's equation.

The interested reader will find more details on the preceding computations in any orbital mechanics text such as [**8**] or [**9**]. For a concise, elementary treatment of computing the orbital elements from $\mathbf{r}$ and $\mathbf{r}''$, including derivations of all of the preceding formulas as well as computational examples of their use, see [**10**].

Gauss's method in computing the orbit of Ceres

The computation of the orbital elements as described above was well known in 1801. The problem, of course, is that through telescopic observations alone we cannot determine the vectors $\mathbf{r}$ and $\mathbf{r}''$; instead, we are only able to determine the geocentric (Earth-centered) longitude and latitude of the planet, and no information whatsoever about distance. Thus, the problem that confronted Gauss in 1801 was the following: from three geocentric observations (longitude and latitude) of a planet, determine two heliocentric vectors approximating the planet's position at two different times. From these two heliocentric vectors, the six orbital elements for the planet can be determined as previously outlined. We now describe the method that Gauss used in the *Summarische Übersicht* to solve this problem.

We begin by introducing two coordinate systems as shown in Fig. 3. The following notation is that of the *Summarische Übersicht*, with some exceptions which will be

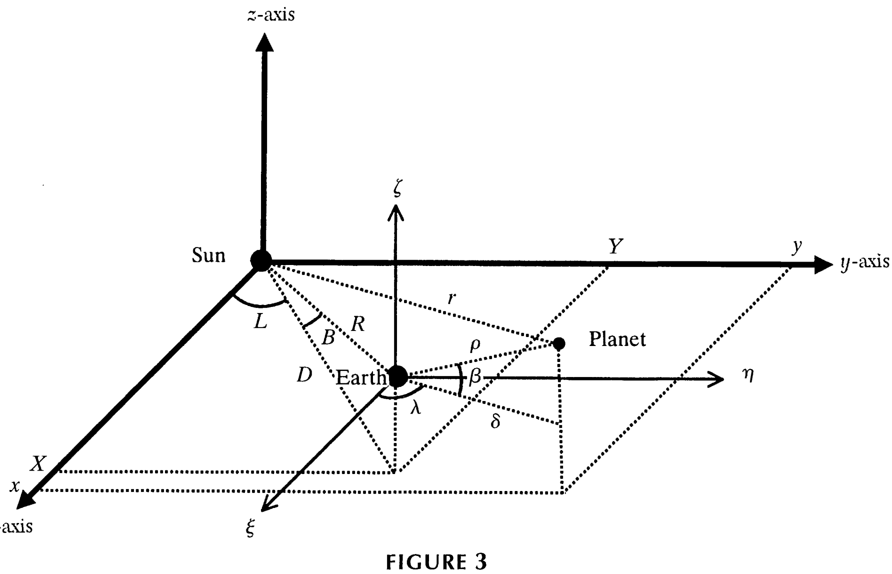


FIGURE 3
The xyz (heliocentric) and $\xi\eta\zeta$ (geocentric) coordinate systems

noted as they occur. Uppercase letters will consistently refer to the Earth, and lowercase letters to the planet. In the heliocentric coordinate system, the positions of the Earth and the planet at time τ are (X, Y, Z) and (x, y, z), respectively. The planet's geocentric coordinates are (ξ, η, ζ), and the two coordinate systems are related by $\xi = x - X$, $\eta = y - Y$, and $\zeta = z - Z$. Alternatively, one can describe the Earth's position using the heliocentric longitude L, latitude B, and distance R; similarly the planet's position is described by the geocentric longitude λ, latitude β, and distance $\rho = \delta \sec \beta$. (It is the values of λ and β that one determines by telescopic observation.) By adding a single prime (X', x', L', λ', etc.) or double prime

(X'', x'', L'', λ'', etc.), these symbols represent the corresponding quantities at times τ' and τ''.

To complete the notation, we let f denote the area of the triangle formed by the sun, the planet at time τ', and the planet at time τ'', and let g denote the area of the corresponding sector of the elliptical orbit. Similarly, $-f'$ and $-g'$ denote the areas of the triangle and sector corresponding to times τ and τ'', and f'' and g'' denote the areas corresponding to times τ and τ'. Finally F, $-F'$, F'', G, $-G'$, and G'' are analogously defined for the Earth.

Though Gauss did not have modern matrix notation at his disposal, it will be convenient for us to set

$$\phi = \begin{pmatrix} x & x' & x'' \\ y & y' & y'' \\ z & z' & z'' \end{pmatrix} \quad \text{and} \quad \mathbf{f} = \begin{pmatrix} f \\ f' \\ f'' \end{pmatrix}.$$

Since the columns $\mathbf{r}$, $\mathbf{r}'$, and $\mathbf{r}''$ of ϕ all lie in the orbital plane, there are constants c_1 and c_2 such that $\mathbf{r} = c_1\mathbf{r}' + c_2\mathbf{r}''$. Then $\mathbf{r} \times \mathbf{r}' = c_2(\mathbf{r}'' \times \mathbf{r}')$ and $\mathbf{r} \times \mathbf{r}'' = c_1(\mathbf{r}' \times \mathbf{r}'')$, with $c_1 > 0$ and $c_2 < 0$. (These cross products are all perpendicular to the orbital plane with $\mathbf{r}'' \times \mathbf{r}'$ directed opposite the others.) Using $\|\mathbf{r}' \times \mathbf{r}''\| = 2f$, $\|\mathbf{r} \times \mathbf{r}''\| = -2f'$, and $\|\mathbf{r} \times \mathbf{r}'\| = 2f''$, we have $f'' = -c_2 f$ and $-f' = c_1 f$, so that $\mathbf{r} = -\frac{f'}{f}\mathbf{r}' - \frac{f''}{f}\mathbf{r}''$. Therefore $f\mathbf{r} + f'\mathbf{r}' + f''\mathbf{r}'' = 0$, or equivalently, $\phi\mathbf{f} = 0$. Analogously, for

$$\Phi = \begin{pmatrix} X & X' & X'' \\ Y & Y' & Y'' \\ Z & Z' & Z'' \end{pmatrix} \quad \text{and} \quad \mathbf{F} = \begin{pmatrix} F \\ F' \\ F'' \end{pmatrix},$$

we have $\Phi\mathbf{F} = 0$, from which it follows that

$$(F + F'')(\phi - \Phi)\mathbf{f} = \Phi((f + f'')\mathbf{F} - (F + F'')\mathbf{f}). \tag{$*$}$$

Next, we transform equation $(*)$ by introducing spherical coordinates. Referring again to Fig. 3, define $\pi = \langle \cos\lambda, \sin\lambda, \tan\beta \rangle$. (Here Gauss presents a severe handicap to the reader: he defines π in his carefully laid out list of symbols just as we previously defined it, i.e., as the longitude of aphelion, but his first use of π is as we are using it here. Let the context distinguish among the various uses in the remainder of this paper!) Then

$$\delta\pi = \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = \begin{pmatrix} x - X \\ y - Y \\ z - Z \end{pmatrix},$$

which is the first column of $\phi - \Phi$. Similarly, the second and third columns are $\delta'\pi'$ and $\delta''\pi''$, respectively. In the same fashion, let $P = \langle \cos L, \sin L, \tan B \rangle$, so that the columns of Φ are DP, $D'P'$, and $D''P''$. If equation $(*)$ is left-multiplied by the 4×3 matrix whose rows are $\delta''\pi'' \times \delta\pi$, $D'P' \times \delta\pi$, $D'P' \times \delta'\pi'$, and $D'P' \times \delta''\pi''$

(a straightforward but tedious task!), the resulting set of four equations is

$$(F+F'')f'\delta'[\pi\pi'\pi''] = (Ff''-F''f)(D[\pi P\pi''] - D''[\pi P''\pi''])$$
$$+((f+f'')F' - (F+F'')f')D'[\pi P'\pi''] \tag{1}$$

$$(F+F'')(f'\delta'[\pi\pi'P'] + f''\delta''[\pi\pi''P']) = (Ff''-F''f)(D[\pi PP'] - D''[\pi P''P']) \tag{2}$$

$$(F+F'')(f\delta[\pi'\pi P'] + f''\delta''[\pi'\pi''P']) = (Ff''-F''f)(D[\pi' PP'] - D''[\pi' P''P']) \tag{3}$$

$$(F+F'')(f\delta[\pi''\pi P'] + f'\delta'[\pi''\pi'P']) = (Ff''-F''f)(D[\pi'' PP'] - D''[\pi'' P''P']). \tag{4}$$

Here Gauss's original notation $[\mathbf{a}\,\mathbf{b}\,\mathbf{c}]$ denotes the determinant of the matrix whose columns are $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$; equivalently, it is the triple scalar product $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$. Common factors of $\delta\delta''$, $\delta D'$, $\delta' D'$, and $\delta'' D'$ have been divided out of equations (1)–(4), respectively.

Some comments are appropriate here. Equations (1)–(4) appear above precisely as in the *Summarische Übersicht*. Gauss does not use equation (3) in the remaining development. His derivation of these equations involves no matrices, and although he uses the determinants $[\pi\pi'\pi'']$ etc., it is interesting to note that much of the modern theory of determinants was developed after Gauss's paper appeared **[1]**. Gauss makes up for the cumbersome notation available to him by simply presenting the main results with little or no clue as to the computations behind them.

Next, Gauss writes, "we now want to examine these four equations, which are precisely true, more closely in order to build the first approximation on them." To this end, he argues that in equations (2) and (4), the left side is $\mathcal{O}(t^3)$, whereas the right side is $\mathcal{O}(t^5)$ (or in Gauss's words, "If we view the intervening times as infinitely small quantities of the first order, ... what stands on the right of the second, third, and fourth equations above is of fifth order..."). Thus, by setting the right side equal to zero, "one can, as the first approximation, set

from 2) $\quad f'\delta'[\pi\pi'P'] = -f''\delta''[\pi\pi''P']$

from 4) $\quad f\delta[\pi\pi''P'] = -f'\delta'[\pi'\pi''P']$."

Solving these equations for δ and δ'', and using Kepler's second law in the form

$$\frac{g}{\tau''-\tau'} = \frac{-g'}{\tau''-\tau} = \frac{g''}{\tau'-\tau},$$

Gauss obtains

$$\delta = \frac{g}{f}\frac{f'}{g'}\frac{\tau''-\tau}{\tau''-\tau'}\frac{[\pi'\pi''P']}{[\pi\pi''P']}\delta' \tag{5}$$

and

$$\delta'' = \frac{g''}{f''}\frac{f'}{g'}\frac{\tau''-\tau}{\tau'-\tau}\frac{[\pi\pi'P']}{[\pi\pi''P']}\delta'. \tag{6}$$

The ratios $\frac{f}{g}$, $\frac{f'}{g'}$, and $\frac{f''}{g''}$ can all be approximated by one, as Gauss apparently did in his earliest work (see **[4]**, p. 156 and **[7]**, pp. 16–17). All other right side quantities are known except δ'. Thus, if δ' can be found, we will have two complete geocentric positions $\langle \delta\cos\lambda, \delta\sin\lambda, \delta\tan\beta \rangle$ and $\langle \delta''\cos\lambda'', \delta''\sin\lambda'', \delta''\tan\beta'' \rangle$, from which two heliocentric positions $\mathbf{r}$ and $\mathbf{r}''$ can easily be obtained.

Next, Gauss develops the main result of his paper from (1). Letting r, r', and r'' represent the lengths of the planet's heliocentric position vectors $\mathbf{r}$, $\mathbf{r}'$, and $\mathbf{r}''$, the polar equation of the elliptical orbit gives us

$$\frac{1}{r} = \frac{1}{k}(1 - e\cos(v - \pi)), \quad \frac{1}{r'} = \frac{1}{k}(1 - e\cos(v' - \pi)), \quad \text{and}$$
$$\frac{1}{r''} = \frac{1}{k}(1 - e\cos(v'' - \pi)).$$

After multiplying these equations by $\sin(v'' - v')$, $\sin(v - v'')$, and $\sin(v' - v)$, respectively, then adding, the result can be rewritten using the triangle area formulas $f = \frac{1}{2}r'r''\sin(v'' - v')$, $f' = \frac{1}{2}rr''\sin(v - v'')$, and $f'' = \frac{1}{2}rr'\sin(v' - v)$ and the identity $\sin\gamma + \sin\psi - \sin(\gamma + \psi) = 4\sin\frac{\gamma}{2}\sin\frac{\psi}{2}\sin\frac{\gamma+\psi}{2}$. The terms of the form $e\cos\gamma\sin\psi$ add to zero, producing

$$\frac{f + f' + f''}{f'} = \frac{-2r'}{k}\frac{\sin\frac{1}{2}(v'' - v')\sin\frac{1}{2}(v' - v)}{\cos\frac{1}{2}(v'' - v)}. \tag{7}$$

Next, using the fact that the planet's elliptical orbit has area $\pi a^{\frac{3}{2}}\sqrt{k}$, Kepler's second law gives us

$$\frac{\pi a^{\frac{3}{2}}\sqrt{k}}{t} = \frac{g}{\tau'' - \tau'} = \frac{g''}{\tau' - \tau},$$

from which Kepler's third law gives

$$\frac{\pi^2 A^3 k}{T^2} = \frac{\pi^2 a^3 k}{t^2} = \frac{gg''}{(\tau'' - \tau')(\tau' - \tau)}. \tag{8}$$

If M, M', and M'' describe the Earth's angular displacement from perihelion at times τ, τ', and τ'' under the assumption of *constant angular velocity*, then

$$\frac{2\pi}{T} = \frac{M' - M}{\tau' - \tau} = \frac{M'' - M'}{\tau'' - \tau'},$$

and so we obtain from (8)

$$k = \frac{T^2 gg''}{\pi^2 A^3(\tau'' - \tau')(\tau' - \tau)} = \frac{4gg''}{A^3(M'' - M')(M' - M)}. \tag{9}$$

(M is actually defined by $M = \frac{2\pi}{T}(\tau - \tau_p)$ and is called the *mean anomaly*.) Equation (9) and the small-angle approximations

$$\cos\tfrac{1}{2}(v'' - v) \approx 1, \quad g \approx r'r''\sin\tfrac{1}{2}(v'' - v'), \quad g'' \approx rr'\sin\tfrac{1}{2}(v' - v), \quad \text{and} \quad rr'' \approx r'^2$$

allow us to transform (7) into the approximation

$$\frac{f + f' + f''}{f'} = -\frac{A^3}{2r'^3}(M' - M)(M'' - M'). \tag{10}$$

Analogously, one obtains the approximation

$$\frac{F+F'+F''}{F'} = -\frac{A^3}{2R'^3}(M'-M)(M''-M'). \tag{11}$$

To complete the method, Gauss now turns to equation (1), arguing that the left side is $\mathscr{O}(t^5)$, while the two terms on the right side are $\mathscr{O}(t^7)$ and $\mathscr{O}(t^5)$, respectively. The first term on the right side of (1) is dropped, and (10) and (11) are used to transform (1) into

$$(F+F'')f'\delta'[\pi\pi'\pi''] = f'F'\frac{A^3}{2}(M'-M)(M''-M')\left(\frac{1}{R'^3}-\frac{1}{r'^3}\right)D'[\pi P'\pi''].$$

Using an approximation which is effectively $\frac{F'}{F+F''} \approx -1$, Gauss rewrites this equation as

$$\frac{[\pi\pi'\pi'']}{[\pi P'\pi'']}\frac{2}{A^3(M'-M)(M''-M')} = -\left(\frac{1}{R'^3}-\frac{1}{r'^3}\right)\frac{D'}{\delta'}, \tag{12}$$

which he describes as "the most important part of the entire method and its first foundation." Finally, by taking the xy-plane to be the ecliptic plane (so that $D' = R'$), approximating A by R', and computing the given determinants, (12) becomes

$$\left(1-\left(\frac{R'}{r'}\right)^3\right)\frac{R'}{\delta'} = \frac{-2}{(M'-M)(M''-M')} \times \frac{\tan\beta'\sin(\lambda''-\lambda) - \tan\beta\sin(\lambda''-\lambda') - \tan\beta''\sin(\lambda'-\lambda)}{\tan\beta\sin(L'-\lambda'') - \tan\beta''\sin(L'-\lambda)}. \tag{13}$$

(Equation (12), as published in the Gauss *Werke*, differs by a minus sign and has an R instead of a D' on the right side. Also, in the *Werke*, (13) differs by a minus sign and has an A^3 in the denominator of the right side. These errors are corrected without comment in [7], where equations equivalent to (12) and (13) are given.)

The right side of (13) is computed from observational data, and the left side can be reduced to the single variable $\frac{R'}{\delta'}$ by means of

$$\frac{R'}{r'} = \frac{R'}{\delta'}\left(1+\tan^2\beta' + \left(\frac{R'}{\delta'}\right)^2 + 2\frac{R'}{\delta'}\cos(\lambda'-L')\right)^{-\frac{1}{2}},$$

which is obtained by applying the law of cosines to the triangle with vertices at $(0,0,0)$, $(x',y',0)$, and $(X',Y',0)$. Solving (13) (numerically) for $\frac{R'}{\delta'}$, and using the known value of R', one obtains a value for δ'; equations (5) and (6) now give values for δ and δ''. The geocentric position vectors $\langle\delta\cos\lambda, \delta\sin\lambda, \delta\tan\beta\rangle$ and $\langle\delta''\cos\lambda'', \delta''\sin\lambda'', \delta''\tan\beta''\rangle$ of the planet can now be added to the Earth's position vectors to obtain two complete heliocentric vectors $\mathbf{r}$ and $\mathbf{r}''$ describing the planet's positions at times τ and τ'', as desired. With this, the problem is solved.

What a magnificent achievement! Though the mathematical tools used are not particularly sophisticated, all sense of the motivating geometry is lost very early in this work, leaving one to wonder what led Gauss through the rather extraordinary computations needed to achieve his goal. Perhaps no less than a "super-terrestrial spirit in a human body" could have done it!

REFERENCES

1. E. T. Bell, *The Development of Mathematics*, McGraw Hill, New York, NY, 1940.
2. W. K. Bühler, *Gauss, A Biographical Study*, Springer-Verlag, New York, NY, 1981.
3. G. W. Dunnington, *Carl Friedrich Gauss: Titan of Science*, Exposition Press, New York, NY, 1955.
4. C. F. Gauss, Summarische Übersicht der Bestimmung der Bahnen der beiden neuen Hauptplaneten angewandten Methoden, *Werke*, v. 6, pp. 148–165.
5. C. F. Gauss, *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*, translation by C. H. Davis, Dover, New York, NY, 1963.
6. T. Hall, *Carl Friedrich Gauss*, MIT Press, Cambridge, MA, 1970.
7. F. Klein and M. Brendel, *Materialien für eine wissenschaftliche Biographie von Gauss*, B. G. Teubner, Leipzig, 1911–1920.
8. J. Prussing and B. Conway, *Orbital Mechanics*, Oxford University Press, New York, NY, 1993.
9. A. E. Roy, *Orbital Motion*, 2nd edition, Adam Hilger, Ltd., Bristol, UK, 1982.
10. D. Teets and K. Whitehead, The computation of planetary orbits, *The College Mathematics Journal* 29 (1998), pp. 397–404.
11. X. von Zach, Fortgesetzte Nachrichten über den Längst Vermutheten Neuen Haupt-Planeten Unseres Sonnen-Systems [Ceres], *Monatliche Correspondenz zur Beförderung der Erd-und Himmels-Kunde* **III** (1801).

A Web-Searchable Database for *Mathematics Magazine*

We are happy to inform readers of the Magazine's useful new Web-searchable database of Articles, Notes, Proofs Without Words, and more. The database can be found through MAA Online, at

`http://www.maa.org`

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

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

The database contains the title, author, and first paragraph (and, in many cases, author's summary) of almost every Article, Note, and Proof Without Words published in the *Magazine*'s inception, in 1927. An easy-to-use search engine allows users to search for specific lists of records, or to browse the full contents of any single issue.

Marriages Made in the Heavens: A Practical Application of Existence

DAN KALMAN
American University
Washington, DC 20016

Introduction

Existence results are familiar to mathematicians, who understand their theoretical significance. But others, including students, are sometimes perplexed by existence results. "What good is it to know something exists," they wonder, "when you have no idea how to find it?" The ultimate answer depends on some intrinsic appreciation of abstract mathematics—not all of mathematics can or should be justified on the basis of practical application. Still, abstract existence results sometimes have practical consequences, as this paper aims to demonstrate. Though the focus is on existence results, the demonstration itself is constructive: I will describe a problem I worked on in the aerospace industry that made thoroughly practical use of an existence result. Specifically, an existence theorem associated with the marriage problem for bipartite graphs was applied to a satellite communications network, matching orbiting satellites with ground stations. A briefer account, which omits most of the mathematical details, can be found in **[2]**.

The allocation problem

The problem setting is a preliminary design study for a satellite communication system. Many variables influence the design of such systems: the number of satellites, the orbits they occupy, details of the communications equipment, power requirements, etc. Normally, the complete set of satellite orbits, referred to as a constellation, is considered in total, rather than focusing on individual orbits one at a time. I was part of a team studying the effects of the constellation design on system performance.

For this preliminary design study we used highly idealized models. The earth is represented by a sphere rotating uniformly once every 24 hours about a fixed axis

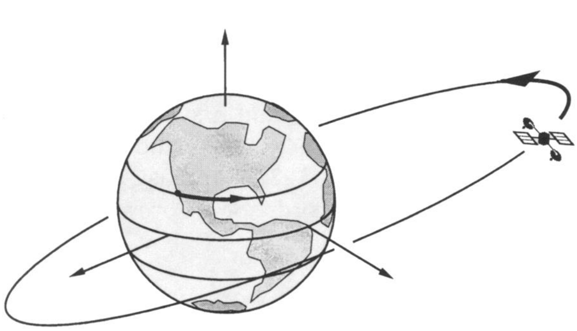


FIGURE 1
Geometrical model.

through the poles. Satellites and radio stations on the ground are points, specified in a Cartesian coordinate frame. The origin of the coordinate frame is at the center of the earth; the equator lies in the xy-plane; the axis of rotation is the z-axis. In this system, satellites travel in Keplerian orbits dictated by an inverse square force law, and we can calculate each satellite's position at any time. The ground-based radio stations move with the rotating earth, so they can also be located at any time. The operational demands on the system are assumed to be constant, with each station handling a fixed volume of message traffic in any 24-hour period.

Visibility The concept of visibility is crucial to the model. A satellite and radio station are *visible* to one another if they can communicate. In the simplest models visibility is interpreted literally, as an unobstructed line of sight. More complicated models take into account the geometry of the station and satellite, as well as physical constraints on radio transmission. These can include signal loss due to atmospheric conditions, and the sensitivity of antennas to the direction of arrival of a signal.

In these more complicated models, visibility is described in geometric terms. For example, we envision the antennas on board satellites and fixed to the ground as having a conical field of view. An arriving signal must fall within the cone to be visible. In more elaborate models the field of view can be more complicated than a simple cone, as shown in exaggerated form in FIGURE 2. To be visible from a station in the

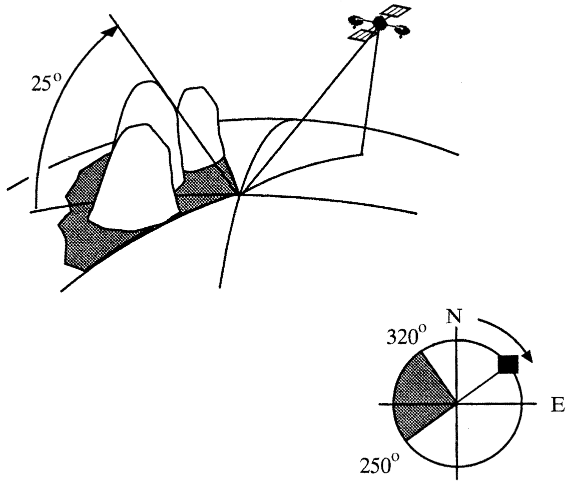


FIGURE 2
Visibility constraints.

figure, a satellite must be within the antenna's cone, but also above the mountains to the west. In this case we can define visibility as follows: Project the line joining the satellite and station onto the local horizontal plane at the station, thereby defining the satellite's heading. Measuring clockwise from due north, for headings between 250 and 320 degrees the satellite's elevation must be at least 25 degrees. For any other heading, the satellite is visible if its elevation is at least 5 degrees.

In all of these models, visibility is determined by simple geometrical relationships. At any specific time, using the instantaneous positions of the satellite and radio station, as well as the geometric constraints, one can calculate using vector analysis whether the satellite is visible to the station. To obtain an overview of the system's behavior, the calculations are repeated for many specific times, defining a *discrete time model*.

Discrete time model Although satellite and earth motions are continuous, it is reasonable to analyze them by considering a discrete set of times. Assume, for instance, that the visibility computations are performed once for each minute of the simulation. Assume, too, that a satellite can communicate for one minute with any radio station that is visible, but only with one station during that minute. On the other hand, each radio station can communicate with several satellites at once. (This asymmetry reflects the fact that we can position multiple receivers and transmitters on the ground at little expense, while resources on satellites are very limited.)

Our model also includes a predefined quota of connect-time for each radio station. These quotas are based on a projected volume of communications traffic at each station and may differ from station to station. The first station may require 90 minutes of connect-time during the simulation, the next station 30 minutes, and so on. Now we can state the fundamental problem:

> *ALLOCATION PROBLEM. At each time step, assign each satellite to one visible radio station so that, over the course of the simulation, each station achieves its quota of connect-time.*

Graph theory formulation

The allocation problem may be reformulated in terms of graph theory. To begin, consider FIGURE 3, which shows a satellite at three different times, and several radio

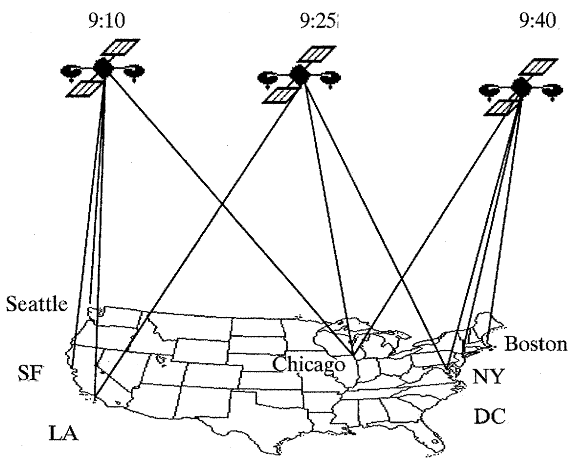


FIGURE 3
Visibility for several time steps.

stations. The lines represent visibility: at 9:10 the satellite is visible to Seattle, San Francisco, Los Angeles, and Chicago. By 9:25 the satellite has lost sight of Seattle and San Francisco, but can now see Washington, DC.

It is natural to abstract away the geographical map, leaving only a *bipartite graph* (FIGURE 4). There are two sets of vertices, or nodes: one set for the satellite at different times and the other for radio stations; each edge joins a vertex in one set to a vertex in the other.

The simple example of FIGURE 4 shows how the allocation problem is reformulated using graph theory. Define a bipartite graph V, the *visibility graph*, as follows: (i)

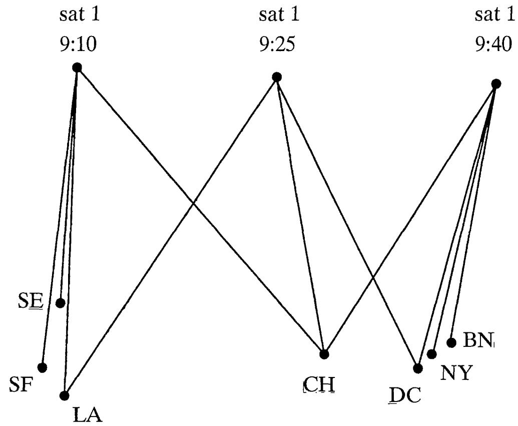


FIGURE 4
Sample visibility graph.

assign one vertex to each radio station; (ii) assign one vertex to each satellite at each of the discrete time steps; (iii) let there be an edge between the vertex for satellite s at time t and radio station r if and only if s and r are visible to one another at time t. This produces a bipartite graph, with vertices divided into two groups: satellite-time (ST)-vertices and radio station (R)-vertices.

A computer simulation is used to determine V. At each time step, positions are computed for each radio station and each satellite, and used to determine which satellites are visible to which radio stations. A typical simulation might involve 24 hours (1440 minutes), 10 satellites, and 10 radio stations. This produces over 14000 ST-vertices, only 10 R-vertices, and up to 140000 edges in the visibility graph—a result far more complicated than FIGURE 4 depicts. FIGURE 5 comes a little closer to the true situation, but it is clearly hopeless to portray anything like the true complexity of the problem.

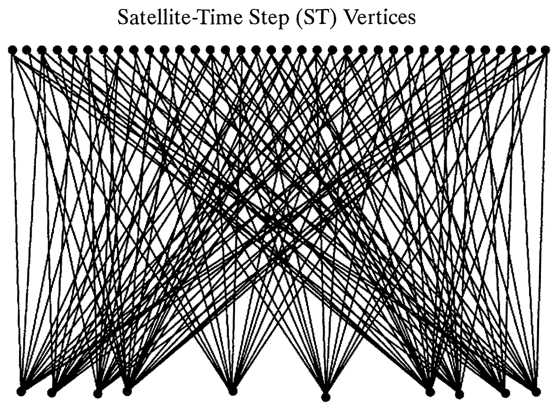


FIGURE 5
Detailed visibility graph.

Assignment subgraphs Part of the allocation problem is to assign each satellite at each time step to just *one* radio station. In the graph-theoretic formulation, that means selecting just one edge emanating from each ST-vertex, thus defining a subgraph, which we call an *assignment subgraph.* The bold lines in FIGURE 6 show one possible assignment subgraph for the graph of FIGURE 5.

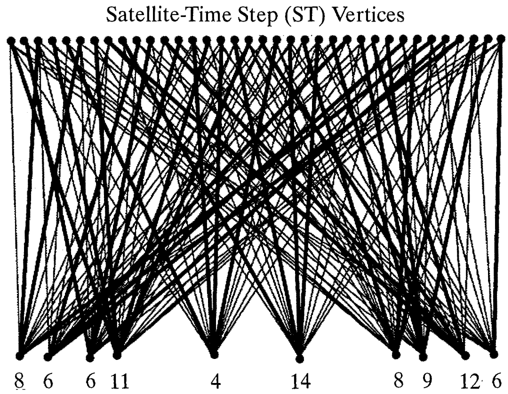


FIGURE 6
Assignment subgraph.

The number of edges in a graph G incident at a vertex v is called the *degree* of v in G, and denoted $\deg(v, G)$. An assignment subgraph A of the visibility graph V is thus characterized by the requirement that $\deg(st, A) \leq 1$ for every ST-vertex st.

The assignment problem has another requirement: each radio station must be connected to satellites for a predetermined quota of minutes. In the assignment subgraph, each edge represents one minute of connect-time. Thus the connect-time quota for a radio station r dictates a minimum number of incident edges (i.e., the degree) at r in the assignment subgraph.

Connect time quotas are shown for each R-vertex in FIGURE 6. The assignment subgraph shown clearly fails to satisfy the allocation problem because, for example, the first R-vertex has 3 incident bold edges—less than the quota of 8.

We can now state the allocation problem in graph-theoretic terms.

ALLOCATION GRAPH PROBLEM: *Given a visibility graph V and a connect-time quota $q(r)$ for each R-vertex, find an assignment subgraph A such that* $\deg(r, A) \geq q(r)$ *for every R-vertex r.*

It is not obvious at the outset whether this kind of problem is solvable. Next we will develop some necessary conditions for solvability. These conditions will lead to an existence theorem—our desired existence result—that characterizes the solvability of allocation graph problems.

Necessary conditions

It is easy to see that the sample assignment subgraph in FIGURE 6 does not solve the allocation problem, but it is less obvious whether *any* solution exists. A closer look at FIGURE 6 reveals that each R-vertex has far too few incident bold lines. This suggests that, no matter how we rearrange the assignments, we will be unable to fill the quota at each R-vertex. This is indeed so, and there is a simple proof. Do you see it?

Because at most one edge in the assignment subgraph can meet each ST-vertex, the total number of edges cannot exceed the number of ST-vertices, 36. The allocation problem requires enough bold edges to match the number beneath each R-vertex. Since the sum of those numbers is much greater than 36, the number of bold edges available, this allocation problem is unsolvable.

The actual simulations that I worked with had much more complicated graphs than that in FIGURE 6, but the same reasoning applies. If the total number of ST-vertices is less than the sum of all the connect-time quotas, then the allocation problem will be unsolvable. This gives a necessary condition for solvability: The sum of all the connect-time quotas $q(r)$ cannot exceed the number of ST-vertices.

Other necessary conditions arise just as naturally. The graph in FIGURE 7, for instance, admits no solution to the allocation problem, because only one edge in V is

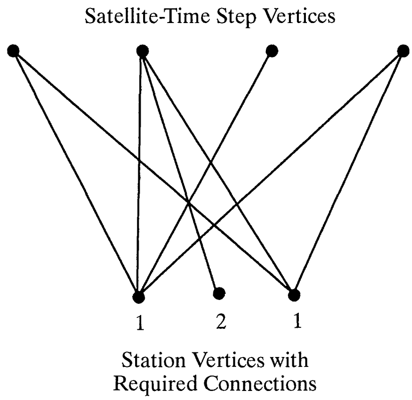


FIGURE 7
Required degree too high at one station.

incident at the second R-vertex, but the allocation problem requires two edges there. This suggests a second necessary condition for the allocation problem to be solvable: $q(r) \leq \deg(r, V)$ for every R-vertex r.

FIGURE 8 illustrates yet another unsolvable allocation problem. As before, a general principle is at work, but this time a little more subtly. To see how it works, consider

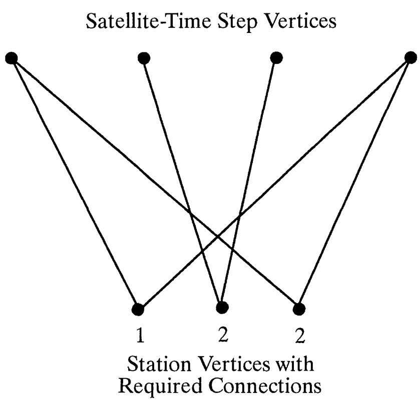


FIGURE 8
Total required degrees too high for a set of stations.

only the part of the graph that involves the first and last R-vertices. If we ignore the middle R-vertex, we may as well ignore the two middle ST-vertices as well. What remains are two ST-vertices, two R-vertices, and 4 edges. The allocation problem is clearly unsolvable in this subgraph, for the sum of the required degrees (3) exceeds the number of ST nodes (2). Since the allocation problem cannot be solved for this subgraph, it cannot be solved for the original graph either.

A unifying principle The preceding example illustrates a more general principle. For any subset E of the R-vertices, consider the subgraph V_E consisting of the edges that touch elements of E, together with the endpoints of these edges. Let $N_{ST}(V_E)$ be the number of ST-vertices in V_E. For the original allocation problem to be solvable, we must have, for every non-empty subset E of R-vertices,

$$\sum_{r\in E} q(r) \leq N_{ST}(V_E). \tag{1}$$

Note that this one principle subsumes both of the preceding examples. If E is the full set of R-vertices, then $V_E = V$, and the corresponding necessary condition is what was presented in the first example. If $E = \{r\}$, for any single R-vertex, then $N_{ST}(V_E) = \deg(r, V)$, and the necessary condition is as in the second example.

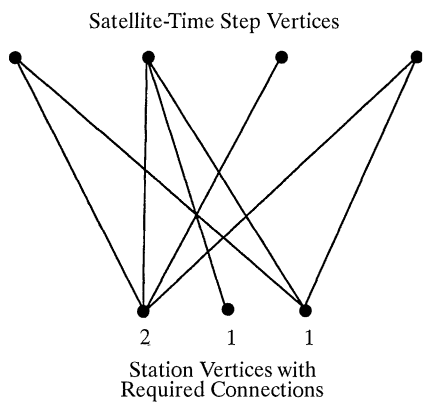


FIGURE 9
Required degrees feasible.

FIGURE 9 shows a graph in which the allocation problem *is* solvable. It is easy to check that the condition (1) holds for every possible E.

The existence theorem

The necessity of the solution condition (1) should now be intuitively clear. What is less clear, but still true, is that the solution condition is also sufficient: If (1) holds for every nonempty set E of R-vertices, then a solution to the allocation problem exists. This result is equivalent to Hall's theorem,[1] a classical result in graph theory. Hall's

[1]According to [3], Philip Hall deduced the result as a theorem in set theory in 1935.

theorem is often cited in the context of the so-called *marriage problem*; in [**4**, p. 159], the result is labeled *Hall's Marriage Theorem.*

The marriage problem is a special case of the allocation problem, with $q(r) = 1$ for every r. In its traditional formulation, the R-vertices represent maidens, the ST-vertices represent bachelors, and an edge indicates that a particular bachelor and maiden are acquainted with each other. The problem is for each bachelor to propose marriage to just one maiden in such a way that each maiden receives at least one proposal. It is assumed that there are just as many bachelors as maidens. Hall's theorem asserts that a solution exists if and only if every set of k maidens is connected to a set of at least k bachelors.

This condition is clearly necessary: if any k maidens are connected to fewer than k bachelors, there will not be enough bachelors to go around. Sufficiency may be proved by induction on the number of bachelors (and maidens). Briefly, the induction step can be handled by considering two cases. In the first case, each set of k maidens is connected to a set of at least $k+1$ bachelors. In this case, we match one bachelor to one maiden, and the induction hypothesis implies that everyone else can also be matched up. In the second case, some set of k maidens is connected to exactly k bachelors. In this case we first argue (by induction) that these k maidens and bachelors can be paired, and then argue (again by induction) that the remaining matches can be made as well.

The satellite allocation problem can be reduced to a marriage problem as follows. Create a new graph V_M by creating $q(r)$ duplicate vertices for each R-vertex r; each of these $q(r)$ vertices is connected by an edge to each ST-vertex that r was connected to in the original graph. If the marriage problem can be solved in V_M, then we will have $q(r)$ ST-vertices matched with the duplicate vertices for r; this specifies how to assign ST-vertices to r in V.

We must also assure that the number of R-vertices equals the number of ST-vertices in V_M. Observe that if there are too few ST-vertices, the original problem is not solvable. If there are too many ST-vertices, we can simply add enough R-vertices to make the two sets compatible, and assume that each of these new R-vertices is connected to every ST-vertex.

Now we see that the condition defined by (1) translates into the necessary and sufficient condition for solvability of the marriage problem. A solution to the marriage problem, however, induces a solution to the satellite allocation problem. It may happen that in formulating the marriage problem, some extra R-vertices were added. In this case, the matches of these extra R-vertices with ST-vertices in the solution of the marriage problem will be discarded in translating back to the satellite allocation problem. But the result will still be an assignment subgraph that meets all the connection-time quotas.

With (1) established as a necessary and sufficient condition for solvability of the allocation graph problem, deciding whether a satellite constellation is capable of meeting its performance objectives is reduced to a computation. Note that this computation does not provide any practical operational guidance: we still do not know how to assign the satellites to ground stations. The existence of a solution has a practical significance in a different direction, providing a metric of system performance. This idea is developed next.

A practical use for existence

The allocation problem arose in the consideration of various satellite constellations for a communications system. The existence result provided a way to compare different constellations. Beyond the simple observation that for some constellations the allocation problem is solvable and for others it is not, the existence result led to the determination of an optimal data transmission rate for each constellation. Here is how it worked.

First, the calculations necessary to check the conditions for solvability were incorporated into the computer simulation. With N R-vertices, there are $2^N - 1$ nontrivial choices for E; for each, it is necessary to tabulate the number of ST-vertices connected to E. Although the dependence on N is exponential, N was small enough in our problems to permit a direct calculation in reasonable time.

There are well-known algorithms for solving graph matching problems by reducing them to flow optimization problems.[2] However, for the visibility graphs encountered in the satellite design problem, these algorithms require more calculations (by several orders of magnitude) than simply checking the necessary and sufficient conditions for solvability. Thus, it was feasible to compute whether a solution existed—but not to find a solution.

Second, to the degree of accuracy of the model, it is reasonable to assume that the values $q(r)$ are inversely proportional to the rate of data transmission. For example, doubling the transmission rate should halve the amount of connect time required.

As a general rule, faster data transmission is more expensive. This is certainly so for computer modems, and it applies even more stringently to satellites. The intrinsic cost of faster data rates is compounded by increased power requirements, which generally translate into greater weight and complexity of the satellite. So it is of interest to estimate the minimal feasible data rate for a satellite constellation, defined as the lowest data rate for which the assignment problem is solvable.

Note here that changing the data rate has no effect on the visibility graph—it simply increases or reduces the values of the $q(r)$. If a very high data rate is set, the values of the $q(r)$ will be low, and it is likely that the allocation problem can be solved. With a low data rate, the allocation problem is harder to solve.

Here, then, is how to estimate the minimal feasible data rate. Run the simulation once to compute V.' Select an initial choice for the data rate and, using the existence result, determine whether the allocation problem is solvable. If it is, lower the transmission rate; otherwise, raise the transmission rate. Then check the conditions for solvability again. Repeating this process a few times will usually establish the minimal feasible transmission rate within a few percent.

Conclusion

Ultimately, the minimal feasible transmission rate became just one of many criteria that were used to compare competing satellite system designs. I spent considerable time running the simulation and computing the optimal transmission rate for many satellite constellations, and that contributed to a much larger tradeoff analysis. In the process, solutions to satellite allocation problems were neither desired nor obtained. Much later in the design process, algorithms would have to be developed to actually assign the satellites to communicate with particular radio stations at particular times, and in all likelihood, these assignments would not be optimal. But at the point of the

[2]See [**1**, **4**]; the first of these includes an analysis of the computational complexity of the algorithms.

design process I have been describing, that level of detail was not required. Rather, the theoretical solvability of the optimal allocation problem was used to determine one measure of system performance. And that is how an existence result was used in a completely practical setting.

REFERENCES

1. Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*, MIT Press, Cambridge, MA, 1990, 587–599.
2. Dan Kalman. When existence is enough, *Math Horizons*, April 1997, 23–26.
3. Oystein Ore and Robin J. Wilson. *Graphs and Their Uses*, New Mathematical Library, Vol. 34, Mathematical Association of America, Washington, DC, 1990, 53–59.
4. Alan Tucker. *Applied Combinatorics*, John Wiley & Sons, New York, NY, 1995, 137–167.

REVIEWS (continued from page 157)

Shulman, Polly, From Muhammad Ali to Grandma Rose, *Discover* 19 (12) (December 1998) 85–89; `http://www.discover.com/decissue/smallworld.html` . Peterson, Ivars, Close connections: It's a small world of crickets, nerve cells, computers, and people, *Science News* 154 (22 August 1998) 124–126. Harris, John M., and Michael J. Mossinghoff, The eccentricities of actors, *Math Horizons* (February 1998) 23–25. Strogatz, Steven H., and Duncan J. Watts, Collective dynamics of "small-world" networks, *Nature* 393 (4 June 1998) 440–442.

Modern folklore claims that everyone on the planet is connected to everyone else through a relatively short chain of acquaintances. This claim was investigated by psychologist Stanley Milgram in the 1960s, concretized in the play and film *Six Degrees of Separation* by John Guare, and popularized in a game of trying to trace connections of actors to the actor Kevin Bacon through joint appearances in films. Mathematicians are familiar with a similar phenomenon, the *Erdős number* of a mathematician. We are talking eccentricity here; in graph theory, the maximum distance from a particular vertex to any other vertex in a graph is the *eccentricity* of the vertex. Harris and Mossinghoff document that the eccentricity of Kevin Bacon in the "Hollywood" graph is 7, which is the minimum eccentricity of any actor, putting Bacon into what is known as the *center* of the graph. Such networks are neither regular (every node has the same small number of links to neighboring points) nor sparse (few connections relative to the number of nodes). Strogatz and Watts showed that introducing a few random connections into a regular graph can greatly decrease the average path length between two nodes. Graphs with a small average path length they call *small-world networks*, and they cite as examples the neural network of the worm *C. elegans*, the power grid of the western U.S., and the Hollywood graph. Small-world networks are important in the spread of disease, the diffusion of trade goods, and the transmission of information (including marketing over the Internet), as Peterson notes.

Petković, Miodrag, *Mathematics and Chess: 110 Entertaining Problems and Solutions*, Dover, 1998; v + 133 pp, \$5.95 (P). ISBN 0-486-29432-3.

This book uses chess and the chessboard as occasions for mostly mathematical puzzles, though some chess puzzles occur too. Naturally, rook polynomials occur, as do knight's tours and their generalizations, plus domino coverings, dissections, and generalized chessboards. The puzzles are fun, solutions are provided, and the reader will see much mathematics applied. There are a few references but only to specific results, rather than to further reading; and lamentably there is no index.

Pianos and Continued Fractions

EDWARD DUNNE
American Mathematical Society
Providence, RI 02904

MARK MCCONNELL
Oklahoma State University
Stillwater, OK 74078-1058

You can't tune a piano perfectly. In some way, the instrument must always be out of tune with itself. The origin of the problem, naturally, is in physics. Once the problem is translated into mathematics, only a small amount of effort reveals that the difficulty is unavoidable. However, mathematics does offer some compromises, many of which music already knew about.

1. The physics

Sound is produced when an object vibrates in air. Let's assume that the vibration is at ν cycles per second. The object will also naturally vibrate at all positive integer multiples $n\nu$ of ν, with the intensity generally decreasing with n. These are the *overtones* of the basic vibration. They give the tone its distinctive color. For instance, we distinguish a violin from an oboe by unconsciously recognizing the different overtones they produce. (In addition, we distinguish instruments by their attack—how they sound at the start of each note.)

Thus, given one tone, Nature, in her guise as physics, picks out tones to go with it. The interval from ν to 2ν is an *octave*. The interval from 2ν to 3ν is a *perfect fifth*. These two intervals, plus a few others, have become the foundation of Western music. Also, notice that intervals correspond to *ratios* of frequencies: the octave is a $2:1$ ratio, and the perfect fifth a $3:2$ ratio. We will expand on this in what follows.

The Pythagorean scale The primacy of certain intervals in our musical system is usually attributed to Pythagoras (ca. 579–520 B.C.). Legend has him listening to the sounds of the hammers of four smiths. The sounds were pleasant. Upon investigation, he found the hammers weighed twelve, nine, eight and six pounds. From these weights, Pythagoras derived the intervals:

Octave	—	$12:6$	$=$	$2:1$		
Perfect fifth	—	$12:8$	$=$	$9:6$	$=$	$3:2$
Perfect fourth	—	$12:9$	$=$	$8:6$	$=$	$4:3$
Whole step	—	$9:8$				

Now it is hard to say what really happened twenty-six centuries ago. But this certainly seems lucky. Perhaps he was sitting in the same bathtub that Archimedes bathed in three centuries later.[1]

Pythagoras attempted to construct "all the notes," i.e., a complete chromatic scale, using only two rules (and their inverses):

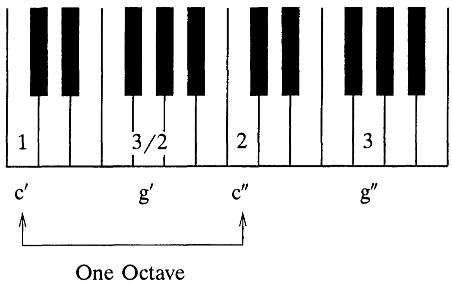
Rule 1 Doubling the frequency moves up an octave,

Rule 2 Multiplying the frequency by $\frac{3}{2}$ moves up a perfect fifth.

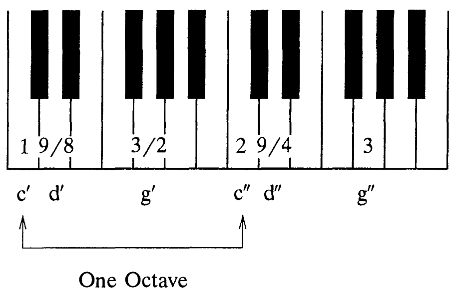
Although this scale will be artificial compared to those actually used by musicians, examining it will quickly lead us to the essential problem in constructing a scale.

We illustrate the construction using the key of C. We adapt our units of time so that the frequency of our starting point, called middle C and denoted c', is 1. This will also avoid the argument over the proper frequency for tuning an orchestra (which is usually waged over where to place A).

Applying Rule 1 to c', we see that the C one octave up from c' has frequency 2; this note is called c''. Applying Rule 2 to c' we obtain g', the perfect fifth above c', with frequency $\frac{3}{2}$. Likewise, g'' is the perfect fifth above c''.



Applying Rule 2 to g', we obtain d'' with frequency $\frac{9}{4}$. This is the D in the next octave up from the octave we are trying to construct. Applying the inverse of Rule 1, we divide the frequency by 2 to obtain d', the D in our basic octave. It has frequency $\frac{9}{8}$.



[1]It is not clear what really matters—the weights of the hammers, the anvils, or the objects struck. Quite possibly, Pythagoras based his scale on investigations of plucked strings. Whatever its merits, the legend of the Pythagorean hammers persists. See Howard W. Eves, *In Mathematical Circles*, Prindle, Weber & Schmidt, 1969, p. 27.

If we continue, we arrive at the following frequencies:

Tonic	Frequency
c'	1
g'	3/2
d'	9/8
a'	27/16
e'	81/64
b'	243/128
$f\sharp'$	729/512
$c\sharp'$	2187/2048
$g\sharp'$	6561/4096
$d\sharp'$	19683/16384
$a\sharp'$	59049/32768
$e\sharp'$	177147/131072
$b\sharp'$	531441/262144

Now comes the problem. Western music assumes there are only twelve tones. This is clear on the piano keyboard: each octave has seven white keys and five black keys, twelve in all. If we cycle through the tones by going up by fifths and dropping back down an octave where necessary, we will have to come back to our starting point sometime. And indeed, B♯ on the piano is the same as C. We also identify certain notes coming from the sharp and flat keys: F♯ = G♭, C♯ = D♭, etc. The cycle of notes, called the *Circle of Fifths*, looks like this.

The Circle of Fifths

C
F G
B♭ D
E♭ A
A♭ E
D♭ B
G♭/F♯

The strange number $\frac{531441}{262144} = \frac{3^{12}}{2^{18}}$ is the frequency of our $b\sharp'$. But if $b\sharp' = c''$, our very first rule says that this number should be 2, which it manifestly is not. If we take half of $\frac{531441}{262144}$, to return to our starting point of c', we obtain

$$\frac{531441}{524288} = 1.013643264770507812 5\ldots, \tag{1}$$

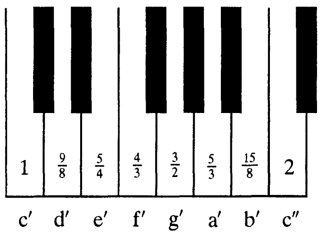
which ought to be 1. This discrepancy is known as the *Pythagorean comma*. It amounts to about 23.5% of a half-step. (See the section "Measuring Tones in Cents" for the precise meaning of percentages of a half-step.)

Helmholtz's scale and the syntonic comma One way to try to get out of our difficulty is to design a scale using more rules. Namely, we could incorporate other basic intervals. After all, it's not only octaves (2 : 1) and fifths (3 : 2) that sound pleasant. The next few intervals between consecutive overtones are these:

interval	**ratio**	**example**
perfect fourth	4 : 3	c'–f'
major third	5 : 4	c'–e'
minor third	6 : 5	c'–$e\flat'$

The perfect fourth is the inverse of the perfect fifth—use Rule 1 followed by the inverse of Rule 2—so it will give us nothing new.

Helmholtz proposed a definition of the major scale using a mixture of these intervals, emphasizing the fifth (3 : 2) and the major third (5 : 4) [**6**, p. 274]. Here is his system:



This scale has many good points. The perfect fifths $c' - g'$, $e' - b'$ and $f' - c''$ are all 3 : 2 ratios, as they should be. The major thirds $c' - e'$, $f' - a'$ and $g' - b'$ are 5 : 4 ratios. Many of the minor thirds are 6 : 5 ratios.

While some things have improved, there is still trouble with Helmholtz's scale. The interval d'–a' is $5/3 \div 9/8 = 40/27$. According to the Circle of Fifths, though, d'–a' should be a perfect fifth, in the ratio 3 : 2. The two ratios differ by the factor

$$\frac{3}{2} \div \frac{40}{27} = \frac{81}{80},$$

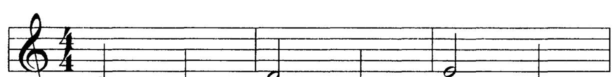
a figure called the *syntonic comma*. It amounts to about 21.5% of a half-step.

If you'd like to hear what the trouble is, try the software provided with Erich Neuwirth's book *Musical Temperaments*, Springer-Verlag, Vienna, 1997, p. 16. This is a wonderful resource for listening to notes and chords in Pythagorean and Helmholtz scales, as well as scales we will look at later (mean-tone and equal temperament). It is amazing how awful that $40/27$ ratio of d'–a' is; it sounds like a bad car horn.

The syntonic comma is actually a much greater problem in Western music than the Pythagorean comma. To encounter a Pythagorean comma, a piece would have to modulate through all twelve keys of the Circle of Fifths; only the most intellectually daring composers, like Bach and Brahms, tend to do this. But only a few modulations bring you to the syntonic comma. Consider the tune

$$\begin{array}{ccccccccccc} c' & & a & & d' & & b & & e' & & c' \\ & \div 6/5 & & \times 4/3 & & \div 6/5 & & \times 4/3 & & \div 5/4 & \end{array}$$

In music notation, this is:



The initial c' (frequency 1) becomes

$$1 \div \frac{6}{5} \times \frac{4}{3} \div \frac{6}{5} \times \frac{4}{3} \div \frac{5}{4} = \frac{80}{81}$$

at the end of the phrase, while its final value should be 1 again. The ratio $\frac{80}{81}$ is—presto—the syntonic comma (really its inverse). The tune (transposed) is the root progression in Measures 3 through 5 of the "Prelude in B♭ Major" in the *Well-Tempered Clavier*, Book I, by J. S. Bach. If the prelude is played or sung perfectly in tune from moment to moment, the notes will go flat by $\frac{80}{81}$ during these three measures alone. Therefore, any performance of Bach's prelude must be out of tune with itself, at least some of the time.

The trouble also occurs in popular music. Figure 1 is an imaginative reconstruction of the hook in the chorus of Madonna's "Borderline," showing how it would sound if the singer and her band were perfectly in tune from moment to moment. Notes held over from one measure to the next must keep the same frequency, because it would be noticeable if they changed, while other notes should adjust themselves to be in tune with the held notes. The D (9/16) is held over from Measure 3 to 4. To make a perfect fifth with the D, the A in Measure 4 must be 27/32. This A is held over into Measure 5, where Madonna must retune her own A to 27/16 to match it. Her A has gone sharp from its usual value of 5/3. In Measure 6, the band must retune itself to match Madonna's A (which is held over), and the damage is complete. When Madonna reaches her C in Measure 6, it is at 81/80, the syntonic comma.

FIGURE 1
Analysis of "Borderline"

The hook is repeated seven times in "Borderline." Since $\frac{81}{80}$ is 21.5% of a half-step, the song should be sharp by about 151% of a half-step at the end. The song is really in D major, but it "should" end in a sickly key halfway between E♭ and E major.

While few musicians analyze chords with fractions, their ears notice something piquant between Measures 4 and 5. In standard music theory, the F♯ in the bass in Measure 4 should rise, especially since it is against the C (a diminished fifth). Yet Madonna pushes the F♯ over the borderline so that it falls to F. While such progressions were in vogue during the 1800's, they would have been scandalous in the 1500's. Of course, Madonna thrives on scandal...

2. The mathematics

What's wrong? The essence of the Pythagorean comma is that twelve perfect fifths do not make seven octaves. To wit:

$$\left(\frac{3}{2}\right)^{12} = \frac{531441}{4096} = 129.746337891\ldots$$

does not equal

$$2^7 = 128.$$

Equivalently,

$$2^{19} \neq 3^{12}.$$

We now explain why this approach to tuning *must* go wrong.

Elementary number theory The fundamental problem is in trying to equate a function based on tripling (fifths) with a function based on doubling (octaves). Phrased mathematically, we are trying to solve an equation of the type

$$2^x = 3^y \tag{2}$$

where x and y are integers. But there is certainly no solution, by the uniqueness of prime factorizations. The numbers 2 and 3 are primes. A non-zero integer (or rational) power of one prime can never equal an integer power of a different prime. (There is no gain if we look for rational solutions $x = p/q$ and $y = r/s$, since we can return to the case of integers by raising both sides of the equation to the power qs.) The syntonic comma arises from the falsity of a similar "equation" involving three primes: $\frac{81}{80} = 1$, or $2^4 \cdot 5 = 3^4$.

Thus there will never be a scale in which all the fifths, or a complete set of fifths and thirds, are correct. The same type of analysis shows that *any* method of constructing a twelve-tone scale by rational numbers is doomed to inconsistency.

Early temperaments To design a scale that sounds good, we must choose where to place the inconsistencies so they are less noticeable. One group of methods in use since the Renaissance is the family of *mean-tone* systems, which involve a sort of averaging of intervals. In these systems, certain keys sound quite good, but at the expense of other keys, which can sound terrible. The Baroque period saw the introduction of *well-tempered* systems, based on a more complicated averaging scheme. Here, although some keys are "better" than others, all the keys are playable [**7**, p. 18].

Many modern organs aim for authenticity by using the old temperaments. The Charles Fisk organ at Stanford University, completed in 1984, allows the organist to choose either a mean-tone or a well-temperament by moving a single lever. The white keys do not change, but each black key plays either of two pipes, one for mean-tone and one for well-temperament. On the compact disc "D. Buxtehude and his Time" (Organa ORA 3208), Harald Vogel plays Baroque pieces in both temperaments, one after the other. The pieces sound good in either temperament, and they differ in character, not absolute quality. The well-tempered notes sound a little less distinct and brilliant, as if they had been recorded by a different audio engineer. During one quiet, sustained chord in mean-tone, it seemed as if an extra pipe had been added in the wrong key. It turned out to be the $n = 7$ harmonic of the bass; this clashes with the other notes in the chord, but sings out loud and clear whenever the tuning is pure enough to let us hear it. At the corresponding moment in the well-tempered version, there was only a touch of the $n = 7$ harmonic; instead, one could hear the beats, as the imperfect intervals went gently in and out of phase with each other.

The most prevalent tuning method today is the most democratic. We describe it in the next section.

Equal temperament Western music has adopted the system of *equal temperament* (also known as *even temperament*). Here the ratio of the frequencies of any two adjacent notes (i.e., half-steps) is constant, and the only interval that is acoustically correct is the octave. It is not clear when this was originally developed. Guitars in Spain were evenly tempered at least as early as the fifteenth century, two hundred years before Bach. And Hermannus Contractus, born 18 July 1013, invented a system of intervalic notation that anticipated equal temperament [**10**].

A much-discussed example is Bach's *Well-Tempered Clavier*, which consists of two series of twenty-four preludes and fugues in each of the twelve major and twelve minor keys. There is an old misconception that Bach used equal temperament on his own instruments, and wrote the *Well-Tempered Clavier* to popularize that system. However, most scholars today believe Bach used a well-tempered system [**3**, vol. 8, p. 379].

Curiously, perhaps because of electronic music, there has recently been a resurgence of interest in temperaments based on rational intervals. The reader is encouraged to visit the Just Intonation Network Web Site, located at
`http://www.dnai.com/~jinetwk/`.

Even temperament spreads the error around in two ways. First, the errors in any particular key are more or less evenly distributed. Equally tempered fifths are flatter than the true $3:2$ ratio by about 2% of a half-step. By contrast, equally tempered major thirds are sharper than the true $5:4$ by 14% of a half-step. What one gains, the other loses. Trained musicians can hear this 14% difference, and it bothers some of them very much.

Second, in equal temperament, no one key is better off than another. With alternative temperament systems, such as Helmholtz's system or mean-tone temperament, roughly four (out of the possible twelve) major keys are clearly better than the others.

We can compute the ratio r between two frequencies that are one half-step apart. Since twelve half-steps make an octave, we must have $r^{12} = 2$, or $r = \sqrt[12]{2} = 2^{1/12}$. Notice the fundamental shift here, moving from rational operations (fractions) to exponential operations.

Measuring tones in cents Acousticians use a logarithmic scale for measuring intervals. They divide the octave into 1200 equal parts, so that each half-step is divided into 100 *cents*. One cent is the ratio $2^{1/1200}$. If I is an interval (a ratio of two frequencies) the number of cents in the interval is $1200 \log_2 I$.

The Pythagorean comma was $\frac{531441}{524288}$. Measured in cents, this is

$$1200 \log_2\left(\frac{531441}{524288}\right) \approx 23.5 \text{ cents.}$$

This is the "23.5% of a half-step" that we mentioned earlier. The syntonic comma is

$$1200 \log_2\left(\frac{81}{80}\right) \approx 21.5 \text{ cents,}$$

the "21.5% of a half-step" mentioned earlier. These intervals are so small that most people can't hear them. As we have said, though, trained musicians can hear them all too well.

Measured in cents, the ideal perfect fifth is

$$1200\log_2\left(\frac{3}{2}\right) \approx 702.0 \text{ cents}.$$

This is only 2.0 cents away from the fifth in equal temperament, which is exactly 700 cents:

$$1200\log_2(2^{7/12}) = 1200\left(\frac{7}{12}\right) = 700.$$

We can also compute the major third in terms of cents:

$$1200\log_2\left(\frac{5}{4}\right) \approx 386.3 \text{ cents}.$$

Except for the octave, all of the acoustic intervals (3:2, 5:4, 6:5,...) will be imperfect in an equally tempered scale.

Continued fractions The heart of our problem with fifths and octaves was the attempt to solve the equation $2^x = 3^y$, where x and y are integers. If we take $\log_2$ of both sides of this troublesome equation, we are left with $x\log_2 2 = y\log_2 3$. Since $\log_2 2 = 1$, the equation reduces to $x = y\log_2 3$, or

$$\frac{x}{y} = \log_2 3.$$

We cannot solve this for integer or rational values of x and y, as we know from the section "Elementary Number Theory" above. The best we can do is to approximate $\log_2 3$ by a rational number. A decimal approximation is 1.5849625007211561 81...

A good (and well-known) way to approximate an irrational number by a rational number is by continued fractions.

A *continued fraction* is an expression of the form:

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cfrac{1}{a_5 + \cdots}}}}}$$

where a_0, a_1, $a_2, \ldots$ are integers. We write $[a_0, a_1, a_2, \ldots]$ for this infinite continued fraction. If we cut off an infinite continued fraction after N terms, we have a finite continued fraction called the N^{th} *convergent* of the original continued fraction, and given by

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\cdots + a_N}}}},$$

which is denoted $[a_0, a_1, a_2, \ldots, a_N]$. This is obviously a rational number, which we write (in reduced form) as $c_N = \dfrac{p_N}{q_N}$. The value of an infinite continued fraction is defined to be the limit of the convergents: $c = \lim_{N\to\infty} c_N$.

It is an exercise to see that any rational number can be expressed as a finite continued fraction. See [**5**, Ch. X, especially §10.9] for a discussion of the uniqueness of such an expression.

There is a classic algorithm for computing the continued fraction expansion of a given real number x. For any number A, let $\lfloor A \rfloor$ denote the integer part of A (e.g., $\lfloor 2.9 \rfloor = 2$ and $\lfloor -2.9 \rfloor = -3$). To compute the continued fraction for x, take $a_0 = \lfloor x \rfloor$. So $x = a_0 + x_0$, where $0 \le x_0 < 1$. Now write

$$\frac{1}{x_0} = a_1 + x_1 \quad \text{with} \quad a_1 = \left\lfloor \frac{1}{x_0} \right\rfloor,$$

$$\frac{1}{x_1} = a_2 + x_2 \quad \text{with} \quad a_2 = \left\lfloor \frac{1}{x_1} \right\rfloor,$$

and so on. If we reach an N such that $x_N = 0$, then $x = [a_0, a_1, \ldots, a_N]$, implying x is rational. On the other hand, if x is irrational, the continued fraction will not terminate.

Examples of continued fractions In what follows, we write $[p, q, n, \dot{m}]$ for the repeating continued fraction $[p, q, n, m, m, m, m, \ldots]$.

- $\sqrt{2} + 1 = [2, 2, 2, \ldots] = [\dot{2}]$ with convergents

$$1, \frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \ldots$$

Proof. $(\sqrt{2} + 1)(\sqrt{2} - 1) = 1$. So $\sqrt{2} + 1 = 2 + (\sqrt{2} - 1) = 2 + \dfrac{1}{\sqrt{2} + 1}$. Apply this process again to the denominator of the fraction, and so on forever. One still has to prove that the limit of the convergents is really $\sqrt{2} + 1$; see Theorem 2 below.

- $\dfrac{1 + \sqrt{5}}{2} = [1, 1, 1, \ldots] = [\dot{1}]$ with convergents

$$1, 2, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \frac{21}{13}, \frac{34}{21}, \frac{55}{34}, \ldots$$

The reader is invited to find a similar proof. Notice that the numerators and denominators of the convergents are consecutive Fibonacci numbers. If you start computing the convergents by hand, you'll soon see why. The ratio of successive Fibonacci numbers approaches, in the limit, the *golden mean* $\dfrac{1 + \sqrt{5}}{2}$. This indicates that this continued fraction does indeed converge to the irrational number it is supposed to converge to.

- $e = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \ldots]$ with convergents

$$2, 3, \frac{8}{3}, \frac{11}{4}, \frac{19}{7}, \frac{87}{32}, \frac{106}{39}, \frac{193}{71}, \frac{1264}{465}, \frac{1457}{536}, \frac{2721}{1001}, \ldots$$

Remarkably, the pattern in the continued fraction expansion for e actually persists, with the consecutive even integers appearing as every third a_i for $i \ge 2$.

- $\pi = [3, 7, 15, 1, 292, 1, 1, 1, 2, \ldots]$ with convergents

$$3, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103993}{33102}, \frac{104348}{33215}, \frac{208341}{66317}, \frac{312689}{99532}, \frac{833719}{265381}, \ldots$$

Unlike the continued fraction expansion for e, the complete expansion for π is unknown.

Theorems on continued fractions Continued fractions are useful because they give the "best" approximations to irrational numbers x by rational p/q. The next theorem makes this precise.

THEOREM 1. [**5**, Theorem 181, p. 151] *Let x be an irrational number, $n \geq 1$, and p_n/q_n be the n^{th} convergent of the continued fraction expansion of x. If $0 < q \leq q_n$ and $p/q \neq p_n/q_n$ with p,q integers, then*

$$\left| x - \frac{p_n}{q_n} \right| < \left| x - \frac{p}{q} \right|.$$

That is to say, the n^{th} convergent provides the best approximation to x among all fractions whose denominator is no greater than q_n. It is common to use the size of the denominator as a measure of the "complexity" of the rational number. In this sense, the n^{th} convergent is optimal for a given complexity.

Another pertinent result is

THEOREM 2. [**5**, Theorem 171, p. 140] *Let x be an irrational number, $n \geq 1$, and let p_n/q_n be the n^{th} convergent of the continued fraction expansion of x. Then*

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}.$$

Since the denominator of the $(n+1)^{\text{st}}$ convergent is strictly larger than the denominator of the n^{th} convergent (and they are all integers), this theorem implies that the continued fraction expansion does indeed converge to the irrational number it is meant to be approximating.

Continued fractions and scales based on perfect fifths What does this say about our musical problem? Recall that we had already moved from the Diophantine equation $2^x = 3^y$ to the problem of approximating $\log_2 3$ by rational numbers p/q. After a little calculating, we find that the continued fraction expansion for $\log_2 3$ is $[1, 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, \ldots]$. The first few convergents are

$$1, 2, \frac{3}{2}, \frac{8}{5}, \frac{19}{12}, \frac{65}{41}, \frac{84}{53}, \frac{485}{306}.$$

Thus, taking the fourth approximation (start counting at zero):

$$3 = 2^{\log_2 3} \approx 2^{19/12}$$

or

$$\frac{3}{2} \approx 2^{7/12}.$$

That is to say, we obtain an approximation to the perfect fifth by dividing the octave into twelve equal intervals and using seven of them, just as on a piano keyboard. Western music has adopted, quite by accident we assume, the fourth best approximation to a Pythagorean scale using equal temperament.

Obviously it is possible to have scales that come from dividing an octave into other than twelve pieces. For instance, the typical Chinese scale has five "notes" to the octave. Western folk music also uses five-tone scales (called *pentatonic* scales), like $c' - d' - e' - g' - a' - c''$. Remarkably, the "five" corresponds to the third convergent of the continued fraction expansion. Pentatonic scales are not tuned in five equal pieces, though.

Going in the other direction, we could use the next most accurate continued fraction approximation of $\log_2 3$. This would lead to an octave consisting of *forty-one parts*. We will call each part a *mini-step*. The perfect fifth is approximated by $\frac{3}{2} \approx 2^{24/41}$, representing an interval of 24 of the 41 mini-steps.

Below is a comparison of what happens to some exact intervals in these three systems.

The twelve-tone scale. In a twelve-tone scale,

- The fifth is $12\log_2(\frac{3}{2}) \approx 7.0196 \approx 7$ basic intervals (half-steps);
- The perfect fourth is $12\log_2(\frac{4}{3}) \approx 4.9804 \approx 5$ basic intervals (half-steps);
- The major third is $12\log_2(\frac{5}{4}) \approx 3.8631 \approx 4?$ basic intervals (half-steps);
- The minor third is $12\log_2(\frac{6}{5}) \approx 3.1564 \approx 3?$ basic intervals (half-steps).

The "?" for the major third and minor third indicate that the rounding to the nearest integer is noticeably inaccurate.

The five-tone scale. If we used a five-tone scale, we would have:

- The fifth is $5\log_2(\frac{3}{2}) \approx 2.9248 \approx 3$ basic intervals;
- The perfect fourth is $5\log_2(\frac{4}{3}) \approx 2.0752 \approx 2$ basic intervals;
- The major third is $5\log_2(\frac{5}{4}) \approx 1.6096 \approx 2??$ basic intervals;
- The minor third is $5\log_2(\frac{6}{5}) \approx 1.3152 \approx 1??$ basic interval.

Thus, the major third and the perfect fourth would not be distinguished in an equal-tempered five-tone scale. The "??" for the major third and minor third indicate that the rounding to the nearest integer is quite rough.

The forty-one tone scale. If we used forty-one mini-steps per octave, we would have:

- The fifth is $41\log_2(\frac{3}{2}) \approx 23.9835 \approx 24$ mini-steps;
- The perfect fourth is $41\log_2(\frac{4}{3}) \approx 17.0165 \approx 17$ mini-steps;
- The major third is $41\log_2(\frac{5}{4}) \approx 13.1991 \approx 13?$ mini-steps;
- The minor third is $41\log_2(\frac{6}{5}) \approx 10.7844 \approx 11?$ mini-steps.

The rounding for the major and minor thirds is again noticeably inaccurate. For instance, the major third is sharper than thirteen mini-steps by 19.9% of a mini-step.

This forty-one tone scale has a fairly good separation of the standard acoustically distinct notes. One would guess that if we regularly used such a scale, our ears would be trained to hear the difference between adjacent mini-steps ($\frac{1}{41}$-sts). This interval is $1200/41 \approx 29.3$ cents, only a few cents more than the Pythagorean comma. We would probably come to believe that the 19.9%-of-a-mini-step error in the 41-tone major third is audible and bad, just as musicians today believe the 14.7%-of-a-half-step error in the 12-tone major third is audible and bad.

The World Wide Web site `http://www.math.okstate.edu/~mmcconn/41tone.html` contains a BASIC program that converts a computer keyboard into a two-octave 41-tone "organ." There is also a composition in the forty-one tone scale by the second author.

3. History

According to Barbour **[1]**, the German mathematician M. W. Drobisch (1855) used continued fractions as the basis for a system for subdividing the octave into intervals. He considered not only fifths, but also thirds, as the ratios whose logarithms were to be approximated. It is not clear whether he was the first to have thought of this approach. For instance, Euler was aware of the mathematical aspects of temperament and also wrote one of the great tracts on continued fractions, *De Fractionibus Continuis* (1737). The "modern" theory of continued fractions is said to have begun in the sixteenth century, as evidenced by the writings of Bombelli and Cataldi.

Around 40 B.C., King Fang, in China, studied scales related to the sixth best approximation given above. He noticed that fifty-three perfect fifths are very nearly equal to thirty-one octaves. This leads to what is sometimes called the *Cycle of* 53. It can be represented by a spiral of fifths, replacing the more usual Circle of Fifths. We don't know whether he actually used continued fractions or some essentially equivalent method to do this.

The use of continued fractions to investigate problems of tunings is still alive and well. A modern investigation of tunings and continued fractions can be found in Blackwood's book **[2]**. There is also the article by Carey and Clampitt **[4]**, which looks at well-formed scales and the continued fraction expansion of $\log_2 3$.

Finally, we would like to mention, if only because it gives us an excuse to cite Nicolas Slonimsky **[10]**, that experiments with "microtonality" in twentieth-century music have tended towards dyadic parts of the half-step (scales with twenty-four tones, forty-eight tones, etc.). The most popular invention seems to be the quarter-tone. In other words, we are still tied to 12. The failure to employ the more natural microtones of the 41-tone scale is clear and compelling evidence of the continuing need for mathematics across the curriculum, especially the need for Number Theory in Music.

Acknowledgments We are grateful to Norman Carey for his pointers to the music literature. We thank Gerry Frank for several helpful comments.

REFERENCES

1. James Murray Barbour, *Tuning and Temperament: A Historical Survey*, second ed., Michigan State College Press, Lansing, MI, 1953.
2. Easley Blackwood, *The Structure of Recognizable Diatonic Tunings*, Princeton Univ. Press, Princeton, NJ, 1985.
3. Eric Blom (ed.), *Grove's Dictionary of Music and Musicians*, fifth ed., St. Martin's Press, Inc., New York, NY, 1955.
4. Norman Carey and David Clampitt, Aspects of well-formed scales, *Music Theory Spectrum* 1 1/2 (1989), 187–206.
5. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, fourth ed., Oxford Univ. Press, Oxford, UK, 1975. (See Chapters X and XI concerning continued fractions.)
6. Hermann L. F. Helmholtz, *On the Sensations of Tone*, second English ed., Dover Publ., New York, NY, 1954.
7. G. C. Klop, *Harpsichord Tuning: Course Outline*, trans. Glen Wilson, Sunbury Press, Raleigh, NC, 1974.
8. Don Michael Randel, *Harvard Concise Dictionary of Music*, Harvard Univ. Press, Cambridge, MA, 1978. (See the sections "Comma" and "Temperament.")
9. Percy Scholes, *The Oxford Companion to Music*, ninth ed., Oxford Univ. Press, Oxford, UK, 1955. See the section "Temperament."
10. Nicolas Slonimsky, *Music Since* 1900, Schirmer Books, New York, NY, 1994. See the section "Dictionary of Terms."

Anatomy of a Circle Map

W. E. BRIGGS
University of Colorado at Boulder (emeritus)
Boulder, CO 80309

WILLIAM L. BRIGGS
University of Colorado at Denver
Denver, CO 80217

Introduction

Circle maps, or *self-maps of the circle*, were first used to model biological rhythms in the 1959 thesis research of the great Russian mathematician V. I. Arnold **[1]**. An excellent review of the subject and an account of why this research was *not* published in Arnold's thesis is given in **[4]**. (Briefly, Arnold's thesis advisor, Kolmogorov, felt "that is not one of the classical problems one ought to work on.")

The general form of a circle map G is

$$G(\phi) = g(\phi) \pmod 1,$$

where g is differentiable on $[0, 1]$. A dynamical system is formed when this map is used to define the iteration $\phi_{n+1} = G(\phi_n)$.

The state variable $0 \le \phi < 1$ often represents the phase of an oscillator, usually with respect to a stimulus or underlying rhythm. As ϕ passes from 0 to 1, the oscillator moves through one complete cycle. Circle maps of this form have been used to model heart beats **[5, 6, 10]**, breathing patterns **[8]**, cell divisions, and the firing of nerve impulses, to name just a few applications. In general, g is a nonlinear function (for example, the sine circle map uses $g(\phi) = b + \phi + a\sin(2\pi\phi)$), in which case the structure of the map is amazingly complex and not fully understood.

In this paper we consider the family of linear maps given by $g(\phi) = a\phi + b$, where $a \in \mathbb{Z}^+$ (the positive integers) and $0 < b < 1$. The constant a is called the *degree* of the map; it counts the number of cycles through which ϕ_{n+1} advances as ϕ_n passes through one complete cycle. If a is an integer, as we will require, then G is continuous as $\phi \to 0^+$ and $\phi \to 1^-$. Linear maps of this form have been studied extensively and their properties mentioned incidentally in many sources. We aim to collect these properties in one place and to apply a variety of mathematical approaches to understand them. In particular, we will classify all rational points on $[0, 1)$, either as periodic points of the map or as preimages of periodic points.

A few preliminary definitions will be helpful. The *lift* of any circle map G is the result of unwrapping the map along the real line: if $G(\phi) = a\phi + b \pmod 1$, then its lift is simply $g(\phi) = a\phi + b$. We will use g^p and G^p to denote the p-fold composition of g and G, respectively. The fact that $a \in \mathbb{Z}^+$ makes compositions of circle maps easy to analyze. In particular, we can postpone the (mod 1) operation until the end of a composition:

$$G^p(\phi) = g^p(\phi) \pmod 1 \quad \text{for} \quad a \in \mathbb{Z}^+. \tag{1}$$

For example,

$$\begin{aligned} G^2(\phi) &= \left[a(\left[a\phi + b\right] \pmod 1) + b\right] \pmod 1 \\ &= a^2\phi + ab + b \pmod 1 \\ &= g^2(\phi) \pmod 1. \end{aligned}$$

Compositions of a map lead to periodic points of the map. A point ϕ^* is *periodic with period* p if $G^p(\phi^*) = \phi^*$. A point ϕ^* is *periodic with primitive period* p if $G^p(\phi^*) = \phi^*$, and this holds for no *smaller* value of p. The distinction between periods and primitive periods will be important; a point with period, say 6, *could* also have primitive period 2 or 3. Finally the *periodic orbit* of a p-periodic point is the set consisting of itself and its $p-1$ images $\{\phi^*, G(\phi^*), G^2(\phi^*), \dots, G^{p-1}(\phi^*)\}$. All points in a periodic orbit have the same periodic orbit.

Maps of degree $a = 1$

The linear circle map of degree 1, $G(\phi) = \phi + b \pmod 1$, is well-understood. It can be visualized as follows. Suppose that two nearby towers have bells and that the Bell A chimes four times for every five chimes of Bell B. We can let ϕ be the phase of Bell B relative to Bell A. Assume that when we first hear the two bells they chime in unison. Bell B next chimes $4/5$ of the way through the cycle of Bell A, then $3/5$ of the way through the cycle of Bell A, and so forth, as shown in FIGURE 1. The fifth

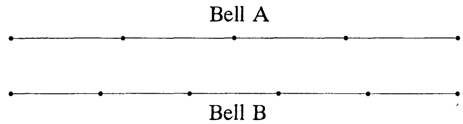


FIGURE 1

The dots represent the chimes of two bells. Bell A chimes four times for every five chimes of the Bell B, setting up a 5:4 phase locking. The phase of Bell B relative to Bell A can be described by a linear circle map of degree one.

chime of Bell B coincides with the fourth chime of Bell A and then the pattern repeats. The phase of Bell B relative to Bell A is described by the circle map

$$\phi_{n+1} = \phi_n + \tfrac{4}{5} \pmod 1.$$

The orbit of zero is $\{0, \frac{4}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}\}$, which has period 5. Any other initial phase would also lead to a period five solution. The two bells are said to be in 5:4 *phase locking*. (Notice that if we measured the phase of Bell A relative to Bell B, the orbit of zero would be $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, which has period 4.)

Let's locate the periodic points of the map of degree 1. The p-periodic points satisfy $G^p(\phi^*) = \phi^*$. By the composition property (1), this condition becomes

$$G^p(\phi^*) = g^p(\phi^*) \pmod 1 = \phi^* + pb \pmod 1 = \phi^*.$$

Notice that if $b = q/p$ where $\gcd(p,q) = 1$, then for any initial point ϕ^*, we have

$$G^p(\phi^*) = \phi^* + pb \pmod 1 = \phi^* + p\frac{q}{p} \pmod 1 = \phi^* + q \pmod 1 = \phi^*.$$

We see that if $b = q/p$ where $\gcd(p,q) = 1$, then all points on the unit interval have period p. In fact, in p iterations, the initial point advances by q cycles. Thus, $b = q/p$ gives the average rotation of the orbit per iteration; this quantity is called the *rotation number* of the map. If b is irrational, then the map has no periodic points; however, the rotation number can still be defined as b (see **[2]**).

Periodic points for $a > 1$

We now investigate the periodic points of maps $G(\phi) = a\phi + b \pmod 1$ with degree $a > 1$. These points satisfy the condition

$$G^p(\phi^*) = g^p(\phi^*) \pmod 1 = \phi^*.$$

In order to solve this equation for ϕ^*, we need to handle the (mod 1) operation. We begin by writing

$$G^p(\phi^*) = g^p(\phi^*) - q \quad \text{where} \quad q \in \mathbb{Z}^+.$$

A short calculation (using a finite geometric series) shows that

$$g^p(\phi^*) = a^p\phi^* + b\frac{a^p - 1}{a - 1}.$$

Therefore, the defining condition for period-p points is

$$a^p\phi^* + b\frac{a^p - 1}{a - 1} - q = \phi^*, \quad \text{where} \quad q \in \mathbb{Z}^+. \tag{2}$$

Solving for ϕ^*, we see that the period-p points of the map G are given by

$$\phi^*_{p,q} = \frac{q}{a^p - 1} - \frac{b}{a - 1} \pmod 1. \tag{3}$$

Notice that there are $a^p - 1$ distinct period-p points, corresponding to $q = 0, 1, 2, \ldots, a^p - 2$. In particular, there are always $a - 1$ period-1 (fixed) points. All periodic points ϕ^* are *unstable* (or *repelling*) in the sense that an iteration beginning with a point arbitrarily close to ϕ^* produces an orbit that drifts away from the orbit of ϕ^*. This can be seen by noting that $g'(\phi^*) = a > 1$.

For example, consider the case $a = 2$ and $b = \frac{1}{2}$. The 15 period-4 points are given by

$$\phi^*_{4,q} = \frac{q}{15} - \frac{1}{2} \pmod 1, \quad \text{where} \quad q = 0, 1, 2, \ldots, 14,$$

and (mod 1) is used when necessary. But notice that not all of these 15 periodic points have *primitive* period 4. In fact, the periodic points can be grouped according to their periods and orbits as follows:

$$\frac{1}{2} \to \frac{1}{2} \quad \text{(period 1)}$$

$$\frac{1}{6} \to \frac{5}{6} \to \frac{1}{6} \quad \text{(period 2)}$$

$$\frac{19}{30} \to \frac{23}{30} \to \frac{1}{30} \to \frac{17}{30} \to \frac{19}{30} \quad \text{(period 4)}$$

$$\frac{7}{10} \to \frac{9}{10} \to \frac{3}{10} \to \frac{1}{10} \to \frac{7}{10} \quad \text{(period 4)}$$

$$\frac{29}{30} \to \frac{13}{30} \to \frac{11}{30} \to \frac{7}{30} \to \frac{29}{30} \quad \text{(period 4)}.$$

We see that of the 15 period-4 points, only 12 have primitive period 4, so there are 3 period-4 orbits. Two of the period-4 points have primitive period 2 and one has primitive period 1.

Counting periodic points and orbits This thinking can be generalized to count the number of periodic points and orbits of all orders. We will let O_p denote the number of orbits with primitive period p and $P_p = pO_p$ denote the number of points with primitive period p. Then

$$P_p = a^p - 1 - \sum_{k|p} P_k \quad \text{and} \quad O_p = \frac{P_p}{p}. \tag{4}$$

If p is prime, then the only non-primitive periodic points are the $a-1$ period-1 points. Therefore, when p is prime, the number of points with primitive period p is $P_p = a^p - a$, suggesting how quickly the number of periodic points grows with p. When p is prime and $\gcd(a, p) = 1$, the fact that p divides P_p is a consequence of Euler's theorem **[9]**. The fact that P_p is divisible by p for *all* p is remarkable.

The sequences $\{O_p\}$ and $\{P_p\}$ can be computed recursively using expression (4) and results in the values shown in Table 1 for $a = 2$ and 3. This sequence first appeared in

TABLE 1. Number of primitive periodic orbits ($a = 2, 3$)

p	1	2	3	4	5	6	7	8
O_p for $a=2$	1	1	2	3	6	9	18	30
O_p for $a=3$	2	3	8	18	48	116	312	810
p	9	10	11	12	13	14	15	16
O_p for $a=2$	56	99	186	335	630	1161	2182	4080
O_p for $a=3$	2184	5880	16,104	44,220	122,640	341,484	956,576	2,690,368

a 1961 paper **[3]** that answered the question: How many distinct (up to translation) arrangements of beads can be made on a necklace of length n using q different kinds of beads? It was also given in a seminal paper **[7]** on periodic points of one-dimensional maps.

The sequences $\{O_p\}$ and $\{P_p\}$ can also be expressed in number-theoretic terms. If we let T_p be the *total* number of orbits of period p, including non-primitive orbits, then we have

$$T_p = a^p - 1 = \sum_{k|p} P_k.$$

The Möbius inversion formula **[9]** can be used to find the individual terms of this sum in terms of T_p. It asserts that

$$P_p = \sum_{k|p} \mu(k) T\left(\frac{p}{k}\right),$$

where $\mu(1) = 1$, $\mu(n) = (-1)^r$ if n is square-free and has r distinct prime factors, and $\mu(n) = 0$ otherwise. Because $\sum_{k|p} \mu(k) = 0$, it follows that

$$O_p = \sum_{k|p} \mu(k)\left(a^{\frac{p}{k}} - 1\right) = \sum_{k|p} \mu(k) a^{\frac{p}{k}}.$$

For example, with $a = 2$,

$$\begin{aligned} P_{12} &= \mu(1)2^{12} + \mu(2)2^6 + \mu(3)2^4 + \mu(4)2^3 + \mu(6)2^2 + \mu(12)2^1 \\ &= 1 \cdot 4096 + (-1) \cdot 64 + (-1) \cdot 16 + 0 \cdot 8 + 1 \cdot 4 + 0 \cdot 2 = 4020. \end{aligned}$$

Therefore, $O_{12} = P12/12 = 335$, as given in Table 1.

Bins and itineraries We will use a specific example to show how *bins* and *itineraries* [2] can be used to organize the periodic points of the map. The fact that b affects the periodic points only by a shift allows us to take $b=0$ without loss of generality. Let's look for the period-4 points of the map $g(\phi)=3\phi$; that is, we will let $a=3$, $p=4$, and $b=0$. The $3^4-1=$ 80 period-4 points, given by (3), are

$$\phi_{4,q}^{*}=\frac{q}{80}\quad\text{where}\quad q=0,1,2,\ldots,79.$$

We know from Table 1 that two of these points have primitive period 1 ($\phi_0^*=0$ and $\phi_{40}^*=\frac{1}{2}$), 6 have primitive period 2, and the remaining 72 have primitive period 4. We can refer to these points not by their value ϕ but simply by the integer $q=0,1,2,\ldots,79$. Then the original map can be expressed in terms of q as

$$q_{n+1}=3q_n \pmod{80}.$$

For example, the point $\phi_4^*=4/80$ corresponds to $q=4$ and has the orbit

$$4\to 12\to 36\to 108\equiv 28\to 84\equiv 4,$$

which constitutes a period-4 orbit.

The set of q's can be partitioned into three *bins*, each with 27 elements:

$$S_1=\{0,1,2,\ldots,26\};\quad S_2=\{27,28,29,\ldots,53\};\quad S_3=\{54,55,56,\ldots,80\}.$$

(Notice that the last element of S_3 corresponds to the first element of S_1.) Any orbit of period-4 points visits these bins in a specific order. For example, the orbit $\{4,12,36,28\}$, visits the bins in the order $\langle S_1,S_1,S_2,S_2\rangle$, which we will abbreviate as $\langle 1,1,2,2\rangle$. This order is called the *itinerary* for the orbit. For example, the orbit

$$13\to 39\to 117\equiv 37\to 111\equiv 31\to 93\equiv 13,$$

has the itinerary $\langle 1,2,2,2\rangle$.

A little counting shows that there are exactly 80 different 4-digit itineraries that can be formed from the symbols 0, 1, 2 (the itinerary $\langle 2,2,2,2\rangle$ is the same as $\langle 0,0,0,0\rangle$, reflecting the identity of $q=0$ and $q=80$). Thus the number of itineraries is identical to the number of periodic points. In fact, there is a one-to-one correspondence: for each itinerary there is a unique periodic point, and vice versa. The cyclic permutations of a given itinerary (for example, 1231, 2311, 3112, and 1123) correspond to the four points of a single orbit.

The period-1 points have the itineraries $\langle 0,0,0,0\rangle$ and $\langle 1,1,1,1\rangle$ (again identifying $\langle 0,0,0,0\rangle$ and $\langle 2,2,2,2\rangle$). The period-2 points have the itineraries $\langle 1,2,1,2\rangle$, $\langle 1,3,1,3\rangle$, and $\langle 2,3,2,3\rangle$, all of which have only two nontrivial cyclic permutations.

This arrangement of bins and itineraries generalizes to all values of p, b, and a. In general, the map $g(\phi)=a\phi+b$ will have a bins, a^p-1 different period-p points, and a^p-1 different p-digit itineraries on a symbols. In each case there is a one-to-one correspondence between itineraries and period points. Notice that itineraries give another perspective on the number of periodic points (Table 1). For a fixed value of a, the number of points with primitive period p is the number of ways that a symbols can be arranged in *irreducible* itineraries of length p. By *irreducible*, we simply mean that the itinerary has p distinct cyclic permutations, and no fewer than p. Finding irreducible itineraries is essentially the necklace problem [3].

Inverse images

Our goal is to classify all rational points on $[0, 1)$ according to how they behave under iteration by the linear circle map. So far, we have identified the periodic points; but these points hardly exhaust the rationals on $[0, 1)$. What other kinds of behavior can a rational point exhibit? We will show that all nonperiodic rational points are *preimages* of periodic points. This means we must explore the inverse of the linear circle map.

FIGURE 2 shows the linear circle map, G, and its inverse in the cases $a = 2$, $b = \frac{1}{2}$ (first row) and $a = 3$, $b = \frac{1}{4}$ (second row). The inverse maps are multiple-valued: to

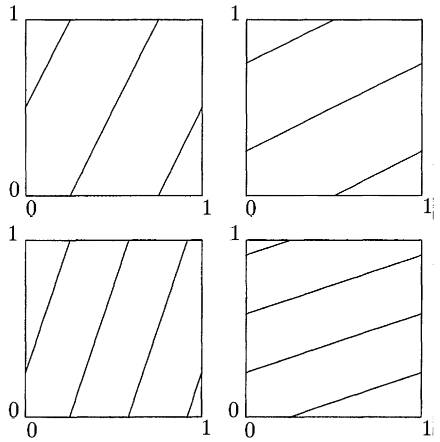


FIGURE 2

The graph of the circle map (left) and its inverse (right) are shown for $a = 2$, $b = 1/2$ (top) and $a = 3$, $b = 1/4$ (bottom).

each value of ϕ correspond a values of $G^{-1}(\phi)$. Formally inverting the forward map, we find that the inverse map is given by

$$G_q^{-1}(\phi) = \frac{\phi - b + q}{a} \pmod 1 \quad \text{for} \quad q = 0, 1, 2, \dots a - 1.$$

It can be shown that for the $q = 1, 2, \dots, a - 1$ branches of the inverse, $0 < G_q^{-1}(\phi) < 1$ and the (mod 1) operation is not necessary. The $q = 0$ branch is generally split, and can be written as

$$G_0^{-1}(\phi) = \begin{cases} \frac{\phi - b}{a} + 1 & \text{for } 0 \le \phi < b \\ \frac{\phi - b}{a} & \text{for } b \le \phi < 1 \end{cases}.$$

Repeating this argument, we can iterate backwards n times and find the nth preimages of any point ϕ. There are a^n of them, given by

$$G_q^{-n}(\phi) = \frac{\phi - \left(\frac{a^n - 1}{a - 1}\right)b + q}{a^n} \pmod 1 \quad \text{for} \quad q = 0, 1, 2, \dots a^n - 1.$$

Let's look at the case $a = 2$ more closely. We see that any point is preceded by an entire tree of preimages (FIGURE 3). If the point in question is a periodic point, those preimages may be other points in the same periodic orbit. But many more of the preimages will be outside the orbit. What happens if we iterate backwards from a given point upward through the tree? Clearly, there are many different paths through the tree, so at each point we must specify which branch of the inverse map we want to use.

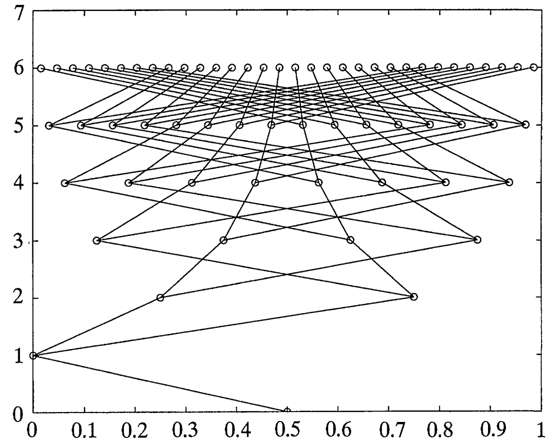


FIGURE 3

The first six generations of preimages of $\phi = 1/2$ in the case $a = 2$ and $b = 1/2$ form a tree. The generations increase upward on the vertical axis.

We will let $b = 0$ without loss of generality because it removes the split in the $q = 0$ branch of the inverse map. Then the two branches of the inverse map are given by

$$G_0^{-1}(\phi) = \frac{\phi}{2} \quad (q = 0) \quad \text{and} \quad G_1^{-1}(\phi) = \frac{\phi + 1}{2} \quad (q = 1).$$

Suppose we iterate backwards and always use the $q = 0$ branch (call this path $0\ldots$). We can find the limit of the sequence generated by looking for a fixed point of G_0^{-1}. Solving

$$G_0^{-1}(\phi) = \frac{\phi}{2} = \phi,$$

we see that this backward iteration converges to $\phi = 0$, the period-1 point of the forward map. Similarly, if we iterate backwards using the $q = 1$ branch at every step (call this path $1\ldots$), the sequence generated converges to $\phi = 1$, or, equivalently, $\phi = 0$.

What happens if we iterate backwards and alternate the two branches, starting with the 0 branch (call this path $01\ldots$)? The limit of this sequence (starting at any initial point) is the fixed point of the composite map

$$G_1^{-1} \circ G_0^{-1} = \frac{\frac{\phi}{2} + 1}{2} = \frac{\phi}{4} + \frac{1}{2},$$

which is $\phi = 2/3$, a period-2 point of the forward map. Similarly, the backward iteration with the path $10\ldots$ converges (starting at any initial point) to $\phi = 1/3$, another period-2 point of the forward map. Notice that the limits of all inverse maps (and compositions of inverse maps) are stable (or attracting) periodic points since the derivative of these iteration functions is less than one in magnitude. Table 2 extends this line of thinking, giving the results of iterating backward on various path itineraries starting with any initial point.

Perhaps the pattern is emerging. We will say that a p-digit path is *irreducible* if there are no repeating patterns in the path numbers of length less than p. If we choose an irreducible p-digit path through the inverse iteration tree, the resulting sequence converges to a period-p point of the forward iteration. Two p-digit path numbers that are cyclic permutations of each other will converge to period-p points *in*

TABLE 2. Backward iteration on various path itineraries ($a = 2$, $b = 0$)

Path number	Limit of iteration sequence	Identity of limit
$0\ldots$	0	period-1 point of G
$1\ldots$	$1 \equiv 0$	period-1 point of G
$01\ldots$	$2/3$	period-2 point of G
$10\ldots$	$1/3$	period-2 point of G
$001\ldots$	$1/7$	period-3 point of G
$010\ldots$	$2/7$	period-3 point of G
$100\ldots$	$4/7$	period-3 point of G
$101\ldots$	$5/7$	period-3 point of G
$011\ldots$	$3/7$	period-3 point of G
$110\ldots$	$6/7$	period-3 point of G
$1110\ldots$ and permutations		period-4 points of G
$1100\ldots$ and permutations		period-4 points of G
$1000\ldots$ and permutations		period-4 points of G

the same orbit. The p-periodic points of a single orbit correspond to the p cyclic permutations of a given p-digit path number. The above observations hold for any integer $a > 1$. Once again, we see that counting the period-p points of the forward map amounts to counting the irreducible arrangements of p symbols.

The fate of all rational points

We have seen that any periodic point has an a-fold branching tree of preimages. If we choose any point in this tree as an initial point of the mapping, then eventually the sequence of iterates falls into a periodic orbit. We can now show that the periodic points and all of their preimages account for all the rational points on $[0, 1)$.

Assume that we choose an initial point $\phi_0 = r/s$, where $\gcd(r, s) = 1$. First observe that all subsequent iterates will be rational. Furthermore, these iterates can take on at most s distinct values. Therefore, a repetition must eventually occur, so ϕ_0 eventually leads to a periodic orbit.

More interesting, suppose we are given an arbitrary rational initial point $\phi_0 = r/s$, where $\gcd(r, s) = 1$. We know that it is the preimage of a periodic point. But what is the period of that periodic point, and how many iterations does it take to arrive at that periodic point?

Case 1: ϕ_0 is a periodic point. Let's first consider a special case. Let ϕ_0 be a rational initial point. We want to find a condition assuring that ϕ_0 is periodic and then find the least value of p such that $G^p(\phi_0) = \phi_0$. Recall the condition (2) for a period-p point:

$$a^p\phi_0 + b\frac{a^p - 1}{a - 1} - q = \phi_0 \quad \text{where} \quad q \in \mathbb{Z}^+.$$

This implies that

$$(a^p - 1)\underbrace{\left(\phi_0 + \frac{b}{a-1}\right)}_{r/s} = q \quad \text{where} \quad q \in \mathbb{Z}^+.$$

If we now let

$$\left(\phi_0 + \frac{b}{a-1}\right) = \frac{r}{s},$$

where $\gcd(r, s) = 1$, then we have a solution provided that

$$(a^p - 1) = \frac{qs}{r} \quad \text{for some } q \in \mathbb{Z}^+.$$

Thus, we are looking for p such that $a^p \equiv 1 \pmod{s}$. If $\gcd(a, s) = 1$, Euler's theorem **[9]** guarantees that $p = \phi(s)$ (here ϕ denotes the Euler phi-function) is a solution; in fact, there may be smaller solutions p' such that p' divides $\phi(s)$. We see that the condition that ϕ_0 is periodic is that $\gcd(a, s) = 1$; in this case, the period of ϕ_0 is the smallest solution of $a^p \equiv 1 \pmod{s}$.

Example. If $a = 2$, $b = 0$ and $\phi_0 = \frac{2}{9}$, we have $\gcd(a, s) = \gcd(2, 9) = 1$ and the period of ϕ_0 satisfies $2^p \equiv 1 \pmod{9}$. The smallest value of p that satisfies this condition is $p = \phi(9) = 6$. Thus, $\phi_0 = \frac{2}{9}$ is a period-6 point. If $a = 4$, $b = \frac{1}{3}$ and $\phi_0 = \frac{2}{5}$, then $\phi_0 + \frac{b}{a-1} = \frac{23}{45}$. We see that $\gcd(a, s) = \gcd(4, 45) = 1$; thus ϕ_0 is a periodic point whose period is the least solution of $4^p \equiv 1 \pmod{45}$. One solution is $p = \phi(45) = 24$; the smallest solution is $p = 6$. Thus, the period of $\phi_0 = \frac{2}{5}$ is 6.

Case 2: ϕ_0 is not a periodic point. In the case that $\gcd(a, s) \neq 1$, we must iterate (multiply by a) often enough to remove all common factors of a and s from s. In this case, we must find the smallest values of m and n such that $\phi_m = \phi_n$, where $m > n$. Then m is the number of iterations required to reach a periodic point and the period of that point is $p = m - n$.

Using (2), the condition that $\phi_m = \phi_n$ becomes

$$a^m\phi_0 + b\frac{a^m - 1}{a-1} = a^n\phi_0 + b\frac{a^n - 1}{a-1} = q \quad \text{where} \quad q \in \mathbb{Z}^+.$$

Rearranging a bit gives us

$$(a^m - a^n)\underbrace{\left(\phi_0 + \frac{b}{a-1}\right)}_{r/s} = q \quad \text{where} \quad q \in \mathbb{Z}^+.$$

Once again, we let $\phi_0 + \frac{b}{a-1} = \frac{r}{s}$, where $\gcd(r, s) = 1$. So we seek the smallest integers m and n that satisfy

$$a^n(a^{m-n} - 1)\frac{r}{s} = q \quad \text{where} \quad q \in \mathbb{Z}^+. \tag{5}$$

Suppose $\gcd(a, s) = p_1^{i_1} \cdots p_m^{i_m}$, $a = p_1^{j_1} \cdots p_m^{j_m} A$, and $s = p_1^{k_1} \cdots p_m^{k_m} S$, where $\gcd(A, S) = 1$. Then condition (5) becomes

$$p_1^{nj_1 - k_1} \cdots p_m^{nj_m - k_m} \frac{rA^n}{S}(a^{m-n} - 1) = q \quad \text{where} \quad q \in \mathbb{Z}^+. \tag{6}$$

Let ν be the smallest positive integer n such that all of the exponents of the p_i's are nonnegative ($nj_i \geq k_i$). Then condition (6) takes the form

$$I\frac{rA^\nu}{S}(a^{m-n^*} - 1) = q \quad \text{where} \quad I \in \mathbb{Z}^+.$$

Note that having determined ν, the quantity $p = m - \nu$ is the period of the resulting orbit. Therefore a solution exists provided $a^p \equiv 1 \pmod{S}$. This equation has a solution since $\gcd(a, S) = 1$. Thus, ν iterations are required to reach a periodic point and the period of that point is the smallest value of p that satisfies $a^p \equiv 1 \pmod{S}$.

Example. For the circle map

$$\phi_{n+1} = 12\phi_n \pmod{1}, \quad \phi_0 = \frac{r}{40}, \quad r \in \mathbb{Z}^+, \quad 1 \le r < 40,$$

we have $a = 12$, $b = 0$, and $\phi_0 = \frac{r}{40}$. It follows that $\gcd(a, s) = \gcd(12, 40) = 2^2$, $a = 2^2 \cdot 3$, $s = 2^3 \cdot 5$, and $S = 5$. The number of iterations, ν, required to reach a periodic point satisfies $2n \ge 3$, which means $\nu = 2$. The period of the resulting orbit satisfies $12^p \equiv 1 \pmod 5$, whose smallest solution is $p = 4$. Thus, for an initial point of the form $\phi_0 = \frac{r}{40}$, it takes two iterations to reach a period-4 point. An example is the sequence

$$\frac{3}{40} \to \frac{9}{10} \to \frac{54}{5} \equiv \frac{4}{5} \to \frac{3}{5} \to \frac{1}{5} \to \frac{2}{5} \to \frac{4}{5} \cdots \quad (\text{period } 4).$$

For the circle map

$$\phi_{n+1} = 6\phi_n + \frac{5}{8} \pmod{1} \quad \phi_0 = \frac{r}{7}, \quad r \in \mathbb{Z}^+, \quad 1 \le r < 7,$$

we have $a = 6$, $b = 5/8$, and $\phi_0 = r/7$. Then $\gcd(a, s) = \gcd(6, 56) = 2$, $a = 2 \cdot 3$, $s = 2^3 \cdot 7$, and $S = 7$. The number of iterations, ν, required to reach a periodic point satisfies $n \ge 3$, which means $\nu = 3$. The period of the resulting orbit satisfies $6^p \equiv 1 \pmod 7$, whose smallest solution is $p = 2$. Thus, for an initial point of the form $\phi_0 = \frac{r}{7}$, it takes three iterations to reach a period-2 point. An example is the sequence

$$\frac{2}{7} \to \frac{19}{56} \to \frac{37}{56} \to \frac{33}{56} \to \frac{9}{56} \to \frac{33}{56} \cdots \quad (\text{period } 2).$$

Conclusion

We have shown that a linear circle map itself has a curious circular structure. Rational periodic points of all possible periods form the backbone of this structure. Any rational non-periodic point is a preimage of some periodic point: iterating on such a non-periodic point produces a sequence that falls into a periodic orbit in a finite number of steps. In particular, if one chooses an initial (rational) point very close to a periodic point, the resulting sequence of iterates will eventually converge to another (usually different) periodic point. Explicit expressions can be given for the periodic points, the preimages of the periodic points, and the ultimate period of any rational point.

Acknowledgment We are grateful to Stan Payne and Dave Fisher of the University of Colorado at Denver Mathematics Department for helpful discussions about this problem.

REFERENCES

1. Arnold, V. I., Cardiac arrhythmias and circle maps, reprinted in *Chaos* 1 (1991), 20–24.
2. Devaney, R. L., *An Introduction to Chaotic Dynamical Systems*, Benjamin/Cummings Publishing, Menlo Park, CA, 1986.
3. Gilbert, E. N. and Riordan, J., Symmetry types of periodic sequences, *Illinois J. of Math* 5 (1961), 657–665.
4. Glass, L., Cardiac arrhythmias and circle maps—a classical problem, *Chaos* 1 (1991), 13–19.
5. Glass, L. and Mackey, M. C., *From Clocks to Chaos: The Rhythms of Life*, Princeton University Press, Princeton, NJ, 1988.
6. Glass, L., Guevara, M.R., Shrier, A., Bifurcation and chaos in a periodically stimulated cardiac oscillator, *Physica* 7D (1983), 89–101.
7. Holmes, P. and Whitely, D. C., Bifurcations of one- and two-dimensional maps, *Phil. Trans. Royal Soc. Lond.* A 311 (1984), 43–84.
8. Lewis, J. E., Glass, L., Bachoo, M., Polosa, C., Phase resetting and fixed-delay stimulation of a simple model of respiratory rhythm generation, *J. Theor. Bio.* 159 (1992), 491–506.
9. Rosen, K. H., *Elementary Number Theory and its Applications*, Addison-Wesley, Reading, MA, 1984.
10. Winfree, A. T., *The Geometry of Biological Time*, Springer-Verlag, New York, NY, 1980.

NOTES

Cyclic Sums for Polygons

G. C. SHEPHARD
University of East Anglia
Norwich NR4 7TJ
England

It is a familiar fact that if O is the centroid of a triangle $[A, B, C]$ (see FIGURE 1) and if AO, BO, CO meet BC, CA, AB in D, E, F respectively, then

$$\frac{|OD|}{|AD|} = \frac{|OE|}{|BE|} = \frac{|OF|}{|CF|} = \frac{1}{3},$$

and therefore

$$\frac{|OD|}{|AD|} + \frac{|OE|}{|BE|} + \frac{|OF|}{|CF|} = 1. \tag{1}$$

It is not so familiar (but very easy to prove) that (1) holds for *all* points O inside the triangle (and not just the centroid). In fact Euler knew of this relation as long ago as

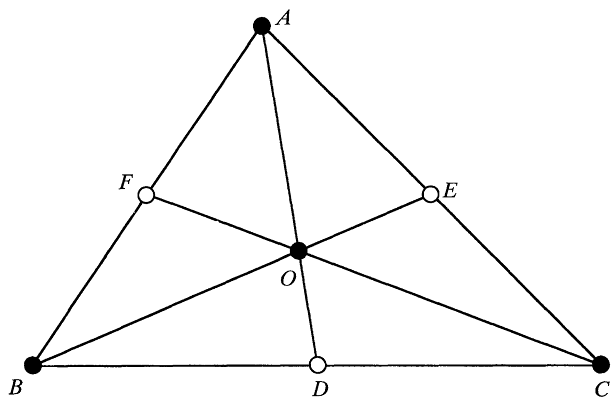


FIGURE 1
The theorem about triangles known to Euler.

1780, see **[1]**. More generally, for every point O (inside the triangle or not),

$$\frac{\|OD\|}{\|AD\|} + \frac{\|OE\|}{\|BE\|} + \frac{\|OF\|}{\|CF\|} = 1 \tag{2}$$

whenever all the points and ratios are defined. The double lines mean that *signed* lengths are to be used: if X, Y, Z are distinct collinear points, then, $\|XY\|/\|YZ\|$ is positive if Y lies between X and Z and is negative otherwise. We may call the left side of (2) a *cyclic sum* since each term corresponds to a vertex of the polygon (triangle). By contrast, the well-known Ceva's theorem concerns a *cyclic product*:

$$\frac{\|BD\|}{\|DC\|} \cdot \frac{\|CE\|}{\|EA\|} \cdot \frac{\|AF\|}{\|FB\|} = 1.$$

This latter relation has been extended in many different ways to more general polygons. (See, for example, any of the papers **[3]**–**[9]**.) However it appears that no analogous generalisations of cyclic sums to polygons with more than three sides are known.

The purpose of this note is to state two such results. One of these (Theorem 2) relates to *affine-regular polygons*, that is, to images of regular polygons under non-singular affine transformations. (An affine transformation is a linear transformation followed by a translation.) We recall that every triangle is affine-regular, and a quadrangle is affine-regular if and only if it is a parallelogram. For $n > 4$, an n-gon is affine regular if its diagonals and sides have the same parallelism and length ratios as in the corresponding regular n-gon. For example (see FIGURE 2) a pentagon $P = [V_0, \dots, V_4]$ is affine-regular if, for each i, the diagonal $V_{i-1}V_{i+1}$ is parallel to the side

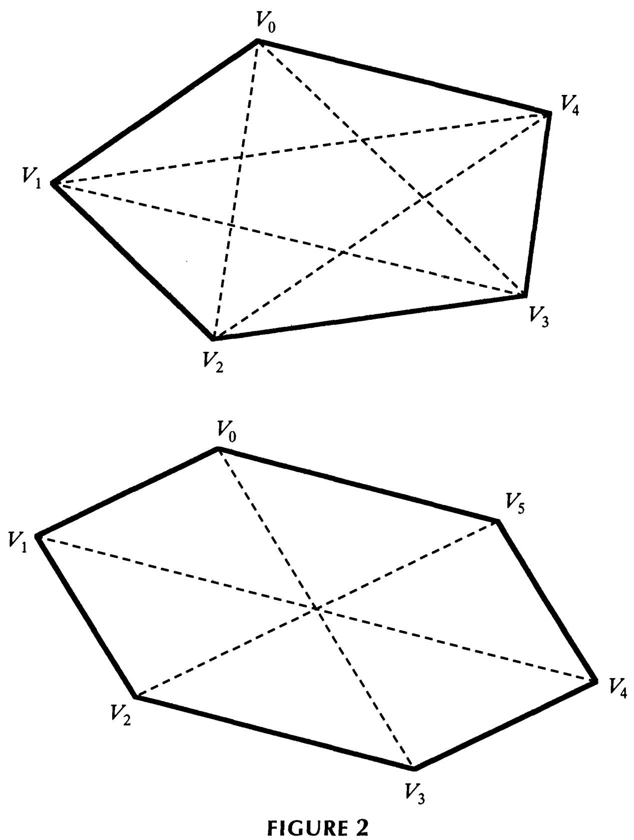


FIGURE 2
Affine-regular polygons.

$V_{i-2}V_{i+2}$ and $|V_{i-1}V_{i+1}| = \tau |V_{i-2}V_{i+2}|$ where τ is the golden section ratio. Here all subscripts are taken modulo 5. For a hexagon $P = [V_0, \dots, V_5]$, affine regularity implies that, for each i, the sides V_iV_{i+1}, $V_{i+3}V_{i+4}$ and the diagonal $V_{i+2}V_{i+5}$ are parallel and

$$|V_{i+2}V_{i+5}| = 2|V_iV_{i+1}| = 2|V_{i+3}V_{i+4}|.$$

Here all subscripts are reduced modulo 6. More generally, when considering an n-gon, we shall always consider subscripts to be reduced modulo n.

If $n=5$ and $P=[V_0,\dots,V_4]$ is an affine-regular pentagon, then, for *any* point C, Theorem 2 implies

$$\frac{\|W_0C\|}{\|W_0V_0\|}+\frac{\|W_1C\|}{\|W_1V_1\|}+\frac{\|W_2C\|}{\|W_2V_2\|}+\frac{\|W_3C\|}{\|W_3V_3\|}+\frac{\|W_4C\|}{\|W_4V_4\|}=2.236068\dots$$

where, for $i=0,\dots,4$, W_i is the point of intersection of V_iC with $V_{i-2}V_{i+2}$, see FIGURE 5(a). The crucial point to observe is that in the above equation (which is the analogue of (2)), the value of the constant on the right depends *only* on n and is independent both of the choice of C and of the affine-regular pentagon P.

If the polygon P is general, that is, not necessarily affine-regular, then a similar relation holds (Theorem 1); here, however, numerical coefficients appear before the terms on the left. For example, if $n=5$, and $P=[V_0,\dots,V_4]$ is a general pentagon, then for any point C, with the same notation as before (see FIGURE 3(a)),

$$\lambda_0\frac{\|W_0C\|}{\|W_0V_0\|}+\lambda_1\frac{\|W_1C\|}{\|W_1V_1\|}+\lambda_2\frac{\|W_2C\|}{\|W_2V_2\|}+\lambda_3\frac{\|W_3C\|}{\|W_3V_3\|}+\lambda_4\frac{\|W_4C\|}{\|W_4V_4\|}=K_1$$

where the coefficients $\lambda_0,\dots,\lambda_4$ and the constant K_1 depend *only* on the pentagon P and are independent of the choice of the point C. For the pentagon in FIGURE 3(a), we may take, approximately, $\lambda_0=6.175$, $\lambda_1=-3.176$, $\lambda_2=6.000$, $\lambda_3=3.506$, $\lambda_4=3.862$, and K_1 is the area of P; it therefore depends on the unit of length which is adopted. Theorem 1 shows how these values can be calculated explicitly for any polygon from the coordinates of its vertices.

The polygons we consider in Theorem 1 are completely general: edges may cross or overlap, vertices may be collinear, non-adjacent vertices may coincide, etc. The only requirement is that the points, lines, and ratios of segment lengths that occur in our assertions should be properly defined. Thus if a point is specified as the intersection of two lines then it will be assumed that these lines are not parallel, and if we require the quotient of the lengths of two line segments, then it will be assumed that the segment in the denominator has non-zero length.

Let $P=[V_0,\dots,V_{n-1}]$ be a given n-gon and r be any positive integer not divisible by n. Write $d=HCF(n,r)$ and let $m=n/d$. Then the *r-stellation* of P, denoted by $P^{(r)}$, is defined as the set of d m-gons

$$\begin{aligned}P^{(r)}=&\left[V_0,V_r,V_{2r},\dots,V_{(m-1)r}\right]\\ &\cup\left[V_1,V_{r+1},V_{2r+1},\dots,V_{(m-1)r+1}\right]\\ &\quad\dots\\ &\cup\left[V_{d-1},V_{r+d-1},V_{2r+d-1},\dots,V_{(m-1)r+d-1}\right].\end{aligned}$$

As subscripts are taken modulo n, every vertex of P occurs exactly once amongst the vertices of the polygons in $P^{(r)}$. It can be shown that the area of $P^{(r)}$ (which is defined as the sum of the areas of its constituent polygons) is given by the simple expression

$$\|P^{(r)}\|=\sum_{i=0}^{n-1}\|CV_iV_{i+r}\|, \tag{3}$$

where C is any fixed point of the plane. The terms on the right are the signed areas of the indicated triangles. The area $\|ABC\|$ of the triangle $[A,B,C]$ is taken to be positive if the vertices A, B, C are listed in a counterclockwise direction, and negative if they are listed in a clockwise direction.

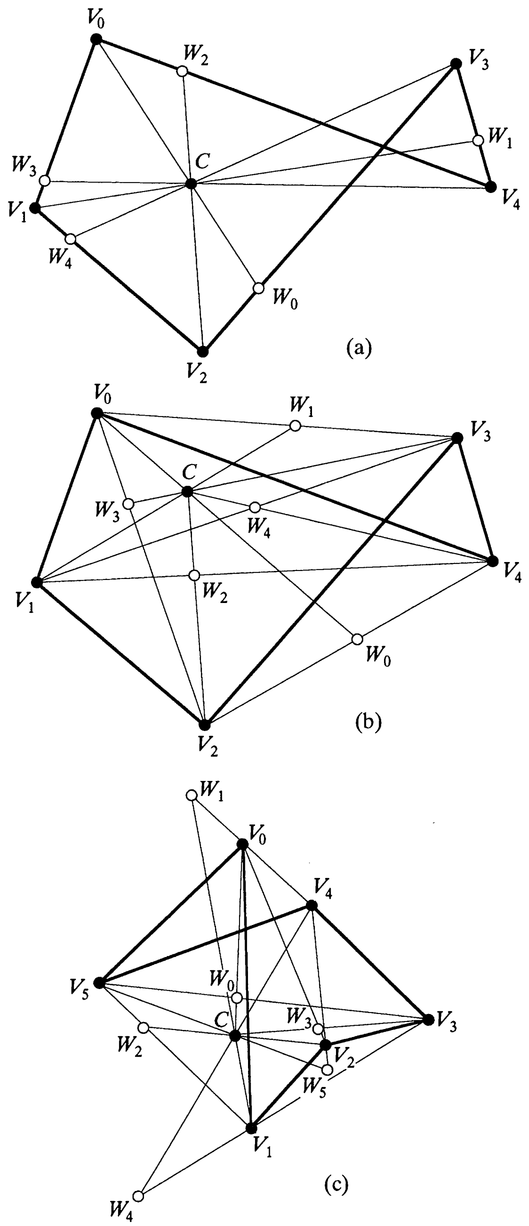


FIGURE 3

Illustrations of Theorem 1: (a) $n=5$, $j=k=2$; (b) $n=5$, $j=1$, $k=2$; and (c) $n=6$, $j=1$, $k=3$. The theorem asserts that in each case

$$\sum_{i=0}^{n-1} \lambda_i \frac{\|W_iC_i\|}{\|W_iV_i\|} = K_1$$

where the constants λ_i depend only on the polygon and the values of j and k, and not on the position of the point C.

THEOREM 1. *Let P be a given n-gon and j, k be positive integers such that $j + k < n$. Then there exist constants $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$ and K_1 with the following property. For an arbitrary point C and for $i = 0, 1, \ldots, n-1$, let W_i be the point of intersection of the line V_iC with the side or diagonal $V_{i-j}V_{i+k}$ of P. Then*

$$\sum_{i=0}^{n-1} \lambda_i \left\| \frac{W_iC}{W_iV_i} \right\| = K_1. \tag{4}$$

We stress that the values of the multipliers λ_i and of the constant K_1 depend *only* on the choice of P, j, and k, and *not* on the position of the point C.

In FIGURE 3 we show examples of the theorem for $n = 5$ and $n = 6$. The proof is very simple. By the area principle **[3]**,

$$\left\| \frac{W_iC}{W_iV_i} \right\| = \frac{\|CV_{i+k}V_{i-j}\|}{\|V_iV_{i+k}V_{i-j}\|},$$

since the triangles on the right have the same base $[V_{i-j}, V_{i+k}]$ and their heights are proportional to the lengths of the line segments $[W_i, C]$ and $[W_i, V_i]$, see FIGURE 4(a).

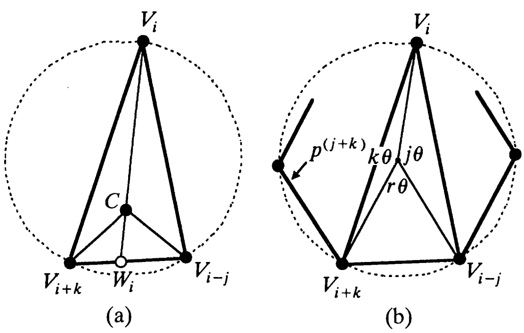


FIGURE 4

(a) The area principle used to show

$$\frac{\|W_iC\|}{\|W_iV_i\|} = \frac{\|CV_{i-j}V_{i+k}\|}{\|V_iV_{i-j}V_{i+k}\|}.$$

(b) The calculation of the constant in Theorem 2. Here $r = n - j - k$.

Hence

$$\sum_{i=0}^{n-1} \|V_iV_{i+k}V_{i-j}\| \cdot \left\| \frac{W_iC}{W_iV_i} \right\| = \sum_{i=0}^{n-1} \|CV_{i+k}V_{i-j}\| = \|P^{(j+k)}\|,$$

where the equality on the right follows from (3). Thus (4) holds with $\lambda_i = \|V_iV_{i+k}V_{i-j}\|$ and $K_1 = \|P^{(j+k)}\|$. These relations enable the values of the coefficients λ_i and constant K_1 to be calculated for any given polygon P and integers j, k.

THEOREM 2. *Let P be an affine-regular n-gon, C a fixed point, and j, k be positive integers such that $j+k<n$. For each $i=0,1,\ldots,n-1$, let W_i be the point of intersection of the line V_iC and the side or diagonal $V_{i-j}V_{i+k}$ of P. Then*

$$\sum_{i=0}^{n-1}\left\|\frac{W_iC}{W_iV_i}\right\| = K_2, \tag{5}$$

where K_2 is a constant, depending on n, j, and k, but independent of the choice of affine regular n-gon P and of the position the point C. The value of K_2 is given by

$$K_2 = \frac{n\sin(j+k)\theta}{\sin(j+k)\theta - \sin j\theta - \sin k\theta}$$

where $\theta = 2\pi/n$.

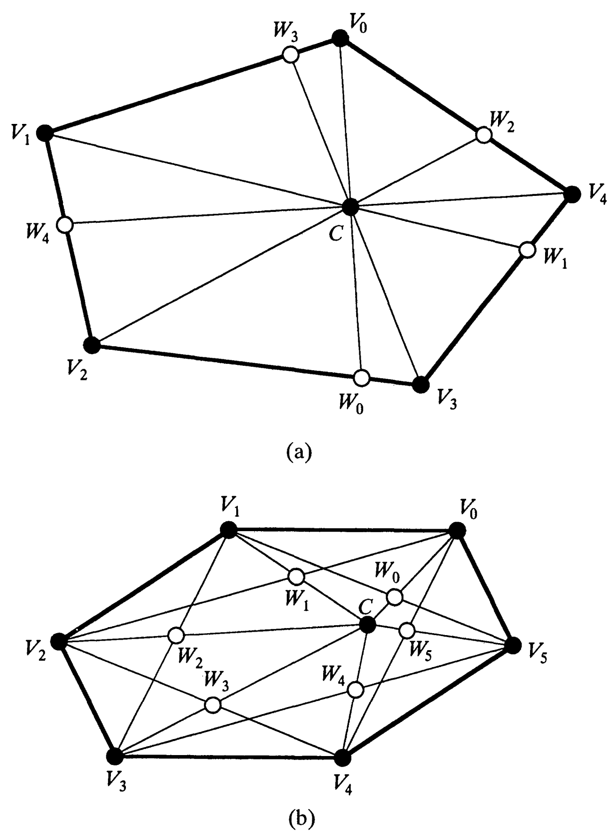


FIGURE 5

Illustrations of Theorem 2: (a) $n=5$, $j=k=2$; and (b) $n=6$, $j=k=1$. The theorem asserts that in each case

$$\sum_{i=0}^{n-1}\frac{\|W_iC\|}{\|W_iV_i\|} = K_2$$

where the constant K_2 depends only on n, j and k and is independent of the affine-regular polygon that is chosen and of the point C.

In FIGURE 5 we show examples of Theorem 2 for $n=5$ and $n=6$. The proof of Theorem 2 follows immediately from Theorem 1 and the observation that

$$\frac{K_1}{\lambda_i} = \frac{\|P^{(j+k)}\|}{\|V_i V_{i+k} V_{i-j}\|} = K_2$$

for all i. To see this, we observe that affine-invariance implies that we may, without loss of generality, take P to be a regular n-gon of circumradius 1. Then, by elementary trigonometry,

$$\|V_1 V_{i+k} V_{i-j}\| = \tfrac{1}{2}(\sin j\theta + \sin k\theta + \sin(n-j-k)\theta)$$

and

$$\|P^{(j+k)}\| = \tfrac{1}{2} n \sin(n-j-k)\theta,$$

see FIGURE 4(b). From these the value of K_2 given in the theorem follows immediately. This completes the proof of Theorem 2.

Finally we remark that, in view of the many investigations of cyclic products during this century, it is remarkable that the simple results on cyclic sums, presented here, were not discovered many years ago.

Acknowledgment I am indebted to Branko Grünbaum and the referees for many helpful comments and also for the references [1] and [2]. The latter may be regarded as a sequel in that it contains a theorem on weighted cyclic sums for polygons which the author acknowledges is based on the results of this paper.

REFERENCES

1. L. Euler, Geometrica et sphaerica quaedam, *Memoires de l'Academie des Sciences de St.-Petersbourg* 5(1812), 96–114. (Received May 1, 1780). = L. Euleri *Opera Omnis*, Series 1, Vol. 26 (1953), 344–358.
2. Branko Grünbaum, Cyclic ratio sums and products, *Crux. Math.* 24 (1998), 20–25.
3. Branko Grünbaum and G. C. Shephard, Ceva, Menelaus, and the area principle, this MAGAZINE 68 (1995), 254–268.
4. Branko Grünbaum and G. C. Shephard, A new Ceva-type theorem, *Math. Gazette* 80 (1996), 492–500.
5. Branko Grünbaum and G. C. Shephard, Ceva, Menelaus, and selftransversality, *Geom. Dedicata* 65 (1997), 179–192.
6. H. Gülicher, Problem 64 and solution, *Math. Semesterberichte* 40 (1993), 91–92 and 42 (1995), 182–185.
7. L. Hoehn, A Menelaus-type theorem for the pentagram, this MAGAZINE 66 (1993), 121–123.
8. L. Hoehn, A concurrency theorem and Geometer's Sketchpad, *College Math. Journal* 28 (1997), 129–132.
9. Guido M. Pinkernell, Identities on point-line figures in the Euclidean plane, this MAGAZINE 69 (1996), 377–382.

When Can One Load a Set of Dice so that the Sum Is Uniformly Distributed?

WILLIAM I. GASARCH
CLYDE P. KRUSKAL
Department of Computer Science
University of Maryland
College Park, MD 20742

Introduction If you toss two fair six-sided dice, you will get a number between 2 and 12. Although each die is fair, the sum is not: the probability of getting a 2 is $\frac{1}{36}$, while the probability of a 7 is $\frac{1}{6}$. Thus the question arises: Can *unfair* (or loaded) dice lead to a fair sum? The answer is no, as was shown by Honsberger **[4]**, using elementary methods. A proof using generating functions is in Hofri's book **[3]**.

In **[1]**, Chen, Rao, and Shreve raised the more general question of what happens with n m-sided dice; they showed that the answer is *still* no. In this note we generalize their result by considering m dice $D_1, \dots, D_m$, where D_i is n_i-sided. We find that there are cases where one gets a fair sum, and we characterize exactly when this happens. Our techniques also lead to a different proof of the theorem of Chen, Rao, and Shreve.

A die is *fair* if all numbers appear with equal probability. A tuple of dice is *fair* if all sums appear with equal probability. For convenience, we will number an n-sided die $0, 1, \dots, n-1$, and use the following definitions:

DEFINITIONS. Let $m, n_1, \dots, n_m \geq 2$, and let $N = \sum_{j=1}^{m} n_j$.

- An n-sided *die* is an ordered n-tuple of numbers $(p_0, \dots, p_{n-1})$ such that (i) for all i, $0 \leq p_i \leq 1$; and (ii) $\sum_{i=0}^{n-1} p_i = 1$. (We think of p_i as the probability of rolling an i.)
- For $1 \leq j \leq m$ let D_j be an n_j-sided die. The m-tuple $(D_1, \dots, D_m)$ is *fair* if, for all i, $0 \leq i \leq \sum_{j=1}^{m}(n_j - 1) = N - m$, the probability of rolling a sum of i is $\frac{1}{N-m+1}$.
- The ordered m-tuple $(n_1, \dots, n_m)$ is *fair* if there exists $(D_1, \dots, D_m)$ such that (i) each D_j is an n_j-sided die; and (ii) $(D_1, \dots, D_m)$ is fair. In this case, we say that $(n_1, \dots, n_m)$ is fair via $(D_1, \dots, D_m)$.

Example. The ordered pair $(2, 3)$ is fair since $((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, 0, \frac{1}{2}))$ is fair. Every sum has probability $1/4$ of being rolled.

In the preceding example, every sum can be rolled in one and only one way. We will prove that, for a pair of dice to be fair, this condition *must* hold.

DEFINITIONS. A die $D = (p_0, \dots, p_{n-1})$ is *symmetric* if, for all i, $p_i = p_{n-1-i}$. D is *nice* if it is symmetric and, for all i, either $p_i = 0$ or $p_i = p_0$.

Note that if a die is nice then $p_0 \neq 0$—otherwise, $p_i = 0$ for all i.

MAIN THEOREM: *If $(D_1, \dots, D_m)$ is fair then each D_i is nice.*

MAIN COROLLARY: *$(D_1, \dots, D_m)$ is fair if and only if each D_i is nice and every sum can be rolled in exactly one way.*

ANOTHER MAIN COROLLARY: *There is a decision procedure that will, given $(n_1, \dots, n_m)$, decide whether the tuple is fair.*

Proving the main theorem In this section we prove the main theorem. Our only tools are generating functions and some rudiments of complex algebra.

If $D = (p_0, \dots, p_{n-1})$ is a die then the polynomial $F_D(z) = \sum_{i=0}^{n-1} p_i z^i$ is the *generating function* for D. The following key observation links tuples of dice and products of generating functions:

> If $(D_1, \dots, D_m)$ is an m-tuple of dice, then the coefficient of z^i in $\prod_{i=1}^{m} F_{D_i}(z)$ is the probability of obtaining a sum of i.

The *kth roots of unity* are the complex solutions of the equation $z^k - 1 = 0$. It is easy to see that all of these roots lie on the complex unit circle. If k is even, then 1 and -1 are the only real roots of unity; if k is odd, then 1 is the only real root of unity. All the roots of unity have multiplicity 1.

LEMMA 1. *If $(n_1, \dots, n_m)$ is fair via $(D_1, \dots, D_m)$ and $N = \sum_{j=1}^{m} n_j$, then the roots of $\prod_{j=1}^{m} F_{D_j}(z)$ are exactly the $(N - m + 1)$th roots of unity except 1. Each root has multiplicity one.*

Proof. Since $(D_1, \dots, D_m)$ is fair the probability of the sum being i is $\frac{1}{N-m+1}$. Hence

$$\prod_{j=1}^{m} F_{D_j}(z) = \sum_{i=0}^{N-m} \frac{z^i}{N-m+1} = \frac{(z^{N-m+1} - 1)}{(N-m+1)(z-1)},$$

and it follows that $z^{N-m+1} - 1 = (N - m + 1)(z - 1)\prod_{j=1}^{m} F_{D_j}(z)$. □

We use Lemma 1 to restrict the kind of dice that can be used in an m-tuple of fair dice.

LEMMA 2. *If $(D_1, \dots, D_m)$ is fair, then each D_j is symmetric.*

Proof. Let $D_j = (p_0, \dots, p_{n-1})$ and let $r_1, \dots, r_{n-1}$ be the roots of $F_{D_j}(z)$. Since $F_{D_j}(z)$ has real coefficients and has roots on the unit circle

$$\{r_1, \dots, r_{n-1}\} = \{\overline{r_1}, \dots, \overline{r_{n-1}}\} = \left\{\frac{1}{r_1}, \dots, \frac{1}{r_{n-1}}\right\}$$

where $\overline{r_i}$ denotes the complex conjugate of r_i. Hence the roots of $F_{D_j}(z)$ are the roots of $F_{D_j}(\frac{1}{z})$ which are also the roots of $z^{n-1}F_{D_j}(\frac{1}{z}) = \sum_{i=0}^{n-1} p_{n-1-i} z^i$. Since $\sum_{i=0}^{n-1} p_i z^i$ and $\sum_{i=0}^{n-1} p_{n-i} z^i$ have the same roots, the same degree, and $\sum_{i=0}^{n-1} p_i = 1$, these polynomials are identical. Hence $p_i = p_{n-1-i}$. □

We will prove our main theorem by induction on the number of dice. To this end, we need a way to combine two dice into one:

DEFINITION. $[D_1, D_2]$ is the die obtained by rolling dice D_1 and D_2 and considering their sum.

The following lemma is key to the proof of the theorem.

LEMMA 3. *If D_1 and D_2 are symmetric and $[D_1, D_2]$ is nice, then D_1 and D_2 are nice.*

Proof. Let $D_1 = (p_0, \dots, p_{n_1-1})$ and $D_2 = (q_0, \dots, q_{n_2-1})$. We assume $n_1 \le n_2$. Since D_1 and D_2 are symmetric, $p_0 \neq 0$ and $q_0 \neq 0$. Let $[D_1, D_2] = (r_0, \dots, r_{n_1+n_2-2})$.

We first prove that, for all i with $1 \leq i \leq n_1 - 1$, either $p_i = 0$ or $q_i = 0$. Since $[D_1, D_2]$ is nice $r_{n_2-1} = r_0$ or $r_{n_2-1} = 0$. Hence

$$p_0 q_{n_2-1} + p_1 q_{n_2-2} + \cdots + p_{n_1-1} q_{n_2-n_1} = a p_0 q_0,$$

where $a \in \{0, 1\}$.

Since $q_i = q_{n_2-1-i}$ for all i, we have $p_0 q_0 + p_1 q_1 + \cdots + p_{n_1-1} q_{n_1-1} = a p_0 q_0$, so

$$p_1 q_1 + \cdots + p_{n_1-1} q_{n_1-1} = (a-1) p_0 q_0 \leq 0.$$

Since all the p_i and q_i are nonnegative, we have, for all i with $1 \leq i \leq n_1 - 1$, either $p_i = 0$ or $q_i = 0$.

We now prove that, for all i with $0 \leq i \leq n_1 - 1$, the following two conditions hold:

(i) either $p_i = 0$ or $p_i = p_0$; (ii) either $q_i = 0$ or $q_i = q_0$.

We prove this by induction on i. For $i = 0$ this is trivial. Assume it holds for all $i' < i$. Since $[D_1, D_2]$ is nice, either $r_i = r_0 = p_0 q_0$ or $r_i = 0$ for all i, so

$$p_0 q_i + p_1 q_{i-1} + \cdots + p_{i-1} q_1 + p_i q_0 = a p_0 q_0,$$

where $a \in \{0, 1\}$. By the induction hypothesis, each term $p_1 q_{i-1}, \ldots, p_{i-1} q_1$ is either 0 or $p_0 q_0$. Hence there exists $b \leq a$ such that $p_0 q_i + b p_0 q_0 + p_i q_0 = a p_0 q_0$, so

$$p_0 q_i + p_i q_0 = (a-b) p_0 q_0. \tag{1}$$

Note that $a - b \in \{0, 1\}$.

Now if $a - b = 1$, then $p_0 q_i + p_i q_0 = p_0 q_0$. If $p_i = 0$ (resp. $q_i = 0$) then $q_i = q_0$ (resp. $p_i = p_0$). By equation (1), either $p_i = 0$ or $q_i = 0$.

If, alternatively, $a - b = 0$, then $p_0 q_i + p_i q_0 = 0$. Since $p_0 \neq 0$ and $q_0 \neq 0$, we have $p_i = q_i = 0$.

It remains to prove that, for all i with $n_1 \leq i \leq n_2 - 1$, either $q_i = 0$ or $q_i = q_0$. This is done by another induction on i, similar to the one just given. □

THEOREM 4. *If $(D_1, \ldots, D_m)$ is fair, then each D_j is nice.*

Proof. We prove this by induction on m. The $m = 1$ case is obvious. Assume the result holds for $m - 1$, and let $(D_1, \ldots, D_m)$ be fair. Then a simple calculation shows that $([D_1, D_2], D_3, D_4, \ldots, D_m)$ is fair. By the inductive hypothesis, each of the dice $[D_1, D_2], D_3, D_4, \ldots, D_m$ is nice. By Lemmas 2 and 3, D_1 and D_2 are nice. □

COROLLARY 5. *The tuple $(D_1, \ldots, D_m)$ is fair if and only if each D_i is nice and every sum can be rolled in exactly one way.*

Proof. Assume that $(D_1, \ldots, D_m)$ is fair, and that, for all j, D_j has n_j sides—we write $D_j = (p_{j0}, p_{j1}, p_{j2}, \ldots, p_{j(n_j-1)})$. Let Prob($a$) denote the probability of rolling an a. Assume, by way of contradiction, that there exist a and distinct $(b_1, \ldots, b_m)$ and $(c_1, \ldots, c_m)$ such that

$$\text{(i)} \quad a = \sum_{j=1}^{m} b_j = \sum_{j=1}^{m} c_j; \quad \text{(ii)} \quad \prod_{j=1}^{m} p_{j,b_j} \neq 0; \quad \text{(iii)} \quad \prod_{j=1}^{m} p_{j,c_j} \neq 0.$$

Since $(b_1, \ldots, b_m)$ and $(c_1, \ldots, c_m)$ are two distinct ways of rolling an a, we have $\prod_{j=1}^{m} p_{j,b_j} + \prod_{j=1}^{m} p_{j,c_j} \leq \text{Prob}(a)$.

Since $(D_1, \ldots, D_m)$ is fair, each die is nice, so $\prod_{j=1}^m p_{j,b_j} = \prod_{j=1}^m p_{j,c_j} = \prod_{j=1}^m p_{j,0}$. Hence

$$2\text{Prob}(0) = \prod_{j=1}^m p_{j,0} + \prod_{j=1}^m p_{j,0} = \prod_{j=1}^m p_{j,b_j} + \prod_{j=1}^m p_{j,c_j} \le \text{Prob}(a) = \text{Prob}(0).$$

This implies that Prob(0) $= 0$, which contradicts $(D_1, \ldots, D_m)$ being fair.

To prove the converse, assume that each D_i is nice and every sum can be rolled in exactly one way. Let a be rolled by $(b_1, \ldots, b_m)$ where, for all i, $p_{i,b_i} \neq 0$. Then the probability of rolling an a is $\prod_{j=1}^m p_{j,b_j} = \prod_{j=1}^m p_{j,0}$. Since this quantity is independent of a, $(D_1, \ldots, D_m)$ is fair. □

COROLLARY 6. *One can determine whether any given tuple $(n_1, \ldots, n_m)$ is fair.*

Proof. Given $(n_1, \ldots, n_m)$, we need only consider dice $(D_1, \ldots, D_m)$ with each D_i nice. There are only a finite number of possibilities; each one can be checked for fairness. □

The number of fair n-sided dice $(p_0, \ldots, p_{n-1})$ is the number of ways to assign values to $p_0, \ldots, p_{n-1}$ such that (1) for all i, $p_i = p_{n-1-I}$, (2) $p_0 \neq 0$, (3) for all i either $p_i = p_0$ or $p_i = 0$, and (4) $\sum_{i=0}^{n-1} = 1$. This is $\sum_{i=0}^{\lceil n/2 \rceil -1} \binom{\lceil n/2 \rceil - 1}{i} = 2^{\lceil n/2 \rceil - 1}$. Thus the number of possibilities that must be considered is bounded by $\prod_{i=1}^m 2^{\lceil n_i/2 \rceil - 1}$.

Curious facts Next we explore some curious facts that follow from our work.

COROLLARY 7. *If a set of dice is fair, then at most one of them has an even number of sides.*

Proof. Assume, by way of contradiction, that $(D_1, \ldots, D_m)$ is fair and that n_i and n_j, $i \neq j$, are both even; then $n_i - 1$ and $n_j - 1$ are odd. Then the polynomials $F_{D_i}(z)$ and $F_{D_j}(z)$ have odd degree; since they also have real coefficients, both $F_{D_i}(z)$ and $F_{D_j}(z)$ must have real roots. Therefore $\prod_{j=1}^m F_{D_j}(z)$ either has at least two distinct real roots or one real root of multiplicity at least 2. The first possibility contradicts Lemma 1—there is at most one real $(N - m + 1)$th root of unity other than 1, where $N = \sum_{j=1}^m n_j$. The second possibility also contradicts Lemma 1, since all roots of $\prod_{j=1}^m F_{D_j}(z)$ have multiplicity 1. □

The next corollary is the main theorem from **[1]**. We give an alternative proof.

COROLLARY 8. *If a set of dice is fair, then no two have the same number of sides.*

Proof. Suppose for the sake of contradiction that $n = n_i = n_j$, with $i \neq j$, and that $(D_1, \ldots, D_m)$ is fair. Let $D_k = (p_{k0}, p_{k1}, p_{k2}, \ldots, p_{k(n_k-1)})$. Let Prob$(n-1)$ be the probability of rolling an $n - 1$. By Lemma 2, each D_k is symmetric, so $p_{i,n-1} = p_{i,0}$ and $p_{j,n-1} = p_{j,0}$. Hence

$$2\,\text{Prob}(0) = 2\prod_{k=1}^m p_{k,0} = \left(p_{i,n-1}\, p_{j,0} \prod_{\substack{1 \le k \le m \\ k \neq i,j}} p_{k,0} \right) + \left(p_{i,0}\, p_{j,n-1} \prod_{\substack{1 \le k \le m \\ k \neq i,j}} p_{k,0} \right)$$

$$\le \text{Prob}(n-1) = \text{Prob}(0).$$

Hence Prob(0) $= 0$, which contradicts $(D_1, \ldots, D_m)$ being fair. □

The next corollary mentions the Euler ϕ-function: for a positive integer n, $\phi(n)$ is the number of positive integers less than n that are relatively prime to n. The proof involves *cyclotomic polynomials*. For a positive integer n, the nth cyclotomic polynomial $\Phi_n(z)$ is a complex polynomial of degree $\phi(n)$; the roots of $\Phi_n(z)$ are the *primitive* nth roots of unity—those for which no lower power than n gives 1. (Curious readers may find more information on cyclotomic polynomials in any abstract algebra textbook; see, e.g., **[2]**.)

COROLLARY 9. *If $(n_1, \dots, n_m)$ is fair and $N = \sum_{j=1}^m n_j$, then $\phi(N-m+1) \le \max_j(n_j - 1)$. Hence if $N-m+1$ is prime, then $(n_1, \dots, n_m)$ is not fair.*

Proof. Assume $(n_1, \dots, n_m)$ is fair via $(D_1, \dots, D_m)$. By Theorem 4, each $F_{D_j}(z)$ has rational coefficients. By Lemma 1, a root of one of the $F_{D_j}(z)$ is a primitive $(N-m+1)$th root of unity. Therefore, $\Phi_{N-m+1}(z)$ divides some $F_{D_j}(z)$, where $\Phi_{N-m+1}(z)$ is the $(N-m+1)$th cyclotomic polynomial. Hence, for some j, $\phi(N-m+1) \le n_j - 1$. $\square$

Examples The following examples illustrate our results concretely.

1. By Corollary 7, no tuple of the form $(2, 2i)$ is fair.
2. All tuples of the form $(2, 2i-1)$, $i \ge 2$, are fair: use dice $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{i}, 0, \frac{1}{i}, 0, \dots, \frac{1}{i}, 0, \frac{1}{i})$. (This produces $(2i)$-sided dice.)
3. All tuples $(i, i+1)$, $i \ge 2$, are fair: use dice $(\frac{1}{i}, \frac{1}{i}, \dots, \frac{1}{i})$ and $(\frac{1}{2}, \overbrace{0, 0, \dots, 0}^{i-1}, \frac{1}{2})$. (This produces $(2i)$-sided dice.)

The preceding examples show that, for $i \ge 3$, fair $(2i)$-sided dice can be produced in at least two different ways. For example, a six-sided die can be produced from $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{3}, 0, \frac{1}{3}, 0, \frac{1}{3})$, and also from $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $(\frac{1}{2}, 0, 0, \frac{1}{2})$.

4. All tuples $(3, 3i-2)$, $i \ge 2$, are fair: use dice $(\frac{1}{i}, 0, 0, \frac{1}{i}, 0, 0, \dots, \frac{1}{i}, 0, 0, \frac{1}{i})$ and $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. (This produces $(3i)$-sided dice.)
5. All tuples $(3, 4i-2)$, $i \ge 1$, are fair: use dice $(\frac{1}{2}, 0, \frac{1}{2})$ and $(\frac{1}{2i}, \frac{1}{2i}, 0, 0, \frac{1}{2i}, \frac{1}{2i}, 0, 0, \dots, \frac{1}{2i}, \frac{1}{2i}, 0, 0, \frac{1}{2i}, \frac{1}{2i})$. (This produces $(4i)$-sided dice.)

The last two examples cover all the fair 2-tuples $(3, i)$, since we have exhausted all combinations of nice 3-sided dice. Some 2-tuples $(3, i)$ are fair in two different ways. For example, $(3, 10)$ is produced from $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $(\frac{1}{4}, 0, 0, \frac{1}{4}, 0, 0, \frac{1}{4}, 0, 0, \frac{1}{4})$, and also from $(\frac{1}{2}, 0, \frac{1}{2})$ and $(\frac{1}{6}, \frac{1}{6}, 0, 0, \frac{1}{6}, \frac{1}{6}, 0, 0, \frac{1}{6}, \frac{1}{6})$.

Finally, observe that one can construct fair dice from arbitrarily long tuples. All tuples of the form $(2^0+1, 2^1+1, 2^2+1, \dots, 2^{m-1}+1)$, $m \ge 2$, are fair: use dice $(\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 0, \frac{1}{2})$, $(\frac{1}{2}, 0, 0, 0, \frac{1}{2}), \dots, (\frac{1}{2}, \overbrace{0, 0, \dots, 0}^{2^{m-1}-1}, \frac{1}{2})$. (This produces (2^m)-sided dice.)

Almost uniform sums When giving a talk on this topic we were asked whether we can get "close to" a uniform sum using *real* dice. In this section, therefore, we assume dice are numbered from 1 to n.

There are several ways to measure how close a distribution is to uniform. We wrote a Matlab program to find, for given n, vectors $(p_1, \dots, p_n)$ and $(q_1, \dots, q_n)$ such that (i) $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$; (ii) for all i, $0 \le p_i, q_i \le 1$; and (iii) if we interpret the vectors as dice, then

$$\sum_{m=2}^{2n} \left(\text{Prob(sum is } m) - \frac{1}{2n-1} \right)^2$$

is minimized. For all n, Matlab produced symmetric dice that were identical to each other. However, Matlab *does not* guarantee that the results are the true optimum, so the question of whether or not optimal dice must be identical and symmetric is interesting and open, even in the cases where we obtained numerical results. Other measures of being "close to uniform" might also be considered.

We provide the statistics, in the $n = 6$ case, first for two (ordinary) fair dice and then for the dice we obtained from the program.
If each die is fair then the following happens.

	prob(the sum is 2) = 0.027778
	prob(the sum is 3) = 0.055556
	prob(the sum is 4) = 0.083333
prob(the die is 1) = 0.166667	prob(the sum is 5) = 0.111111
prob(the die is 2) = 0.166667	prob(the sum is 6) = 0.138889
prob(the die is 3) = 0.166667	prob(the sum is 7) = 0.166667
prob(the die is 4) = 0.166667	prob(the sum is 8) = 0.138889
prob(the die is 5) = 0.166667	prob(the sum is 9) = 0.111111
prob(the die is 6) = 0.166667	prob(the sum is 10) = 0.083333
	prob(the sum is 11) = 0.055556
	prob(the sum is 12) = 0.027778

The two dice obtained by the Matlab program were unfair but identical, and had the following properties.

	prob(the sum is 2) = 0.059479
	prob(the sum is 3) = 0.067058
	prob(the sum is 4) = 0.076768
prob(the die is 1) = 0.243883	prob(the sum is 5) = 0.090488
prob(the die is 2) = 0.137480	prob(the sum is 6) = 0.113753
prob(the die is 3) = 0.118637	prob(the sum is 7) = 0.184909
prob(the die is 4) = 0.118637	prob(the sum is 8) = 0.113753
prob(the die is 5) = 0.137480	prob(the sum is 9) = 0.090488
prob(the die is 6) = 0.243883	prob(the sum is 10) = 0.076768
	prob(the sum is 11) = 0.067058
	prob(the sum is 12) = 0.059479

Acknowledgment The first author acknowledges NSF grant CCR-97-32692 for supporting this work. Both authors thank the referees for many helpful comments.

REFERENCES

1. G. Chen, M. Rao, and W. Shreve, Can one load a set of dice so that the sum is uniformly distributed? this MAGAZINE 70 (1997), 204–207.
2. I. N. Herstein, *Topics In Algebra*, Wiley, New York, NY, 1975.
3. M. Hofri, *Analysis of Algorithms*, Oxford University Press, Oxford, UK, 1995.
4. R. Honsberger, *Mathematical Morsels*, The Mathematical Association of America, Washington, DC, 1978.

When Is $U(n)$ Cyclic? An Algebraic Approach

DAVID R. GUICHARD
Whitman College
Walla Walla, WA 99362

Early in a typical abstract algebra course we learn that the set $U(n) = \{0 < x \le n \mid \gcd(x, n) = 1\}$ is a group under multiplication mod n for every $n \ge 1$. This first appears as example 11 in Chapter 2 of Gallian's excellent text **[2]**, for instance. These groups are particularly nice: it is not hard to see, but not immediately obvious, that they *are* groups; they are important in some modern cryptographic applications; and they figure prominently in elementary number theory.

Some of the groups $U(n)$ are cyclic and some are not, and the two categories can be completely characterized by the form of the prime factorization of n. If $U(n)$ is cyclic then we can write $U(n) = \langle g \rangle$ for some $g \in \mathbb{Z}_n$, relatively prime to n. In number theory g is known as a *primitive root* modulo n; we will call the characterization of those n with primitive roots the *Primitive Root Theorem*, or *PRT*.

I recently taught an abstract algebra course using Gallian's text, and I wanted to prove the PRT for the class. Though this result is standard in elementary number theory books (see, e.g., **[3]**), the number-theoretic notation and proofs would have led me farther afield than I cared to go. I failed to find an algebraic proof of the result, but put one together by mining the proof in **[3]** for hints. The proof uses many results and exercises from **[2]**; this made it a satisfying conclusion to my course. Most of the proof requires only group theory, though some field theory and experience with polynomial rings is required at the very end.

This proof should be accessible to students who have been through any standard undergraduate course. I will refer explicitly to theorems and exercises in **[2]**.

Here is what we are shooting for:

THEOREM. (*Primitive Root Theorem*) *$U(n)$ is cyclic if and only if n is* 1, 2, 4, p^k, *or* $2p^k$, *where p is an odd prime and* $k \ge 1$.

Preliminaries First, we need some facts from number theory. The number of elements in $U(n)$ is commonly denoted by $\phi(n)$, the *Euler phi-function* or *totient function*. When p is prime, $\phi(p) = p - 1$, because every number in $\{1,2,\dots,p-1\}$ is relatively prime to p. Also, $\phi(p^k) = p^k - p^{k-1}$ for prime p, because precisely p^{k-1} of the p^k integers in $\{1,2,\dots,p^k\}$ are multiples of p, and all other integers in that range are relatively prime to p^k. Note for future reference that if p is an odd prime, or if $p = 2$ and $k \ge 2$, then $p^k - p^{k-1}$ is even. (This is all we will need about ϕ, but it is also true that if m and n are relatively prime then $\phi(mn) = \phi(m)\phi(n)$. Thus $\phi(n) = \prod_{i=1}^{k}(p_i^{a_i} - p_i^{a_i-1})$ if $n = \prod_{i=1}^{k} p_i^{a_i}$ is the prime factorization of n.)

It is easy to check the primitive root theorem for $n = 1, 2, 4$ directly. (Don't take my word for it—do it!)

Recall that every cyclic group has exactly one subgroup of order d for each d that divides the order of the group. Thus we may show that $U(2^k)$ is not cyclic for $k > 2$ by showing that $U(2^k)$ contains two distinct elements of order 2, each of which generates a subgroup of order 2. We leave this as an exercise; it is number 54 in chapter 4 of Gallian.

In Chapter 8, *External Direct Products*, Gallian characterizes the direct products that are cyclic groups:

If $G \cong G_1 \oplus \cdots \oplus G_m$, then G is cyclic if and only if the G_i are cyclic and their orders are pairwise relatively prime.

Gallian also proves (modulo some exercises left to the reader) that if $m = n_1 n_2 \cdots n_k$, and the n_i are pairwise relatively prime, then $U(m) \cong U(n_1) \oplus \cdots \oplus U(n_k)$. It is now not hard to see that $U(n)$ is not cyclic if n is divisible by two distinct odd primes or by 4 and an odd prime, using the fact (mentioned earlier) that $p^k - p^{k-1}$ is even when p is an odd prime or $p = 2$ and $k \geq 2$, together with $U(\prod_{i=1}^k p_i^{a_i}) \cong U(p_1^{a_1}) \oplus \cdots \oplus U(p_k^{a_k})$. (Exercise 46 of Chapter 8 is essentially this result.)

Now we know that the only groups that might be cyclic are $U(p^k)$ and $U(2p^k)$. (A different algebraic proof of this much appeared in **[1]**.)

In what follows, p always denotes an odd prime. Since $U(2p^k) \cong U(2) \oplus U(p^k) \cong U(p^k)$, we need only show that $U(p^k)$ is cyclic. We will show that, if $U(p)$ is cyclic, then $U(p^2)$ is cyclic; that this implies that $U(p^k)$ is cyclic for $k > 2$; and, finally, that $U(p)$ is cyclic.

If G is a finite group, every $g \in G$ has an *order*, denoted $|g|$, which is the smallest positive integer m such that $g^m = e$ (e is the identity of the group). Recall that $g^k = e$ if and only if $|g|$ divides k, and that Lagrange's theorem tells us that $|g|$ divides $|G|$.

For the first step, we suppose that $U(p)$ is cyclic and show that $U(p^2)$ is cyclic. Let $U(p) = \langle g \rangle$, $g \in \{1, 2, \ldots, p-1\}$. We will show that either g or $g + p$ generates $U(p^2)$. Let h_t be the order of $g + tp$, $t = 0$ or 1, so that $(g + tp)^{h_t} \equiv 1 \pmod{p^2}$; then $(g + tp)^{h_t} \equiv 1 \pmod{p}$ as well. Now

$$\begin{aligned} 1 \equiv (g+tp)^{h_t} &= g^{h_t} + \binom{h_t}{1} g^{h_t - 1} tp + \binom{h_t}{2} g^{h_t - 2} (tp)^2 + \cdots + (tp)^{h_t} \\ &\equiv g^{h_t} \pmod{p}, \end{aligned}$$

so the order of g in $U(p)$ divides h_t, that is, $(p-1)$ divides h_t. Since h_t is the order of an element of $U(p^2)$, we also know that h_t divides $|U(p^2)| = p(p-1)$. Thus, $h_t = p - 1$ or $h_t = p(p-1)$; we want to show that the latter is true for at least one of $t = 0$ or $t = 1$. Suppose not, so that

$$g^{p-1} \equiv (g+p)^{p-1} \equiv 1 \pmod{p^2}.$$

Then

$$(g+p)^{p-1} = g^{p-1} + \binom{p-1}{1} g^{p-2} p + \binom{p-1}{2} g^{p-3} p^2 + \cdots + p^{p-1},$$

or, modulo p^2,

$$1 \equiv 1 + (p-1) g^{p-2} p, \quad \text{so} \quad 0 \equiv (p-1) g^{p-2} p.$$

But p^2 does not divide $(p-1)g^{p-2}p$. This contradiction implies that either g or $g + p$ has order $p(p-1)$, and generates $U(p^2)$.

Now we suppose that g generates $U(p^2)$ and show that g generates $U(p^k)$, $k \geq 2$. We proceed by induction. Suppose that g generates $U(p^2)$ and $U(p^i)$, for all i such that $2 \leq i \leq k$, where $k \geq 2$. In particular, in $U(p^k)$, the order of g is $p^{k-1}(p-1)$ and in $U(p^{k-1})$ the order of g is $p^{k-2}(p-1)$. Let h denote the order of g in

$U(p^{k+1})$; we want to show that $h = p^k(p-1)$. Since $g^h \equiv 1 \pmod{p^{k+1}}$, it is also true that $g^h \equiv 1 \pmod{p^k}$. This means that the order of g in $U(p^k)$ divides h, that is, $p^{k-1}(p-1)$ divides h. Also, h divides $|U(p^{k+1})|$, that is, h divides $p^k(p-1)$, because h is the order of an element of $U(p^{k+1})$. Thus $h = p^k(p-1)$ or $h = p^{k-1}(p-1)$; we need to show that the latter is not possible. It suffices to show that $g^{p^{k-1}(p-1)} \not\equiv 1 \pmod{p^{k+1}}$. We know that $g^{p^{k-2}(p-1)} \not\equiv 1 \pmod{p^k}$ and $g^{p^{k-2}(p-1)} \equiv 1 \pmod{p^{k-1}}$, by the induction hypothesis (or by direct verification if $k=2$). Thus $g^{p^{k-2}(p-1)} = 1 + bp^{k-1}$ for some b not divisible by p. Then

$$\begin{aligned} g^{p^{k-1}(p-1)} &= \left(1 + bp^{k-1}\right)^p \\ &= 1 + \binom{p}{1} bp^{k-1} + \binom{p}{2} b^2 p^{2k-2} + \cdots \\ &\quad + \binom{p}{p-1} b^{p-1} p^{(p-1)(k-1)} + b^p p^{pk-p}. \end{aligned}$$

Since $pk - p \geq k+1$, p^{k+1} divides the last term in this sum. The binomial coefficient $\binom{p}{i}$ is divisible by p when $1 \leq i \leq p-1$, because p is prime and

$$\binom{p}{i} = \frac{p!}{i!(p-i)!},$$

with every factor in the denominator $i!(p-i)!$ less than p. Together with the fact that $2k - 2 \geq k$, this means that p^{k+1} divides every term in the preceding sum except the first two, so

$$g^{p^{k-1}(p-1)} \equiv 1 + bp^k \pmod{p^{k+1}}.$$

Since p does not divide b,

$$g^{p^{k-1}(p-1)} \not\equiv 1 \pmod{p^{k+1}},$$

which is what we were after—now we know that g generates $U(p^{k+1})$.

Completing the cycle Finally, it all comes down to $U(p)$. We need to know that some $g \in U(p)$ has order $m = p-1$. Pick g to have order m in $U(p)$, with m as large as possible. If h is any element of $U(p)$, then $|h|$ divides $|g|$, for suppose not. Then we may write $|h| = q^r a$ and $|g| = q^s b$, where q is prime, $r > s$, and q does not divide either a or b. That is, if $|h|$ does not divide $|g|$, it must be because some prime q appears more often in the factorization of $|h|$ than in the factorization of $|g|$. Now in $U(p)$,

$$(h^a g)^{q^r b} = \left(h^{q^r a}\right)^b \left(g^{q^s b}\right)^{q^{r-s}} = 1,$$

so $|h^a g|$ divides $q^r b$. Thus the order of $h^a g$ must be $q^t c$, where $t \leq r$ and $c | b$. If $t < r$, then

$$1 = (h^a g)^{q^{r-1} b} = \left(h^{q^{r-1} a}\right)^b \left(g^{q^{r-1} b}\right) = \left(h^{q^{r-1} a}\right)^b,$$

so $|h|$ divides $q^{r-1} ab$, a contradiction. On the other hand, if $c < b$, then

$$1 = (h^a g)^{q^r c} = \left(h^{q^r a}\right)^c g^{q^r c} = g^{q^r c},$$

so $|g|$ divides $q^r c$, another contradiction. Thus $t = r$ and $c = b$, and the order of $h^a g$ is $q^r b > q^s b = |g|$, yet another contradiction, since g was chosen to have largest possible order. Hence, $|h|$ divides $|g| = m$.

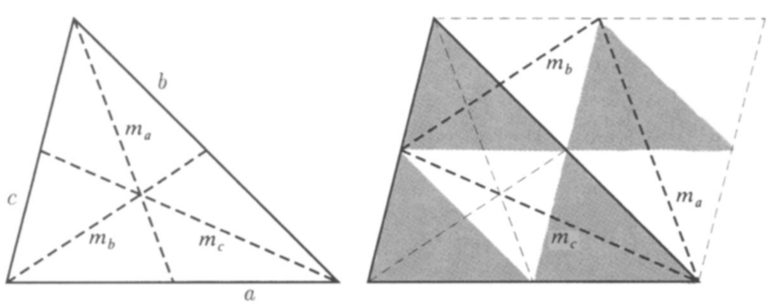
Now we need a bit of field theory and we're done. For every $h \in U(p)$, $h^m = 1$, that is, h is a root of the polynomial $x^m - 1$, so $x^m - 1$ has $p - 1$ roots in $U(p)$ and in $\mathbb{Z}_p$. But since $\mathbb{Z}_p$ is a field, $x^m - 1$ can have at most m roots. Thus $p - 1 \leq m$, so in fact the order of g is $p - 1$ and $U(p)$ is cyclic.

Acknowledgment Many thanks to the referee for suggesting the proof used here that $U(p)$ is cyclic.

REFERENCES

1. David J. DeVries, The group of units in $\mathbb{Z}_m$, this MAGAZINE 62 (1989), 340.
2. Joseph A. Gallian, *Contemporary Abstract Algebra*, Fourth Edition, Houghton Mifflin, Boston, MA, 1998.
3. Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery, *An Introduction to the Theory of Numbers*, Fifth Edition, John Wiley & Sons, Inc., New York, NY, 1991.

Proof Without Words: The Triangle of Medians Has Three-Fourths the Area of the Original Triangle



$$\text{area}(\Delta m_a m_b m_c) = \frac{3}{4}\text{area}(\Delta abc)$$

—NORBERT HUNGERBÜHLER
MAX-PLANCK INSTITUTE FOR MATHEMATICS IN THE SCIENCES
INSELSTRASSE 22–26
04103 LEIPZIG
GERMANY

so $|g|$ divides $q^r c$, another contradiction. Thus $t = r$ and $c = b$, and the order of $h^a g$ is $q^r b > q^s b = |g|$, yet another contradiction, since g was chosen to have largest possible order. Hence, $|h|$ divides $|g| = m$.

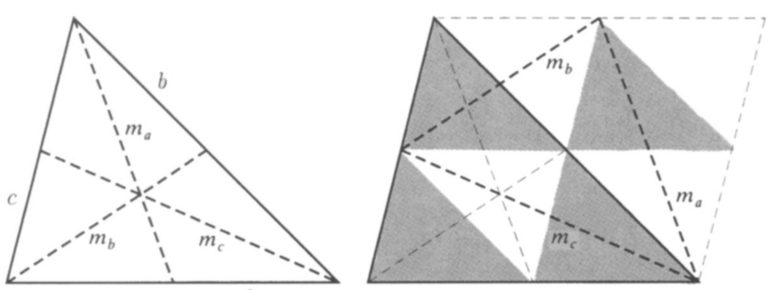
Now we need a bit of field theory and we're done. For every $h \in U(p)$, $h^m = 1$, that is, h is a root of the polynomial $x^m - 1$, so $x^m - 1$ has $p - 1$ roots in $U(p)$ and in $\mathbb{Z}_p$. But since $\mathbb{Z}_p$ is a field, $x^m - 1$ can have at most m roots. Thus $p - 1 \leq m$, so in fact the order of g is $p - 1$ and $U(p)$ is cyclic.

Acknowledgment Many thanks to the referee for suggesting the proof used here that $U(p)$ is cyclic.

REFERENCES

1. David J. DeVries, The group of units in $\mathbb{Z}_m$, this MAGAZINE 62 (1989), 340.
2. Joseph A. Gallian, *Contemporary Abstract Algebra*, Fourth Edition, Houghton Mifflin, Boston, MA, 1998.
3. Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery, *An Introduction to the Theory of Numbers*, Fifth Edition, John Wiley & Sons, Inc., New York, NY, 1991.

Proof Without Words: The Triangle of Medians Has Three-Fourths the Area of the Original Triangle



$$\operatorname{area}(\Delta m_a m_b m_c) = \frac{3}{4}\operatorname{area}(\Delta abc)$$

—NORBERT HUNGERBÜHLER
MAX-PLANCK INSTITUTE FOR MATHEMATICS IN THE SCIENCES
INSELSTRASSE 22–26
04103 LEIPZIG
GERMANY

Mutual Multiples in $\mathbb{Z}_n$

ZUN SHAN
Nanjing Normal University
Nanjing, Jiangsu
People's Republic of China

EDWARD T.H. WANG
Wilfrid Laurier University
Waterloo, Ontario N2L 3C5
Canada

The following problem appeared in [**1**]:

PROBLEM: *Let R be a commutative ring with unit element 1. Prove or disprove: If $a, b \in R$ are multiples of one another, then they are* unit *multiples of one another; that is, there is an invertible element $u \in R$ such that $a = ub$.*

The given statement is false for a general commutative ring R (see [**3**]). We show here, however, that it is true for $\mathbb{Z}_n$.

In what follows we will use ϕ to denote the Euler phi-function; $\tau(n)$ will denote the number of positive divisors of n. The following definition will be convenient:

DEFINITION. Let $\mathbb{Z}_n$ be the commutative ring of integers modulo n, where $n > 1$ is a given natural number. Two elements a and b of $\mathbb{Z}_n$, not necessarily distinct, are said to form an MM pair (mutual multiple pair) if there exist $i, j \in \mathbb{Z}_n$ such that $a = ib$ and $b = ja$ in $\mathbb{Z}_n$; that is, $a \equiv ib \pmod{n}$ and $b \equiv ja \pmod{n}$. In this case, i and j are called *multipliers*.

Note that 0 cannot form an MM pair with any element of $\mathbb{Z}_n$ but itself. Also, if $n = p$ is a prime, then clearly any two non-zero elements of Z_p form an MM pair, and the multipliers i and j are both unique.

Following is a more meaty example.

Example: In $\mathbb{Z}_6$, the numbers 2 and 4 form an MM pair since $4 = 2 \times 2$ and $2 = 5 \times 4$. Similarly, 1 and 5 form an MM pair since $5 = 5 \times 1$ and $1 = 5 \times 5$. On the other hand, 3 and 4 do not form an MM pair since $3j \equiv 0$ or $3 \pmod{6}$ depending on whether j is even or odd.

Observe that when a and b form an MM pair, the multipliers i and j need not be unique in general even if $a \neq 0$ and $b \neq 0$; e.g., in $\mathbb{Z}_6$ we could also write $4 = 5 \times 2$ and/or $2 = 2 \times 4$.

LEMMA 1. *Let $a, b \in \mathbb{Z}_n$. Then a and b form an MM pair if and only if* $\gcd(a, n) = \gcd(b, n)$.

Proof. Suppose a and b form an MM pair. Then there exist $i, j \in \mathbb{Z}_n$ such that $a = ib$ and $b = ja$. Since $\gcd(a, n) | a$ implies $\gcd(a, n) | ja$, we have $\gcd(a, n) | b$ and so $\gcd(a, n) | \gcd(b, n)$. Similarly, $\gcd(b, n) | \gcd(a, n)$ and thus $\gcd(a, n) = \gcd(b, n)$.

Conversely, suppose $\gcd(a, n) = \gcd(b, n) = d$. Let $a = da'$ and $b = db'$. Then either $a = b = 0$ or $\gcd(a', n) = \gcd(b', n) = 1$. Since $\gcd(a', n/d) = \gcd(b', n/d) = 1$ there exist $i, j \in \mathbb{Z}_n$ such that $a' \equiv b'i$ and $b' \equiv a'j \pmod{n/d}$. Hence $a \equiv ib$ and $b \equiv ja \pmod{n}$. This completes the proof.

THEOREM 1. *Suppose $a, b \in \mathbb{Z}_n$ form an MM pair. Then there exists an invertible element $u \in \mathbb{Z}_n$ such that $a = ub$.*

Proof. As in the proof of the lemma, let $\gcd(a, n) = \gcd(b, n) = d$, $a = da'$, and $b = db'$.

Then there exists $i \in \mathbb{Z}_n$ such that $a' \equiv b'i \pmod{n/d}$. Clearly $\gcd(i, n/d) = 1$ as $\gcd(a', n/d) = 1$. By the celebrated theorem of Dirichlet, there are infinitely many primes in the sequence $\{i + k(n/d)\}_{k=0}^{\infty}$, and hence, *a fortiori*, there are primes in this sequence that exceed n. Thus there exists $k_0 \in \mathbb{N}$ for which $i + k_0(n/d)$ is such a prime, and so $\gcd(i + k_0(n/d), n) = 1$. If we let u denote the least positive residue of $i + k_0(n/d)$ modulo n, then $u \in \mathbb{Z}_n$ is such that

$$\gcd(u, n) = 1 \quad \text{and} \quad ub' \equiv ib' \equiv a' \pmod{n/d};$$

it follows that $a \equiv ub \pmod{n}$, which completes the proof.

Remark 1. The key to the preceding proof is the existence of an integer in the sequence $\{i + k(n/d)\}_{k=0}^{\infty}$ that is coprime with n. This result, which is a consequence of Dirichlet's theorem, appeared in [**4**, p. 12, Ex. 3] with an elementary proof.

As we explored MM pairs in $\mathbb{Z}_n$, we were led to wonder how many there are. We found the following answer:

THEOREM 2. *Let $f(n)$ denote the number of unordered MM pairs in $\mathbb{Z}_n$. Then $f(n) = \frac{1}{2}[n + \sum_{d|n} \phi(d)^2]$; the summation is over all positive divisors d of n.*

Proof. For each divisor d of n and for any $a \in \mathbb{Z}_n$, note that $\gcd(a, n) = d$ if and only if $a = da'$ for some $a' \in \mathbb{Z}_{n/d}$ such that $\gcd(a', n/d) = 1$. Hence if we let $\mathbb{Z}_{n/d}^* = \{m \in \mathbb{Z}_{n/d} | \gcd(m, n/d) = 1\}$, then, by Lemma 1, any two elements of $\mathbb{Z}_{n/d}^*$ would form an MM pair and no elements of $\mathbb{Z}_{n/d}^*$ can form an MM pair with elements not in the set. Since $|\mathbb{Z}_{n/d}^*| = \phi(n/d)$, we have

$$f(n) = \sum_{d|n} \left[\phi(n/d) + \binom{\phi(n/d)}{2} \right] = \sum_{d|n} \left[\phi(d) + \binom{\phi(d)}{2} \right]$$

$$= \frac{1}{2} \sum_{d|n} \left[\phi(d) + \phi(d)^2 \right] = \frac{1}{2} \left[n + \sum_{d|n} \phi(d)^2 \right],$$

where the last equality holds because $\sum_{d|n} \phi(d) = n$ (see, e.g., [**2**, Thm. 6.7, p. 212]).

Remark 2. Since there is no known closed form expression for $\sum_{d|n} \phi(d)^2$, the only way to find the exact value of $f(n)$ is to compute $\phi(d)$ for all divisors d of n. A corollary, however, gives a lower bound for $f(n)$.

COROLLARY: $f(n) \geq \frac{n}{2}\left(\frac{n}{\tau(n)} + 1\right)$.

Proof. By the Cauchy–Schwarz inequality,

$$\sum_{d|n} \phi(d)^2 \sum_{d|n} 1^2 \geq \left(\sum_{d|n} \phi(d) \right)^2 = n^2,$$

so $\sum_{d|n} \phi(d)^2 \geq n^2/\tau(n)$. Substituting this into the formula from Theorem 2 completes the proof.

Acknowledgment This paper was written when the first author was visiting the Department of Mathematics at Wilfrid Laurier University, December 1996–August 1997. The hospitality of WLU is greatly appreciated. The authors would like to thank the referee for many constructive suggestions, which substantially improved this paper.

REFERENCES

1. Benkart, G. M., et. al., Problem #600, *College Math. Journal* 28:2 (1997), 146.
2. Rosen, Kenneth H., *Elementary Number Theory and Its Applications*, 3rd ed., Addison Wesley, Reading, MA, 1993.
3. Shan, Zun and Wang, Edward T. H., Solution to Problem #600 (Unit Multiples), *College Math. Journal*, 29:2 (1998), 156–157.
4. Sierpinski, W., *Elementary Theory of Numbers*, North-Holland, Amsterdam, The Netherlands, 1988.

Unevening the Odds of "Even Up"

ARTHUR T. BENJAMIN
Harvey Mudd College
Claremont, CA 91711

JENNIFER J. QUINN
Occidental College
Los Angeles, CA 90041

The Disclaimer The authors take no responsibility for any gambling hustles or scams based on applications of the principles contained in this note.

The Game "Even Up" is a game of solitaire played with 40 cards from a standard deck that has its jacks, queens, and kings removed. The cards are shuffled and dealt in a row. If a consecutive pair of cards adds to an even number, then that pair can be removed. The object of the game is to remove all of the cards.

More generally, we can play Even Up with $2n$ cards, x of them being odd and $2n-x$ being even. We require the number of cards to be even since the game cannot be won with an odd number of cards. In fact, the game cannot be won when x is odd since odd valued cards are removed in pairs. Harkleroad **[1]** showed that the game involves no skill, in that the outcome is predetermined by the original order of the $2n$ cards, and that the probability of winning is $p(2n,x)=\binom{n}{x/2}^2\Big/\binom{2n}{x}$. Thus the probability of winning the original game is $p(40,20)=0.248$.

A few remarks about $p(2n,x)$ are called for. Clearly $p(2n,0)=1=p(2n,2n)$. By comparing $p(2n,x)$ with $p(2n,x-2)$, one sees that for fixed n the probability of winning is minimized when $x=n$. When n is large, we can use Stirling's formula ($n!\approx(n/e)^n\cdot\sqrt{2\pi n}$) to obtain $p(2n,n)\approx 2/\sqrt{\pi n}$.

For our purposes, any arrangement of $2n$ cards can be represented as the product of a's and b's with a's denoting odd cards and b's denoting even cards. The rules of Even Up reduce to the two multiplications $a^2=1$ and $b^2=1$. Every game simplifies to exactly one string of the form $(ab)^z$, where $-n\le z\le n$ and $(ab)^{-z}=(ba)^z$. Winning games occur when $z=0$. Letting $f(2n,x,z)$ denote the number of arrange-

Acknowledgment This paper was written when the first author was visiting the Department of Mathematics at Wilfrid Laurier University, December 1996–August 1997. The hospitality of WLU is greatly appreciated. The authors would like to thank the referee for many constructive suggestions, which substantially improved this paper.

REFERENCES

1. Benkart, G. M., et. al., Problem #600, *College Math. Journal* 28:2 (1997), 146.
2. Rosen, Kenneth H., *Elementary Number Theory and Its Applications*, 3rd ed., Addison Wesley, Reading, MA, 1993.
3. Shan, Zun and Wang, Edward T. H., Solution to Problem #600 (Unit Multiples), *College Math. Journal*, 29:2 (1998), 156–157.
4. Sierpinski, W., *Elementary Theory of Numbers*, North-Holland, Amsterdam, The Netherlands, 1988.

Unevening the Odds of "Even Up"

ARTHUR T. BENJAMIN
Harvey Mudd College
Claremont, CA 91711

JENNIFER J. QUINN
Occidental College
Los Angeles, CA 90041

The Disclaimer The authors take no responsibility for any gambling hustles or scams based on applications of the principles contained in this note.

The Game "Even Up" is a game of solitaire played with 40 cards from a standard deck that has its jacks, queens, and kings removed. The cards are shuffled and dealt in a row. If a consecutive pair of cards adds to an even number, then that pair can be removed. The object of the game is to remove all of the cards.

More generally, we can play Even Up with $2n$ cards, x of them being odd and $2n-x$ being even. We require the number of cards to be even since the game cannot be won with an odd number of cards. In fact, the game cannot be won when x is odd since odd valued cards are removed in pairs. Harkleroad [1] showed that the game involves no skill, in that the outcome is predetermined by the original order of the $2n$ cards, and that the probability of winning is $p(2n, x) = \binom{n}{x/2}^2 / \binom{2n}{x}$. Thus the probability of winning the original game is $p(40, 20) = 0.248$.

A few remarks about $p(2n, x)$ are called for. Clearly $p(2n, 0) = 1 = p(2n, 2n)$. By comparing $p(2n, x)$ with $p(2n, x-2)$, one sees that for fixed n the probability of winning is minimized when $x = n$. When n is large, we can use Stirling's formula ($n! \approx (n/e)^n \cdot \sqrt{2\pi n}$) to obtain $p(2n, n) \approx 2/\sqrt{\pi n}$.

For our purposes, any arrangement of $2n$ cards can be represented as the product of a's and b's with a's denoting odd cards and b's denoting even cards. The rules of Even Up reduce to the two multiplications $a^2 = 1$ and $b^2 = 1$. Every game simplifies to exactly one string of the form $(ab)^z$, where $-n \le z \le n$ and $(ab)^{-z} = (ba)^z$. Winning games occur when $z = 0$. Letting $f(2n, x, z)$ denote the number of arrangements of x a's and $2n-x$ b's that reduce to $(ab)^z$, Harkleroad gives two proofs that

$$f(2n, x, z) = \binom{n}{(x+z)/2}\binom{n}{(x-z)/2}. \tag{1}$$

One proof used induction and the other used a complicated summation. Here we provide a direct combinatorial explanation of equation (1).

The Proof Consider a string of $2n$ symbols as n ordered pairs, where each pair is either (a,a), (b,b), (a,b), or (b,a). Such a string will reduce to $(a,b)^z$, where z is the number of (a,b) pairs minus the number of (b,a) pairs. If we let the four pairs have scores of $0,0,1$, and -1 respectively, then z denotes the total score. Another way to calculate the score is to assign a value to each symbol as follows: a beginning a in an ordered pair gets a score of 1, an ending a in an ordered pair gets a score of -1, and all b's get a score of 0. Hence the total score is equal to the number of beginning a's minus the number of ending a's. If x' denotes the number of beginning a's $(0 \le x' \le x)$, then the score $z = x' - (x - x') = 2x' - x$. Thus to achieve a score of z, we must choose $x' = (x+z)/2$ of our n pairs to begin with an a and $x - x' = (x-z)/2$ of our n pairs to end with an a. This completes the proof of equation (1).

The probability that a game with x a's and $2n-x$ b's reduces to $(ab)^z$ is $f(2n, x, z)/\binom{2n}{x}$. Notice that $z = 2x' - x$ will always have the same parity as x. (This is reflected in (1) since $\binom{n}{k/2}$ is 0 when k is odd.) Thus, we see again that a game with an *even* number of cards but *odd* number of odd cards is impossible to win.

The Scam After explaining the game of Even Up to your mark, challenge him to a duel. Wager that you can win the game in strictly fewer attempts than he can. He can shuffle your cards before every deal. You play the game until you win. Say it takes you 4 attempts. Now it's his turn to play. How can you be sure that it will take him at least 5 attempts? When *you* play the game, use 10 even cards and 10 odd cards. Each attempt has a 34% chance of success. After you win, shuffle the cards, but secretly add (or remove, if you prefer) one even and one odd card to his deck. This will make his winning probability 0.

REFERENCE

1. Leon Harkleroad, Lowering the odds for "Even up", this MAGAZINE 63 (1990), 160–163.

Math Bite: Constructing Efficient Particle Accelerators Is as Easy as 1 + 1 = 2 (Thanks to Vladimir Visnjic)

The difficulty in "applied math" is usually not the "math" part (it is often trivial), but the "applied" part, i.e., *realizing* that math can help, and separating the mathematical wheat from the scientific chaff.

A dramatic illustration of the above is Vladimir Visnjic's ingenious construction of two local waves on the particle beam in an accelerator that promises to save the American taxpayer millions of dollars (but, rather unfairly, will not earn its discoverer even a penny).

Sources at positions $x = x_1, x_2, \ldots, x_n$ with amplitudes $A_1, A_2, \ldots, A_n$, produce waves of two types, $A_i \sin(x - x_i)$ and $f(x_i) A_i \sin((x - x_i)/2)$. Here $f(x)$ is a certain function (specific to the accelerator) of which we only need to know that it is nonnegative and has zeros. The number of sources, their positions, and amplitudes can be chosen freely up to the constraint $\sum A_i = 0$. The objective is to find the minimal set of sources which creates local waves of both types at the same position.

Visnjic's construction hinges on the following immediate corollary of the celebrated identities $1 + 1 - 2 = 0$ and $1 - 1 + 0 = 0$.

PROPOSITION. *Let $H(x)$ be the Heaviside function (i.e., $H(x) = 1$ if $x > 0$, $H(x) = 0$ otherwise). Then*

$$H(x) \sin(x) + H(x - 2\pi) \sin(x - 2\pi) - 2H(x - 4\pi) \sin(x - 4\pi)$$

vanishes outside $0 < x < 4\pi$. If $f(0) = f(2\pi)$ and $f(4\pi) = 0$, then

$$f(0) H(x) \sin\left(\frac{x}{2}\right) + f(2\pi) H(x - 2\pi) \sin\left(\frac{x}{2} - \pi\right) - 2f(4\pi) H(x - 4\pi) \sin\left(\frac{x}{2} - 2\pi\right)$$

vanishes outside $0 < x < 2\pi$.

REFERENCE

1. Vladimir Visnjic, A novel focusing element for accelerators, *Phys. Rev. Letters* 21 (1994), 2860.

—DORON ZEILBERGER
TEMPLE UNIVERSITY
PHILADELPHIA, PA 19122

PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

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

Proposals

To be considered for publication, solutions should be received by September 1, 1999.

1569. *Proposed by Ismor Fischer, Department of Biostatistics, University of Wisconsin, Madison, Wisconsin.*

Given three concentric circles in the plane, prove that (up to rotation and reflection) there exists a unique triangle of maximum area having exactly one vertex on each circle, respectively.

1570. *Proposed by Ice B. Risteski, Skopje, Macedonia.*

Solve the differential equation

$$\left(\frac{dy}{dx}\right)^{n+1} + axy^{2n}\frac{dy}{dx} + ay^{2n+1} = 0, \qquad a \neq 0, n \in \mathbb{N}.$$

1571. *Proposed by Michael H. Brill, Sarnoff Corporation, Princeton, New Jersey.*

Let the "real world" be those convex three-dimensional solids whose surfaces are smooth. Let the "digital world" be a three-dimensional tiling of tiny identical cubes, which we call "voxels," analogous to a two-dimensional digital image of square pixels. Each "digital-world" object X' is a maximal subset of these voxels that lies inside the corresponding "real-world" object X.

What is the maximal ratio of the amount of paint needed to cover X' to the amount needed to cover X taken over all convex X whose boundary is a smooth surface and all possible X' as the orientation and size of the voxels vary? In other words, the problem is to find the supremum of the ratio of the exposed surface area of X' to that of X.

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

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

Solutions and new proposals should be mailed to George T. Gilbert, Problems Editor, Department of Mathematics, Box 298900, Texas Christian University, Fort Worth, TX 76129, or mailed electronically (ideally as a LaTeX file) to `g.gilbert@tcu.edu`. *Readers who use e-mail should also provide an e-mail address.*

1572. *Proposed by Western Maryland College Problems Group, Westminster, Maryland.*

Let $b_0 = 1$ and b_1 satisfy $0 < b_1 < 1$. For $n \geq 1$, define b_{n+1} by

$$b_{n+1} = \frac{2b_n b_{n-1} - b_n^2}{3b_{n-1} - 2b_n}.$$

Show that $(b_n)_{\{n \geq 0\}}$ converges, and compute its limit in terms of b_1.

1573. *Proposed by Jiro Fukuta, Professor Emeritus, Gifu University, Gifu-ken, Japan.*

Given $\triangle ABC$, let AD be a cevian to the side BC, and let E be on segment AD. The circumcircle of $\triangle ACD$ intersects the line BE at points M and N, and the circumcircle of $\triangle ABD$ intersects the line CE at points P and Q. Prove that the points M, N, P, and Q lie on a common circle and its center is on the line perpendicular to the side BC at the point D.

Quickies

Answers to the Quickies are on page 155

Q889. *Proposed by Michael McGeachie (student) and Stan Wagon, Macalester College, St. Paul, Minnesota.*

Characterize the set of positive integers n such that, for all integers a, the sequence $a, a^2, a^3, \ldots$ is periodic modulo n.

Q890. *Proposed by Ira Rosenholtz, Eastern Illinois University, Charleston, Illinois.*

Heron's formula expresses the area of a triangle in terms of the lengths of its sides.

(a) Is there an analogous formula for the volume of a tetrahedron in terms of the areas of its faces?

(b) Is there an analogous formula for the volume of a tetrahedron in terms of the lengths of its edges that is symmetric in the edges?

Remark. The Cayley–Menger determinant formula gives the volume of a tetrahedron with vertices A, B, C, and D in terms of the lengths of its six edges that is not symmetric in those lengths:

$$V = \frac{1}{2^{3/2}3!}\left|\det\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & AB^2 & AC^2 & AD^2 \\ 1 & AB^2 & 0 & BC^2 & BD^2 \\ 1 & AC^2 & BC^2 & 0 & CD^2 \\ 1 & AD^2 & BD^2 & CD^2 & 0 \end{pmatrix}\right|^{1/2}.$$

Solutions

First 1 in a Recursion — April 1998

1544. *Proposed by 1996 MathCamp Students, University of Washington.*

Let a_0 be a positive integer and let

$$a_{k+1} = \begin{cases} a_k + 1 & \text{if } a_k \text{ is odd,} \\ a_k/2 & \text{if } a_k \text{ is even.} \end{cases}$$

Find a nonrecursive expression in terms of a_0 for the smallest positive integer k such that $a_k = 1$.

Solution by Seth Zimmerman, Evergreen Valley College, San Jose, California.

If $a_0 = 2^n$, then the iteration takes n steps to reach 1. Otherwise, expand a_0 into its binary representation b_0. Then the number of steps is one more than the sum of the number of digits of b_0 and the number of zeros to the left of the rightmost one in b_0.

To see this, assume that a_0 is not a power of 2, that case being obvious. Division of an even a_0 by 2 corresponds to erasing the rightmost zero of b_0. Adding one to an odd a_0 corresponds to changing the rightmost string of ones in b_0 to zeros and changing the zero to the left of this string to a one. Because the leading one will eventually be converted to a 10, the number of steps of the first type equals the number of digits in b_0 and the number of the second type is one more than the number of zeros to the left of the rightmost 1 in b_0.

Comments. Seth Zimmerman also notes that the number of different integers requiring n steps is the nth Fibonacci number. Lenny Jones and Jedd Beall remark that generalizing the result from denominator 2 to denominator d, the two cases being dividing by d if possible or going to the next higher multiple of d if not, requires a total number of iterations equal to $\lceil \log_d a_0 \rceil$ plus the number of nonzero digits in the base d representation of $d^{\lceil \log_d a_0 \rceil} - a_0$. When $d = 2$ it gives an alternate way of expressing the above.

Also solved by Rajesh K. Barnwal, Rich Bauer, J. C. Binz (Switzerland), Mansur Boase (student, England), Jeffrey Clark, Charles R. Diminnie, Daniele Donini (Italy), Robert L. Doucette, Tim Flood, Marty Getz and Dixon Jones, Jerrold W. Grossman, Danrung Huang, Joel Iiams, Lenny Jones and Jedd Beall, Harris Kwong, Kee-Wai Lau (China), Kathleen E. Lewis, Gao Peng (graduate student), Philip D. Straffin, Stephen Swiniarski, Gillian Valk, Peter Vanden Bosch, Western Maryland College Problems Group, Bilal Yurdakul (student, Turkey), and the proposers.

A Divisibility Condition on Consecutive Terms of a Sequence April 1998

1545. *Proposed by Erwin Just, Professor Emeritus, Bronx Community College, Bronx, New York.*

Let k be a positive integer. Prove that there exists an infinite, monotone increasing sequence of integers (a_n) such that a_n divides $a_{n+1}^2 + k$ and a_{n+1} divides $a_n^2 + k$ for all $n \geq 1$.

I. *Solution by John Christopher, California State University, Sacramento, California.*

For k a positive integer, set $a_0 = 1$ and construct a sequence as follows:

$$a_1 = 1 \quad \text{and} \quad a_{n+1} = \frac{a_n^2 + k}{a_{n-1}} \quad \text{for } n \geq 1.$$

The sequence $(a_n)_{n \geq 1}$ is clearly monotonic. Since $a_{n-1}a_{n+1} = a_n^2 + k$, it suffices to show that a_n is a positive integer for all n. We use induction to prove that a_n is an integer and that $(a_n, ka_{n-1}) = 1$. Calculation verifies the claims for $n = 1$, $n = 2$, and $n = 3$. Assume the claims for $1 \leq j \leq n$, with $n \geq 3$. Write a_{n+1} in lowest terms as

p/q. The definition of a_{n+1} shows that q divides a_{n-1}. Furthermore,

$$\frac{a_n^2+k}{a_{n-1}} = \frac{\left(\dfrac{a_{n-1}^2+k}{a_{n-2}}\right)^2+k}{a_{n-1}}$$

$$= \frac{a_{n-1}^3+2ka_{n-1}+k\dfrac{k+a_{n-2}^2}{a_{n-1}}}{a_{n-2}^2}$$

$$= \frac{a_{n-1}^3+2ka_{n-1}+ka_{n-3}}{a_{n-2}^2}.$$

Thus, q also divides a_{n-2}^2. Since $(a_{n-1}, a_{n-2}) = 1$, it follows that $q = 1$ so that a_{n+1} is an integer. Moreover, since (a_{n+1}, a_n) is a factor of k and $(a_n, k) = 1$, we have $(a_{n+1}, ka_n) = 1$, completing the proof.

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

We show by induction that the monotone increasing sequence (which ends up the same as that above) defined by

$$a_0 = a_1 = 1, \quad a_n = (k+2)a_{n-1} - a_{n-2} \quad \text{for } n \geq 2$$

satisfies $a_{n-1}^2 + k = a_{n-2}a_n$ for $n \geq 2$, and therefore possesses the desired properties.

The initial case, $n = 2$, is easily verified. Now suppose that $a_{n-1}^2 + k = a_{n-2}a_n$ for all $n \leq m$. We have

$$a_m^2 + k = \left[(k+2)a_{m-1} - a_{m-2}\right]a_m + k$$

$$= (k+2)a_{m-1}\frac{a_{m+1}+a_{m-1}}{k+2} - a_{m-2}a_m + k$$

$$= a_{m-1}a_{m+1} + a_{m-1}^2 - a_m a_{m-2} + k = a_{m-1}a_{m+1},$$

completing the induction.

Also solved by Brian D. Beasley, Robert E. Bernstein, J. C. Binz (Switzerland), Mansur Boase (student, England), Sabin Cautis (Canada), Jeffrey Clark, Thomas Dence and Gordon Swain, Charles R. Diminnie, Daniele Donini (Italy), Gerald A. Heuer, Lenny Jones and Jedd Beall, Kathleen E. Lewis, Norman F. Lindquist, Gao Peng (graduate student), Stephen Swiniarski, and the proposer.

An Extremal Polynomial — April 1998

1546. *Proposed by Benjamin G. Klein, Davidson College, Davidson, North Carolina, and Arthur L. Holshouser, Charlotte, North Carolina.*

Given $y > 1$, let P be the set of all real polynomials $p(x)$ with nonnegative coefficients that satisfy $p(1) = 1$ and $p(3) = y$. Prove there exists $p_0(x) \in P$ such that

(i) $\{p(2): p(x) \in P\} = (1, p_0(2)]$;
(ii) if $p(x) \in P$ and $p(2) = p_0(2)$, then $p(x) = p_0(x)$.

Solution by Roger Zarnowski and Charles Diminnie, Angelo State University, San Angelo, Texas.

Let $S := \{p(2): p(x) \in P\}$. If k is the unique positive integer for which $3^{k-1} < y \leq 3^k$, define

$$p_0(x) := \frac{y - 3^{k-1}}{3^k - 3^{k-1}} x^k + \frac{3^k - y}{3^k - 3^{k-1}} x^{k-1}.$$

It is easily verified that $p_0(x) \in P$.

We first prove assertion (ii) and show that $p_0(2) = \max S$. Consider $p(x) \in P$ and set $q(x) := p(x) - p_0(x)$. Since $q(x)$ has at most two negative terms and these involve consecutive powers of x, its coefficients can have at most two changes of sign (when $q(x)$ is written in descending powers of x). By Descartes' Rule of Signs, either $q(x) \equiv 0$ or $q(x)$ can have at most two positive roots. If $q(x) \not\equiv 0$, then it follows from $q(1) = q(3) = 0$ that $q(x)$ has exactly two positive roots, which are simple. In particular $p(2) = p_0(2)$ implies $q(2) = 0$, so that $q(x) \equiv 0$, proving (ii). Also, the leading coefficient of $q(x)$ must be positive in order to have two sign changes. If $p(2) > p_0(2)$, then $q(2) > 0$ and hence $q(x) > 0$ for all $x \in (1,3)$. However, since the leading coefficient of $q(x)$ is positive, $\lim_{x \to \infty} q(x) = \infty$, a contradiction. Therefore, $p(2) \leq p_0(2)$ for all $p(x) \in P$, and $p_0(2) = \max S$.

We next show that $\inf S = 1$ and that $1 \notin S$. If $p(x) \in P$, then, since all coefficients of $p(x)$ are nonnegative and $\deg p(x) > 0$, we have $p'(x) > 0$ for all $x \in [1,3]$. This implies that $p(2) > p(1) = 1$ for all $p(x) \in P$, hence 1 is a lower bound for S and $1 \notin S$. For each positive integer n such that $3^n \geq y$, define

$$p_n(x) = \frac{y-1}{3^n - 1} x^n + \frac{3^n - y}{3^n - 1}.$$

It is easily shown that $p_n(x) \in P$ for all such n. Furthermore, $\lim_{n \to \infty} p_n(2) = 1$. Therefore, $1 = \inf S$.

We complete the proof of (i) by proving that $S = (1, p_0(2)]$. Note that for $\alpha \in [0, 1]$,

$$p(x) := \alpha p_n(x) + (1 - \alpha) p_0(x) \in P.$$

If $1 < t \leq p_0(2)$, then $p(2) = t$ for n sufficiently large and suitable choice of α. Therefore, $S = (1, p_0(2)]$.

Also solved by Tewodros Amdeberhan, Daniele Donini (Italy), W. R. Smythe, Western Maryland College Problems Group, and the proposers. There were two incorrect solutions.

Non-periodicity of an Iterative Sequence Mod N **April 1998**

1547. *Homer White, Georgetown College, Georgetown, Kentucky, and Robert Bailey, Lexington, Kentucky.*

Consider the function

$$f(n) = \begin{cases} \dfrac{3n}{2} & \text{if } n \text{ is even,} \\[2ex] \dfrac{3n+1}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Let $N > 1$ be a positive integer, and define $a_n(i)$ to be the remainder when the ith iterate of f, $f^{(i)}(n)$, is divided by N. Prove that, for any $n \geq 1$, the sequence $(a_n(i))_{i \geq 0}$ is not periodic.

Solution by Gao Peng, physics graduate student, University of Oklahoma, Norman, Oklahoma.

We first prove that it is impossible for $(a_n(i))_{i\geq 0}$ to be periodic when $N=2$. Suppose the sequence has period t. The parity of $f^{(i)}(n)$, represented by $a_n(i)$, determines which case of the definition of f applies, a pattern that repeats with period t. Thus there exists a nonnegative integer c, such that for every positive integer k,

$$f^{(kt)}(n)=\left(\frac{3}{2}\right)^t f^{((k-1)t)}(n)+\frac{c}{2^t}.$$

An easy induction implies that

$$f^{(kt)}(n)=\left(\frac{3}{2}\right)^{kt} f^{(0)}(n)+\frac{\left(\frac{3}{2}\right)^{kt}c-c}{3^t-2^t}=\left(\frac{3}{2}\right)^{kt}\left(n+\frac{c}{3^t-2^t}\right)-\frac{c}{3^t-2^t}.$$

However, $f^{(kt)}(n)$ cannot remain an integer as k increases to ∞, a contradiction.

Now consider general N. Suppose $(a_n(i))_{i\geq 0}$ has period t. Then $f^{(i)}(n)\equiv f^{(i+t)}(n) \pmod N$, which implies

$$3f^{(i)}(n)\equiv 3f^{(i+t)}(n) \pmod N. \tag{1}$$

Write $f^{(i+1)}(n)=(3f^{(i)}(n)+\epsilon_i)/2$, where $\epsilon_i=0$ or 1. From $f^{(i+1)}(n)\equiv f^{(i+1+t)}(n) \pmod N$, we obtain

$$3f^{(i)}(n)+\epsilon_i\equiv 3f^{(i+t)}(n)+\epsilon_{i+t} \pmod N.$$

Furthermore, if $f^{(i)}(n)\not\equiv f^{(i+t)}(n) \pmod 2$, then

$$3f^{(i)}(n)\equiv 3f^{(i+t)}(n)\pm 1 \pmod N,$$

contradicting equation (1). Therefore, $f^{(i)}(n)\equiv f^{(i+t)}(n) \pmod 2$ for all i, contradicting the $N=2$ case above.

Also solved by the proposer.

The Reach of a Rope **April 1998**

1548. *Ken Richardson, Texas Christian University, Fort Worth, Texas.*

Let D be a convex domain in the plane, and suppose that its boundary curve α is piecewise C^2. Imagine that a fence is built along the boundary, and that a rope of length L is attached to the outside of the fence at a point along the boundary. By pulling on the rope so that it is taut but constrained to remain outside D, a new curve β is traced out by the end of the rope. Assuming that L is at least half of the length of the curve α, is it true that the curve β determines the curve α?

Solution by the proposer.

Yes, the curve β determines the curve α.

There are three types of points on α. First are the attachment point and a finite number of "corner" points, points where the curve is not C^2. Because D is convex, for either the attachment point or any corner point on α, some part of the curve β follows the arc of a circle with this point at its center. Therefore, the attachment point and corner points of α are determined by β. Moreover, the attachment point corresponds to the circular arc with the greatest radius (L), so that the location of the attachment point and the length L of the rope are determined by β.

Now, recall the following facts about the geometry of plane curves, which may be found in vector calculus books as well as elementary differential geometry books. If γ is an oriented, C^2 curve that is parametrized so that it has unit speed, then $T = \gamma'$ is the unit tangent vector of the curve. The equation $T' = kN$ defines the curvature $k \geq 0$ and, if $k > 0$, the unit normal vector N of the curve. If $k > 0$, it follows that $N' = -kT$. The corresponding formulas for curves that do not have unit speed are $T = \gamma'/s$, $T' = skN$, and $N' = -skT$, where $s = \|\gamma'\|$ is the speed. The vectors T, N, and the function k depend only on the image point on the curve and are independent of the choice of parametrization. The function k satisfies $k = 1/r$, where r is the radius of curvature of the curve.

The second type of point on α consists of those that are not of the first type and that have positive curvature k. Starting with the circular arc corresponding to the attachment point, follow β in a counterclockwise direction. This defines an orientation of the unknown curve α as well. Suppose that α is parametrized so that it has unit speed, $\alpha(0)$ is the attachment point, and $\alpha(t)$ follows the counterclockwise orientation. Let $t \in (0, L)$ be chosen so that $\alpha(t)$ has positive curvature. The endpoint $\beta(t)$ of the rope is therefore

$$\beta(t) = \alpha(t) + (L - t)\alpha'(t).$$

That is, suppressing the variable t in the functions,

$$\beta = \alpha + (L - t)T_\alpha. \tag{1}$$

Differentiating, we have

$$\beta' = T_\alpha - T_\alpha + (L - t)T'_\alpha = (L - t)k_\alpha N_\alpha.$$

Therefore,

$$\begin{aligned} s_\beta &= (L - t)k_\alpha, \\ T_\beta &= N_\alpha. \end{aligned} \tag{2}$$

Differentiating again, we have

$$T'_\beta = N'_\alpha = -k_\alpha T_\alpha = s_\beta k_\beta N_\beta.$$

Then $N_\beta = -T_\alpha$ and, using equation (2), $r_\beta = 1/k_\beta = L - t$. Then equation (1) becomes

$$\beta = \alpha - r_\beta N_\beta,$$

or

$$\alpha = \beta + r_\beta N_\beta. \tag{3}$$

If the opposite orientation of the curves α and β were chosen, the identical equation would result. In addition, equation (3) remains valid for the points of β that are arcs of circles, determining either the attachment point or the corner point. Thus, using the fact that L is at least half the length of α, equation (3) implies that β determines the points of α that are either corner points, or points of nonzero curvature.

Finally, the third type of points on α is comprised of up to countably many line segments with zero curvature. An entire line segment will map to a single point on β. However, since D is convex and α is piecewise C^2, these points of α are completely determined by the points of α of the other two types.

Answers

Solutions to the Quickies on page 149

A889. The assertion holds for precisely those n that are squarefree. Suppose p^2 divides n, where p is prime. Let $a = p$ and suppose the given sequence is periodic with period d. Then $p \equiv p^{d+1} \equiv 0$, (mod p^2), a contradiction. Conversely, suppose n is squarefree. It suffices to show that $a^{\phi(n)+1} \equiv a \pmod p$ for any prime p dividing n, where $\phi(n)$ is Euler's phi function, for then the congruence will hold modulo n, and the congruence implies that the sequence of powers is periodic. If p divides a, then both sides are 0 (mod p). Otherwise $a^{\phi(n)} \equiv a^{\phi(p)\phi(n/p)} \equiv a^{(p-1)\phi(n/p)} \equiv 1 \pmod p$ by Fermat's Little Theorem, and we may multiply by a to get the desired congruence.

A890. There is no such formula in either case.

(a) When each face is equilateral with sides of length 2, each face has area $\sqrt{3}$ and the tetrahedron has volume $2\sqrt{2}/3$. When the three pairs of disjoint edges have lengths 1, $\sqrt{12}$, and $\sqrt{13}$, each face is a right triangle again having area $\sqrt{3}$. The altitude to one face is at most 1, so this tetrahedron has volume at most $(1/3)\sqrt{3}\cdot 1 = 1/\sqrt{3}$. Thus, there is no formula in terms of face areas.

(b) A tetrahedron with one face equilateral with sides of length 2 and the other three sides of length 3 has volume $\sqrt{23}/3$. One with with one face equilateral with sides of length 3 and the other three sides of length 2 has volume $3\sqrt{3}/4$. Thus, there is no symmetric formula in terms of edge lengths.

---

60 Years Ago in *Mathematics Magazine*

From *Mathematical World News*, January 1939 issue:

> The following note concerns Princeton University: To meet the growing demand for men trained in mathematical statistics, the department, in cooperation with the department of economics, now gives three one-term courses: Introduction to Statistics, Elementary Mathematical Statistics, and Statistical Inference. "With this arrangement of courses and the opportunities for further study of the subject in the independent reading of junior and senior years," Dean Eisenhart says, "we will have a program which should qualify a student very well in this field at the undergraduate level."

From *Mathematical World News*, May 1939 issue:

> The W.P.A. of New York City has as one of its projects the computation of certain mathematical tables, such as a table of exponential functions to fifteen decimal places and a table of the first ten posers of the integers from 1 to 1000. Dr. Arnold N. Lowan is in charge. This is one part of a general project directed by Dr. Lyman Briggs, of the National Bureau of Standards.

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.

Enzensberger, Hans Magnus, *Drawbridge Up: Mathematics—A Cultural Anathema* [*Zugbrücke außer Betrieb: Die Mathematik in Jenseits der Kultur; Eine Außenansicht*]. Translated by Tom Artin; German and English translation on facing pages. A K Peters, 1999; 48 pp, $5 (P). ISBN 1-56881-099-7.

This booklet is the text of an address at the International Congress of Mathematicians in Berlin last August, given after I had to leave for the start of term at Beloit. I had very much wanted to hear this talk, because Enzensberger for 50 years has been a leading intellectual and cultural force in Germany, and because only a few weeks earlier I had examined an utterly charming book that he wrote for children to counter their math anxiety (*Der Zahlenteufel: Ein Kopfkissenbuch für alle, die Angst vor der Mathematik haben*, Carl Hanser Verlag, München, 1997, now available in translation as *The Number Devil: A Mathematical Adventure*, Henry Holt and Co., New York, 1998). Enzensberger takes to task the "intellectual castration" that "exclusion of mathematics from the cultural sphere amounts to." Are mathematicians to blame? "It is simply implausible to lay the blame on a handful of experts, while the overwhelming majority of mankind happily renounces the acquisition of a cultural capital of immense significance and enormous charm." Enzensberger disputes Hardy's paean to the uselessness of pure mathematics but cites the difficulty of perceiving connections between pure and applied mathematics, plus a "cultural time lag" far larger than any other field—"Popular consciousness trails research by centuries." Enzensberger hones in on schools, wondering if the first five years "even include such a thing as mathematical instruction." He notes the utility of knowing arithmetic algorithms but rightfully asserts that "they have nothing to do with mathematical thinking." Older students are insufficiently challenged; teachers must teach the curriculum forced from on high. (David Mumford, President of the International Mathematical Union, relates in the Preface that he could not get the teachers at his children's high schools to "go outside during geometry and find out how tall the oak in the yard really is.") Enzensberger closes on a hopeful note, citing increasing numbers of "interpreters" of mathematics-speak into natural language. For them and for us, he repeats some advice from master interpreter Ian Stewart: "Lie a bit. ... [B]e prepared to bend [the truth] a little, if it helps people understand what you're doing." (This essay, which will become a classic, would be well supplemented in reprintings by some footnotes citing references or giving explanatory detail; for example, mention of "the Sokal affair" will arouse even less recognition within a couple of years.)

Peterson, Ivars, *The Mathematical Tourist: New and Updated Snapshots of Modern Mathematics*, W.H. Freeman, 1998; xvii + 266 pp, $14.95 (P). ISBN 0-7167-3250-5.

Peterson has updated the marvelous original edition of 1988, adding material on crystal structures, cellular automata, spacetime, the nature of proof (including Fermat's Last Theorem), finding large primes, factoring using quantum computers, and more. May it sell a million copies!

Hersh, Reuben, *What Is Mathematics, Really?*, Oxford University Press, 1997; xviii + 343 pp. ISBN 0-19-511368-3.

Hersh's answer: Not what you used to think it was, by any means. You think your favorite mathematical objects have independent existence? Sorry, platonism is not intellectually respectable today. You think theorems possess indubitable truth? That would "privilege" (not Hersh's verb) mathematics over other realms of human endeavor. You seek a philosophical foundation for mathematics? You're deluding yourself to think there could be one, and you don't need one anyway. According to Hersh, mathematics is a social entity, mathematical knowledge is fallible, and mathematical objects are cultural artifacts. "There's no need to look for a hidden meaning or definition of mathematics beyond its social-historic-cultural meaning. Social-historic is all it needs to be. Forget foundations, forget immaterial, inhuman 'reality.'" A counter-perspective might be that social-historic is all that current intellectual culture will allow mathematics to be. This is an immensely compelling book, a definitive argument. It is difficult not to be persuaded by Hersh's descriptive philosophy of mathematics, whose cultural humanism exorcises the last eternal truths from the mind of God. In a debate, he could run rings around you or me. The writing is boldly incisive, direct, and highly distilled (perhaps due in part to short sentences); there is not a dull paragraph in the book. In the preface, Hersh summarizes his perspective: "[D]isputing the nature and meaning of mathematics is ideological." For him, it certainly is. But should wisdom proceed from ideology, or the other way round?

Johnson, George, Genius or gibberish? The strange world of the math crank, *New York Times* (9 February 1999) (City Ed.) F1, F5; (National Ed.) D1, D5; `http://www.nytimes.com/library/national/science/020999sci-math-crank.html` .

Not every day does a photo of a former associate editor of this MAGAZINE appear spread across two columns in the *New York Times*. Woody Dudley, for many years one of the dozen people who bring you this MAGAZINE and now editor of the *College Mathematics Journal*, is probably the world expert on mathematical cranks. "I've been at this for a decade and still can't pin down exactly what it is that makes a crank a crank. They are usually men, old men. All are humorless. None of them are fat. It's like obscenity—you can tell a crank when you see one." Occasionally—but only occasionally—behind the unsolicited letter from a manual typewriter ("cranks are always about one level of technology behind") is there a self-taught near-genius. [The statement in the article accusing Gauss of neglecting and belittling Abel is downright wrong. Cauchy neglected refereeing the paper in question, despite protests from Legendre, but finally gave a "hasty, nasty, and superficial" report two and a half years later, after the death of Abel. See Bruno Belhoste, *Augustin-Louis Cauchy: A Biography*, pp. 55–60; Springer-Verlag, 1991.]

Mathews, Robert, Don't get even, get mad, *New Scientist* (10 October 1998) 26–31. Howard, Nigel, N-person soft games, *Journal of the Operational Research Society* 49 (1998) 144ff. Cooperation—or conflict [drama theory Web site]. `http://www.nhoward.demon.co.uk/drama.htm` .

The games of Prisoner's Dilemma and Chicken offer difficulties for game theorists, since they do not possess a unique Nash equilibrium (a strategy in which neither player can win more by individually choosing a different strategy). Now comes "drama theory," which suggests that the impasse comes from the assumption that the players act rationally. Drama theory tries to incorporate "game-changing emotions," which derive from past histories of the players. Nigel Howard's paper (and his book, appearing chapter by chapter at the Web site) shows how to take account of the "credibility paradoxes" at the root of such games and transform the games into others free of paradox.

(continued on p. 103)

design process I have been describing, that level of detail was not required. Rather, the theoretical solvability of the optimal allocation problem was used to determine one measure of system performance. And that is how an existence result was used in a completely practical setting.

REFERENCES

1. Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*, MIT Press, Cambridge, MA, 1990, 587–599.
2. Dan Kalman. When existence is enough, *Math Horizons*, April 1997, 23–26.
3. Oystein Ore and Robin J. Wilson. *Graphs and Their Uses*, New Mathematical Library, Vol. 34, Mathematical Association of America, Washington, DC, 1990, 53–59.
4. Alan Tucker. *Applied Combinatorics*, John Wiley & Sons, New York, NY, 1995, 137–167.

REVIEWS (continued from page 157)

Shulman, Polly, From Muhammad Ali to Grandma Rose, *Discover* 19 (12) (December 1998) 85–89; `http://www.discover.com/decissue/smallworld.html` . Peterson, Ivars, Close connections: It's a small world of crickets, nerve cells, computers, and people, *Science News* 154 (22 August 1998) 124–126. Harris, John M., and Michael J. Mossinghoff, The eccentricities of actors, *Math Horizons* (February 1998) 23–25. Strogatz, Steven H., and Duncan J. Watts, Collective dynamics of "small-world" networks, *Nature* 393 (4 June 1998) 440–442.

Modern folklore claims that everyone on the planet is connected to everyone else through a relatively short chain of acquaintances. This claim was investigated by psychologist Stanley Milgram in the 1960s, concretized in the play and film *Six Degrees of Separation* by John Guare, and popularized in a game of trying to trace connections of actors to the actor Kevin Bacon through joint appearances in films. Mathematicians are familiar with a similar phenomenon, the *Erdős number* of a mathematician. We are talking eccentricity here; in graph theory, the maximum distance from a particular vertex to any other vertex in a graph is the *eccentricity* of the vertex. Harris and Mossinghoff document that the eccentricity of Kevin Bacon in the "Hollywood" graph is 7, which is the minimum eccentricity of any actor, putting Bacon into what is known as the *center* of the graph. Such networks are neither regular (every node has the same small number of links to neighboring points) nor sparse (few connections relative to the number of nodes). Strogatz and Watts showed that introducing a few random connections into a regular graph can greatly decrease the average path length between two nodes. Graphs with a small average path length they call *small-world networks*, and they cite as examples the neural network of the worm *C. elegans*, the power grid of the western U.S., and the Hollywood graph. Small-world networks are important in the spread of disease, the diffusion of trade goods, and the transmission of information (including marketing over the Internet), as Peterson notes.

Petković, Miodrag, *Mathematics and Chess: 110 Entertaining Problems and Solutions*, Dover, 1998; v + 133 pp, \$5.95 (P). ISBN 0-486-29432-3.

This book uses chess and the chessboard as occasions for mostly mathematical puzzles, though some chess puzzles occur too. Naturally, rook polynomials occur, as do knight's tours and their generalizations, plus domino coverings, dissections, and generalized chessboards. The puzzles are fun, solutions are provided, and the reader will see much mathematics applied. There are a few references but only to specific results, rather than to further reading; and lamentably there is no index.

NEWS AND LETTERS

Mathematics Magazine Editor Search

The Mathematical Association of America (MAA) seeks to identify candidates to succeed Paul Zorn as Editor of *Mathematics Magazine* when his term expires in December 2000. The Search Committee must make a recommendation by May 1999, so that the new editor can be approved by the Board of Governors and begin handling all new manuscript submissions in January 2000.

Questions about the nature of the position and its workload can be addressed to Paul Zorn (`zorn@stolaf.edu`); questions about MAA support for the editor's work can be addressed to the MAA's Director of Publications, Don Albers (`dalbers@maa.org`).

Applicants should submit a resumé, names of references, and a statement of interest containing their ideas about the journal. These can be e-mailed with attachments as Word, WordPerfect, or plain-text TeX documents, to the chair of the Search Committee, Jim Daniel (`jimdaniel@mail.utexas.edu`). Alternatively, they can be mailed to

Mr. Don Albers
Mathematical Association of America
1529 Eighteenth St., NW
Washington, DC, 20036

for distribution to the Search Committee. Nominations are also welcome. Applications are welcome until the position is filled, although—because of time pressure—preference will be given to applications received by mid-March.

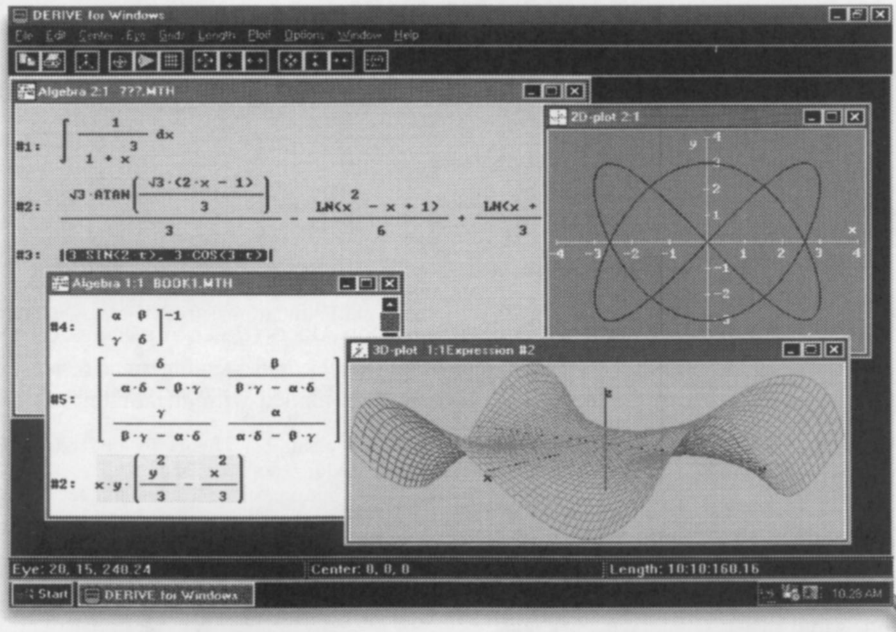
Letter to the Editor

Dear Editor:

Richard Thompson's article on the Global Positioning System in the October 1998 issue of the *Magazine* presents and interesting discussion of how a GPS receiver determines position. As is stated, the calculation relies on time estimates based on received signals from the satellites. Unmentioned is that for the Standard Positioning Service these signals are Gold codes, which are sequences constructed and analyzed using finite fields. Each satellite uses a separate Gold code. Thus, GPS offers a far-from-unique example of an engineering system blending aspects of both continuous and discrete mathematics.

Joseph J. Rushanan, Lead Mathematician
Signal Processing Section
The MITRE Corporation
Bedford, Massachusetts 01730-1420

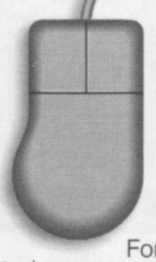
NEW! *Site Licenses and Student Pricing. See www.derive.com*



DERIVE® for Windows

DERIVE is the trusted mathematical assistant relied upon by students, educators, engineers, and scientists around the world. It does for algebra, equations, trigonometry, vectors, matrices, and calculus what the scientific calculator does for numbers — it eliminates the drudgery of performing long and tedious mathematical calculations. You can easily solve both symbolic and numeric problems and see the results plotted as 2D or 3D graphs.

For everyday mathematical work *DERIVE* is a tireless, powerful, and knowledgeable assistant. For teaching or learning mathematics, *DERIVE* gives you the freedom to explore different mathematical approaches better and more quickly than by using traditional methods.



System Requirements: Windows 95, 3.1x or NT running on a computer with 8 megabytes of memory.

Suggested Retail Price: $250. Educational pricing available.

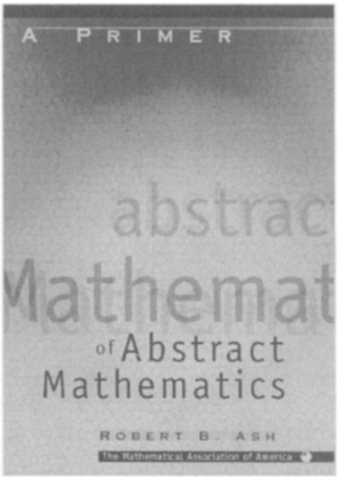
For product information and list of dealers, fax, email, write, or call Soft Warehouse, Inc. or visit our website at http://www.derive.com.

The Easiest just got Easier.

Soft Warehouse INC
HONOLULU • HAWAII

© 1996 Soft Warehouse, Inc. *DERIVE* is a registered trademark of Soft Warehouse, Inc. Other trademarks are the property of their respective owners.

Soft Warehouse, Inc. • 3660 Waialae Avenue
Suite 304 • Honolulu, Hawaii, USA 96816-3259
Telephone: (808) 734-5801 after 10:00 a.m. PST
Fax: (808) 735-1105 • Email: swh@aloha.com.



THE MATHEMATICAL ASSOCIATION OF AMERICA

A Primer of Abstract Mathematics

Robert Ash

Series: Classroom Resource Materials

A Primer of Abstract Mathematics prepares the reader to cope with abstract mathematics, specifically abstract algebra. It can serve as a text for prospective mathematics majors, as well as for those students taking or preparing to take a first course in abstract algebra, or those in applied fields who need experience in dealing with abstract mathematical ideas.

Learning any area of abstract mathematics involves writing formal proofs, but it is equally important to think intuitively about the subject and to express ideas clearly and cogently. The author aids intuition by keeping proofs short and as informal as possible, using concrete examples which illustrate all the features of the general case, and by giving heuristic arguments when a formal development would take too long. The text can serve as a model on how to write mathematics for an audience with limited experience in formalism and abstraction.

Ash introduces several expository innovations in *A Primer of Abstract Mathematics*. He presents an entirely informal development of set theory that gives students the basic results that they will need in algebra. The chapter which presents the theory of linear operators introduces the Jordan Canonical Form right at the beginning, with a proof of existence at the end of the chapter.

Contents:

Logic and Foundations: Truth Tables, Quantifiers, Proofs, Sets, Functions, Relations.

Counting: Fundamentals, The Binomial and Multinomial Theorems, The Principle of Inclusion and Exclusion, Counting Infinite Sets.

Elementary Number Theory: The Euclidean Algorithm, Unique Factorization, Algebraic Structures, Further Properties of Congruence Modulo m, Linear Diophantine Equations and Simultaneous Congruences, Theorems of Euler and Fermat, The Möbius Inversion Formula.

Some Highly Informal Set Theory: Well-Orderings, Zorn's Lemma and the Axiom of Choice, Cardinal Numbers, Addition and Multiplication of Cardinals.

Linear Algebra: Matrices, Determinants and Inverses, The Vector Space F^n, Linear Independence and Bases, Subspaces, Linear Transformations, Inner Product Spaces, Eigenvalues and Eigenvectors.

Theory of Linear Operators: Jordan Canonical Form, The Minimal and Characteristic Polynomials, The Adjoint of a Linear Operator, Normal Operators, The Existence of the Jordan Canonical Form.

Appendix: An Application of Linear Algebra.

Catalog Code: BMA/JR
188 pp., Paperbound, 1998
ISBN 0-88385-709-X
List: $27.95   MAA Member: $21.95

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
______	BMA/JR	______	______

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

Shipping & handling ______

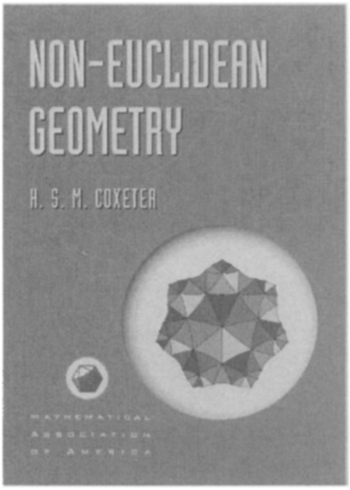
TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Non-Euclidean Geometry

Sixth Edition

H. S. M. COXETER

Series: Spectrum

H. S. M. Coxeter's classic book on non-Euclidean geometry was first published in 1942, and enjoyed eight reprintings before it went out of print in 1968. The MAA is delighted to be the publisher of the sixth edition of this wonderful book, updated with a new section 15.9 on the author's useful concept of inversive distance.

Throughout most of this book, non-Euclidean geometries in spaces of two or three dimensions are treated as specializations of real projective geometry in terms of a simple set of axioms concerning points, lines, planes, incidence, order and continuity, with no mention of the measurement of distances or angles. This synthetic development is followed by the introduction of homogeneous coordinates, beginning with Von Staudt's idea of regarding points as entities that can be added or multiplied. Transformations that preserve incidence are called collineations. They lead in a natural way to elliptic isometries or "congruent transformations". Following a recommendation by Bertrand Russell, continuity is described in terms of order. Elliptic and hyperbolic geometries are derived from real projective geometry by specializing an elliptic or hyperbolic polarity which transforms points into lines (in two dimensions) or planes (in three dimensions) and vice versa.

An unusual feature of the book is its use of the general linear transformation of coordinates to derive the formulas of elliptic and hyperbolic trigonometry. The area of a triangle is related to the sum of its angles by means of an ingenious idea of Gauss. This treatment can be enjoyed by anyone who is familiar with algebra up to the elements of group theory.

Catalog Code: NEC/JR
320 pp., Paperbound, 1988
ISBN 0-88385- 522-4
List: $30.95 MAA Member: $24.50

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
________	NEC/JR	________	________

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

Shipping & handling ________

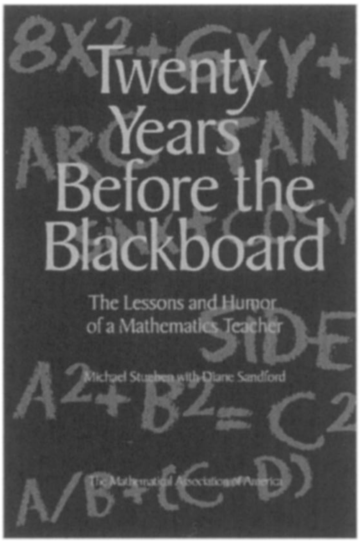
TOTAL ________

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ____________ Expires ___/___

Signature ____________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Twenty Years Before the Blackboard

The Lessons and Humor of a Mathematics Teacher

Michael Stueben with Diane Sandford

Series: Spectrum

A perfect gift for the new teacher... or for anyone interested in the teaching of mathematics.

This book is the legacy of twenty years of mathematics teaching. During this time, the author searched for motivation techniques, mnemonics, insightful proofs, and serious applications of humor to aid his teaching. The result is this book: part philosophy, part humor, and part biography. Readers will be amused and enlightened on every page.

Mr. Stueben shows how he has used humor and wordplay to motivate his students. The book is filled with wonderful problems and proofs, as well as the author's insights about how to approach teaching problem solving to high school students. Sections of the book also treat the use of calculators and computers in the classroom. A section on mnemonics shows how teachers can use memory aids to help their students learn and retain material.

All in all, *Twenty Years Before the Blackboard* provides a goldmine of ideas for the classroom teacher. Although Mr. Stueben taught at the high school level, his book is an excellent "methods" book for mathematics teachers at all levels.

Read what Martin Gardner has to say about this fascinating book:

It's been decades since I read so entertaining a book about mathematics. The book is a treasure-trove of mathematical jokes, rhymes, anecdotes, word play, mnemonics, and beautiful proofs. For teachers there is an abundance of wise advice based on the author's twenty years in high school teaching. Mathematicians at all levels, from amateurs to college professors will not only chuckle over its gems, but learn much they did not know before. —Martin Gardner

Catalog Code: TYB/JR98
174 pp., Paperbound, 1998, ISBN 0-88385-525-9
List: $29.50 MAA Member: $23.50

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
______	TYB/JR98	______	______

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

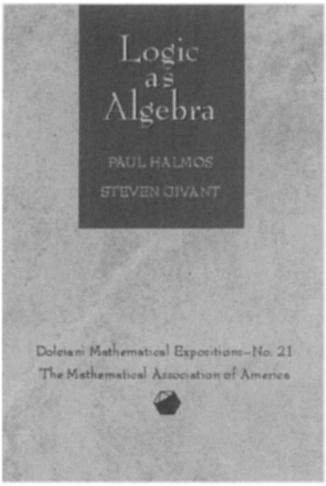
Shipping & handling ______

TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ____________ Expires ___/___

Signature ____________________



THE MATHEMATICAL ASSOCIATION OF AMERICA

Logic as Algebra

Paul Halmos and Steven Givant

Series: Dolciani Mathematical Expositions

This book is based on the notes of a course in logic given by Paul Halmos. This book retains the spirit and purpose of those notes, which was to show that logic can (and perhaps should) be viewed from an algebraic perspective. When so viewed, many of its principal notions are seen to be old friends, familiar algebraic notions that were "disguised" in logical clothing. Moreover, the connection between the principal theorems of the subject and well-known theorems in algebra becomes clearer. Even the proofs often gain in simplicity.

Propositional logic and monadic predicate calculus—predicate logic with a single quantifier— are the principal topics treated. The connections between logic and algebra are carefully explained. The key notions and the fundamental theorems are elucidated from both a logical and algebraic perspective. The final section gives a unique and illuminating algebraic treatment of the theory of syllogisms—perhaps the oldest branch of logic, and a subject that is neglected in most modern logic texts.

The presentation is aimed at a broad audience—mathematics amateurs, students, teachers, philosophers, linguists, computer scientists, engineers, and professional mathematicians. Whether the reader's goal is a quick glimpse of modern logic or a more serious study of the subject, the book's fresh approach will bring novel and illuminating insights to beginners and professionals alike. All that is required of the reader is an acquaintance with some of the basic notions encountered in a first course in modern algebra. In particular, no prior knowledge of logic is assumed. The book could serve equally well as a fireside companion and as a course text.

Contents: **What is Logic?:** To count or to think; A small alphabet; A small grammar; A small logic; What is truth?; Motivation of the small language; All mathematics. **Propositional Calculus:** Propositional symbols; Propositional abbreviations; Polish notation; Language as an algebra; Concatenation; Theorem schemata; Formal proofs; Entailment; Logical equivalence; Conjunction; Algebraic identities. **Boolean Algebra:** Equivalence classes; Interpretations; Consistency and Boolean algebra; Duality and commutativity; Properties of Boolean algebras; Subtraction; Examples of Boolean algebras. **Boolean Universal Algebra:** Subalgebras; Homomorphisms; Examples of homomorphisms; Free algebras; Kernels and ideals; Maximal ideals; Homomorphism theorem; Consequences; The representation theorem. **Logic via Algebra:** Pre-Boolean algebras; Substitution rule; Boolean logics; Algebra of the propositional calculus; Algebra of proof and consequence. **Lattices and Infinite Operations:** Lattices; Non-distributive lattices; Infinite operations. **Monadic Predicate Calculus:** Propositional functions; Finite functions; Functional monadic algebras; Functional quantifiers; Properties of quantifiers; Monadic algebras; Free monadic algebras; Modal logics; Monadic logics; Syllogisms.

Catalog Code: DOL-21/JR98
152 pp., Paperbound, 1998, ISBN 0-88385-327-2
List: $27.00 MAA Member: $21.95

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-21/JR98	______	______

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

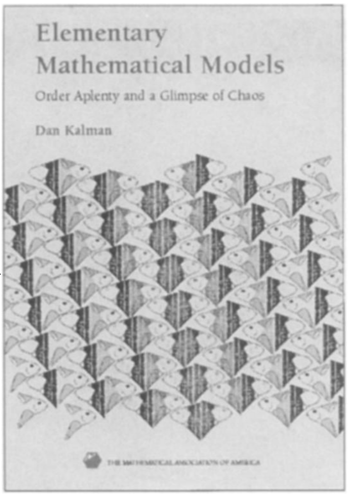
Shipping & handling ________

TOTAL ________

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ____________ Expires ___/___

Signature ____________________



THE MATHEMATICAL ASSOCIATION OF AMERICA

Elementary Mathematical Models

Order Aplenty and a Glimpse of Chaos

Series: Classroom Resource Materials

Dan Kalman

New in the Classroom Resource Materials Series...suitable as a text in college algebra, finite mathematics, precalculus, and liberal arts mathematics courses

Elementary Mathematical Models *(EMM)* claims a middle ground between college algebra and liberal arts mathematics. Like the college algebra course, *EMM* emphasizes the elementary functions of analysis: linear, quadratic, polynomial, and rational functions; square roots; exponentials and logarithms. These functions are the building blocks for the simple models that appear in first courses in the physical, life, and social sciences. And while *EMM* does not stress algebraic manipulation as an end in itself, it does recognize how important algebra is. Moreover, it provides students with the opportunity to see for themselves why algebra is needed, and what it contributes to formulating and analyzing models.

Like the liberal arts mathematics course, *EMM* makes a concerted effort to convey something of the scope, power, and fascination of mathematics, to students who may never study mathematics again. Each mathematical topic evolves naturally in formulating simple discrete models for inherently interesting contexts. For example, exponential functions emerge from the study of models exhibiting geometric growth—defined as growth that increases by equal proportions in equal periods of time. Throughout the course, a recurring theme is the evolution from simple recursive hypotheses (e.g., the population next year will be 10% greater than this year), to difference equations ($p_{n+1} = 1.1\, p_n$), to solutions ($p_n = p_0\,(1.1^n)$), to qualitative behavior of models. This theme appears repeatedly as the students encounter a series of increasingly sophisticated growth models, starting with arithmetic growth and ending with logistic growth. The course climaxes with an exploration of the chaotic behavior that can occur in logistic growth models.

The presentation is accessible even to students with a weak algebraic background. Throughout the book, numerical, graphical, and symbolic approaches are used systematically. There is a rich collection of examples and exercises. Reading comprehension exercises in each chapter provide a strong emphasis on reading and writing about mathematical concepts.

Contents: Overview; Sequences and Differences Equations; Arithmetic Growth; Linear Graphs, Functions and Equations; Quadratic Growth Models; Quadratic Graphs, Functions and Equations; Polynomials and Rational Functions; Fitting a Line to Data; Geometric Growth; Exponential Functions; More on Logarithms; Geometric Sums and Mixed Models: Logistic Growth; and Chaos in Logistic Models.

Catalog Code: EMM/JR
360 pp., 1997, Paperbound, ISBN 0-88385-707-3
List: $32.50 MAA Member: $25.95

Supplementary problem collection available with adoption orders. Call us toll free at 1-800-331-1622 for more details.

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
______	EMM/JR	______	______
	Shipping & handling		______
	TOTAL		______

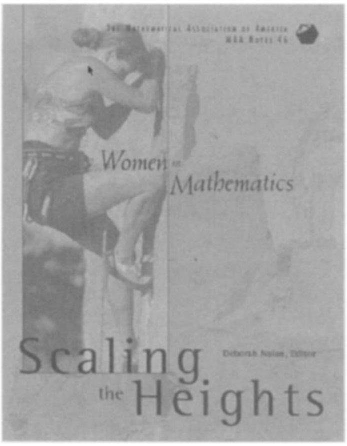
All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Women in Mathematics

Scaling the Heights

Deborah Nolan, Editor

Series: MAA Notes

Motivate your students to study advanced mathematics

Women in Mathematics: Scaling the Heights will provide you with a wealth of ideas and examples you can use to develop programs and courses that will motivate undergraduates who want to study advanced mathematics. The articles in this collection make a unique and valuable contribution to upper-division undergraduate mathematics education. They show us what talented undergraduates can do.

The heart of this book presents the insights of eight individuals who have taught at the Summer Mathematics Institute at Mills College. They share their course materials and give pedagogical tips on how to teach topics in mathematics that are not ordinarily part of the undergraduate curriculum, and in ways not often found in the undergraduate classroom. Although the courses described here were designed to encourage talented undergraduate women to pursue advanced degrees in mathematics, the good ideas found in them are gender free and can be used equally well with male as well as female students.

Exercises, class handouts, lists of research projects, and references are included. Topics covered are algebraic coding theory, hyperplane arrangements, p-adic numbers, quadratic reciprocity, stochastic processes, and linear optimization.

The book rounds out the material presented by the Summer Mathematics Institute instructors, with perspectives from mathematicians who have been active in the promotion of women in the field. Results from a survey of undergraduate mathematics majors in which they tell us what they think about the major and their future in mathematics complements these essays.

Catalog Code: NTE-46/JR
146 pp., Paperbound, 1997
ISBN 0-88385-156-3
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
______	NTE-46/JR	______	______

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

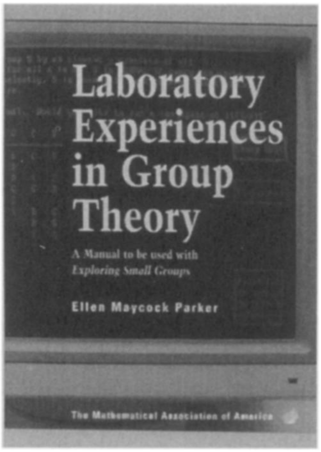
Shipping & handling ________

TOTAL ________

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________



THE MATHEMATICAL ASSOCIATION OF AMERICA

Laboratory Experiences in Group Theory

A Manual to be Used with *Exploring Small Groups*

Ellen Maycock Parker

Series: Classroom Resource Materials

A lab manual with software for introductory courses in group theory or abstract algebra

Laboratory Experiences in Group Theory is a workbook of 15 laboratories designed to be used with the software *Exploring Small Groups* as a supplement to the regular textbook in an introductory course in group theory or abstract algebra. Written in a step-by-step manner, the laboratories encourage students to discover the basic concepts of group theory and to make conjectures from examples that are easily generated by the software. The labs can be assigned as homework or can be used in a structured laboratory setting. Since the software is user-friendly and the laboratories are complete, students and faculty should have no difficulty in using the labs without training.

Most students find that the laboratories provide an enjoyable alternative to the "theorem-proof-example" format of a standard abstract algebra course. At the end of the semester, one student wrote in his evaluation of the course:

I am truly grateful for the laboratory component...Work on the computer helped to make the abstract theory more concrete... One of the best things about the labs was that we formed our own conjectures about the patterns we saw...I believe that the progression of (1) lab, (2) conjecture, (3) class discussion, and (4) proof was highly beneficial in gaining understanding of the abstract material of the course.

Table of Contents: 1. Groups and Geometry; 2. Cayley Tables; 3. Cyclic Groups and Cyclic Subgroups; 4. Subgroups and Subgroup Lattices; 5. The Center and Commutator Subgroups; 6. Quotient Groups; 7. Direct Products; 8. The Unitary Groups; 9. Composition Series; 10. Introduction to Endomorphisms; 11. The Inner Automorphisms of a Group; 12. The Kernel of an Endomorphism; 13. The Class Equation; 14. Conjugate Subgroups; 15. The Sylow Theorems; Appendix A. Table Generation Menu of *Exploring Small Groups (ESG)*; Appendix B. Sample Library of *ESG*; Appendix C. Group Library of *ESG*; Appendix D. Group Properties Menu

Exploring Small Groups, the software packaged with this lab manual, is on a 3½" DD PC compatible disk. This is a DOS program that can be run in Windows. The software was developed by Ladnor Geissinger, University of North Carolina at Chapel Hill.

Catalog Code: LABE/JR
112 pp., Paperbound, 1996, ISBN 0-88385-705-7
List: $26.50 MAA Member: $21.00

ORDER FROM:
THE MATHEMATICAL ASSOCIATION OF AMERICA
PO Box 91112, Washington, DC 20090-1112
1-800-331-1622 (301) 617-7800 FAX (301) 206-9789

Membership Code: ____ ____ ____ ____ ____ ____

Name ______________________

Address ______________________

City ______________________

State ____________ Zip ____________

Qty.	Catalog Code	Price	Amount
____	LABE/JR	____	____
____	____________	____	____
		TOTAL	____

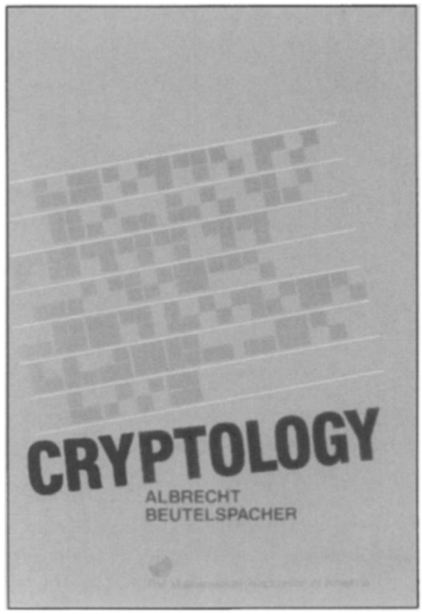
Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

CRYPTOLOGY

Albrecht Beutelspacher



This fascinating little book is eminently readable, and it is a great deal of fun to peruse... the book is a real treat. We need more books like this, crafted by expert hands yet crafted so that the general reader can enjoy them.
—Bulletin of The Institute of Combinatorics and Its Applications

This excellent and entertaining book is suitable for a first course in cryptology for mathematical enthusiasts. An abundance of exercises and an excellent list of related references are included.
—The Mathematics Teacher

In spite of the light-hearted style in which the book is written throughout, it is a serious—and successful—attempt to explain the basis of coding and decoding messages...I can strongly recommend this book to anyone who wants a brief but comprehensive, eminently readable, and up-to-date introduction to this increasingly popular topic.
— The Mathematical Gazette

All of cryptology is covered in this work...Occupying a niche in the halls of the ivory tower of pure mathematics for nearly two millennia, number theory now forms a pillar of modern society. This book is the best explanation available today of how that pillar was constructed.
— Charles Aschbacher

A model to follow in order to make mathematics better known and understood. Accessible to a broad audience. Have fun reading this book, while you are getting a better understanding of cryptology.
— Bulletin of the Belgian Mathematics Society

How can messages be transmitted secretly? How can one guarantee that the message arrives safely in the right hands exactly as it was transmitted? Cryptology—the art and science of "secret writing"—provides ideal methods to solve these problems of data security.

The book is fun to read, and the author presents the material clearly and simply. Many exercises and references accompany each chapter.

176 pp., Paperbound, 1994
ISBN 0-88385-504-6
List: $34.00
MAA Member: $26.95
Catalog Code: CRYPT/JR

ORDER FROM:
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW Washington, DC 20036
1-800-331-1622 (301) 617-7800 FAX (202) 265-2384

Membership Code: ____ ____ ____ ____ ____ ____

Name ______________________

Address ______________________

City ______________________

State ____________ Zip ____________

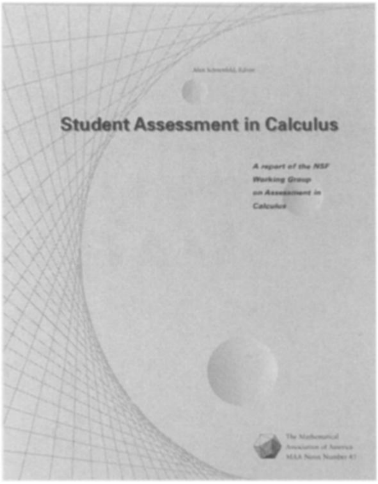
QTY.	CATALOG CODE	PRICE	AMOUNT
______	CRYPT/JR	______	______
______	______	______	______
		TOTAL	______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires __/__

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Student Assessment in Calculus

A Report of the NSF Working Group on Assessment in Calculus

ALAN SCHOENFELD, EDITOR

Series: MAA Notes

If you teach calculus, you should read this book. If you want to know what mathematics your students understand, or if you want to know how to find out what they understand, this book contains essential information for you.

It doesn't matter whether you teach a reform or traditional course, whether you have large or small sections, or whether you use lectures or laboratories. The bottom line is the same: When all is said and done, what counts is what our students understand. And that's what *Student Assessment in Calculus* is about.

Over the last ten years calculus instruction has changed in numerous ways. Whether they were trying on new ideas or following the more traditional routes towards conceptual understanding, both individual faculty and departments needed to know if their instruction was effective. To help deal with that issue, the National Science Foundation brought together a Working Group of experts in students' mathematical thinking, in assessment, and in calculus reform. The goals of their work were to:

- develop a framework to tailor calculus instruction to the students' needs;
- establish an agenda for further research on student understanding;
- describe how to make use of a range of techniques to test what students know, such as multiple-choice tests or short essay questions, student portfolios and "clinical" interviews;
- summarize major goals of the reform movement and describe the challenges faced by those who are taking a closer look at how students learn;
- illustrate the ways in which calculus projects attempt (via exams, papers, projects, etc.) to find out what their students have learned.

This book is the result of those efforts. If you teach calculus, if you want to see examples of useful assessment techniques, or if you are interested in issues of how to measure student learning in mathematics, then there is a lot for you here.

Catalog Code: NTE-43/JR97
122 pp., Paperbound, 1997
ISBN 0-88385-152-0
List: $34.95 MAA Member: $29.00

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
______	NTE-43/JR97	______	______

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

Shipping & handling ________

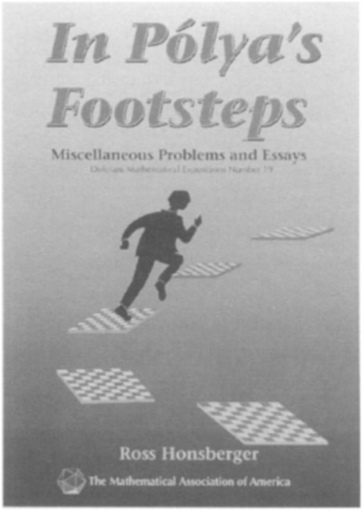
TOTAL ________

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



In Pólya's Footsteps

Miscellaneous Problems and Essays

Series: Dolciani Mathematical Expositions

Ross Honsberger

Another elegant collection of problems from Ross Honsberger

The study of mathematics is often undertaken with an air of such seriousness that it doesn't always seem to be much fun at the time. However, it is quite amazing how many surprising results and brilliant arguments one is in a position to enjoy with just a high school background. This is a book of miscellaneous delights, presented not in an attempt to instruct but as a harvest of rewards that are due good high school students and, of course, those more advanced — their teachers, and everyone in the university mathematics community. Admittedly, they take a little concentration, but the price is a bargain for such gems.

A half dozen essays are sprinkled among some hundred problems, most of which are the easier problems that have appeared on various national and international Olympiads. Many subjects are represented — combinatorics, geometry, number theory, algebra, probability, ... The sections may be read in any order. The book concludes with twenty-five exercises and their detailed solutions.

Something to delight will be found in every section — a surprising result, an intriguing approach, a stroke of ingenuity — and the leisurely pace and generous explanations make them a pleasure to read.

The inspiration for many of the problems came from the Olympiad Corner of *Crux Mathematicorum,* published by the Canadian Mathematical Society.

Catalog Code: DOL-19/JR
328 pp., Paperbound, 1997
ISBN 0-88385-326-4
List: $31.95 MAA Member: $24.95

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-19/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 ______________________________

CONTENTS

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