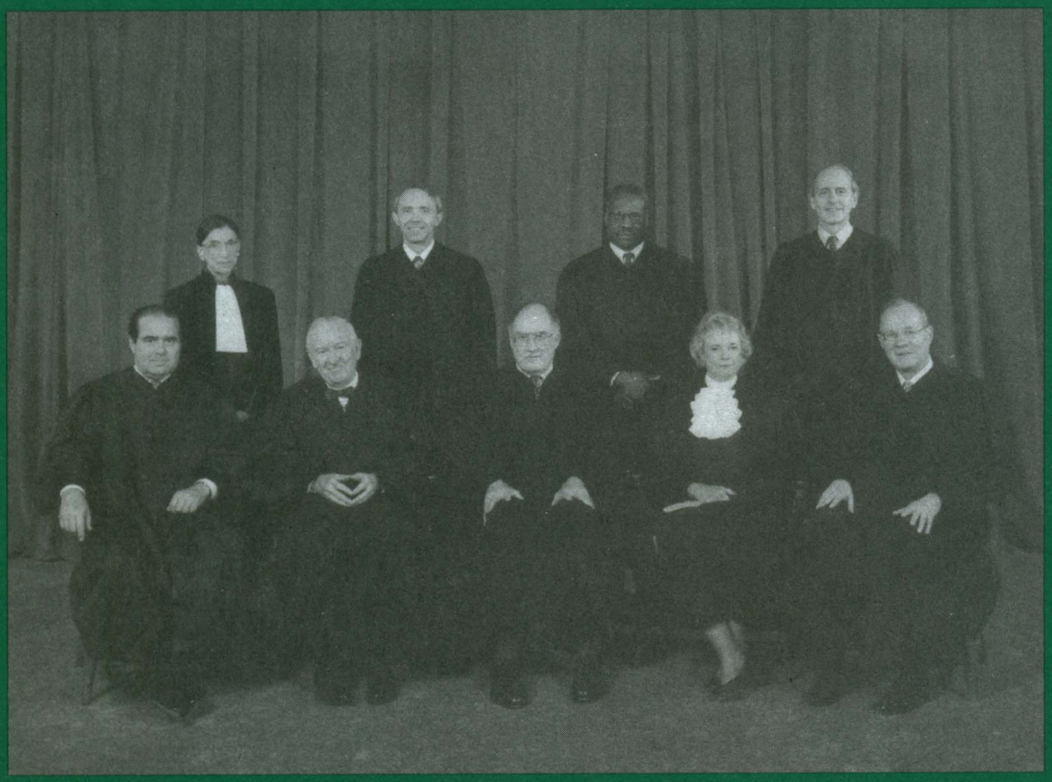
Vol. 79, No. 5, December 2006

MATHEMATICS MAGAZINE



The Rehnquist Court

- Spectral Analysis of the Supreme Court
- Surprising Dynamics From a Simple Model
- Counting Train Track Layouts
- Dials and Levers and Glyphs, Oh My!
  Linear Algebra Solutions to Computer Game Puzzles

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 74, pp. 75–76, and is available from the Editor or at www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Submit new manuscripts to Allen Schwenk, Editor, Mathematics Magazine, Department of Mathematics, Western Michigan University, Kalamazoo, MI, 49008. Manuscripts should be laser printed, with wide line spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should mail three copies and keep one copy. In addition, authors should supply the full five-symbol 2000 Mathematics Subject Classification number, as described in *Mathematical Reviews*.

The **cover image** shows the 2003 Supreme Court Justices photographed by Richard Strauss, Smithsonian Institution. Collection, The Supreme Court Historical Society.

AUTHORS

Brian Lawson is an assistant professor of political science at the University of Cincinnati. He received a BS in applied mathematics and a PhD in political science, both from UCLA. He has worked in political offices at the national, state, and local level for various elected officials. His research focuses on statistical analysis of voting in deliberative bodies and on political corruption and professionalization in the United States. In his free time he enjoys Tai Chi.

Michael Orrison is an assistant professor of mathematics at Harvey Mudd College. He received his BA from Wabash College and his PhD from Dartmouth College. His research interests include algebra, graph theory, voting theory, and computational noncommutative harmonic analysis. When he is not teaching or doing research, he enjoys spending time with his wife and children, plunking around on a piano, and watching too much television.

David Uminsky is a third year graduate student in mathematics at Boston University. He received his BS in mathematics at Harvey Mudd College, where Michael Orrison advised his senior thesis work. His primary research interests include partial differential equations, dynamical systems, and complex dynamics. On his days off from being a graduate student, David enjoys biking around Boston, drinking too much coffee, and being outrun on the soccer field.

Jim Walsh receiveed his PhD degree from Boston University in1991. Since that time he hs been happily ensconced at Oberlin College, even though he recently began a term as Chair of the Department. His research area is dynamical systems, with a recent foray into iteration of bimodal interval maps. He enjoys bringing the innherent wonders of dynamical systems to students at all levels. His nonmathematical wonders most often arise in connection with his wife and two children.

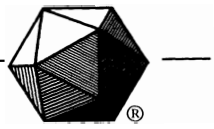
James Beaman is a statistical consultant for the Clinical Immunology Laboratory at the Rosalind Franklin University of Medicine and Science. He received his bachelor's degree from Carthage College (Cum Laude) in 2005, majoring in mathematics. He was a member of Pi Mu Epsilon, and was awarded a Faculty Honors Scholarship and the David J. Dorak Memorial Scholarship for scholar athletes. James was also a three-year captain of the swim team and twice named to the All-Conference team. He studied train track layouts as a January-Term undergraduate research project with Mark Snavely.

Erin Beyerstedt graduated from Carthage College as a mathematics major, and was named the Outstanding Senior Mathematics Major for 2004. She is currently a graduate student at Western Washington University and hopes to teach college mathematics. Erin studied train track layouts as a summer research student under the guidance of Mark Snavely.

Mark Snavely received his BS degree from Grove City College in 1985, and his PhD in dynamical systems from Northwestern University in 1990. He joined the Carthage College mathematics faculty in 1990, and is currently Chair of the Department. He enjoys guiding undergraduate research projects in discrete mathematics, dynamical systems, and mathematical modeling, as well as playing with trains with his sons, Brian and Matthew.

Jessica Sklar is a mathematics professor at Pacific Lutheran University in Tacoma, WA. She received a BA in mathematics and English from Swarthmore College, and her MS and PhD in mathematics from the University of Oregon, where she specialized in noncommutative ring theory. She thinks abstract algebra is the cat's meow, and is especially interested in making it accessible to the general public. In her spare time, she enjoys going bowling and doing crossword puzzles in pen.

Vol. 79, No. 5, December 2006



MATHEMATICS MAGAZINE

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for *MATHEMATICS MAGAZINE* to an individual member of the Association is $131. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 20% dues discount for the first two years of membership.)

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

MAA Advertising
c/o Marketing General, Inc.
209 Madison Street Suite 300
Alexandria VA 22201

Phone: 866-821-1221
Fax: 866-821-1221
E-mail: rhall@marketinggeneral.com

Further advertising information can be found online at www.maa.org

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

Surprising Dynamics From a Simple Model

JAMES A. WALSH
Oberlin College
Oberlin, OH 44074-1019
jim.a.walsh@oberlin.edu

You are knee-deep in the craziness of a typical semester. There are (at least) two major obligations vying for your valuable time, be they teaching-related issues versus revising that paper, preparing for a committee meeting versus going to the gym, or perhaps playing softball with your daughter and soccer with your son. How do you decide how much time to devote to each of these activities in a given day?

Fully aware that "there is no best model, only better ones" [**1**, p. xv], I nevertheless tout a particularly simple model of this decision-making process as being exemplary [**7**]. It is simple enough to include in an undergraduate dynamical systems or modeling course, yet sophisticated enough to capture a variety of possible behaviors. The model yields insights both valuable and mathematically interesting. Above all, perhaps, this versatile model also arises in economics when studying price fluctuations in a single commodity market [**10**].

After introducing a queueing model of this decision-making process [**7**], I will show how a similar family of mappings arises in the cobweb model of adaptive price expectations from economic dynamics [**10**]. This will be followed by an investigation into the dynamics of this model, which include both the well-known period-doubling route to chaos and the less well-known "period-halving route to stable equilibrium." I will conclude with several observations.

A queueing model of a two-task decision-making process

Suppose that each day (or, more generally, each time period), you must decide how much effort to devote to two activities, or *jobs*, denoted A and B. Begin by scaling time so that the time period is 1. To further simplify matters, assume the input rates α and β of jobs A and B, respectively, are constant. Referring to FIGURE 1, think of these jobs as flowing into queues, with queue volumes corresponding to the amount of each job waiting to be "served" by you.

Let ϕ_A and ϕ_B be *decision functions*, denoting the rate at which you work on jobs A and B, respectively. (Equivalently, the rate at which queues A and B would empty, absent any input.) As the time period is 1, note that $\alpha + \beta$ and $\phi_A + \phi_B$ represent the total volume inflow and total voume outflow, respectively, for the system over one time period.

Finally, assume the system is closed in the sense that your total capacity $\phi_A + \phi_B$ equals the total input $\alpha + \beta$ over each time period, and normalize so that $\phi_A + \phi_B = \alpha + \beta = 1$.

Let x_n and y_n denote the volumes of queues A and B, respectively, at the nth time period. With an eye toward arriving at a function of a single variable, assume the

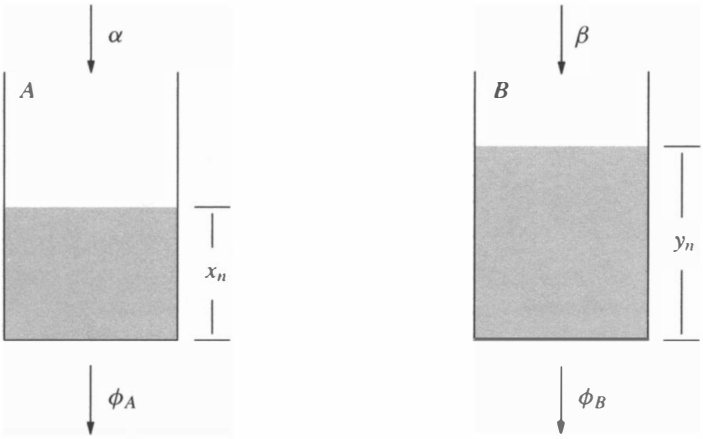


Figure 1 The queueing model

functions ϕ_A and ϕ_B depend on the difference $x_n - y_n$ in queue volumes. The discrete evolution of queue volumes over time is then given by the pair of equations

$$x_{n+1} = x_n + \alpha - \phi_A(x_n - y_n) \tag{1}$$

and

$$y_{n+1} = y_n + \beta - \phi_B(x_n - y_n). \tag{2}$$

To reduce to one recurrence relation containing a single variable, start by adding equations (1) and (2). Recalling that $\alpha + \beta = \phi_A + \phi_B$, the sum of the queue volumes is then constant for all n:

$$x_{n+1} + y_{n+1} = x_n + y_n = C. \tag{3}$$

Solving (3) for y_n and substituting into (1) yields

$$x_{n+1} = x_n + \alpha - \phi_A(2x_n - C). \tag{4}$$

In analyzing model behavior, we can thus focus on the size of queue A (or, equivalently via (3), the size of queue B) over time.

This model will be complete once we choose the decision function ϕ_A, now simply denoted ϕ. The choices presented here lead to a model in which the longer queue is served with higher priority. To that end, the simplest such ϕ is the function ϕ_1 graphed in FIGURE 2. The model (4) then encompasses an all-or-nothing strategy: If, for ex-

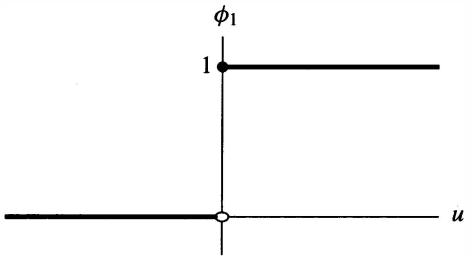


Figure 2 The all-or-nothing decision function ϕ_1

ample, $x_n > y_n$ (so that $x_n - y_n > 0$ and $x_n > C/2$), the entire time period is devoted to job A.

A second possibility for ϕ is the logistic-type function ϕ_2 graphed in FIGURE 3. In this case most, though not all, of the time is devoted to the activity having the longer queue. The graph of ϕ_2 is called an *S-shaped curve*, meaning that it has a unique inflection point which is a maximum of the derivative function $\phi_2'(u)$.

As is often the case in discrete dynamical systems a parameter will be included, which, in this case, will tune the steepness of the graph of ϕ_2. For example, each map in the one-parameter family

$$\phi_2(u) = \frac{1}{1 + e^{-\lambda u}}, \quad \lambda > 0 \tag{5}$$

has an S-shaped graph. For future reference, also note that, for $u \neq 0$, $\phi_2(u) \to \phi_1(u)$ as $\lambda \to \infty$. With this choice of ϕ_2, equation (4) becomes

$$x_{n+1} = x_n + \alpha - \frac{1}{1 + e^{-\lambda(2x_n - C)}}. \tag{6}$$

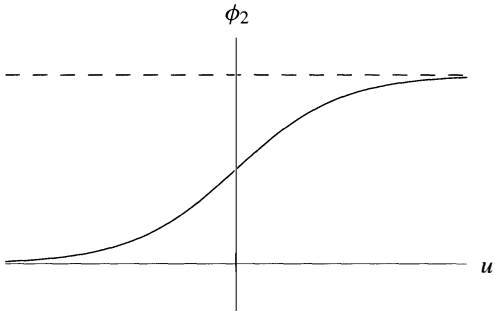


Figure 3 The decision function ϕ_2

The problem then is to understand the behavior of the sequence $\{x_n\}$ generated by (4) as $n \to \infty$. To avoid negative queue volumes, I will assume that the constant C is sufficiently large and that λ is chosen appropriately. As we will see (FIGURE 9, for example), for $C \geq 2$ all interesting behavior involves λ-values for which neither queue is ever empty. I will thus set $C = 2$ in all that follows. Equation (6) then becomes

$$x_{n+1} = x_n + \alpha - \frac{1}{1 + e^{-2\lambda(x_n - 1)}}. \tag{7}$$

The function $f(x) = x + \alpha - \phi_2(2x - 2), 0 < \alpha < 1$, which determines the evolution of x_n via (7), is the sum of a linear function and the negation of an S-shaped curve. Before investigating the dynamics of this model, I will show how a similar family of maps arises in a well-known model from economic dynamics.

The cobweb model from economic dynamics

One of the basic supply and demand models from economics, the cobweb model concerns the price dynamics in a single commodity market with a one year lag in supply.

As the market is typically agricultural (and I live in Ohio!), I will assume the commodity is corn.

In order to determine how much corn to plant in year $n-1$, to be harvested the following year, the farmer must estimate the price at which corn will sell in year n. This estimate is called the *expected price* and is denoted π_n. Let p_n denote the actual selling price of corn in year n. Then the demand for corn will be a function of its current price p_n, while the supply is a function of the expected price π_n. Note that while many calculus books present price as a dependent variable, here it is an independent variable.

Let $q_n^d = D(p_n)$ and $q_n^s = S(\pi_n)$ denote the demand and supply of corn at year n. Assume each of D and S is continuous, with D strictly decreasing and S strictly increasing (FIGURE 4). An explicit assumption of the cobweb model states that $q_n^d = q_n^s$ for each n, representing a so-called temporary equilibrium. (In spirit this follows the basic supply-and-demand model in which, once the supply is specified, prices adjust so that demand equals supply [**12**, p. 72].) Moreover, the expected price is assumed to be a weighted average of each of the previous year's expected price π_n and price p_n:

$$\pi_{n+1} = (1-w)\pi_n + wp_n, \quad 0 < w \leq 1. \tag{8}$$

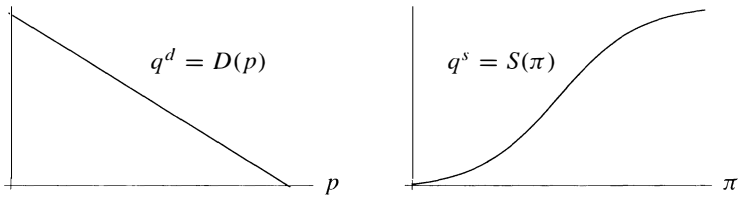


Figure 4 Demand and supply curves for the cobweb model

As in the previous section, this model will be reduced to one recurrence relation containing a single variable. The equation $q_n^d = q_n^s$ or, equivalently, $D(p_n) = S(\pi_n)$, implies $p_n = D^{-1}S(\pi_n)$. Substituting this expression for p_n into (8) yields

$$\pi_{n+1} = (1-w)\pi_n + wD^{-1}S(\pi_n). \tag{9}$$

Note that for $w = 1$, (8) reduces to $\pi_{n+1} = p_n$, implying the expected price is simply the current year's price. In this case (9) yields $\pi_{n+1} = D^{-1}S(\pi_n)$. As the function of a single real variable $D^{-1}S$ is continuous and monotonic, the possible types of long-term behavior for the sequence $\{\pi_n\}$ are very limited.

PROPOSITION 1. *Let $w = 1$ and pick any initial π_0. As $n \to \infty$, the sequence $\{\pi_n\}$ defined by* (9) *either*

(i) *converges to an equilibrium point,*

(ii) *converges to a period-2 oscillation, or*

(iii) *satisfies $|\pi_n| \to \infty$.*

The proof of Proposition 1, a particularly nice exercise for students, is left to the reader [**6**, p. 23, problem 7]. In the following I will assume $0 < w < 1$.

In the simplest setting, each of D and S would be linear functions. In this case, (9) has the form $\pi_{n+1} = r\pi_n + s$, for some constants r and s. Again, for $r \neq 0$ the function f defined by $\pi_{n+1} = f(\pi_n) = r\pi_n + s$ is continuous and monotonic, so nothing more complicated than a period-2 oscillation can occur. (More can be said in this linear case [**5**, chapter 1].)

Following Hommes, I will thus assume that while D remains a linear function, the supply function S is nonlinear [**10**]. Set $D(p) = a - bp$, $b > 0$. Recall that S is continuous and monotonically increasing. What else might plausibly be assumed about the function S? If prices are low, the supply would likely increase slowly due to fixed production costs and, perhaps, start-up costs. If prices are high, the supply would again increase slowly due to capacity constraints. It is then reasonable to assume that the graph of S is an S-shaped curve, akin to that sketched in FIGURE 3 but shifted to the right. To that end, I will set

$$S(u) = \frac{1}{1 + e^{-2\lambda(u-1)}}, \quad \lambda > 0.$$

With these choices for D and S, equation (9) simplifies to

$$\pi_{n+1} = (1 - w)\pi_n + \frac{aw}{b} - \frac{w}{b}\frac{1}{1 + e^{-2\lambda(\pi_n - 1)}}. \tag{10}$$

Note the similarities between equations (7) and (10); in particular, (10) has the form $\pi_{n+1} = g(\pi_n)$, where g is the sum of a linear function and the negation of an S-shaped curve. It is natural to expect then that, as parameters are varied appropriately, the asymptotic behavior of the size of queue A and that of the expected price of corn would have much in common. This is indeed the case. Leaving the investigation of the cobweb model to the reader (you might set $w = b$ to simplify (10)), I will illustrate this behavior via the queueing model.

Surprising model dynamics

Terminology Given $\lambda > 0$, consider the one-parameter family of maps

$$f_\alpha : \mathbb{R} \to \mathbb{R}, \quad f_\alpha(x) = x + \alpha - \frac{1}{1 + e^{-2\lambda(x-1)}}, \quad \alpha \in (0, 1). \tag{11}$$

I will set $h(x) = 1/(1 + e^{-2\lambda(x-1)})$, so that $f_\alpha(x) = x + \alpha - h(x)$. Given x_0, let $x_{n+1} = f_\alpha(x_n)$, $n \geq 0$. Of interest is the long-term behavior of the sequence $\{x_n\}$.

Letting f_α^n denote n-fold composition, note that $x_n = f_\alpha^n(x_0)$ (by convention, $f_\alpha^0(x_0) = x_0$). The function f_α^n is called the nth *iterate* of f_α, and the sequence $\{x_n\}$ is the *orbit* of x_0 under f_α. The task can then be rephrased as, "What can be said about the behavior of orbits under f_α?"

If $f_\alpha^n(p) = p$ for some p, where n is the smallest such positive integer, then p is called a *periodic point of period n* or, more simply, a *period-n point*. The orbit of p is a *periodic orbit of period n*, or an *n-cycle*. A period-1 point is called a *fixed point*; periodic orbits provide the simplest types of orbit behavior.

A periodic point p of period n is *attracting* if there is a neighborhood U of p such that, for any $x \in U$, $f_\alpha^{jn}(x) \to p$ as $j \to \infty$. A proof of the fact that p is an attracting period-n point if $|(f_\alpha^n)'(p)| < 1$ can be found in Devaney [**6**].

If p is an attracting fixed point, then there exists an interval of initial conditions whose orbits are asymptotically trivial in that they simply converge to p (FIGURES 7(a), 8(a), 8(d)). Searching for fixed points is often the first step in the study of orbit behavior in discrete dynamical systems.

PROPOSITION 2. *Let f_α be given as in* (11).

(i) *For each $\alpha \in (0, 1)$, f_α has a unique fixed point $x = p_\alpha$.*

(ii) *If* $\lambda > 4$, *let* $\alpha_1 = 0.5(1 - \sqrt{1 - 4/\lambda})$ *and let* $\alpha_2 = 0.5(1 + \sqrt{1 - 4/\lambda})$. *For each* $\alpha \in (0, \alpha_1) \cup (\alpha_2, 1)$, f_α *has an attracting fixed point.*

Proof. (i) Note that $f_\alpha(p) = p$ if and only if $h(p) = \alpha$. It is easy to check that h is continuous and strictly increasing, with $\lim_{x\to-\infty} h(x) = 0$ and $\lim_{x\to\infty} h(x) = 1$. As $\alpha \in (0, 1)$, the equation $h(x) = \alpha$ therefore has a unique solution, which we denote p_α.

(ii) If p_α is the unique fixed point of f_α, then

$$\alpha = h(p_\alpha) = \frac{1}{1 + e^{-2\lambda(p_\alpha - 1)}}, \quad \text{and} \quad e^{-2\lambda(p_\alpha - 1)} = \frac{1 - \alpha}{\alpha}. \tag{12}$$

Using (12) and simplifying a bit yields the result that $f'_\alpha(p_\alpha) = 1 + 2\lambda(\alpha^2 - \alpha)$. Thus, $|f'_\alpha(p_\alpha)| < 1$ is equivalent to $-2 < 2\lambda(\alpha^2 - \alpha) < 0$, or $-1/\lambda < \alpha^2 - \alpha < 0$. Now, $\alpha^2 - \alpha < 0$ for $\alpha \in (0, 1)$ and, for $\lambda > 4$, $\alpha^2 - \alpha > -1/\lambda$ precisely when $\alpha < \alpha_1$ or $\alpha > \alpha_2$. ■

Proposition 2 implies that for $\lambda > 4$ and α sufficiently close to 0 or 1, the queueing system approaches a steady state over time. This means that the size of queue A (as well as that of queue B) approaches a fixed volume, so that the amount of activities in queue A is the same at the beginning of each time period. Put another way, order has descended upon your daily schedule!

What can be said, however, for $\alpha \in [\alpha_1, \alpha_2]$? Though the fixed point $x = p_\alpha$ still exists, it is no longer attracting for α in this range. Searching for period-n points is problematic in that the equation $f_\alpha^n(x) = x$ is difficult to solve. I will thus first proceed with a numerical investigation into orbit behavior.

Let us draw what is called an *orbit diagram*. Fix $\lambda > 4$, and place $\alpha \in (0, 1)$ on the horizontal axis. For each α, select x_0 and plot the asymptotic behavior of the orbit of x_0 under f_α. (For example, compute the first $2m$ elements in the orbit of x_0, and plot numbers $m + 1$ to $2m$; in this article $m = 100$.) Via Proposition 2, the diagram will both begin (α near 0) and end (α near 1) with a curve of attracting fixed points p_α. This implies, in particular, that the orbit diagram for f_α is very much different from the well-known orbit diagram for the logistic family $L_k(x) = kx(1 - x)$ shown in FIGURE 5. Indeed, for $\lambda = 5$, the f_α-orbit diagram is sketched in FIGURE 6.

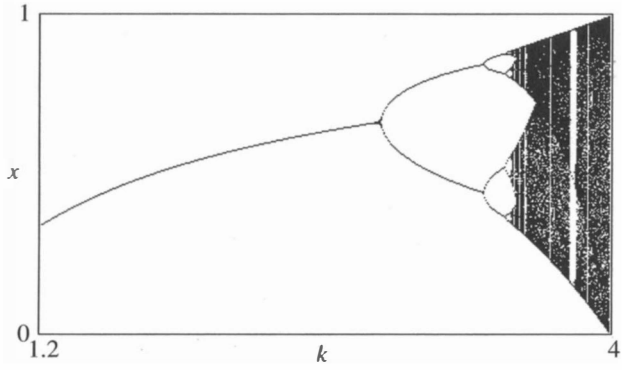


Figure 5 The orbit diagram for $L_k(x) = kx(1 - x)$

Evidently, for $\lambda = 5$ the attracting fixed point for f_α has given way to an attracting 2-cycle for $\alpha \in (\alpha_1, \alpha_2) = (0.27, 0.73)$. The size of queue A thus oscillates between two values for these α. This is somewhat more interesting, but it remains a simple type

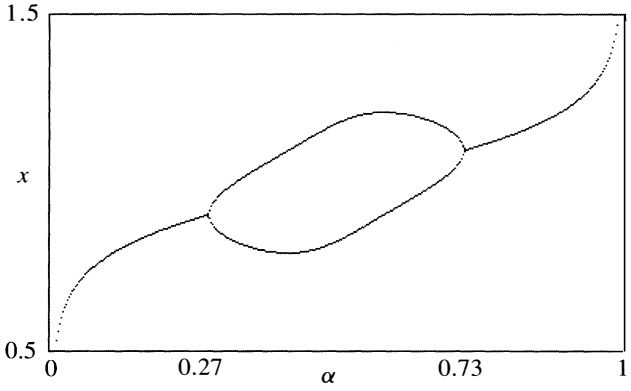


Figure 6 The orbit diagram for f_α with $\lambda = 5$

of asymptotic behavior. Note, however, that the two values in the 2-cycle emerge from the fixed point at $\alpha_1 = 0.27$, spread apart as α increases, and coalesce back to the fixed point at $\alpha_2 = 0.73$. These are examples of *bifurcations* in discrete dynamical systems, which I will investigate before constructing orbit diagrams for other λ-values.

Bifurcations The family f_α appears to undergo a period-doubling bifurcation at $\alpha = \alpha_1$ and a period-halving bifurcation at $\alpha = \alpha_2$. Note that if the derivative at the fixed point satisfies $f'_\alpha(p_\alpha) = -1$ then, by the chain rule, $(f_\alpha^2)'(p_\alpha) = f'_\alpha(f_\alpha(p_\alpha))f'_\alpha(p_\alpha) = (f'_\alpha(p_\alpha))^2 = 1$. This implies that the graph of $f_\alpha^2(x)$ is tangent to the line $y = x$ at $x = p_\alpha$. As the fixed point set for $f_\alpha^2(x)$ corresponds to the intersection of the graphs of $y = f_\alpha^2(x)$ and $y = x$, a small change in α may then result in a change in the number of fixed points of $f_\alpha^2(x)$, that is, of period-2 points for $f_\alpha(x)$.

To prove that this is in fact the case, I will rely on the following fundamental result [**9**, section 3.4]. Property (I) in the theorem is a type of nondegeneracy condition, while (II) relates to the cubic term in the Taylor series expansion of $f_\alpha^2(x) - x$ about the fixed point and determines the stability and direction of the bifurcation of period-2 points.

THEOREM. *Let $f_\mu : \mathbb{R} \to \mathbb{R}$ be a one-parameter family of mappings such that f_{μ_0} has a fixed point x_0 with $\partial f/\partial x = -1$ at (x_0, μ_0). Assume that*

(I) $$\left(\frac{\partial f}{\partial \mu}\frac{\partial^2 f}{\partial x^2} + 2\frac{\partial^2 f}{\partial x \partial \mu}\right) \neq 0 \text{ at } (x_0, \mu_0);$$

(II) $$Q = \left(\frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}\right)^2 + \frac{1}{3}\left(\frac{\partial^3 f}{\partial x^3}\right)\right) \neq 0 \text{ at } (x_0, \mu_0).$$

Then there is a smooth curve of fixed points of f_μ passing through (x_0, μ_0), the stability of which changes at (x_0, μ_0). There is also a smooth curve γ passing through (x_0, μ_0) so that $\gamma - \{(x_0, \mu_0)\}$ is the union of period 2 orbits. The curve γ has quadratic tangency with the line $\mathbb{R} \times \{\mu_0\}$ at (x_0, μ_0).

When the hypothesis of the theorem are satisfied, the family f_μ is said to undergo a *period-doubling bifurcation* at $\mu = \mu_0$. Furthermore, if $Q > 0$ then the period-2 branches in the orbit diagram open to the right at $\mu = \mu_0$, and the corresponding 2-cycles are attracting. The claim here is that the family f_α given in (11) undergoes a period-doubling bifurcation at $\alpha_1 = 0.5(1 - \sqrt{1 - 4/\lambda})$ for $\lambda > 4$.

To prove this, recall that $f'_\alpha(p_\alpha) = 1 + 2\lambda(\alpha^2 - \alpha)$. As is easily checked,

$$\alpha_1^2 - \alpha_1 = -1/\lambda, \tag{13}$$

so $f'_{\alpha_1}(p_{\alpha_1}) = -1$ as desired.

As for hypothesis (I), clearly $\partial f/\partial\alpha = 1$ and $\partial^2 f/\partial x\partial\alpha = 0$ for any (x, α). Thus, if $\partial^2 f/\partial x^2 \neq 0$ at (p_{α_1}, α_1), (I) is satisfied. A direct computation shows that for any (x, α), $\partial^2 f/\partial x^2 = 0$ if and only if $x = 1$. Note that $p_{\alpha_1} = 1$ implies $\alpha_1 = 1/2$ by (12), which in turn implies $\lambda = 4$ by (13), a contradiction. Hence hypothesis (I) is satisfied.

A bit more elbow grease is required to check (II). Using the equalities in (12) and equation (13), a lengthy computation gives $\partial^2 f/\partial x^2 = 4\lambda(2\alpha_1 - 1)$ and $\partial^3 f/\partial x^3 = -8\lambda^2(6\alpha_1^2 - 6\alpha_1 + 1)$ at (p_{α_1}, α_1). The sum Q in condition (II) then simplifies to $\frac{16}{3}\lambda^2(3\alpha_1^2 - 3\alpha_1 + 1)$, which is strictly positive for all values of α_1. This completes the proof of the following.

PROPOSITION 3. *For $\lambda > 4$, the family f_α undergoes a period-doubling bifurcation at $\alpha_1 = 0.5(1 - \sqrt{1 - 4/\lambda})$.*

Note that $Q > 0$ implies that the period-2 branches open to the right at $\alpha = \alpha_1$, precisely as in FIGURE 6.

What can be said about the apparent period-halving bifurcation at $\alpha = \alpha_2$? Fortunately, extensive computation can be avoided in this case. Note the symmetry in the orbit diagram (FIGURE 6) about the point $(x, \alpha) = (1, 0.5)$. This symmetry is indeed present for f_α, as you are invited to check that $f_\alpha(x) = -f_{1-\alpha}(2 - x) + 2$ for any x. This implies, in particular, that there is orbit symmetry as indicated by the following lemma.

LEMMA. *Given $x \in \mathbb{R}$, $\alpha \in (0, 1)$, and $n \geq 0$,* $f_\alpha^n(x) + f_{1-\alpha}^n(2 - x) = 2.$ (14)

Remark. If z is a limit point of the orbit of x_0 under f_α, then by the lemma, $2 - z$ is a limit point of the orbit of $2 - x_0$ under $f_{1-\alpha}$. Hence, if $x_0 = 1$, reflecting the orbit diagram for $\alpha \in (0, 0.5]$ horizontally about the line $\alpha = 0.5$ and vertically about the line $x = 1$ will produce the orbit diagram for $\alpha \in [0.5, 1)$. In particular, if f_α has a period-doubling bifurcation at α_1, then it must have a period-halving bifurcation at $1 - \alpha_1 = \alpha_2$.

Proof of the lemma. The proof proceeds by induction. The $n = 1$ case follows from the simple calculation already mentioned. Suppose (14) holds for some $n \geq 1$, and let $x \in \mathbb{R}$. As $f_\alpha(x) \in \mathbb{R}$, $f_\alpha^n(f_\alpha(x)) + f_{1-\alpha}^n(2 - f_\alpha(x)) = 2$ holds by the inductive hypothesis. By the base case, $2 - f_\alpha(x) = f_{1-\alpha}(2 - x)$. Thus, $f_\alpha^n(f_\alpha(x)) + f_{1-\alpha}^n(f_{1-\alpha}(2 - x)) = 2$, or $f_\alpha^{n+1}(x) + f_{1-\alpha}^{n+1}(2 - x) = 2$. ∎

The logistic family $L_k(x) = kx(1 - x)$ undergoes a sequence of period-doubling bifurcations (FIGURES 5 and 7). As the parameter k increases, an attracting fixed point is followed by a 2-cycle, a 4-cycle, and so on. As can be seen in FIGURE 7 via graphical analysis, the image of an x_0 near the critical point under L_k increases with k. As $L_k(x_0)$ gets nearer to 1, the shape of the graph implies that $L_k^2(x_0)$ is brought back to the left of x_0.

Contrast this phenomenon with that illustrated in FIGURE 8. As α increases, for x_0 near the left critical point, $f_\alpha(x_0)$ indeed increases with α. Due to the bimodal nature of the graph of f_α, however, $f_\alpha^2(x_0)$ need not be less than x_0 for α large enough. Rather than approach a 4-cycle, the orbit of x_0 in the third graph in FIGURE 8 ($\alpha = 0.6$) will approach the attracting 2-cycle to which the right critical point also converges. In the fourth graph ($\alpha = 0.75$) the orbit of x_0 converges to the fixed point.

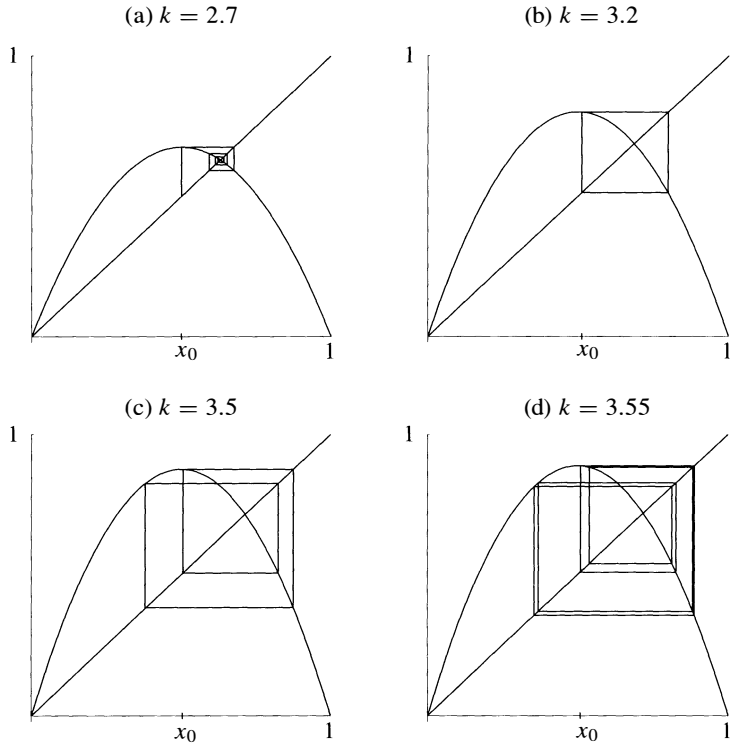


Figure 7 Period-doubling for the logistic family $L_k(x) = kx(1 - x)$

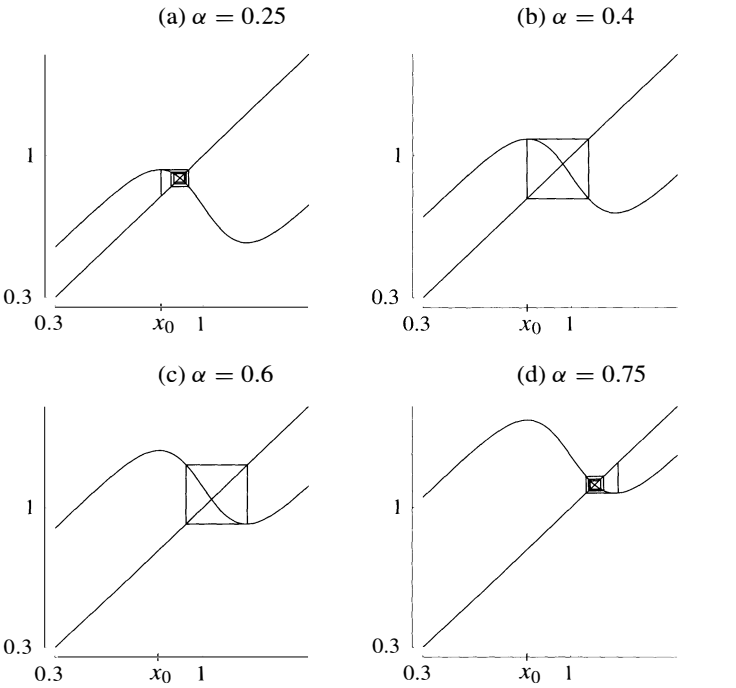


Figure 8 Period-doubling, period-halving for f_α ($\lambda = 5$)

Chaos Does the size of queue A ever vary from either a fixed volume or a period-2 oscillation asymptotically? Provided λ is large enough, the answer is a resounding yes.

Several f_α-orbit diagrams are sketched in FIGURE 9 as λ is increased. Recall that, by symmetry, for each period-doubling bifurcation there is a corresponding period-halving bifurcation. The apparent existence of a 4-cycle and an 8-cycle (seen in FIGURE 9(a)) would imply that both the second iterate f_α^2 and the fourth iterate f_α^4 have undergone period-doubling bifurcations. Consequently, both f_α^4 and f_α^2 also undergo period-halving bifurcations, so that f_α moves from having an 8-cycle to a 4-cycle to a 2-cycle with increasing α.

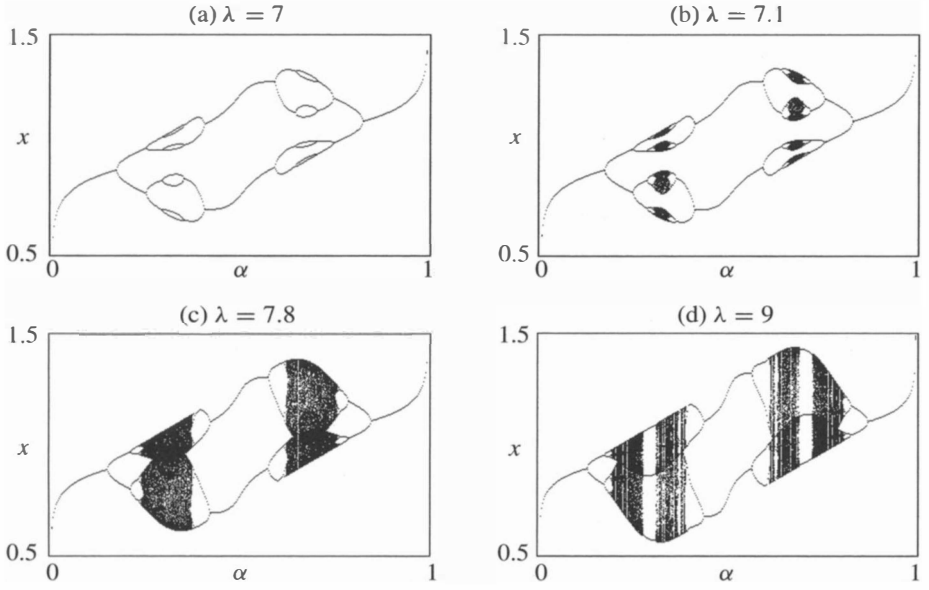


Figure 9 f_α-orbit diagrams for increasing λ

Interestingly, period-doubling bifurcations are also evident in FIGURE 9 for α slightly larger than 0.5 which, by symmetry, imply the existence of period-halving bifurcations for α slightly less than 0.5.

Orbit behavior becomes increasingly complex as λ increases. For λ sufficiently large, the left part of the orbit diagram shows an infinite sequence of period-doubling bifurcations, while an infinite sequence of period-halving bifurcations appears in the right portion. In this case, what can be said for α-values nearer the middle?

Note the clear period-3 windows in FIGURE 9(d). It is a remarkable fact that the mere existence of a 3-cycle for a continuous map $g : I \to I$, I an interval, implies extraordinarily complex orbit behavior. For example, it follows that g must have an n-cycle for *every* integer $n > 0$. Moreover, there is an uncountably infinite subset S of I containing points whose orbits are not asymptotically periodic and on which g exhibits *sensitive dependence on initial conditions* [**6**, section 1.8]. These bounded, nonperiodic orbits, displaying sensitivity to initial conditions, comprise what Li and Yorke first called *chaos* in 1975 [**11**].

More technically, for (α, λ) pairs for which f_λ has a 3-cycle as in FIGURE 9(d), the Li–Yorke definition of chaos implies that:

(i) For each $n \geq 0$, there exists a period-n point.

(ii) There is an uncountable set S and an $\epsilon > 0$ such that

(a) For every $x, y \in S$, $x \neq y$,

$$\limsup_{n\to\infty} |f_\alpha^n(x) - f_\alpha^n(y)| > \epsilon \quad \text{and} \quad \liminf_{n\to\infty} |f_\alpha^n(x) - f_\alpha^n(y)| = 0,$$

(b) For every $x \in S$ and any periodic point y,

$$\limsup_{n\to\infty} |f_\alpha^n(x) - f_\alpha^n(y)| > 0.$$

In these parameter ranges, f_α thus has an infinite number of periodic points with distinct periods. Property (ii)(a) implies that f_α both stretches and folds the interval under iteration, with the stretching accounting for the sensitive dependence on initial conditions. Condition (ii)(b) implies that the orbit of any $x \in S$ does not converge to a periodic orbit (such orbits are called *aperiodic*).

Strikingly different models in the limit I will briefly highlight one significant difference between the cobweb (10) and queueing (7) models, letting the reader place the working out of most of the details in her or his queue. I will set $w = b$ in (10) for this discussion. Recall that, for $u \neq 0$, the function ϕ_2 in (5) converges to ϕ_1 (FIGURE 3) as $\lambda \to \infty$, that is, as the steepness of the S-shaped curve grows without bound. For the cobweb model this leads to consideration of the 2-parameter family

$$g = g_{a,w} : \mathbb{R} \to \mathbb{R}, \quad g(x) = \begin{cases} (1-w)x + a, & x < 1 \\ (1-w)x + a - 1, & x \geq 1, \end{cases} \quad w \in (0, 1).$$

As g is piecewise linear, with $1 - w \in (0, 1)$ so that each piece of g contracts, it is reasonable to expect the asymptotic behavior of g-orbits to be trivial. One can show this is indeed true for $a - w \geq 1$ or $a - w \leq 0$, as for these parameter values there exists a point $p = p(a, w)$ such that for any $x \in \mathbb{R}$, $g^n(x) \to p$ as $n \to \infty$ (p is a fixed point unless $a = w$; this is most easily seen if you draw a few graphs and use graphical analysis). Interestingly, for $a - w \in (0, 1)$ the story is quite different.

Let $a - w \in (0, 1)$ and let $I = [a - w, a - w + 1)$. Then $g : I \to I$, and for any $x \in \mathbb{R}$ there exists $k \geq 0$ with $g^k(x) \in I$ (see FIGURE 10(a)). It suffices then to consider orbits for $x_0 \in I$. Note that the discontinuity allows for the possibility of recurrent or cyclic behavior as some points move first right and then left under iteration, although g does contract on each linear piece. This type of function, called a *contracting interval exchange transformation*, has only recently been studied [**2**]. It turns out

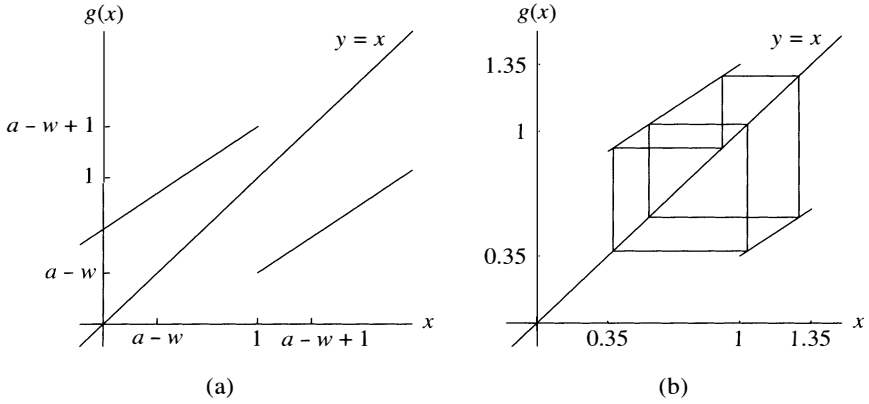


Figure 10 The function $g(x)$ for $a - w \in (0, 1)$

that, if w is fixed, then for almost all a (with respect to Lebesgue measure), there is a unique periodic orbit that attracts all other orbits [**3**]; a period-5 attractor is sketched in FIGURE 10(b) for $a = 0.65$, $w = 0.3$.

Contrast this with the family

$$f = f_\alpha : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} x + \alpha, & x < 1 \\ x + \alpha - 1, & x \geq 1, \end{cases} \quad 0 < \alpha < 1$$

(letting $\lambda \to \infty$ in (7)). Note that each linear piece of f neither contracts nor expands. Letting $J = [\alpha, \alpha + 1)$, as above one can show $f : J \to J$ and all f-orbits land in J. By identifying the endpoints of J, f can be viewed as a rotation by α on a circle of circumference 1. This leads nicely into the theory of the dynamics of circle homeomorphisms, an appropriate and rich topic for advanced undergraduates [**13**]. For the map f above, if $\alpha = p/q \in \mathbb{Q}$, with p and q having no common factors, then every orbit is periodic with period q. For α irrational, not only is there no periodic behavior, but the closure of every orbit (the union of the orbit and its limit points) equals J. What a lovely contrast to the dynamics of the map g with $a - w \in (0, 1)$!

Discussion

There is no denying that these two simple models exhibit dynamical behavior that belies their simplicity. Through introduction of one of these models into a modeling or dynamical systems course, students can see classical dynamics topics such as periodic and aperiodic behavior, bifurcations, chaos, and even the dynamics of piecewise linear interval maps, all motivated by the real world. They will also be exposed to dynamical similarities and differences between unimodal and bimodal maps.

To return to the modeling aspect, it is true that much of queueing theory concerns probabilistic models, as opposed to deterministic models such as (7). However, deterministic queueing models are increasingly used to estimate the gross behavior of queueing systems [**7**]. In these cases the arrival and/or departure process is treated as a continuous fluid rather than a discrete time system flow. The queueing model in [**7**] and investigated here is an example of a sampled fluid model for a simple dynamically routed closed queueing network [**4, 14**].

The cobweb model is a fundamental model in economic dynamics. Though the model originally focused on linear supply and demand functions, more recent attention has been given to monotonic but nonlinear supply and demand. It is not clear if this very simple cobweb model is capable of realistically explaining the price fluctuations in an independent market [**10**]. Given the limited types of behavior exhibited by the cobweb model with linear supply and demand, the nonlinear version does exhibit vastly many more types of asymptotic behavior. This indicates that the nonlinear cobweb model has more potential for application; indeed, refinement of this model continues [**8**].

If you are interested in a discrete-time model that evolves into iteration of a nonunimodal map, either for your own investigations or for use in the classroom, one yielding rich dynamical behavior, this exemplary model will serve you well.

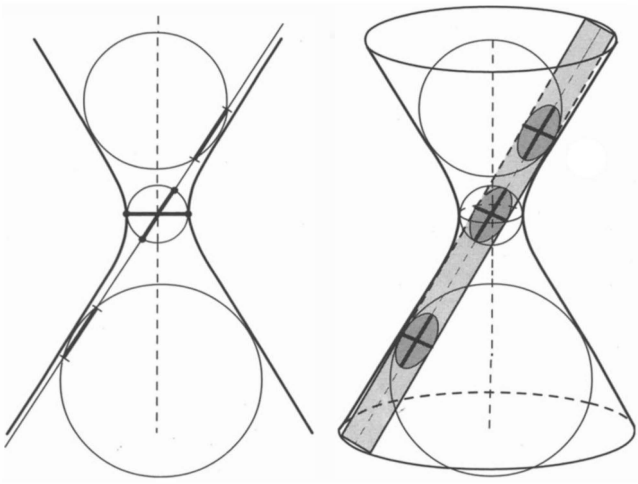
REFERENCES

1. E. Beltrami, *Mathematics for Dynamic Modeling*, Academic Press, San Diego, CA, 1987.
2. J. Brémont, Contracting interval exchange transformations, preprint.
3. Y. Bugeaud and J.-P. Conze, Dynamics of some contracting linear functions modulo 1, in *Noise, Oscillators and Algebraic Randomness* (Chapelle des Bois, 1999), M. Planat (ed.), Springer-Verlag, New York, NY, 2000.

4. C. Chase, J. Serrano and P. Ramadge, Periodicity and chaos from switched flow systems: Contrasting examples of discretely controlled continuous systems, *IEEE Transactions on Automatic Control* **38** (1) (1993), 70–83.
5. G. Cobb et al, *Laboratories in Mathematical Experimentation*, Springer-Verlag, New York, NY, 1997.
6. R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd edition, Westview Press, Boulder, CO, 2003.
7. G. Feichtinger, C. Hommes, and W. Herold, Chaos in a simple deterministic queueing system, *ZOR – Mathematical Methods of Operations Research* **40** (1994), 109–119.
8. J. Goeree and C. Hommes, Heterogeneous beliefs and the nonlinear cobweb model, *J. of Economic Dynamics & Control* **24** (2000), 761–798.
9. J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York, NY, 1983.
10. C. Hommes, Dynamics of the cobweb model with adaptive expectations and nonlinear supply and demand, *J. of Economic Behavior and Organization* **24** (1994), 315–335.
11. T.-Y. Li and J. Yorke, Period three implies chaos, *Amer. Math. Monthly* **82** (10) (1975), 985–992.
12. J. Sandefur, *Discrete Dynamical Systems Theory and Applications*, Clarendon Press, Oxford, UK, 1990.
13. J.A. Walsh, The dynamics of circle homeomorphisms: a hands-on introduction, this MAGAZINE **72** (1) (1999), 3–13.
14. J.A. Walsh, G.R Hall and B. Elenbogen, Computer protocol and torus maps, *Dynamics and Stability of Systems* **11** (3) (1996), 239–263.

Proof Without Words: Surprising Property of Hyperbolas

For each member of the family of hyperbolas with the same pair of vertices, every circle tangent to both branches intersects each asymptote along a chord of constant length, regardless of the location of the circle and the angle between the asymptotes. This constant length is the distance between vertices.



—TOM M. APOSTOL AND MAMIKON A. MNATSAKANIAN
Project Mathematics!
253-37 CALTECH
PASADENA, CA 91125

Spectral Analysis of the Supreme Court

BRIAN L. LAWSON
Department of Political Science
University of Cincinnati
Cincinnati, OH 45221
brian.lawson@uc.edu

MICHAEL E. ORRISON
Harvey Mudd College
Claremont, CA 91711
orrison@hmc.edu

DAVID T. UMINSKY
Boston University
Boston, MA 02215
duminsky@math.bu.edu

Introduction

Imagine a survey in which 110 people were given a set $\{A, B, C, D, E\}$ of five items, and were asked to vote for their two favorite items from the set. The data has been presented to you as vector $f \in \mathbb{R}^{10}$ where

$$f = \begin{bmatrix} 2 \\ 12 \\ 11 \\ 6 \\ 17 \\ 8 \\ 4 \\ 24 \\ 20 \\ 6 \end{bmatrix} \begin{matrix} AB \\ AC \\ AD \\ AE \\ BC \\ BD \\ BE \\ CD \\ CE \\ DE \end{matrix}$$

with the numbers corresponding to the votes for the unordered pairs on the right. For example, two people voted for items A and B, while 24 people voted for items C and D.

A relatively new starting point for analyzing this sort of data is (generalized) *spectral analysis*. This is a nonmodel-based approach to the exploratory analysis of data associated with sets, like the set of unordered pairs above, that have a fair amount of symmetry. It was initially pioneered by Diaconis in [**1, 2**]. See also [**4**].

Should you decide that spectral analysis is worth looking into (as we hope to convince you), then you will be happy to know that there are efficient algorithms for doing spectral analysis [**5, 6, 7**]. Perhaps more interestingly, at least from a mathematical perspective, these algorithms involve an intriguing mixture of ideas and techniques from linear algebra, abstract algebra, numerical analysis, and graph theory.

The focus of this paper is the linear algebraic framework in which the spectral analysis of voting data like that above is carried out. As we will show, this framework can be used to pinpoint voting coalitions in small voting bodies like the United States Supreme Court. Our goal is to show how simple ideas from linear algebra can come together to say something interesting about voting. And what could be more simple than where our story begins—with counting.

From counting to orthogonal subspaces

There are 110 people involved in the survey above. The sum of the entries in the vector f is therefore 110, so the average number of votes given to each of the ten pairs of items is simply $110/10 = 11$. In the long run, this information may or may not be useful, but it seems like a reasonable place to start. After all, if each of the entries in f had been near the average, we could summarize the data by saying that each pair seems just as likely to have been chosen.

After computing the average, our next step might be to compute the number of times an individual item, such as A or C, was chosen. This would help us to see if there was an item that was particularly popular or unpopular, regardless of with which item it was paired. Since we would want to do this for each item, we could use a matrix for this calculation:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 12 \\ 11 \\ 6 \\ 17 \\ 8 \\ 4 \\ 24 \\ 20 \\ 6 \end{bmatrix} \begin{matrix} AB \\ AC \\ AD \\ AE \\ BC \\ BD \\ BE \\ CD \\ CE \\ DE \end{matrix} = \begin{bmatrix} 31 \\ 31 \\ 73 \\ 49 \\ 36 \end{bmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \tag{1}$$

This matrix-vector product shows, for example, that A was chosen $2 + 12 + 11 + 6 = 31$ times, while C was chosen $12 + 17 + 24 + 20 = 73$ times. For reference, we will refer to the 5×10 matrix used in (1) as T_1. In other words,

$$T_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

But why stop with one matrix? For example, the average could have been computed using the matrix

$$T_0 = \begin{bmatrix} \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{bmatrix}$$

since the product of T_0 and our data vector is precisely the sum of the entries in the vector divided by 10. In fact, we could even construct the matrix T_2 that computes the number of times that each pair was chosen. Of course, this turns out to just be the 10×10 identity matrix since the data was originally defined in terms of pairs of items. Nonetheless, the matrices T_0, T_1, and T_2 seem to be just the ticket when it comes to counting. But wait, there's more!

If we define N_i to be the nullspace of T_i, then we get the chain

$$N_0 \supset N_1 \supset N_2.$$

By definition, however, all of the nonzero counting information is actually contained in the orthogonal complements of the N_i, which are also the row spaces of the T_i. This leads to the "complementary" chain

$$N_0^{\perp} \subset N_1^{\perp} \subset N_2^{\perp}.$$

This chain of subspaces makes sense. After all, if we know how many times every pair was chosen, we can figure out the number times each individual item was chosen. This information can then be used to compute the average.

We will refer to the 1-dimensional subspace $N_0^\perp$ as the *mean effects* space; it contains all of the information needed to compute the average. The 5-dimensional subspace $N_1^\perp$ is the *first order effects* space; it contains all of the information needed to compute the number of times a particular item was chosen. Lastly, the 10-dimensional subspace $N_2^\perp$ is the *second order effects* space; it is simply the original vector space containing all of the information associated with pairs of items.

Given the chain $N_0^\perp \subset N_1^\perp \subset N_2^\perp$ of subspaces, we could next ask about the effect that each subspace has on our data vector f. For example, we could compute the parts of f that are contained in $N_0^\perp$, $N_1^\perp$, and $N_2^\perp$. Because these subspaces form a chain, however, it is more instructive to compute the parts of f that are introduced as we move up the chain, i.e., to compute what we need to build the vector f as we move from $N_0^\perp$ to $N_1^\perp$ to $N_2^\perp$. This gives rise to an orthogonal decomposition

$$\mathbb{R}^{10} = M_0 \oplus M_1 \oplus M_2$$

of the original vector space, where $N_0^\perp = M_0$, $N_1^\perp = M_0 \oplus M_1$, and $N_2^\perp = M_0 \oplus M_1 \oplus M_2$. The vector f can therefore be written uniquely as a sum $f = f_0 + f_1 + f_2$ where $f_i \in M_i$. For our example, we have

$$f = \begin{bmatrix} 2 \\ 12 \\ 11 \\ 6 \\ 17 \\ 8 \\ 4 \\ 24 \\ 20 \\ 6 \end{bmatrix} \quad f_0 = \begin{bmatrix} 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \end{bmatrix} \quad f_1 = \begin{bmatrix} -26/3 \\ 16/3 \\ -8/3 \\ -7 \\ 16/3 \\ -8/3 \\ -7 \\ 34/3 \\ 7 \\ -1 \end{bmatrix} \quad f_2 = \begin{bmatrix} -1/3 \\ -13/3 \\ 8/3 \\ 2 \\ 2/3 \\ -1/3 \\ 0 \\ 5/3 \\ 2 \\ -4 \end{bmatrix} \begin{matrix} AB \\ AC \\ AD \\ AE \\ BC \\ BD \\ BE \\ CD \\ CE \\ DE \end{matrix}$$

The 1-dimensional space M_0 is still just the mean effects space. The 4-dimensional subspace M_1, however, can be thought of as the space of *pure* first order effects, since we have removed the mean effects contained in $N_0^\perp$. Likewise, the 5-dimensional subspace M_2 can be thought of as the space of *pure* second order effects, since we have removed the mean and first order effects from $N_2^\perp$.

Now that we have isolated f_0, f_1, and f_2, the next step might be to explore more deeply the way in which these vectors contribute to the data vector f. We could begin by comparing the squared norms of the f_i. One reason for doing this lies in the fact that, since the M_i are orthogonal to each other, we know that

$$||f||^2 = ||f_0||^2 + ||f_1||^2 + ||f_2||^2.$$

By comparing the squared norms, we can therefore get a sense for where the data is concentrated. In our case, $||f_0||^2 = 1210$, $||f_1||^2 \approx 422.67$, and $||f_2||^2 \approx 53.33$. Now the norm of f_0 captures nothing more than the number of people voting. The relatively large size of f_1 in comparison to f_2, however, suggests that the first order effects are contributing heavily to this data.

Before we attempt to pinpoint which item or items from the set $\{A, B, C, D, E\}$ might actually be contributing to f_1, notice that even though $\dim M_1 = 4$, there are five natural effects to consider, namely the individual effects of each of the five items. In other words, there are too many items to just find a basis vector in M_1 for each and to

then write f_1 in terms of that basis. As noted in [1], however, there is a straightforward way around this that makes use of inner products.

For each item x, consider the function g_x which is defined on the unordered pairs of items, and whose value at a pair is 1 if x is in the pair, and 0 otherwise. These functions correspond to the rows of T_1. We can project each of these functions into M_1, normalize the projection, and then compute their inner products with a normalized version of the projection f_1. The resulting numbers, all of which are between -1 and 1, measure how much the directions of f_1 and the components of the g_x in M_1 agree. For our data vector f, this approach leads to the numbers

A	B	C	D	E
−0.41	−0.41	0.91	0.16	−0.25

which suggest that the respondents in the survey really liked item C but were slightly averse to choosing items A and B. Indeed, a quick glance back at the original data confirms this.

We could also compute similar inner products for the pure second order effects. Here the natural functions to consider correspond to the original pairs and the rows of T_2, with a 1 in just one position and zeros elsewhere. The resulting numbers

AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
−0.06	−0.84	0.52	0.39	0.13	−0.06	0.00	0.32	0.39	−0.77

suggest, for example, that the 12 votes for the pair $\{A, C\}$ are due mostly to C's popularity, not the popularity of the pair.

Now for such a small data set, it may seem as though we went to a lot of trouble to end up only saying that "people seem to really like item C." In fact, you may have already come to that conclusion when you first saw the data, or after you saw how many people chose a pair containing C. The point we want to make is that this approach applies to any survey in which people are asked to choose their top k items from a list of size n. In fact, if we assume that $0 \leq k \leq n/2$ (we will ask them to choose their least favorites if we must), then we get an orthogonal decomposition

$$M = M_0 \oplus M_1 \oplus \cdots \oplus M_{k-1} \oplus M_k$$

where M is the underlying $\binom{n}{k}$-dimensional vector space of real-valued functions defined on the k-element subsets of an n-element set. The subspace M_i captures the *pure ith order effects* of the voting data. The projections of a data vector $f \in M$ into each of the M_i can be computed with the hope of uncovering hidden large-scale structure. Subsequent inner product calculations can then lead to the uncovering of hidden small-scale structure.

As we will see in the next section, this simple approach to untangling survey data can also be applied to voting data that arises when committees vote "yea" or "nay" on several issues. The trick, perhaps to the delight of committee members everywhere, is to let the issues do all of the voting!

From surveys to the Supreme Court

When the members of a committee are asked to vote "yea" or "nay" on an issue, and none of the members abstain from the vote, the result is a splitting of the committee into two groups—the winners (or majority) and the losers (or minority). Once we know

either group, we automatically know the other, so for convenience, we will focus on the minority members for each issue.

Now although the committee members are really doing all of the voting, we can turn the tables by pretending as though the issues are actually voting on the subset of members that it wants to make up the minority when it comes before the committee. In this way, we can use the issues and their "votes" to try to pinpoint coalitions in the committee. Moreover, to analyze the resulting data, we separate it into different functions, one for each of the possible number of members in the minority. We then analyze each of the functions using the techniques described in the earlier sections. (See [3] for more details.)

As a proof of concept, consider the well-studied nine member "committee" of justices on the United States Supreme Court, say from 1994 to 1998. For each case (issue) in which there are no abstentions, there can be zero, one, two, three, or four justices that form the minority. We limit our analysis to non-unanimous cases, and for the Supreme Court buffs out there, we have also limited our analysis to the cases in which a signed opinion was issued. Our data comes from the database maintained by Spaeth [8], and the results of our analysis are summarized in Table 1.

TABLE 1: Rehnquist Court 1994–1998, 192 non-unanimous cases

split	subspace	norm^2	four largest (using absolute value) inner products							
8–1	M_1	703	St	.996	Br	−.165	Gi	−.165	Sc	−.125
7–2	M_2	183	ThSc	.732	StGi	.476	StBr	.354	StSc	−.330
7–2	M_1	59	St	.695	O'	−.452	So	−.348	Ke	−.348
6–3	M_3	72	ReThSc	.647	O'ThSc	.345	BrThSc	−.321	ReO'Sc	−.290
6–3	M_2	105	ThSc	.626	StGi	.345	GiTh	−.309	ReTh	.237
6–3	M_1	16	Th	.656	Sc	.540	Ke	−.501	Br	−.270
5–4	M_4	316	StGiBrSo	.954	KeReThSc	.344	O'ReThSc	.341	StBrSoSc	−.269
5–4	M_3	199	StBrSo	.379	StGiSo	.368	GiBrSo	.293	StGiBr	.293
5–4	M_2	360	StSo	.315	StBr	.301	ThSc	.282	StGi	.265
5–4	M_1	22	Ke	−.646	So	.418	O'	−.380	Gi	.380

St Stevens Gi Ginsburg Br Breyer So Souter Ke Kennedy
O' O'Connor Re Rehnquist Th Thomas Sc Scalia

As we will see, after looking at the results in the table, it is easy to go back and find the information that supports it. We want to stress, however, that finding the most important coalitions in a committee would be labor intensive and unsystematic without something like spectral analysis. This would be especially true if you were starting with raw data and knew essentially nothing about the committee members. Moreover, as described in [5, 6], by using a combination of ideas and techniques from introductory courses in linear algebra, abstract algebra, numerical analysis, and graph theory, the results presented in Table 1 can be computed *very* efficiently.

Cases with 8–1 splits There were 37 cases which split 8–1. The lone dissenter was Stevens 29 times, Thomas wrote 3 such dissents, neither Ginsburg nor Breyer ever

dissented on their own, and the remaining five justices each wrote one lone dissent. Not surprisingly, the first line of the table show that the 8–1 data points strongly in the direction of Stevens dissenting.

Cases with 7–2 splits There are first and second order effects for the cases which split 7–2. In this case, the squared norm of the projection onto M_2 is 183, while the squared norm of the projection onto M_1 is 59. This portion of the data is therefore dominated by the pure second order effect.

The pure second order effect points in the direction of Thomas-Scalia dissenting, but also has noticeable components in the Stevens-Ginsburg and Stevens-Breyer directions. The pure first order effect, although not as strong as the pure second order effect, points in the direction of Stevens dissenting.

In this portion of the data, there are 48 cases with 7–2 splits, where 11 of them are Thomas-Scalia, 9 are Stevens-Ginsburg, and 8 are Stevens-Breyer dissenting. Of the 48 cases with pairs dissenting, Stevens is a dissenter in 24 of them. In addition to Ginsburg and Breyer, Stevens dissents with Souter, Kennedy, and Thomas.

Cases with 6–3 splits For the 6–3 cases, the second order effect is the largest with a squared norm of 105. The largest inner product for the pure second order effects corresponds to Thomas-Scalia dissenting. Moreover, the other projections for the 6–3 cases also point in this general direction.

The largest pure first order effect is Thomas dissenting, and the next largest is Scalia dissenting. The two largest third order effects are Thomas-Scalia dissenting joined first by Rehnquist, and then by O'Connor. The negative sign on the Breyer-Thomas-Scalia triple suggests that these three justices seldom dissent together in a 6–3 split.

Again, when we examine the data we find that this is a good summary. Of the 44 cases with a 6–3 split, Thomas and Scalia join together in dissent in 22 of them. They are joined ten times by Rehnquist, seven times by O'Connor, three times by Kennedy, and once each by Souter and Stevens.

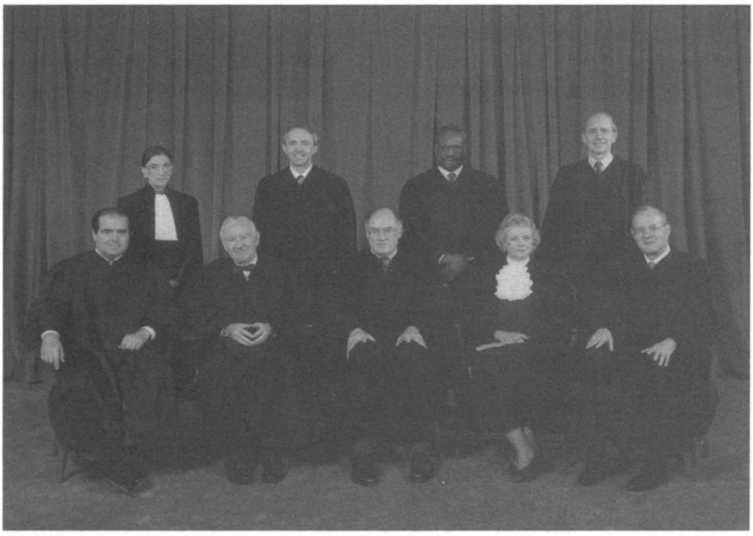
Cases with 5–4 splits The pure second order effect is the largest for the 5–4 cases with a squared norm of 360. The five largest inner products for the pure second order effect are relatively close to each other. They correspond to "liberal" dissents (Stevens-Souter, Stevens-Breyer, and Stevens-Ginsburg) and "conservative" dissents (Thomas-Scalia and Rehnquist-Scalia, the latter of which is not in the table).

The pure fourth order effect is the next largest with a squared norm of 316. The pure fourth order effect is in the "liberal" dissenting direction (Stevens-Ginsberg-Breyer-Souter) with smaller components in the "conservative" dissenting directions (Rehnquist-Thomas-Scalia-Kennedy and Rehnquist-Thomas-Scalia-O'Connor). This seems to fit the data well. For the 63 cases with 5–4 splits, 28 cases are dissents by Stevens-Ginsberg-Breyer-Souter, eight are Rehnquist-Thomas-Scalia-O'Connor, and seven are Rehnquist-Thomas-Scalia-Kennedy.

Now although the pure first order effects are the smallest, they have an interesting interpretation. The "swing" voter Kennedy has a large significant negative value ($-.646$) suggesting that he rarely ends up in the minority in 5–4 splits, which is what would be expected of swing voters. The case for O'Connor, while smaller, is similar.

Conclusion

Spectral analysis is a powerful tool for doing exploratory data analysis, and given that efficient algorithms for doing this type of analysis exist **[5, 6, 7]**, political scientists,



Credit: Collection, The Supreme Court Historical Society.

economists, and market research analysts seem to have every incentive to include it in their arsenal. As we noted in the introduction, spectral analysis was initially pioneered by Diaconis in [**1, 2**]. The interested reader is strongly encouraged to delve into these sources, both of which are teeming with tantalizing open questions and deep ideas.

Acknowledgment. We gratefully acknowledge a Harvey Mudd College Beckman Research Award which supports research by undergraduates with faculty. We would like to thank the Charles Phelps Taft Fund at the University of Cincinnati. Special thanks also to Jon Jacobsen and Francis Su for helpful comments and suggestions.

REFERENCES

1. P. Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics, Hayward, CA, 1988.
2. P. Diaconis, *A generalization of spectral analysis with application to ranked data*, Ann. Statist. **17** (1989), no. 3, 949–979.
3. M. Iwasaki, *Spectral analysis of multivariate binary data*, J. Japan Statist. Soc. **22** (1992), no. 1, 45–65.
4. J. Marden, *Analyzing and modeling rank data*, Monographs on Statistics and Applied Probability, vol. 64, Chapman & Hall, London, 1995.
5. D. Maslen, M. Orrison, and D. Rockmore, *Computing isotypic projections with the Lanczos iteration*, SIAM Journal on Matrix Analysis and Applications **25** (2004), no. 3, 784–803.
6. M. Orrison, *An eigenspace approach to decomposing representations of finite groups*, Ph.D. thesis, Dartmouth College, 2001.
7. D. Rockmore, *Some applications of generalized FFTs*, Groups and computation, II (New Brunswick, NJ, 1995), Amer. Math. Soc., Providence, RI, 1997, pp. 329–369.
8. H. Spaeth, UNITED STATES SUPREME COURT JUDICIAL DATABASE, 1953–1997 TERMS [Computer file]. 9th ICPSR version. East Lansing, MI: Michigan State University, Dept. of Political Science [producer], 1998. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1999.

Counting Train Track Layouts

JAMES D. BEAMAN
Gurnee, IL 60031
jbeaman1983@gmail.com

ERIN J. BEYERSTEDT
Tulane University
New Orleans, LA 70118
ebeyerstedt@math.tulane.edu

MARK R. SNAVELY
Carthage College
Kenosha, WI 53140-1994
snavely@carthage.edu

Young Brian Snavely has a train set with many kinds of track: straight pieces, curves, bridges, tunnels, X-pieces (to make figure-8s), and switches. Brian likes them all, but it was the switches that led us to an interesting mathematical problem. Switches are the structures that allow Engineer Brian to choose between two (or more) paths to guide the train along. Without switches, you either make a circuit or a one-way path; when switches are included, potential layouts become fascinatingly complex. In this paper, we will study tours on a layout that start at a point on the layout and move in only one direction.

A variety of types of switches are available for purchase, as shown in FIGURE 1. We found 2-way switches, 3-way switches, and 5-way switches. For our study, we use n-way switches where $n \in \mathbb{N}$, as we can manufacture an n-way switch by attaching switches together. In FIGURE 1, we demonstrate how to make a 6-way switch out of a 3-way switch and three 2-way switches.

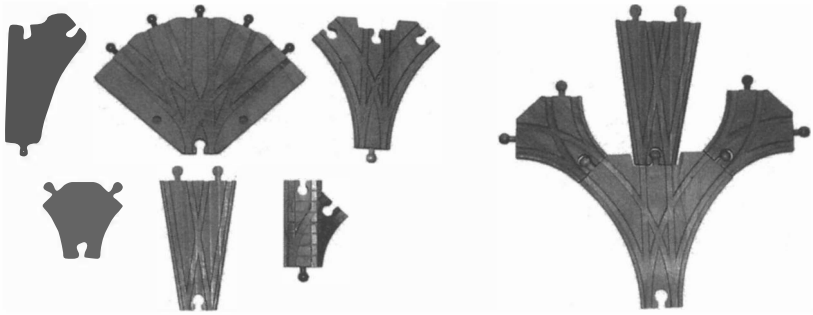


Figure 1 Different Types of Switches

Our question is this: How many different layouts can be made from a track set that has exactly one n-way and one m-way switch? There are actually two different, but reasonable, interpretations of the word *different*, one combinatorial, the other more topological. We consider them both and deliver some answers, using such tools as two-term recurrences, directed graphs, and eigenvalues of incidence matrices. We only present a few of the proofs completely, assuming that the interested reader will be able to work through any details we omit.

Train track terminology

Before we delve into the study of track layouts, we introduce terminology to describe some basic structures. First, we call every point on a piece of track to which other pieces can be attached a *node*. A standard piece of track has two nodes, one at each end. An *n-way switch* is a piece of track with $n + 1$ nodes that can be separated into two sets: the n *options* and one *bottom*. The bottom is the unique node on the switch with the property that, if we enter the switch through the bottom, we exit the switch through one of the n options. The options are the nodes with the property that if we enter the switch through an option, we exit the switch through the bottom. The most important aspect of a switch is that if a train enters the switch through an option, it *must* leave through the bottom, thus limiting the paths a train can traverse. A *length of track* is a part of the layout that attaches one node on a switch to a node on another (not necessarily different) switch.

A *layout* is a collection of switches with track connected to the nodes on the switches. A layout is *connected* if an engine driving on the layout can travel (moving forward or backward) from any point on the layout to any other point on the layout.

Dead ends, or *sidings*, on a layout are not interesting, so we will consider layouts with as few sidings as possible. If we use exactly two n-way switches, the total number of nodes in the layout will be even and there will be no sidings. However, if the total number of nodes is odd, there will be one node that we cannot attach to any other node, and thus we are left with a siding.

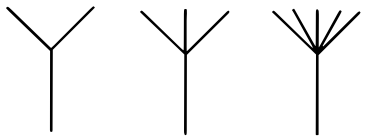


Figure 2 Switch representations

A *layout diagram* is a sketch of a setup using arrows to represent lengths of track, with switches represented as follows: We draw an n-way switch as a fork with n tines, always showing the bottom pointing downward. FIGURE 2 shows representations of 2-way, 3-way, and 5-way switches. Notice again that if a train enters a switch through one of the options, it *must* exit the switch through the bottom. We label each length of track, and give each an arbitrary orientation. If a length of track is labeled p, then traveling along that segment of track in the direction of the arrow is denoted p, while $\overline{p}$ denotes traveling in the opposite direction. FIGURE 3 shows two layouts and their corresponding layout diagrams. The reader may wish to imagine some tours on each layout.

To create a layout diagram, we begin with a given number of switches and connect every node to another node in the layout. This results in a variety of types of connections, which we name according to their function or appearance, as shown in FIGURE 4.

- A connection from one option on a given switch to another option on that same switch will be called a *turnaround*.
- A connection from one option on a given switch to an option on a different switch will be called a *bridge*.
- A connection from an option on a given switch to the bottom of the same switch will be called an *ear*.

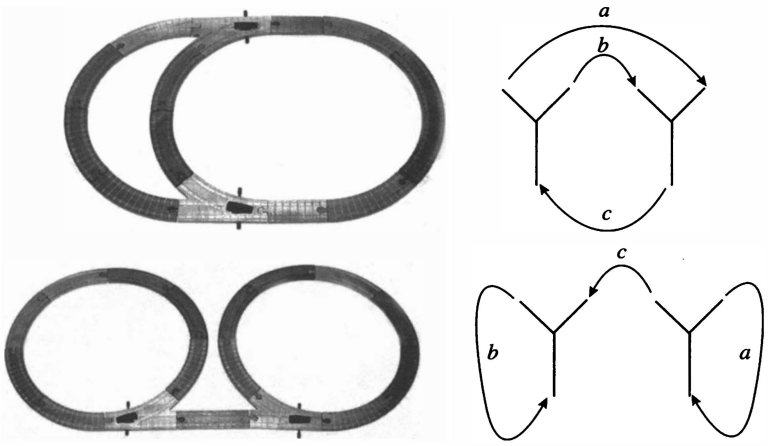


Figure 3 Layouts with their layout diagrams (2 bridges and a link; 1 bridge and 2 ears)

- A connection from the bottom of a given switch to an option on a different switch will be called a *snake*.
- A connection from the bottom of a given switch to the bottom of a different switch will be called a *link*.

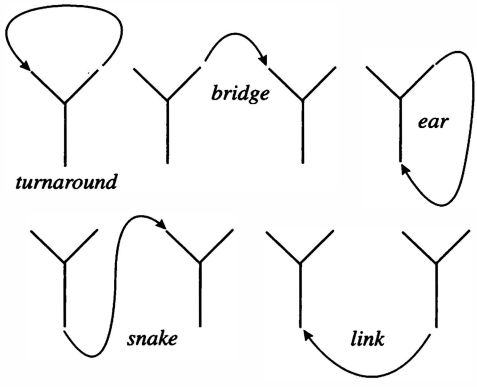
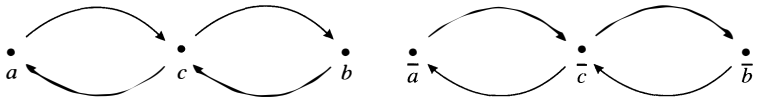


Figure 4 Layout connections

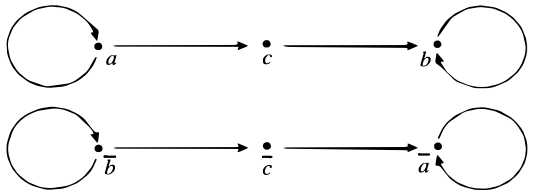
Tour digraphs and matrices Using a layout diagram as a reference, we construct a digraph (directed graph) to represent the structure of the layout, and from that construct an adjacency matrix. We call the digraph the *tour digraph* for a layout, and we call the adjacency matrix the *tour matrix.*

To construct the digraph, let each vertex represent a length of track traversed in a specific direction. If p is a length of track on a layout diagram, p generates *two* vertices in the tour digraph: one for p and another for $\overline{p}$. We connect two vertices v_1 and v_2 with a directed edge if one can drive from track segment v_1 to track segment v_2 through a switch. The tour digraphs for the layouts in FIGURE 3 are shown in FIGURE 5.

These two examples seem to indicate that the tour digraphs are always disconnected. This is not the case. FIGURE 6 demonstrates that if a layout allows us to turn around, the digraph is connected.



The digraph for the upper layout



The digraph for the lower layout

Figure 5 The digraphs for the layouts in FIGURE 3

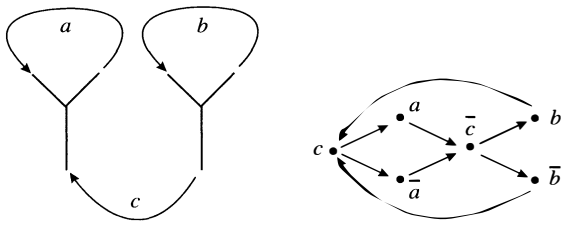


Figure 6 A layout with a connected digraph

From the digraph, we construct the tour matrix M by letting

$$M_{i,j} = \begin{cases} 1 & \text{if there is an edge in the tour digraph from } v_i \text{ to } v_j, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

For clarity, we include the vertex labels on the left side of the matrix. The tour matrices for the upper and lower layouts in FIGURE 3 are

$$M_u = \begin{matrix} a \\ \bar{a} \\ b \\ \bar{b} \\ c \\ \bar{c} \end{matrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad M_l = \begin{matrix} a \\ \bar{a} \\ b \\ \bar{b} \\ c \\ \bar{c} \end{matrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

respectively.

Counting the layouts

How many different layouts exist if we use two 2-way switches? That depends on what we mean by *different*. There are $\lceil (n+m+2)/2 \rceil$ lengths of track in a two-switch layout containing an n-way switch and an m-way switch, where $\lceil x \rceil$ is the ceiling function. Indeed, there are $n+m+2$ nodes, and each length of track is connected to two of them, unless $n+m+2$ is odd. In that case, there will be one node left over and we have a siding connected to only one node. This implies that if we have

a layout with an n-way switch and an m-way switch, the tour digraph will have $v = 2\lceil (n + m + 2)/2 \rceil$ vertices, and the tour matrix will be have v rows and v columns.

As an example, let us count the total number of ways to construct a layout using two 2-way switches. The layout contains three lengths of track connected to six nodes. There are $\binom{6}{2} = 15$ ways to connect the first piece of track to nodes. Similarly, there are $\binom{4}{2} = 2$ ways to connect the second piece of track to nodes, and only one way to connect the last piece. Because the first two pieces of track are placed in order, we divide by 2 to learn that there are 15 distinct layouts. However, many of these are in fact equivalent in some sense. FIGURE 7 shows two layouts where only the roles of the nodes and some orientations are reversed. Note that arrows in the layout may intersect, because we have overpasses to allow such crossings.

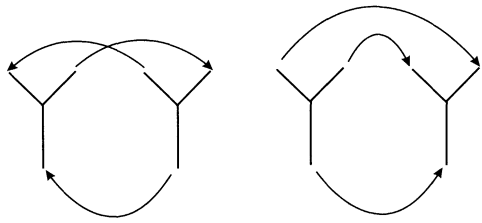


Figure 7 Two equivalent layouts

The constructions developed earlier will aid us in defining one notion of equivalent layouts. Recall from graph theory [**3**, page 7] that two digraphs G_1 and G_2 are *isomorphic* if there is a one-to-one correspondence f between the vertices of G_1 and the vertices of G_2, such that there is an edge in G_1 from v_1 to v_2 if and only if there is an edge in G_2 from $f(v_1)$ to $f(v_2)$. In essence, this means that the graphs are identical except for the placement and labeling of the vertices and edges. We use this concept to develop one definition of equivalent layouts.

DEFINITION. Two layouts are *tour equivalent* if their corresponding tour digraphs are isomorphic.

Using an exhaustive search, we found that there are 5 distinct layouts using two 2-way switches. We have already seen three of them in FIGURES 3 and 6. The remaining two are shown in FIGURE 8.

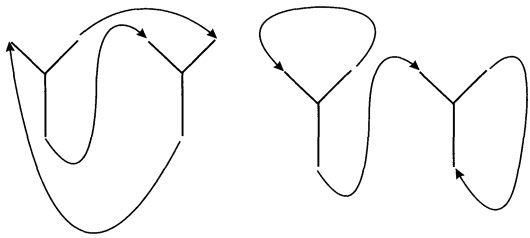


Figure 8 The remaining layouts using two 2-way switches

How many distinct layouts can we construct using two 3-way switches? For convenience, we let $L(n, m)$ represent the number of distinct layouts we can construct using an n-way switch and an m-way switch. We already know that $L(2, 2) = 5$.

Consider a layout using two three-way switches. If there is at least one bridge, remove one of the bridges to obtain a layout with two 2-way switches. Thus, the layouts using two 3-way switches that have at least one bridge are in one-to-one correspondence with the layouts using two 2-way switches. Therefore,

$$L(3, 3) = L(2, 2) + (\text{the number of 3-way switch layouts with no bridges}).$$

How many layouts using two 3-way switches have no bridges? Suppose that a layout has a link. Both bottom nodes are used, leaving three unused nodes on each switch, necessitating at least one bridge. If we have exactly one ear in the layout, one bottom node is unused and we must also have a snake in the layout. Once again, we are left with an odd number of unused nodes on each switch, leading to a bridge. A similar argument demonstrates that we cannot have exactly one snake. Thus, to have no bridges in the layout, we must have either two snakes or two ears, and there are only two such layouts. These are shown in FIGURE 9. Notice that the layout with two ears is not connected, and that we *do* allow this. In general, if n is any odd number, there are two layouts using two n-way switches with no bridges: one has two ears, and the other has two snakes.

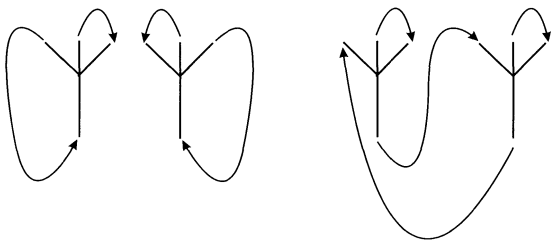


Figure 9 The layouts using two 3-way switches having no bridges

Moving on to layouts with two n-way switches where n is even, an argument similar to the above shows that we need to determine the number of layouts using two n-way switches with no bridges. Again there are only two. This time, one has an ear and a snake, and the other has a link and the rest of the nodes connected by turnarounds.

We have learned that in all cases,

$$L(n + 1, n + 1) = L(n, n) + 2.$$

We already know that $L(2, 2) = 5$, so a straightforward argument shows that $L(n, n) = 2n + 1$. We have verified the first part of the following theorem.

THEOREM. *Let $L(n, m)$ be the number of distinct track layouts using an n-way switch and an m-way switch with $n \geq m \geq 2$. Then*

$$L(n, m) = \begin{cases} 2m + 1 & \text{if } n = m; \\ 2m + 2 & \text{if } n - m > 0 \text{ and } n - m \text{ is even;} \\ 6m + 4 & \text{if } n - m > 1 \text{ and } n - m \text{ is odd;} \\ 6m + 3 & \text{if } n - m = 1. \end{cases}$$

Rather than proving this theorem in detail, we will describe the ideas behind the proof. If $n - m$ is even, only one layout does not have at least $(n - m)/2$ turnarounds on the n-way switch. This layout has all nodes on the m-way switch connected to options on the n-way switch, an ear on the n-way switch, and $(n - m)/2 - 1$ turnarounds

on the n-way switch. In all other layouts, counting the number of options on the n-way switch reveals that we must have at least $(n - m)/2$ turnarounds on that switch. Removing those turnarounds from the n-way switch leaves a layout using two m-way switches, so these layouts match up in one-to-one correspondence with the layouts using two m-way switches, of which there are $2m + 1$. Thus, there are of $2m + 2$ layouts using an n-way switch and an m-way switch if $n - m$ is even.

Why are there so many more layouts if $n - m$ is odd? The answer is *sidings*. We claim that, if $m > 2$,

$$L(n, m) = L(n - 1, m) + L(n, m - 1) + L_B(n, m) \tag{1}$$

where $L_B(n, m)$ is the number of layouts using an n-way switch and an m-way switch that have a siding on the bottom of one of the switches. The first two terms in the sum in (1) correspond to layouts having the siding on an option of the n-way switch and the m-way switch respectively. We leave it to the reader to verify that $L_B(3, 2) = 6$, and then show that if $n - m \geq 1$ is odd,

$$L_B(n, m) = L_B(m + 1, m) = L_B(m, m - 1) + 2.$$

From this we learn that $L_B(n, m) = 2m + 2$.

Thus

$$L(n, m) = (2m + 2) + 2m + (2m + 2) = 6m + 4$$

if $n - 1 \neq m \geq 2$, and one fewer if $n - 1 = m$.

The case where $n - m > 1$ is odd and $m = 2$ is unique and must be handled separately. In the previous cases, removing a siding led to a layout already counted. When one option on the two-way switch is a siding, this does not happen. There are four layouts in this situation, and examples of these for $n = 5$ can be found in the fourth and fifth rows of FIGURE 10. Because we use the orientation of a track segment primarily to produce the tour digraph and matrix, arrows are omitted from these layout diagrams.

Using this fact, we compute

$$L(n, 2) = L(n - 1, 2) + 4 + L_B(n, 2) = 6 + 4 + 6 = 16$$

if $n \neq 3$ and one fewer if $n = 3$. The 16 layouts using a 5-way switch and a 2-way switch are shown in FIGURE 10. Except for a few details, we have proved our theorem.

The tour matrix

Let us now turn our attention to the tour matrix for layouts with an n-way switch and an m-way switch, assuming no sidings on the layout. In particular, we will determine the eigenvalues of the tour matrix. The reason we do not consider layouts with sidings is that sidings tend to introduce dead ends (which produce zero eigenvalues) and they do not add interest to the layout.

This analysis can exclude a matrix from the set of possible tour matrices. If the eigenvalues of the matrix do not fit one of the forms described below, that matrix cannot be the transition matrix for a train track layout. In addition, the specific eigenvalues present give us insights into the structure of the layout. For instance, an eigenvalue of 1 indicates the presence of a length of track that a train can traverse twice consecutively. The presence of -1 as an eigenvalue signifies a two-cycle in the layout: two lengths of track that together make a cycle and can be traversed indefinitely.

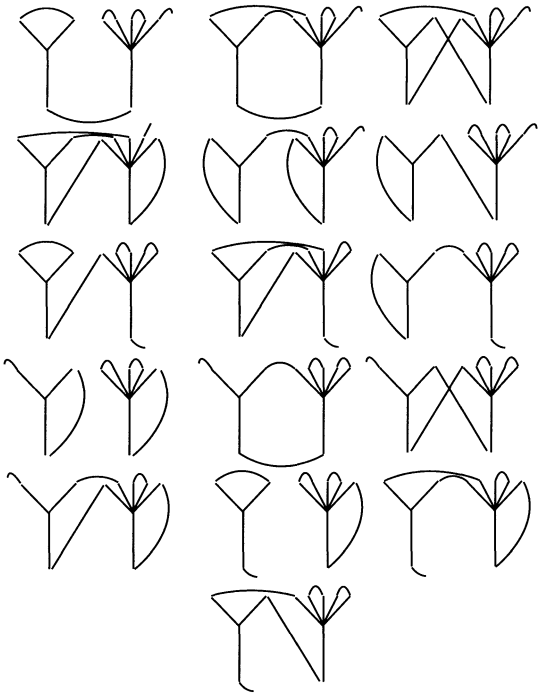


Figure 10 The 16 layouts using a 5-way switch and a 2-way switch

In this part of our study, the layouts with exactly two switches fall into four categories: layouts with one ear and one snake, layouts with two ears, layouts with two snakes, and layouts with a link. These categories are disjoint, and every two-switch layout falls into one of them.

A summary of our results is presented in TABLE 1. When a link is present, the eigenvalues depend on the number of bridges, r, and the number of turnarounds on each switch, t_1 and t_2. Note that in the last line of the chart, $r + t_1 + t_2 = \lfloor (n+m)/2 \rfloor$; where $\lfloor x \rfloor$ is the standard floor function, and that r, t_1, and t_2 may be zero.

TABLE 1: The eigenvalues of the tour matrix

Structure	*Characteristic Polynomial*	*Nonzero Eigenvalues*
one ear, one snake	$\lambda^{n+m}(\lambda - 1)^2$	1, 1
two ears	$\lambda^{n+m-2}(\lambda - 1)^4$	1, 1, 1, 1
two snakes	$\lambda^{n+m-2}(\lambda - 1)^2(\lambda + 1)^2$	1, 1, −1, −1
one link, r bridges, t_1 turnarounds on one switch, t_2 turnarounds on the other	$\lambda^{n+m-2}[\lambda^4 - 2r\lambda^2 + (r^2 - 4t_1t_2)]$	$\sqrt{r - 2\sqrt{t_1t_2}}$, $-\sqrt{r - 2\sqrt{t_1t_2}}$, $\sqrt{r + 2\sqrt{t_1t_2}}$, $-\sqrt{r + 2\sqrt{t_1t_2}}$

The computations necessary to obtain these results can be somewhat tedious, so let us analyze some specific layouts to illustrate the techniques and leave the generaliza-

tions to interested readers. The idea in the first few cases is to attain a simple form in the tour matrix by ordering vertices as we might encounter them on a tour.

Consider a layout with one ear, like the layout in FIGURE 11.

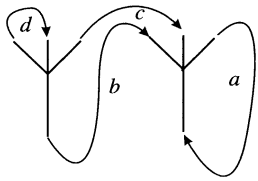


Figure 11 A layout with one ear

We order the vertices in the tour matrix by putting a in the first position, then $\overline{b}$ and $\overline{c}$, then d and $\overline{d}$, then b, c, and finally $\overline{a}$. With the vertices in this order, the tour matrix is

$$M = \begin{matrix} a \\ \overline{b} \\ \overline{c} \\ d \\ \overline{d} \\ b \\ c \\ \overline{a} \end{matrix} \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and M has one eigenvalue 1 with multiplicity 2, and the other eigenvalue 0 with multiplicity 6. A similar technique can be used to verify the result for a layout having two ears.

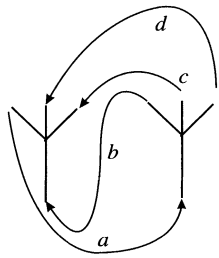


Figure 12 A layout with two snakes

To find the eigenvalues of the tour matrix for a layout with two snakes (FIGURE 12 shows an example), we order the vertices as we see them on a tour that begins with a snake.

$$M = \begin{matrix} a \\ b \\ c \\ d \\ \overline{c} \\ \overline{d} \\ \overline{b} \\ \overline{a} \end{matrix} \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

In this form, the eigenvalues can be computed immediately: 1 appears with multiplicity 2, as does -1, and all remaining eigenvalues are 0.

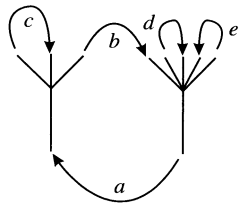


Figure 13 A layout with a link

As the formula suggests, links complicate the analysis considerably. Consider the layout in FIGURE 13. We begin ordering the vertices by placing b, $\overline{b}$, a, and $\overline{a}$ in the first four positions. Then we list the rest of the vertices in no particular order. Putting the remaining vertices in tour order is not necessary. This yields the tour matrix

$$M = \begin{array}{c} b \\ \overline{b} \\ a \\ \overline{a} \\ c \\ \overline{c} \\ d \\ \overline{d} \\ e \\ \overline{e} \end{array} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & 1 & 0 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Focus on the entries shown in bold. To find the eigenvalues of M, we will use row and column operations to eliminate the 1s to the right of this block. We conjugate by shear matrices (matrices with 1s on the diagonal and 0s elsewhere, except for a single 1 off the diagonal), which subtracts one row from another, then adds corresponding columns. Performing this demonstrates that M is similar (actually similar over the integers) to the matrix

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{4} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

which has $\pm\sqrt{1+2\sqrt{2}}$ and $\pm\sqrt{1-2\sqrt{2}}$ as its eigenvalues.

In a more general setting, the vertices representing the link and one of the bridges are listed first. If there are no bridges, choose one turnaround (one direction only) from each switch. The rest of the vertices are then listed in any order. If we have r bridges in the layout, t_1 turnarounds on one switch, and t_2 turnarounds on the other, then the

tour matrix is similar to a matrix with the 4×4 block

$$\begin{pmatrix} 0 & 0 & r & 2t_1 \\ 0 & 0 & 2t_2 & r \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

in the upper left corner and 0s in all columns beyond the fourth. This matrix has the eigenvalues claimed in TABLE 1.

A topological approach

One of us noticed that some layouts that were not tour equivalent seemed very much the same. For example, consider the lower layout in FIGURE 3 and the layout on the left in FIGURE 8. In both cases, if we choose a point on the layout and start driving an engine on the track, either we are stuck in a cycle we cannot exit, or we travel a cycle we can exit until we exit it, at which point we find ourselves stuck in a cycle we cannot exit. Topologically, these layouts have have similar properties, and resemble in nature the layout using one 3-way switch shown in FIGURE 14. Because we are not generating digraphs or matrices in this analysis, we do not need to give each segment of track an orientation.

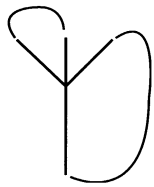


Figure 14 A layout using one 3-way switch

Are there operations that can determine equivalence in this sense? Are there other layouts that are equivalent to these? To answer such questions, we define five operations on track diagrams.

Contracting sidings: If there is a siding on a layout, remove it. Sidings do not add any cycles to a layout.
Rearranging options: We may permute the connections to the options on a switch. All options are essentially equivalent.
Contracting snakes: At the beginning of this paper, we described how to manufacture a 6-way switch from a 3-way switch and three 2-way switches. If we have a snake from the bottom of an n-way switch to an option on an m-way switch, we can remove that length of track and attach the n-way switch to the m-way switch, making one $(m + n - 1)$-way switch. This transforms the layout from a two-switch layout to a one-switch layout.
Identifying ears: If two switches (an n-way and an m-way) both have ears, they are joined only by bridges (perhaps zero bridges if the layout is not connected) and no link. We can fold the layout diagram over, placing one switch on top of the other. Each existing turnaround from either switch remains on the new combined switch. Each bridge becomes a turnaround. This also transforms the layout from a two-switch layout to a one-switch layout.

Eliminating redundant turnarounds: The discussion of the eigenvalues of the tour matrix indicates that turnarounds only add to the complexity of the layout if there is a link. Otherwise, the turnarounds only move us from one cycle to another and do not add to the number of cycles present in the layout. Therefore, *on one switch layouts only*, we remove all turnarounds except one.

DEFINITION. Two layouts are *homeomorphic* if their layout diagrams can be transformed to the same diagram using the five operations outlined above.

If two layouts are tour equivalent, they are homeomorphic. Using these operations, we see immediately that the layout diagrams for the lower layout in FIGURE 3 and the layout on the left in FIGURE 8 can be transformed into the diagram in FIGURE 14, and they are therefore homeomorphic. How many distinct layouts use an n-way switch and an m-way switch, under this definition of equivalence? Because we eliminate all sidings, we only discuss layouts with no sidings.

Consider any two-switch layout with no links, so there are only bridges, turnarounds, snakes, and ears. If there is exactly one snake, contract it. There has to be an ear on the other switch, so now we have a one-switch layout with an ear. Eliminate the extra turnarounds, and we have the diagram in FIGURE 14. If there are two snakes, contract one of them. The other snake has become an ear on the remaining switch. Eliminate the extra turnarounds and we again obtain the diagram in FIGURE 14. If there are no snakes, then each switch has an ear, so perform an ear identification and eliminate the extra turnarounds. In every case, the process ends with the diagram in FIGURE 14. This argument can be modified to show that *every* layout with no links, regardless of the number of switches, is homeomorphic to the layout described in FIGURE 14. This is one key insight in the proof of the following result.

THEOREM. *There are* $\lfloor m/2 \rfloor + 1$ *homeomorphism classes when we use an* n*-way switch and an* m*-way switch with* $2 \leq m \leq n$.

Let us sketch the proof: We already know that if a layout does not have a link, it is homeomorphic to the layout shown in FIGURE 14. How many distinct two-switch layouts exist with links? A study of the tour equivalence classes tells us that there are two such layouts with two 2-way switches, and two such layouts using two 3-way switches. Similarly, there are three distinct layouts using two 4-way or 5-way switches. In general, there are $\lfloor n/2 \rfloor$ distinct layouts that have a link using two n-way switches, and thus $\lfloor n/2 \rfloor + 1$ distinct homeomorphism classes. A similar argument shows that there are $\lfloor m/2 \rfloor + 1$ homeomorphism classes when we use an n-way switch and an m-way switch.

Future work There are many unanswered questions that our research group would like to address in the future. First, if we continue our study of tour equivalence, how many distinct layouts are there if we add more switches to the layout? We found 34 distinct layouts using three 2-way switches. Of these, 24 are connected. We hope to continue our analysis to include larger numbers of switches with more options.

We also intend to explore the notion of homeomorphic layouts for larger layouts. How many homeomorphism classes are there for layouts with more switches? Of the 24 connected layouts involving three 2-way switches, 10 have a siding on the bottom. Of the remaining 14, all but three are homeomorphic to the layout in FIGURE 14 using the operations described in this paper. Should some classes of layouts with links be considered homeomorphic? If so, what operation(s) on the track diagrams will allow us to determine the homeomorphism classes? Carthage student Ben Burch is approaching this question by treating train track layouts as dynamical systems and examining the symbol sequences generated by taking tours on the layout. He intends to find con-

ditions on the symbol sequences that produce a definition of equivalence that agrees with the notion of homeomorphism in the known cases.

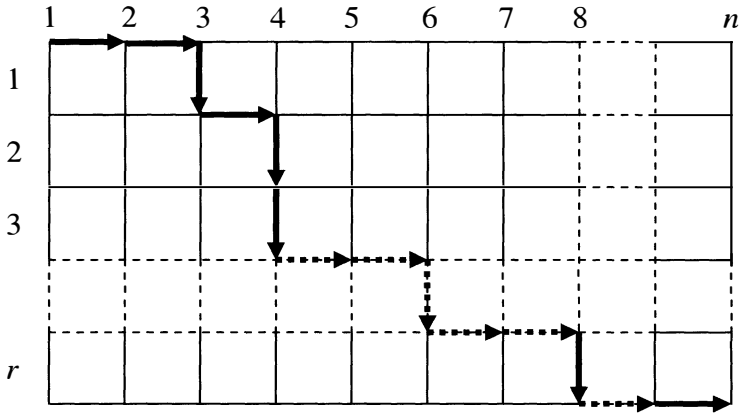
Acknowledgment. The authors would like to thank Brian Snavely, whose love of Thomas the Tank Engine provided his dad with the idea for this research project.

REFERENCES

1. D. Lay, *Linear Algebra, 2nd edition*, Addison-Wesley, Reading, MA, 2000.
2. M. Newman, *Integral Matrices*, Academic Press, New York, NY, 1972.
3. D. West, *Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, NJ, 1996.

Proof Without Words: The Number of Unordered Selections with Repetitions

THEOREM. *The number of unordered selections of r objects chosen from n types with repetitions allowed is* $\binom{n-1+r}{r}$, *the same as the number of paths of length* $n-1+r$ *from top-left to lower-right in the diagram.*



Selection 3, 4, 4, . . . , 6, . . . , 8

—DEREK CHRISTIE
WAIKATO INSTITUTE OF TECHNOLOGY
HAMILTON, NEW ZEALAND
DEREK.CHRISTIE@WINTEC.AC.NZ

Dials and Levers and Glyphs, Oh My! Linear Algebra Solutions to Computer Game Puzzles

JESSICA K. SKLAR
Pacific Lutheran University
Tacoma, WA 98447-0003
sklarjk@plu.edu

Let the games begin!

I am a gaming geek. Not necessarily the kind of gaming geek you might expect a mathematician to be: I get beaten at chess by 12-year-olds, and I refuse to even go near the game of Go. The few times I've tried role-playing games, my mismanagement of my posse invariably caused my men to starve to death before they even encountered a monster. No, my games of choice are computer adventure games. *Myst*, *Riven*, *The Longest Journey*: you know the type. Lush visuals. Extravagant music. Marvelous machines, strange creatures, often haunting storylines, and the best part: puzzles. Ah, the glory of cracking codes, of damming rivers, of finally getting that hovercraft to work!

So you can imagine my excitement when during a cold and gloomy winter vacation, as I sat curled up on the couch with one of my beloved games, I got the following email from a student I'd taught in the fall. Chris wrote: "I was playing [*Myst*] and it had a riddle. Suppose you have 3 rows of numbers... Also, suppose there are 2 levers... Anyway, the game asked me to give a certain combination of numbers... Rather than just play the game, I wanted to play math."

My heart started beating faster.

He continued: "This reminded me of... our dihedral groups, but not exactly. Just messing around, I did find that every combination of levers had an inverse." Keep in mind that the puzzle Chris was describing was not obviously mathematical (at least, to a non-mathematician). It wasn't solvable by elementary arithmetic or geometry: its mathematical solution required the use of linear and abstract algebra. Upon reading this email, I glowed with pride: here was my student, from my very first abstract algebra class, recognizing that the puzzle's levers could be thought of as elements of a group. I immediately flashed back to a similar puzzle my friend, Vanessa, and I had encountered in a game. In that case, Vanessa was the one to interpret the problem mathematically. All of a sudden she and I were madly solving a system of equations and—ta da!—our puzzle was solved.

Mulling over these two puzzles, I began to wonder: were there other puzzles like that out there? The answer was a resounding **yes**. In this paper, I'll explore merely a few of the myriad linear and abstract algebra problems that masquerade as lever, dial, and slide puzzles in computer games. I urge you to play along at home as you read. My descriptions of situations will likely be more understandable that way; plus, hearing chimes that indicate you've solved a puzzle can be much more exciting than just reading about such phenomena.

Clock arithmetic

Let's start by discussing the puzzle in *Myst* about which Chris was writing. Near the beginning of the game, you encounter a contraption in a clock tower, which focuses

mightily on the number 3. The contraption consists, in part, of three levers; from left to right, we'll call these levers A, B, and C. The first time we encounter the contraption, we also see three number 3s facing us, vertically aligned. Basic experimentation with the levers shows us that, in fact, each number 3 marks a face of a "dial" which has exactly two other faces, marked with the numbers 1 and 2, respectively. So we have three dials, each with three faces, marked with the numbers 1, 2, and 3. Here is where a math-friendly person might get excited: regarding these numbers as integers modulo 3, the set of numbers on each dial can be identified with the set $\mathbb{Z}_3 = \{0, 1, 2\}$ (where a dial's number 3 is identified with 0 in $\mathbb{Z}_3$). Since there are three dials, what we have represented here is the set $\mathbb{Z}_3^3$. Let us identify the numbers facing us at any moment on the top, middle, and bottom dials with, respectively, the first, second, and third coordinates of an element in $\mathbb{Z}_3^3$: so, for instance, if the numbers facing us at a certain time read 2, 1, 3 from top to bottom, we'll identify this situation with the vector (2, 1, 0).

Now, if we are clever, we can deduce from facts learned elsewhere in the game that we must somehow rotate these dials so that the numbers facing us, from top to bottom, are 2, 2, and 1 (the proof of this is non-mathematical, and is left to the reader). The mathematical question is: how can this (efficiently) be done?

Well, first, more careful experimentation is needed. Let's say we start with the dials in the positions associated with vector (x_1, x_2, x_3) in $\mathbb{Z}_3^3$. It is not hard to discover that pulling lever A leaves the top dial alone, while rotating the middle and bottom dials so that their resulting visible faces show numbers that are 1 more (modulo 3) than the numbers they previously showed. That is, the resulting situation will be associated with vector $(x_1, x_2 + 1, x_3 + 1)$, where addition is done modulo 3. Lever B, on the other hand, leaves the bottom dial alone, while rotating the top two dials so that (x_1, x_2, x_3) becomes $(x_1 + 1, x_2 + 1, x_3)$. Finally, Lever C merely resets all the dials to their initial position, with three 3s facing us. Lever C is merely an aid so that struggling puzzle-solvers can start from scratch, and thus we may essentially ignore it for the rest of this discussion.

We are now at a point where we can translate this problem entirely into mathematical terms. What we in fact have going on here is a group action: in particular, the action of a subgroup of $\mathbb{Z}_3^3$ on $\mathbb{Z}_3^3$ via left translation. Pulling lever A adds (0, 1, 1) to any element of $\mathbb{Z}_3^3$, while pulling lever B adds (1, 1, 0) to any element. We begin with element (0, 0, 0) in $\mathbb{Z}_3^3$, and want to pull each of levers A and B a particular number of times so that we obtain the element (2, 2, 1): that is, so that we add (2, 2, 1) to (0, 0, 0). This corresponds to writing (2, 2, 1) as a linear combination

$$(2, 2, 1) = \lambda_1(0, 1, 1) + \lambda_2(1, 1, 0),$$

where $\lambda_1, \lambda_2 \in \mathbb{Z}^+$, and where our addition takes place in $\mathbb{Z}_3^3$; we can then pull lever A repeatedly λ_1 times and lever B repeatedly λ_2 times to solve the puzzle. (Notice that it does not matter in which order we pull the levers, as $\mathbb{Z}_3^3$ is an abelian group.) Thus, our immediate goal is to find positive integral solutions to this system of congruences modulo 3:

$$\begin{aligned} \lambda_2 &\equiv 2 \pmod 3 \\ \lambda_1 + \lambda_2 &\equiv 2 \pmod 3 \\ \lambda_1 &\equiv 1 \pmod 3. \end{aligned}$$

Which is all well and good, except that this system clearly has no solutions, since if $\lambda_1 \equiv 1 \pmod 3$ and $\lambda_2 \equiv 2 \pmod 3$, then $\lambda_1 + \lambda_2$ must of necessity be congruent to 0 (mod 3).

So we return to the puzzle and mess with the levers some more; or, if we are impatient, we consult a walk-through of the game. Either way, one will discover the following sneaky fact: holding down lever A or lever B for a beat adds (in $\mathbb{Z}_3^3$) the vector (0, 1, 0) to the currently represented vector; furthermore, copies of (0, 1, 0) are continually added as long as the lever is held down. Thus, by holding lever A or B down for an appropriate amount of time, one can add the vector $(0, \lambda_3, 0)$ (for any $\lambda_3 \in \mathbb{Z}^+$) to the currently represented vector. Thus, our revised mathematical goal is to find positive integers λ_1, λ_2 and λ_3 so that in $\mathbb{Z}_3^3$ we have

$$(2, 2, 1) = \lambda_1(0, 1, 1) + \lambda_2(1, 1, 0) + \lambda_3(0, 1, 0).$$

So the system of congruences for which we really need to find positive integral solutions is

$$\begin{aligned} \lambda_2 &\equiv 2 \pmod 3 \\ \lambda_1 + \lambda_2 + \lambda_3 &\equiv 2 \pmod 3 \\ \lambda_1 &\equiv 1 \pmod 3. \end{aligned}$$

It is easy to see that this system has solution

$$\lambda_1 = 1, \quad \lambda_2 = 2, \quad \text{and} \quad \lambda_3 = 2$$

(where these solution values are unique modulo 3). Turning to the game, we first make sure the dials are in their initial position (pulling lever C to put them there, if need be). Then we can solve the puzzle by pulling lever A once and lever B twice, and holding B down after its second pull exactly long enough to add the vector (0, 1, 0) twice. And voilà, we hear a grinding noise and the gear at the bottom of the contraption opens. Moreover, we have now gained access to a book that will allow us to travel a different world. Nice, huh?

A stasis gun and skulls

We next turn to one of my favorite games, *Timelapse*. A number of puzzles in this game can be interpreted mathematically; we focus on two such puzzles. The first puzzle that we'll discuss is the second of those two that we encounter in the game: it is the stasis tube gun puzzle. (You should likely save your game before exploring the mechanisms of this puzzle, as if you don't solve it quickly enough you will be conquered by a robot and lose the game: not good.) The puzzle consists of six tricolored circles and six yellow triangles. (See Figure 1.) In the center of the puzzle, there is a small, red hexagonal button. At first, clicking on the button merely makes a small noise; the button appears to be currently inactive. Let C_1 be the circle in the upper right-hand corner of the puzzle, and let $C_2, C_3, \ldots, C_6$ be the puzzle's other circles, in clockwise order from C_1. Next, for $i = 1, 2, \ldots, 6$, let T_i be the triangle nearest to C_i. Each circle is divided into red, green and blue sectors, of equal size, in clockwise order in the circle. We'll say that a circle is in state 0 (respectively, states 1 or 2) if its red (respectively, blue or green) sector faces the center of the puzzle. We will soon see that clicking on any of the triangles rotates several of the circles clockwise by multiples of 120°; the set of all orientations of the circles can therefore be identified with a subset of $\mathbb{Z}_3^6$, where the ith entry of an element $x \in \mathbb{Z}_3^6$ is the state of C_i in orientation x of the circles. When you first encounter the puzzle, the orientation of the circles corresponds to the vector $(1, 0, 2, 1, 0, 2) \in \mathbb{Z}_3^6$.

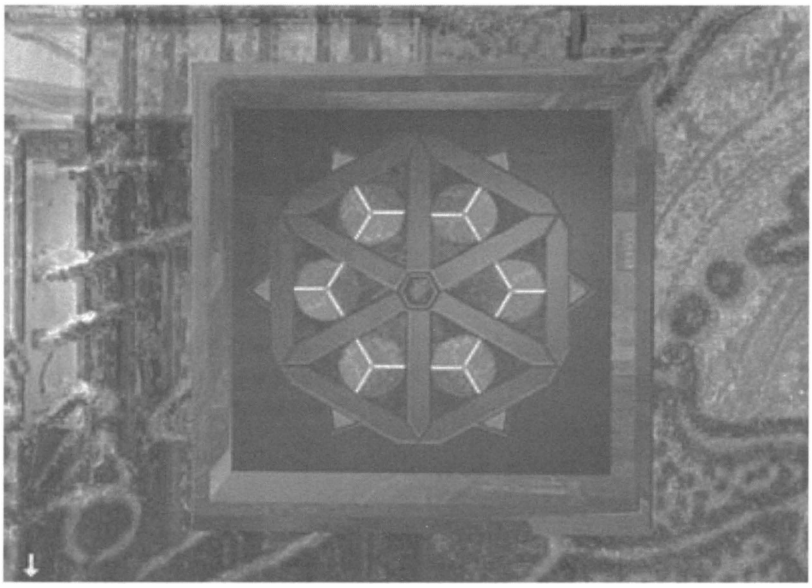


Figure 1 The stasis tube gun puzzle in *Timelapse*

At this point, one might begin to wonder if this puzzle's solution is mathematically similar to that of the clock tower puzzle; one will discover that, indeed, it is. Experimentation shows that clicking on any given yellow triangle rotates exactly three of the circles clockwise by either 120° or 240°, while leaving the other circles alone. For instance, clicking on T_1 rotates C_1 and C_6 clockwise by 120° and rotates C_2 clockwise by 240°, but does nothing to the other circles, while clicking on T_4 rotates C_3 and C_5 clockwise by 240° and rotates C_2 clockwise by 120°, while leaving the other circles alone. Thus, clicking on a triangle causes the states of exactly three circles to change. For instance, clicking on T_1 causes the states of C_1 and C_6 to increase by 1 (mod 3) and the state of C_2 to increase by 2 (mod 3). Mathematically, this corresponds to adding a certain vector to the vector associated with an orientation of the circles. Specifically, for each $i = 1, 2 \ldots, 6$, clicking on triangle T_i adds v_i to the puzzle's current orientation vector, where

$$v_1 = (1, 2, 0, 0, 0, 1), \quad v_2 = (2, 2, 1, 0, 0, 0), \quad v_3 = (0, 2, 1, 1, 0, 0)$$
$$v_4 = (0, 0, 2, 1, 2, 0), \quad v_5 = (0, 0, 0, 2, 1, 1), \quad \text{and } v_6 = (1, 0, 0, 0, 2, 1).$$

Now, what's our goal for this puzzle? Well, recall the inactive red hexagonal button in the center of the puzzle; chances are we want to activate it. Moreover, in adventure-game language, its shape and color suggest that we will be able to do this by rotating the circles so that each circle's red sector faces the puzzle's center: that is, so that each circle is in state 0. Thus, mathematically, we begin with the element $(1, 0, 2, 1, 0, 2)$ in $\mathbb{Z}_3^6$, and want to obtain the element $(0, 0, 0, 0, 0, 0)$: that is, we want to add $(2, 0, 1, 2, 0, 1)$ to $(1, 0, 2, 1, 0, 2)$. Since what we can do is limited to the actions performed by the T_i, we want to find positive integers $\lambda_1, \lambda_2, \ldots, \lambda_6$ so that

$$(2, 0, 1, 2, 0, 1) = \sum_{i=1}^{6} \lambda_i v_i.$$

Looks familiar, huh? Only this time, this corresponds to solving a system of six congruences in six unknowns, which is not so nice to do by hand. An alternate way of solving this system is to use slightly more sophisticated linear algebra. Specifically, we can represent this system by the matrix equation

$$A\mathbf{x} = \mathbf{b},$$

where A is the matrix whose ith column is v_i, $\mathbf{x}$ is the column vector whose ith entry is λ_i, and $\mathbf{b}$ is the transpose of (2, 0, 1, 2, 0, 1). Note that we are thinking of all these matrices as being over $\mathbb{Z}_3$. Solving for $\mathbf{x}$, we obtain

$$\mathbf{x} = A^{-1}\mathbf{b};$$

so we're done if we can invert A. If we don't wish to invert a 6×6 matrix over $\mathbb{Z}_3$ by hand, we can do the inversion using, for instance, the GAP system for computational discrete algebra. We can also then multiply A^{-1} by $\mathbf{b}$ using that system, and obtain the solution

$$\mathbf{x}^T = (A^{-1}\mathbf{b})^T = (2, 1, 0, 0, 1, 1).$$

(Again, these solution values are unique modulo 3.) Thus, to solve the puzzle it should suffice to click on T_1 twice, and each of T_2, T_5 and T_6 once. Sure enough, this works! Nothing happens immediately, but now clicking on the red button in the puzzle's center causes the button to disappear as the red sectors of the circles come together to form a hexagon. Further, we now have access to a gun, and can shoot the robot (assuming we have sufficiently good non-intuitive aim)!

We now turn to another *Timelapse* puzzle. A computer adventure game without a Mayan world seems to be almost as rare as an even prime number; in *Timelapse* you encounter a Mayan calendar puzzle. (Before you experiment with this puzzle, you may want to save your game: one of our examples will assume that we start with the puzzle in its original state, and unlike many other puzzles, this puzzle does not reset to its original state when you back away from it and then return.)

The puzzle contains three rings: two on its left side, and one on its right. We'll call the inner left ring R_1, the outer left ring R_2, and the right ring R_3. Each of these rings displays symbols. At any given time, three symbols, one from each of the three rings, are aligned; basic experimentation yields that you can change the symbols that are aligned by turning the rings (we will later discuss the turning of the rings in more detail). You will need to use this puzzle to obtain access to four temples (respectively associated with skulls, jaguars, monkeys, and lizards); each temple's access requires a different combination of symbols be aligned. It is straightforward to discover that R_1, R_2, and R_3 display 8, 12, and 16 distinct symbols, respectively. The original aligned symbols are shown in FIGURE 2; we'll identify each of these symbols with the number 0 (on their respective rings). We number the remaining symbols clockwise on each ring. Clearly, then, we can identify the set of all possible alignments of symbols with the set $\mathbb{Z}_8 \times \mathbb{Z}_{12} \times \mathbb{Z}_{16}$.

Now it's time to explore more carefully the movements of the rings. We can rotate the rings in this puzzle either clockwise or counterclockwise. (This does not constitute a fundamental difference between this puzzle and the other puzzles we've discussed, but does allow us slightly more flexibility in executing our solution, as in this case our solution can involve negative integers.) Now, suppose the alignment of symbols at a certain time is represented by an element x in $\mathbb{Z}_8 \times \mathbb{Z}_{12} \times \mathbb{Z}_{16}$. If we rotate R_1 one position counterclockwise, this also turns R_2 one position counterclockwise while leaving R_3 unchanged, so we've added $v_1 = (1, 1, 0)$ to x. If we instead rotate R_2 one

Figure 2 The Mayan Calendar in *Timelapse*

position counterclockwise, this also turns R_1 one position counterclockwise and R_3 one position clockwise, so we've added $v_2 = (1, 1, 15)$ to x. Finally, if we rotate R_3 one position counterclockwise, this also rotates R_1 one position counterclockwise and doesn't move R_2, so we've added $v_3 = (1, 0, 1)$ to x.

Since we must obtain four different alignments of symbols to gain access to every temple, and since this puzzle does not reset itself to its original state when we leave it and come back, solving this puzzle efficiently requires we treat it a little differently than we did the other puzzles. To solve our previous puzzles, we merely had to write one element of our set as a linear combination of specific vectors; here, we need to write at least four elements of our set as linear combinations of the v_i (we may need to write more than four this way if we mess up at some point, since the puzzle doesn't reset itself). Thus, we will try to answer the question: given any alignment x and any element (a, b, c) of $\mathbb{Z}_8 \times \mathbb{Z}_{12} \times \mathbb{Z}_{16}$, how can we add (a, b, c) to x by turning the rings? Well, let $(a, b, c) \in \mathbb{Z}_8 \times \mathbb{Z}_{12} \times \mathbb{Z}_{16}$. It suffices to find $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{Z}$ so that in $\mathbb{Z}_8 \times \mathbb{Z}_{12} \times \mathbb{Z}_{16}$ we have

$$(a, b, c) = \sum_{i=1}^{3} \lambda_i v_i;$$

that is, we want to find λ_i that solve the system of congruences

$$\begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &\equiv a \pmod 8 \\ \lambda_1 + \lambda_2 &\equiv b \pmod{12} \\ 15\lambda_2 + \lambda_3 &\equiv c \pmod{16}. \end{aligned}$$

Note that, for this puzzle, the λ_i can be negative: if a λ_i is negative, this simply means we should rotate the corresponding ring by $|\lambda_i|$ positions in the *clockwise*, rather than the counterclockwise, direction. It's straightforward to show that if a_1, b_1 and c_1 are any integers with $a_1 \equiv a \pmod 8$, $b_1 \equiv b \pmod{12}$, and $c_1 \equiv c \pmod{16}$, then a

solution to our system is given by

$$\lambda_1 = 2b_1 + c_1 - a_1, \quad \lambda_2 = a_1 - b_1 - c_1, \quad \text{and} \quad \lambda_3 = a_1 - b_1.$$

Let's now put our work to the test. Starting in our original state (0, 0, 0), we'll attempt to open one of the temples. (Before we start, it's important to know that in order to be able to open a particular temple using the puzzle, the crystal ball across from the puzzle must be in a specific position; we will not discuss this further here, but it's something of which to be aware when playing the game.) The order in which we open the temples doesn't matter (though it will affect our computations). Let's open the Skull Temple first. We learn elsewhere in the game that this requires the attainment of state (4, 2, 4). This means we want to add (4, 2, 4) to our current state, so, in this case, $a = 4$, $b = 2$, and $c = 4$. Thus, one solution to our puzzle is given by $\lambda_1 = 4$, $\lambda_2 = -2$, and $\lambda_3 = 2$. So we turn R_1 4 positions counterclockwise, R_2 2 positions clockwise, and R_3 2 positions counterclockwise. Ah, success—we hear a lovely chime, and future exploration will show us that the Skull Temple is now unlocked.

But we're not yet done with the puzzle: we need to obtain three more elements of $\mathbb{Z}_8 \times \mathbb{Z}_{12} \times \mathbb{Z}_{16}$. Let's open the Jaguar Temple next. In order to open this temple, it turns out that we must obtain state (2, 0, 2). Were we starting in our original state, (0, 0, 0), a solution to our puzzle could be obtained using $a = 2$, $b = 0$, and $c = 2$: thus, a solution would be $\lambda_1 = 0$, $\lambda_2 = 0$, and $\lambda_3 = 2$. But since we are currently in state (4, 2, 4) (the state which opened the Skull Temple), this solution won't work: we instead wish to add (6, 10, 14) to our current state. In this case, then, $a = 6$, $b = 10$, and $c = 14$. Using these values would yield relatively large absolute values for the λ_i, requiring lots of turning of rings (for instance, we'd have to turn R_1 28 positions counterclockwise); instead, using the valid values $a_1 = b_1 = c_1 = -2$, we obtain solution $\lambda_1 = -4$, $\lambda_2 = 2$, and $\lambda_3 = 0$. Hence, we can open the Jaguar Temple simply by turning R_1 four positions clockwise and R_2 two positions counterclockwise. We leave the opening of the remaining two temples to the reader.

We end this section by pointing out another way in which this puzzle differs from our other puzzles: our solutions for this puzzle are not unique modulo the moduli in our system. For instance, we mentioned above that starting in initial state (0, 0, 0), to open the Jaguar Temple we must obtain state (2, 0, 2), and can do that using $\lambda_1 = 0$, $\lambda_2 = 0$, and $\lambda_3 = 2$. But we can also do it using $\lambda_1 = 8$, $\lambda_2 = 16$, and $\lambda_3 = 2$; note that our new λ_1 and λ_2 values are not congruent to their previous values modulo 12, one of the moduli in our system. So for this puzzle, as a result of our equations involving different moduli, we do not have the same kind of uniqueness of solutions as we had for our previous puzzles.

Reflection

Of course, these are all variations on the same puzzle. In fact, when one is attuned to this type of thing, one notices this Puzzle everywhere: I was watching a friend play *SpongeBob SquarePants : Battle for Bikini Bottom*, when he encountered the Puzzle (I muttered, "Ah, this time it's $\mathbb{Z}_2^8$..."). But don't let this distract you from the other math puzzles that are out there. Computer games contain both obvious and subtle mathematical puzzles: ones where you need only translate a foreign number system's numerals and apply basic arithmetic, and ones, such as the Puzzle, that may not originally appear to be mathematical at all.

And one of the most beautiful things about computer games is that they allow you to travel to worlds to which we can't physically go in real life: including mathematical worlds. I'll end this discussion with a teaser. At one point in the old text adven-

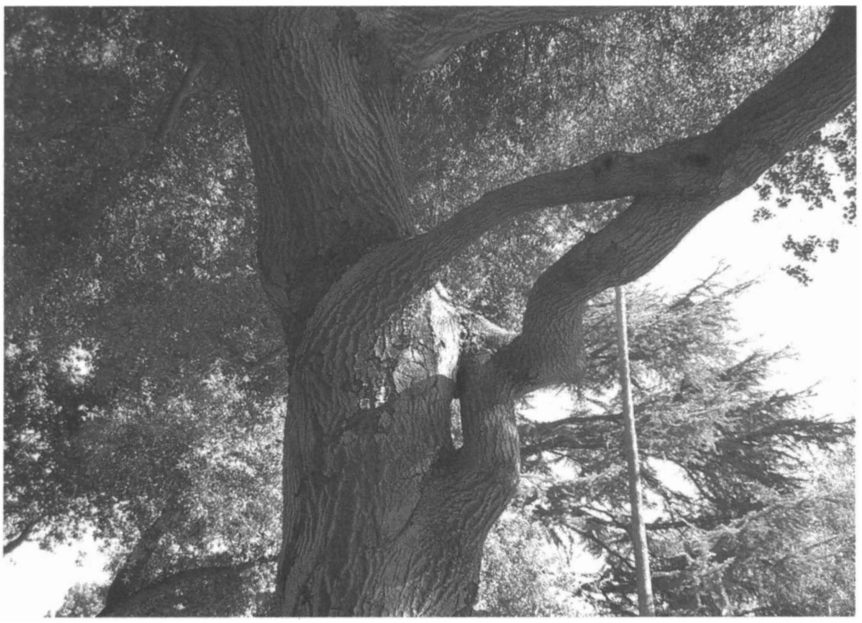
ture game *Trinity*, you have a gnomon you must screw into a hole—but the threads in the hole are going the wrong way. Wandering about, you encounter an "abstract sculpture," inscribed with the words *Felix Klein 1849–1925*. Nearby are strange leafy tunnels. When playing this game, my friend Jen and I suddenly grabbed each other in excitement, as we realized what was going to happen, and how that would solve the puzzle. And with that, I leave you to enjoy the mathematical labyrinths of games.

Acknowledgment. The author thanks Chris Ellison for the inspiration for this paper, and Aaron Malver for his keen eye and his support.

REFERENCES

1. John B. Fraleigh, *A First Course in Abstract Algebra (7th ed.)*, Addison Wesley, 2003.
2. Kenneth Hoffman and Ray Kunze, *Linear Algebra (2nd ed.)*, Prentice-Hall Inc., 1971.
3. *The Longest Journey*, Funcom, 2000.
4. *Myst: 10th Anniversary DVD Edition*, Ubisoft, Inc., 2003. (Includes *Myst* and *Riven*.)
5. *SpongeBob SquarePants: Battle for Bikini Bottom*, THQ Inc., 2003.
6. *Timelapse*, GTE Entertainment, 1996.
7. *Trinity*, Infocom, 1986.

A Tree That's Not a Tree



A graph is a tree if it is connected and has no cycles; so this is not a tree.

—Ramin Naimi
Occidental College
Los Angeles, CA 90041-3314
rnaimi@oxy.edu

NOTES

Disentangling Topological Puzzles by Using Knot Theory

MATTHEW HORAK
University of Wisconsin–Stout
Menomonie, WI 54751
horakm@uwstout.edu

Mathematical puzzles have been a source of entertainment and inspiration throughout the ages, and many puzzles have contributed to the development of large fields of mathematics. For example, a puzzle from the 18th century known as the *Königsberg bridge problem* asks the solver whether or not it is possible to trace a path around the city of Königsberg that crosses each of the city's seven bridges between two islands and the banks of the Pregel River exactly once. Some experimentation easily convinces the solver that there is no possible path, but formally proving this is more difficult. In 1736, the mathematician Leonhard Euler wrote a paper in which he formally proved that solving the Königsberg bridge problem is impossible. Euler's development of the techniques he used in this paper is considered to be the beginning of the field of graph theory.

The Königsberg bridge problem and many other combinatorial puzzles like it have been thoroughly analyzed because they can easily be modelled by relatively simple mathematical objects. In this note, I propose to describe a method of analysis that can be applied to the somewhat less combinatorial class of puzzles known as disentanglement puzzles. This method often will not give a complete solution to a given puzzle, but it will provide insight about the possible solutions, and it will lead through a tour of much interesting mathematics along the way. By necessity, there will be an interplay between rigorous mathematical arguments and physical observations about the puzzle itself. In this respect, analyzing disentanglement puzzles can be thought of as akin to applied mathematics in which a mathematical model is compared against physical observations both during and after its development.

Disentanglement puzzles A *disentanglement puzzle* consists of two or more "parts" entangled with each other that the solver is required to separate, subject to several geometric constraints. The pieces are usually made of some combination of metal, wood and string, and the constraints are that the solver cannot perform any manipulation that bends the wire or breaks the string or wood. Topologically, the pieces that require separation are not linked together, so that if all of the parts were arbitrarily deformable, solving the puzzle would usually be trivial. However, the geometric constraints involving the sizes of the pieces and how they can physically bend prevent the solver from performing the desired topological manipulations. This is the source of the difficulty in analyzing these puzzles mathematically; because the goal of a disentanglement puzzle is topological but the constraints are geometric, it is often difficult to capture all of the important aspects of such a puzzle in a single abstract mathematical object.

By the use of the Quattro puzzle described below, I will illustrate a method by which, to some extent, such abstraction is possible. The method will involve knot theory and the idea of viewing manipulations of the pieces of the puzzle as small "moves" performed on an appropriate initial diagram representing the puzzle. With this interpretation, the topological goal is stated in terms of transforming the initial diagram into a given final diagram, and the geometric constraints are expressed by restricting the allowable moves. This analysis is similar to that used by Louis Kauffman in finding a bound on what he calls the number of "red-blue crossing changes" required to solve the Chinese Rings puzzle [**2**].

Quattro The well-known disentanglement puzzle "Quattro" consists of four identical pieces, each of which is a loop of string tied around a wire ring. All of the wire rings are rigid and of equal size, so an obvious constraint is that the rings cannot move through each other. The strings are entangled as shown in the Quattro link of FIGURE 1, in which the metal rings are represented by the thick circles and the strings represented by the thin curves. The goal is to separate the four pieces.

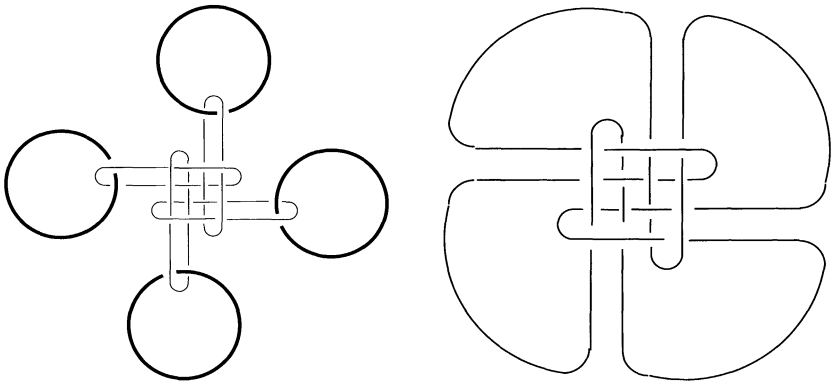


Figure 1 The Quattro link (left) and knot (right)

If one string is long enough to fit around an adjacent ring then solving the puzzle is very simple, as shown in FIGURE 2. In the actual model, each string can certainly move around its own ring, but each string also seems just barely too short to fit around any other ring. I propose to show that if this puzzle is to be solved without passing a string around a ring other than its own, then at some time we must make use of the fact that a string can pass around its own ring. At first glance, this observation seems not to help much, since one way of moving a string around its ring simply results in the

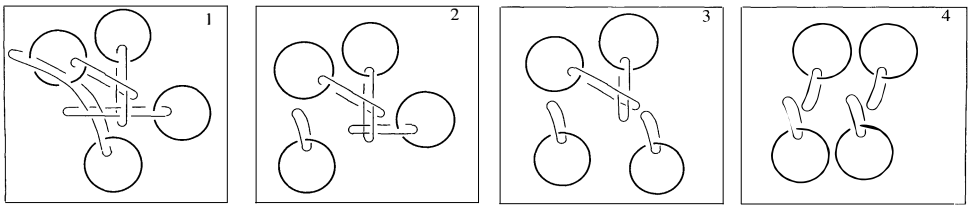


Figure 2 Solving Quattro

string coming back to its original position with a twist. However, moving the "far" end of a string loop around its ring will result in a more significant change to the puzzle. This move is somewhat complicated, but a similar move is used in a solution to the puzzle, "Duplo," which is illustrated in FIGURE 3.

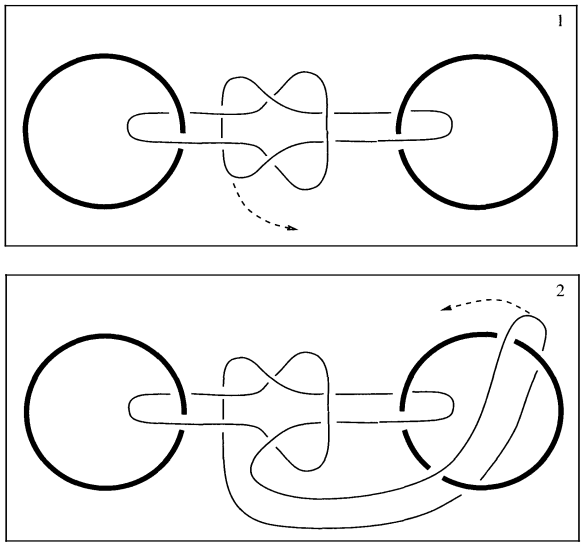


Figure 3 The first move in solving Duplo

This piece of knowledge about the possible solutions can be used to guide the solver on the right track towards a solution. In fact, passing a string around its own ring is a crucial step in the solution to Quattro (Puzzle Q) presented at,

`http://www.gamesandpuzzles.co.uk/reference/eureka_solutions.htm.`

Knot theory A knot is simply an embedding of a circle into three-dimensional Euclidean space. Two knots are equivalent if one can be deformed into the other by a smooth deformation of the ambient Euclidean space. (Technically, the definition involves a smooth one-parameter family of diffeomorphisms of $\mathbb{R}^3$ taking one knot to the other, but the intuitive idea of slowly deforming one knot to the other will suffice here.) We usually view equivalent knots as the same and we speak as though two equivalent knots are actually the same knot. Thus we talk about the "figure eight knot," when we really have a single particular knot in mind and what we mean is "the class of knots equivalent to this particular knot." This abuse of terminology rarely causes a problem in practice.

A link is a disjoint union of one or more knots, each component of which may or may not be linked in various ways to the other components. The *diagram* of a given knot or link is simply its projection onto the xy-plane with gaps wherever two strands cross to indicate which strand is physically above the other. With some slight manipulations of the link, we can always assume that the curves in a diagram intersect transversally and that no crossing involves more than two curves. For a more thorough discussion of knot theory and a method for using "polygonal" knots to remove the requirement of a diffeomorphism in the definition of equivalence, the reader could consult the first chapters in Charles Livingston's book, *Knot Theory* [**3**].

We can abstractly view the Quattro puzzle as an 8-component link by regarding all strings and wires as components of the link, which we will call the *Quattro link*. The link in FIGURE 1 is a diagram of the Quattro link. There are many other possible diagrams of this link, and in fact the manipulations of FIGURE 2 prove that the Quattro link is equivalent to the four unlinked pairs of linked circles shown in the last diagram of the figure. These manipulations are captured abstractly as Reidemeister moves.

The Reidemeister moves are local moves that are performed on a small part of a knot diagram. There are only three Reidemeister moves, all of which are shown in FIGURE 4. In that figure, each picture represents a small part of a larger diagram, and for each move the only change to the diagram in question is in the small neighborhood shown. Each Reidemeister move on a diagram corresponds to an obvious manipulation of the components of the link that changes it to an equivalent link. Conversely, a fundamental theorem of knot theory implies that two diagrams represent equivalent links if and only if one can be transformed into the other by a finite sequence of Reidemeister moves or deformations of the diagram that do not destroy or create crossings as they are performed [**1**, Section 1.3].

Type 1 Type 2 Type 3

Figure 4 The Reidemeister moves

The Quattro Knot Moving back to the puzzle, we wish to show that if the strings are in fact too short to fit around each others' rings, then any solution of the puzzle must involve passing a loop around its own ring. We begin by using a closely related knot to show that the Quattro puzzle cannot be unlinked without a Reidemeister move involving a wire component. Consider the Quattro knot, whose diagram is shown in FIGURE 1. We can form this knot by cutting the strings of the Quattro puzzle, removing the rings and attaching each cut end of string to its neighbor. This operation is an example of the procedure of closing a tangle, which is commonly used in the study of knots and links.

A *tangle* is an object similar to a knot, but which consists of a number of arcs inside of a hollow sphere. The arcs intersect the sphere only in their endpoints and at equal intervals along the equator of the sphere. As with knots, two tangles are equivalent if one can be smoothly deformed to the other, but for tangles we require that the endpoints of the arcs not move during the deformation and that the deformation never take any portion of an arc outside of the sphere. Also as with knots, a tangle is usually represented by its diagram, which is its projection to the xy-plane with over and under crossings marked by gaps in the arcs. The left illustration of FIGURE 5 shows the diagram of the

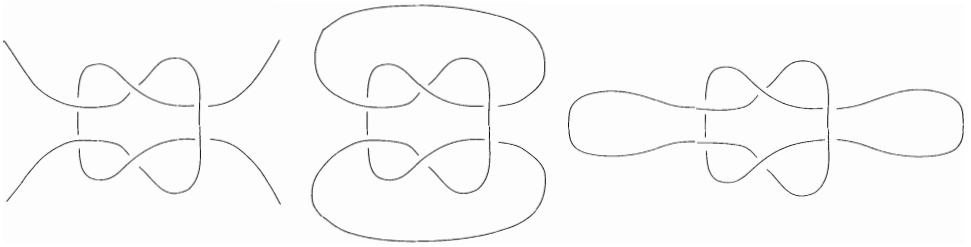


Figure 5 A 2-string tangle and its two closures

tangle that results from enclosing in a sphere the knotted part of the link representing the Duplo puzzle. Since this tangle consists of two separate arcs, it is called a *2-string tangle*.

A tangle can be *closed* by moving along the equator of the sphere and joining pairs of adjacent endpoints with arcs travelling outside of the sphere. This is done in such a way that the arcs do not cross when viewed in the diagram. For any tangle, there are exactly two ways of performing the closure, depending on which two endpoints are joined first, as shown in the second two illustrations of FIGURE 5. Closing a tangle results in a knot or a link that displays much of the same knotting as the original tangle, and in this capacity tangles and their closures are used in the study and classification of knots and links.

The Quattro knot can be constructed as a closure of the tangle produced by surrounding the knotted portion of the Quattro link by the sphere illustrated on the left in FIGURE 6. It displays the exact same crossings as the Quattro link inside dashed circles in FIGURE 6, and it will be an important tool in our analysis of Quattro. Any Reidemeister move performed inside the dashed circle on the puzzle diagram has a corresponding move on the knot diagram. Additionally, it makes sense to speak of entire sequences of Reidemeister moves occurring within these circles, because if we perform one such move, then all of the arcs resulting from that move lie within the dashed circle on the new diagram.

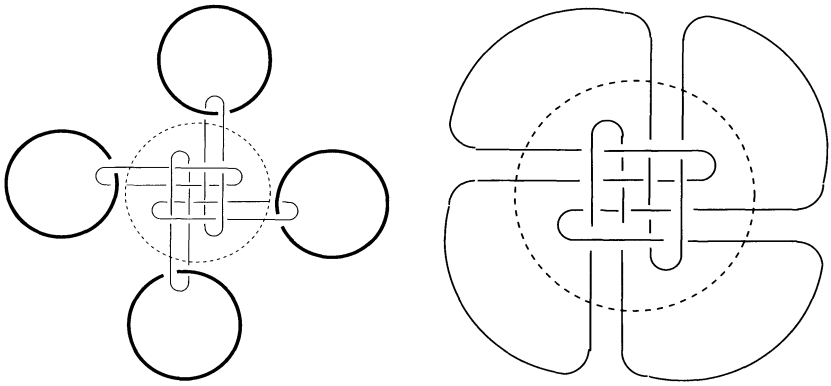


Figure 6 The same knotting

Now, if the physical puzzle can be unlinked without any moves involving a wire strand, then by the constraints on string size, the corresponding sequence of Reidemeister moves on the diagram of FIGURE 6 can be taken to occur entirely within the dashed circle. Performing the corresponding moves on the Quattro knot diagram would unknot it. Therefore, in order to show the necessity of a wire Reidemeister move in the solution to Quattro, all that remains to show is that the Quattro knot is not equivalent to the unknot (the knot consisting of a circle without any knotting).

Tricolorability Showing directly that a given knot is not the unknot is usually difficult. Indeed how could one possibly prove that there is no sequence of Reidemeister moves that unknots the Quattro knot? One way to do this is by using a property called "tricolorability." We say that a link diagram is *tricolorable* if each segment in the diagram can be colored by one of three colors in such a way that,

1. At each crossing, either all three colors meet or only one color is present.
2. At least two of the three colors are used.

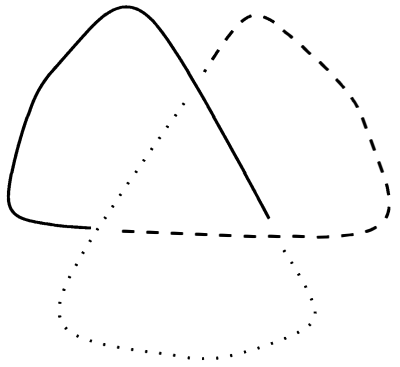


Figure 7 A tricoloring of the trefoil knot

FIGURE 7 shows a tricoloring of a diagram for the trefoil knot. The most important fact about the property of tricolorability is that if diagrams D and D' represent equivalent links, then either both D and D' are tricolorable or neither is, as the following theorem asserts (see [**1**], Section 1.5). The theorem allows us to regard tricolorability as a property of links themselves, and we say that a link is tricolorable if one (and therefore all) of its diagrams is tricolorable.

THEOREM 1. *If D and D' are diagrams corresponding to equivalent links, then D is tricolorable if and only if D' is tricolorable.*

The proof of Theorem 1 relies on the fact that since D and D' represent equivalent links, they are related by a finite sequence of diagrams,

$$D = D_0 \to D_1 \to D_2 \to \cdots \to D_{n-1} \to D_n = D'$$

with each D_i differing from D_{i+1} by a single Reidemeister move. If we can show that a diagram differing by a single Reidemeister move from a tricolorable diagram is itself tricolorable, then starting off with D tricolorable would imply that every diagram in the above sequence, and in particular D', is tricolorable. Similarly, if we started with D' tricolorable, then running the sequence of Reidemeister moves backwards would show that D is tricolorable also.

Proving that performing a Reidemeister move on a tricolorable diagram results in another tricolorable diagram requires the analysis of several cases. One takes the Reidemeister moves one at a time. For each move, one must consider all of the possible ways that the segments involved in the move could be colored in a tricoloring of the initial diagram. Then, one must produce a tricoloring of the resulting diagram. Let us take Type 2, and indicate the colors by solid, dashed, and dotted shading of the strands. The left hand diagram of FIGURE 8 shows one possible way the strands may be col-

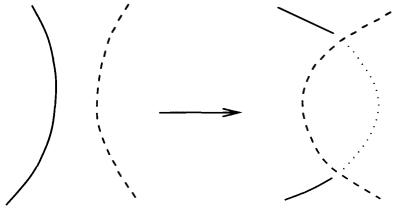


Figure 8 Tricoloring preserved

ored as a result of a tricoloring. The right hand diagram shows a candidate for how to color the strands resulting from the Reidemeister move, and we must show that it is actually a tricoloring.

First, we see that this assignment of color actually gives a coloring of the diagram, because the strands leaving the modified section of the diagram after the move have the same color as the ones leaving that section before the move. Secondly, the new crossings of the diagram satisfy condition (1) of a tricoloring, and since there are no other additional crossings in the new diagram, the entire coloring of the new diagram satisfies tricoloring condition (1). Finally, since there are two colors visible in the diagram, the proposed coloring satisfies requirement (2) and is therefore a tricoloring.

There are other ways in which the initial diagram could be tricolored, and in the proof of the theorem each one must be considered in the above fashion. To complete the proof, we would have to analyze a Type 2 move going in the other direction (eliminating the crossings) and we would have to analyze both directions of the other two Reidemeister moves. In each case, we would start with a tricoloring of the initial diagram and then show that it is possible to tricolor the resulting diagram. This would finish the proof that Reidemeister moves cannot destroy tricolorability. As explained, this finishes the proof that two diagrams representing equivalent links are either both or neither tricolorable.

To use tricolorability to show that the Quattro knot is not equivalent to the unknot, we first note that the unknot is not tricolorable. This is because the unknot possesses a diagram without any crossings at all. This diagram is definitely not tricolorable, since it can not be colored by using more than one color. On the other hand, FIGURE 9 gives a tricoloring of a diagram for the Quattro knot. If the Quattro knot were equivalent to the unknot, then Theorem 1 would imply that the unknot is tricolorable, which is a contradiction. Therefore, the Quattro knot is not equivalent to the unknot, which finishes the proof that any sequence of Reidemeister moves unlinking the physical puzzle of FIGURE 1 must have a move that involves a wire strand.

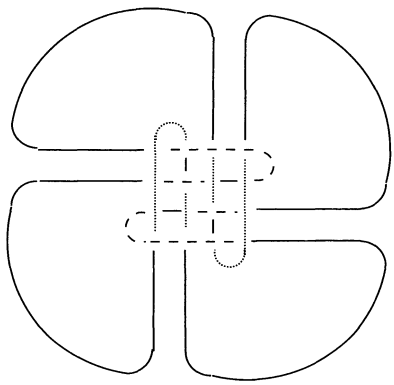


Figure 9 A tricoloring the Quattro knot

Solving Quattro The previous sections focused on our abstract model of the puzzle, the Quattro link, and we found that a Reidemeister move involving wire is necessary to unlink it. We again move back to the physical puzzle and see what we can conclude about it from this fact about its abstract model. Suppose that the rings of the puzzle are too large to pass completely through each others' string loops. Doing nothing but passing rings partially through loops other than their own in the physical puzzle leads immediately to a dead end with no slack string left for any movement at all. Passing rings over and under strings without passing through any loops likewise either knots

the puzzle or has no effect on the knotted strings in the middle. So, if a solution of the puzzle is to include a wire-string move, then it must include at least one move of a ring partially or completely through its own loop of string.

Solving the puzzle without any wire-string moves would give an unlinking of the Quattro link whose corresponding Reidemeister moves involve no wire strands. Since this is impossible, any solution of the puzzle must involve a wire-string move. As mentioned in the previous paragraph, the physical constraints of the puzzle lead to the conclusion that any solution must involve passing a ring at least partially through its own loop. While this information certainly does not solve the puzzle, it at least guides the solver very strongly down the path towards a correct solution, as a crucial move of the solution is to pass a loop around its own ring while manipulating the other strings around that ring just enough to create sufficient slack to complete the move.

Conclusion The above discussion is one example of how topological disentanglement puzzles can be analyzed by interpreting the geometric constraints as constraints on which Reidemeister moves are allowable on a diagram of the puzzle. Ideally, this allows one to rephrase the problem as a problem in ordinary knot theory about determining whether or not two knots or links are equivalent. This part of the problem is usually resolved with one or another of the available knot or link invariants such as tricolorability. As we saw in the analysis of the Quattro puzzle, the physical constraints apparent in the actual puzzle come to play in deciding which constraints to impose on the Reidemeister moves as well in determining what the conclusions about the abstract model tell about the physical puzzle. This interplay seems unavoidable in disentanglement puzzles in which some of the constraints are formed by the physical sizes and flexibilities of the pieces of the puzzle.

There exists a wealth of disentanglement puzzles that could benefit from this kind of analysis. One that comes to mind is a more difficult version of Duplo that has two metal rings and three loops of string. Another possible application of this method might be in answering Louis Kauffman's question of how many wire-string crossing changes are actually necessary to solve the Chinese Rings puzzle [2].

Acknowledgment. I thank Stan Isaacs for a useful discussion about this puzzle and its solutions, and I thank the referees of this paper for pointing out a website posting the solution to Quattro and for suggesting the use of tricolorability.

REFERENCES

1. Colin C. Adams, *The knot book*, American Mathematical Society, Providence, RI, 2004.
2. Loius H. Kauffman, Tangle complexity and the topology of the Chinese rings, *Mathematical approaches to biomolecular structure and dynamics* (Minneapolis, MN, 1994), IMA Vol. Math. Appl., vol. 82, Springer, New York, 1996, pp. 1–10.
3. Charles Livingston, *Knot Theory*, Carus Mathematical Monographs, vol. 24, Mathematical Association of America, Washington, DC, 1993.

Trigonometric Series via Laplace Transforms

COSTAS J. EFTHIMIOU
Department of Physics
University of Central Florida
Orlando, FL 32816
costas@physics.ucf.edu

In another Note in this MAGAZINE [**2**], I presented a method that uses the Laplace transform to find exact values for a large class of convergent series of rational terms. Recently, also in the MAGAZINE, Lesko and Smith [**3**] revisited the method and demonstrated an extension of the original idea to additional infinite series. My intention in this note is to illustrate the power of the technique in the case of trigonometric series.

Trigonometric series Trigonometric series play a vital role in mathematics and physics. Many results are known but most of them can be obtained only via Fourier analysis. For example

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n} = -\ln\left(2\sin\frac{x}{2}\right), \qquad 0 < x < 2\pi, \tag{1}$$

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n^2} = \frac{3x^2 - 6\pi x + 2\pi^2}{12}, \qquad 0 \le x \le 2\pi, \tag{2}$$

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n} = \frac{\pi - x}{2}, \qquad 0 < x < 2\pi, \tag{3}$$

$$\sum_{n=1}^{\infty} (-1)^n \frac{\cos(nx)}{n} = -\ln\left(2\cos\frac{x}{2}\right), \qquad -\pi < x < \pi, \tag{4}$$

$$\sum_{n=1}^{\infty} (-)^n \frac{\cos(nx)}{n^2} = \frac{3x^2 - \pi^2}{12}, \qquad -\pi \le x \le \pi, \tag{5}$$

$$\sum_{n=0}^{\infty} \frac{\cos((2n+1)x)}{(2n+1)} = -\frac{1}{2}\ln\tan\frac{x}{2}, \qquad 0 < x < \pi, \tag{6}$$

$$\sum_{n=0}^{\infty} \frac{\cos((2n+1)x)}{(2n+1)^2} = \frac{\pi^2 - 2\pi x}{8}, \qquad 0 \le x \le \pi. \tag{7}$$

Davis [**1**] offers these and many additional results. Given a function, expanding it in a trigonometric series is relatively straightforward. However, it is almost impossible to guess the function that would generate a given trigonometric series as its Fourier series. For example, given the series

$$\sum_{n=0}^{\infty} (-1)^n \frac{\sin((2n+1)x)}{(2n+1)^2}, \qquad -\frac{\pi}{2} \le x \le \frac{\pi}{2},$$

it is not easy to guess that the function

$$f(x) = \frac{\pi x}{4}$$

gives rise to it. Proving that the guess is correct requires additional work. On the other hand, the method I presented [**2**], with the extension of Lesko and Smith [**3**], yields the results in a straightforward manner with no ad hoc guessing.

The method As Lesko and Smith pointed out, the original method can be applied to series of the form $\sum_{n\in I} u_n v_n$ where I is a subset of the integers. In series of this form, it is often convenient to write only one of the factors, say v_n, as a Laplace transform of a function $f(t)$

$$v_n = \int_0^{+\infty} e^{-nt} f(t)\, dt.$$

Then

$$\sum_{n\in I} u_n v_n = \sum_{n\in I} u_n \int_0^{+\infty} e^{-nt} f(t)\, dt.$$

Assuming that the order of the operations of summation and integration can be exchanged

$$\sum_{n\in I} u_n v_n = \int_0^{+\infty} \left(\sum_{n\in I} u_n e^{-nt}\right) f(t)\, dt.$$

In this note we shall always exchange the order of the two operations, assuming that the reader knows how to justify this step. Details on this may be found in the original papers [**2, 3**].

If one can find an explicit function $h(t) = \sum_{n\in I} u_n e^{-nt}$, then this leads to a simple integral representation of the initial series:

$$\sum_{n\in I} u_n v_n = \int_0^{+\infty} h(t)\, f(t)\, dt.$$

If, furthermore, the integration can be performed, then analytic expressions for the sums of the initial series are obtained.

Elementary trigonometric sums To apply the method to trigonometric series, we need to be able to handle series of the form

$$S = \sum_{n\in I} \sin(nx)\, e^{-nt},$$
$$C = \sum_{n\in I} \cos(nx)\, e^{-nt}.$$

These summations are performed easily using complex notation:

$$C + iS = \sum_{n\in I} e^{inx}\, e^{-nt}.$$

In particular, when $I = \mathbb{N}^* = \{1, 2, \dots\}$, assuming that x is a real number and $t > 0$, we find

$$\sum_{n=1}^{\infty} \sin(nx)\, e^{-nt} = \frac{e^{-t} \sin x}{1 - 2\cos x\, e^{-t} + e^{-2t}},$$

$$\sum_{n=1}^{\infty} \cos(nx)\, e^{-nt} = \frac{e^{-t}(\cos x - e^{-t})}{1 - 2\cos x\, e^{-t} + e^{-2t}}.$$

Similarly we can compute other sums. Also, it is possible to start with these sums and, by changes in the argument x and simple manipulations, derive formulæ for new sums, such as sums over the even or odd integers only. The reader may wish to experiment with this idea.

Trigonometric series via the Laplace transform We are now ready to find exact sums for more complicated trigonometric series. We shall demonstrate the method with two examples, namely (1) and (3). The other formulas may be obtained similarly. Alternatively, they may be computed using algebraic and integral operations on (1) and (3). For example, equation (2) may be proved by integrating (3) and using the well known sum $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$. The last sum, in turn, may be obtained easily using my original method [**2**] or various other techniques.

The approach described here, however, allows us to derive equation (2) without reference to any other sum, assuming that one can integrate the necessary functions, perhaps using a table of integrals. Since tables of integrals are widely available and they are quite extensive (for instance, [**4**]), the method seems to be quite effective and straightforward.

- We start with the series

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n^{\nu}}$$

where $\nu \in \mathbb{N}^*$ and $0 < x < 2\pi$, if $\nu = 1$, or $0 \le x \le 2\pi$, if $\nu > 1$. Using

$$\frac{1}{n^{\nu}} = \frac{1}{(\nu - 1)!} \int_0^{\infty} e^{-nt}\, t^{\nu - 1}\, dt,$$

we write

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\cos(nx)}{n^{\nu}} &= \frac{1}{(\nu - 1)!} \sum_{n=1}^{\infty} \cos(nx) \int_0^{\infty} e^{-nt}\, t^{\nu - 1}\, dt \\
&= \frac{1}{(\nu - 1)!} \int_0^{\infty} \left(\sum_{n=1}^{\infty} \cos(nx) e^{-nt} \right) t^{\nu - 1}\, dt \\
&= \frac{1}{(\nu - 1)!} \int_0^{\infty} \frac{e^{-t}(\cos x - e^{-t})}{1 - 2\cos x\, e^{-t} + e^{-2t}}\, t^{\nu - 1}\, dt.
\end{aligned}$$

With the change of variables $u = e^{-t}$, the integral assumes a more compact form

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n^{\nu}} = \frac{(-1)^{\nu - 1}}{(\nu - 1)!} \int_0^1 \frac{\cos x - u}{1 - 2\cos x\, u + u^2} (\ln u)^{\nu - 1}\, du.$$

For $\nu = 1$

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n} = -\frac{1}{2} \int_0^1 \frac{d(1 - 2\cos x\, u + u^2)}{1 - 2\cos x\, u + u^2}$$

$$= -\frac{1}{2} \ln(1 - 2\cos x\, u + u^2)\Big|_0^1 = -\ln\left(2\sin\frac{x}{2}\right).$$

- Following the same steps for the series

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n^{\nu}},$$

we can arrive at the integral representation

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n^{\nu}} = \frac{(-1)^{\nu-1}}{(\nu-1)!} \sin x \int_0^1 \frac{(\ln u)^{\nu-1}}{1 - 2\cos x\, u + u^2}\, du.$$

In particular, for $\nu = 1$

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n} = \sin x \int_0^1 \frac{1}{(u - \cos x)^2 + \sin^2 x}\, du$$

$$= \tan^{-1} \frac{u - \cos x}{\sin x}\Big|_0^1 = \tan^{-1} \frac{\sin x}{1 - \cos x} = \frac{\pi - x}{2}.$$

Conclusion Although the results presented in this paper are not new, we hope that readers will appreciate the ease and transparency of the method. Given *any* series such as those described in the original articles of Efthimiou [**2**], Lesko and Smith [**3**], and this note, the steps are well-defined and require no special tricks. However, traditional methods do vary from series to series, ingenious tricks may be necessary to be introduced, and some (or a lot) guessing might be required. We invite the reader to verify our claim by using the Laplace transform technique to find exact sums for her favorite series (of one of the types under discussion) and then compare with the traditional methods.

Acknowledgment. While this article was under review, we received a message from Harvey J. Hindin, who pointed out that Albert D. Wheelon had also used the idea of integral transformations to compute exact values for infinite series [**5**].

We are grateful to one of the referees who read the article with great care and dedication. Although typesetting had introduced many typographical errors into the document, the penetrating and thorough reading of the referee helped very much in detecting them.

REFERENCES

1. H.F. Davis, *Fourier series and orthogonal functions*, Dover, New York, 1989.
2. C. Efthimiou, Finding exact values for infinite sums, this MAGAZINE **72** (1999), 45–51.
3. J.P. Lesko and W.D. Smith, A Laplace transform technique for evaluating infinite series, this MAGAZINE **76** (2003), 394–398.
4. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products, Corrected and Enlarged Edition*, Academic Press, New York, 1980.
5. A.D. Wheelon, Tables of Summable Series and Integrals Involving Bessel Functions, Holden-Day Publishers, San Francisco, 1968.

Two Methods for Approximating π

CHIEN-LIH HWANG
Department of Mathematics
National Taiwan University
Taipei, Taiwan
hcl@math.ntu.edu.tw

Dedicated to Professor Darío Castellanos (December 4th, 1937–November 23rd, 1995) of Facultad de Ingeniería, Universidad de Carabobo, Valencia, Venezuela.

It is not difficult to obtain formulas that approximate π. For instance, from the series

$$\sin x = x - \frac{x^3}{6} + \frac{x^5}{120} - \cdots \quad \text{and} \quad \cos x = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \cdots,$$

we get, neglecting terms of order five or higher,

$$3\sin x - x\cos x \approx 2x, \quad \text{so that}$$
$$x \approx \frac{3\sin x}{2+\cos x}. \tag{1}$$

This formula was given by Cardinal Nicolaus Cusanus, and later by the Dutch mathematician and physicist Willebrord Snellius [**1**], most famous for the law of refraction now known as Snell's law.

Combining the first two series with

$$\frac{1}{2}\sin 2x = \sin x \cos x = x - \frac{2}{3}x^3 + \frac{2}{15}x^5 - \cdots,$$

we get, neglecting terms of order seven or more,

$$14\sin x - 6x\cos x + \sin x\cos x \approx 9x \quad \text{or}$$
$$x \approx \sin x \frac{14+\cos x}{9+6\cos x} \tag{2}$$

a formula due to Newton.

These formulas give better approximations for smaller angles. Using $x = \pi/12$, we get $\pi \approx 3.141509994$ from (1), while Newton's formula gives $\pi \approx 3.141592169$, the true value being, of course, 3.141592653589....

Ways of the master The method that led to (1) and (2) is due to Newton [**2**]. Let us extend his procedure, following the adage that much is to be learned by treading the ways of the great master.

To this end we consider the expression

$$x \approx A_1 \sin x - A_2 x\cos x + A_3 \sin 2x - A_4 x\cos 2x$$
$$+ \cdots + A_{2s-1}\sin sx - A_{2s}x\cos sx, \tag{3}$$

where the A_ks are constants to be determined and the $\approx$ sign is to be interpreted in the sense that the Maclaurin expansions of both sides agree up through the power x^{4s-1} (and thus up through the power x^{4s} since all the functions involved in (3) are odd).

Expanding the right-hand side by use of the appropriate Maclaurin series, and equating corresponding powers of x, one finds a system of $2s$ equations with $2s$ unknowns for the determination of the A_ks:

$$\begin{aligned} A_1 - A_2 + 2A_3 - A_4 + \cdots + sA_{2s-1} - A_{2s} &= 1 \\ A_1 - 3A_2 + 2^3A_3 - 3 \cdot 2^2A_4 + \cdots + s^3A_{2s-1} - 3s^2A_{2s} &= 0 \\ &\vdots \\ A_1 - (4s-1)A_2 + 2^{4s-1}A_3 - 3 \cdot 2^2A_4 + \cdots & \\ + s^{4s-1}A_{2s-1} - (4s-1)s^{4s-2}A_{2s} &= 0 \end{aligned}$$

Solving for x in (3) we find

$$x \approx \frac{A_1 \sin x + A_3 \sin 2x + A_5 \sin 3x + \cdots + A_{2s-s} \sin sx}{1 + A_2 \cos x + A_4 \cos 2x + A_6 \cos 3x + \cdots + A_{2s} \cos sx}.$$

For $s = 1$, we find $A_1 = 3/2$, $A_2 = 1/2$, which gives (1).

To list expressions for additional values of s we will use the identities $\sin 2x = 2 \sin x \cos x$, $\cos 2x = \cos^2 x - \sin^2 x$, and so on, so that all trigonometric functions will be evaluated with the same argument.

The value $s = 2$ gives

$$x \approx \frac{5}{3} \sin x \, \frac{16 + 5 \cos x}{17 + 16 \cos x + 2 \cos^2 x}. \tag{4}$$

Using $s = 3$ gives

$$x \approx \frac{7}{5} \sin x \, \frac{92 + 66 \cos x + 7 \cos^2 x}{82 + 111 \cos x + 36 \cos^2 x + 2 \cos^3 x} \tag{5}$$

and $s = 4$ gives

$$x \approx \frac{1}{35} \sin x \, \frac{91{,}648 + 103{,}511 \cos x + 28{,}544 \cos^2 x + 1{,}522 \cos^3 x}{1{,}667 + 2{,}944 \cos x + 1{,}560 \cos^2 x + 256 \cos^3 x + 8 \cos^4 x}. \tag{6}$$

Setting $x = \pi/6$ yields interesting approximations for π. Formulas (4), (5), and (6) give successively

$$5 \cdot \frac{32 + 5\sqrt{3}}{37 + 16\sqrt{3}} = 3.141592229\ldots, \tag{7}$$

$$\frac{21}{5} \cdot \frac{389 + 132\sqrt{3}}{436 + 255\sqrt{3}} = 3.14159265346\ldots, \tag{8}$$

$$\frac{3}{70} \cdot \frac{45{,}224 + 209{,}305\sqrt{3}}{5{,}683 + 3{,}136\sqrt{3}} = 3.141592653589754\ldots; \tag{9}$$

the correct value of π being, we recall, $3.141592653589793238\ldots$.

It is interesting to note that (8) and (9) improve Ramanujan's formula [5]

$$\frac{63}{25} \cdot \frac{17 + 15\sqrt{5}}{7 + 15\sqrt{5}} = 3.14159265380\ldots. \tag{10}$$

Though for similarly-sized coefficients (10) is a better approximation than (7), it is worth emphasizing that Ramanujan's method, based on the theory of elliptic modular functions, is considerably more involved than the simple procedure we have used here.

In this method, we did not justify one assumption: How do we know that we can always solve the system of equations for the coefficients A_k? How do we know that the matrix of the coefficients is always nonsingular? The answer is interesting and not too obvious.

The problem amounts to finding n functions with the property that some linear combination $k(x)$ of them agrees with a given smooth function $f(x)$ in the sense that $f^{(j)}(0) = k^{(j)}(0)$ for $j = 0, 1, \ldots, n-1$. To show that this is always possible, consider n independent solutions of some nth order homogeneous linear differential equation. Let $k(x)$ be the linear combination of these n functions that satisfies the initial conditions $f^{(j)}(0) = k^{(j)}(0)$, $j = 0, 1, \ldots, n-1$. Such a k exists and is unique on account of the usual existence and uniqueness theorem. Furthermore, if $n = 2s$ and s of the independent solutions to the differential equation are even and s are odd, then if f is odd (even) it can be expressed as a linear combination of the s odd (even) functions. In particular, the differential equation

$$(D^2 + 1^2)^2(D^2 + 2^2)^2 \cdots (D^2 + s^2)^2 y = 0$$

is of order $4s$ and has $2s$ odd solutions, $\sin x$, $x \cos x$, $\sin 2x$, $x \cos 2x$, $\cdots$ $\sin sx$, $x \cos sx$. Thus, choosing the odd function $y = x$ we may write $x \approx$

$$A_1 \sin x - A_2 x \cos x + A_3 \sin 2x - A_4 x \cos 2x + \cdots + A_{2s-1} \sin sx - A_{2s} x \cos sx,$$

which is (3).

It is more interesting to show that the determinant of this matrix is nonzero. Define

$$H(u_1, \ldots, u_s) = \begin{vmatrix} u_1 & -1 & u_2 & -1 & \cdots & u_s & -1 \\ u_1^3 & -3u_1^2 & u_2^3 & -3u_2^2 & \cdots & u_s^3 & -3u_s^2 \\ u_1^5 & -5u_1^4 & u_2^5 & -5u_2^4 & \cdots & u_s^5 & -5u_s^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_1^{4s-1} & (1-4s)u_1^{4s-2} & u_2^{4s-1} & (1-4s)u_2^{4s-2} & \cdots & u_s^{4s-1} & (1-4s)u_s^{4s-2} \end{vmatrix}.$$

Then by row and column operations, we can show

$$H(u_1, \ldots, u_s) = -2u_1^3(u_2^2 - u_1^2)^4 \cdots (u_s^2 - u_1^2)^4 H(u_2, \ldots, u_s),$$

which implies

$$H(u_1, \ldots, u_s) = (-2)^s (u_1 \cdots u_s)^3 \prod_{1 \le i < j \le s} (u_j^2 - u_i^2)^4.$$

Taking $u_j = j$ for each j we see that the coefficient matrix is nonsingular. This type of determinant is sometimes called a confluent Vandermonde determinant. A similar one (with odd and even powers), which dates back to a paper of D. Besso in 1882, appeared as problem 10601 in the *American Mathematical Monthly* several years ago, with solution on pp. 688–689 of volume **106** (1999). The reference to Besso is in Sir Thomas Muir's *The Theory of Determinant in the Historical Order of Development*, volume 4, p. 155.

Transcendental approximations to π Let us consider now the expression

$$x \approx A_1 \sin x - A_2 \sin 2x + A_3 \sin 3x - \cdots + (-1)^{s+1} A_s \sin sx, \tag{11}$$

where the Maclaurin expansions on the right-hand side agree with that of x up through the power x^{2s-1}. Using the appropriate sine series and equating corresponding powers of x, we find the system

$$\begin{aligned}
A_1 - \ 2A_2 \ + \ 3A_3 \ - \cdots + \ (-1)^{s-1} s A_s \ &= 1 \\
A_1 - \ 2^3 A_2 \ + \ 3^3 A_3 \ - \cdots + \ (-1)^{s-1} s^3 A_s \ &= 0 \\
&\ \ \vdots \\
A_1 - 2^{2s-1} A_2 + 3^{2s-1} A_3 - \cdots + (-1)^{s-1} s^{2s-1} A_s &= 0
\end{aligned}$$

of s equations which we hope will determine the s unknowns, the A_ks.

To show that the matrix of the coefficients is nonsingular, consider the differential equation

$$(D^2 + 1^2)(D^2 + 2^2) \cdots (D^2 + s^2) y = 0$$

which is of order $2s$ and has s odd solutions: $\sin x$, $\sin 2x, \ldots,$ and $\sin sx$.

Just as before, we can use the determinant to show this matrix always has a unique solution. Define

$$G(v_1, \ldots, v_s) = \begin{vmatrix}
v_1 & v_2 & v_3 & \cdots & v_s \\
v_1^3 & v_2^3 & v_3^3 & \cdots & v_s^3 \\
v_1^5 & v_2^5 & v_3^5 & \cdots & v_s^5 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
v_1^{2s-1} & v_2^{2s-1} & v_3^{2s-1} & \cdots & v_s^{2s-1}
\end{vmatrix}$$

Then by row and column operations we can show

$$G(v_1, \ldots, v_s) = v_1 \cdots v_s \prod_{1 \le i < j \le s} (v_j^2 - v_i^2).$$

This is essentially a Vandermonde determinant; taking $v_j = (-1)^{j-1} j$ for each j we see that the coefficient matrix is nonsingular.

Suppose now that P is an approximation to π. Write $x = P - \pi$, so that x is the error involved in the approximation. If we replace x by $P - \pi$ in (11), we get

$$P - \pi \approx \sum_{k=1}^{s} (-1)^{k-1} A_k \sin kP = -\sum_{k=1}^{s} A_k \sin kP,$$

using $\sin k(P - \pi) = (-1)^k \sin kP$. When we rearrange to remove the minus signs, we get

$$\pi \approx P + A_1 \sin P + A_2 \sin 2P + A_3 \sin 3P + \cdots + A_s \sin sP.$$

Solving the system for $s = 1$, we find

$$\pi \approx P + \sin P, \tag{12}$$

an approximation given by D. Shanks [**3**]. If P is an approximation of π with n decimals correct, then (12) will have $3n$ decimals correct.

Note that the right-hand side of (12) may be written as

$$\pi + x - \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots\right) = \pi + \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots,$$

which shows that the error x involved in the approximation P is reduced to $x^3/3!$ in (12).

Solving the system of equations for larger values of s we get the successively better approximations

$$\pi \approx P + \frac{4}{3}\sin P + \frac{1}{6}\sin 2P \tag{13}$$

with $5n$ decimals correct,

$$\pi \approx P + \frac{3}{2}\sin P + \frac{3}{10}\sin 2P + \frac{1}{30}\sin 3P \tag{14}$$

with $7n$ decimals correct,

$$\pi \approx P + \frac{8}{5}\sin P + \frac{2}{5}\sin 2P + \frac{8}{105}\sin 3P + \frac{1}{140}\sin 4P \tag{15}$$

with $9n$ decimals correct, and so on. For instance, if we take $P = 3.1$, (12) gives $\pi = 3.1415806$, (13) gives $\pi = 3.1415926494$, (14) gives $\pi = 3.141592653588$, and (15) gives $\pi = 3.1415926535897922$.

It is interesting to compare this idea with the Fourier expansion

$$x = 2\sum_{n=1}^{\infty} \frac{(-1)^{n-1}\sin nx}{n}, \quad -\pi < x < \pi. \tag{16}$$

Replace x by $P - \pi$ and use $\sin(k(P - \pi)) = (-1)^k \sin kP$ as before to obtain

$$\pi = P + 2\sum_{n=1}^{\infty} \frac{\sin nP}{n}, \quad 0 < P < 2\pi. \tag{17}$$

Now, when one studies Fourier series [**4**], one learns that among all possible linear combinations of elements of a system $S = \{\phi_0, \phi_1, \phi_2, \ldots\}$ orthonormal on $[a, b]$, the partial sums s_n of the Fourier series of a function f (assumed to be Riemann-integrable on $[a, b]$) yield the best possible approximation to f, in the sense that the mean-square error is least. Formulas (13) to (17) are linear combinations of sine functions that yield increasingly better approximations to the odd function $f(x) = \pi$, $0 < x < \pi$, and are periodic with period 2π. These approximations are *not* truncated Fourier series inasmuch as the coefficients change as s changes. Yet, they are of much better quality than the partial sums of (17) whose convergence is very poor.

Acknowledgment. The author wishes to express his gratitude to the anonymous referees for their careful reading of the manuscript and for many valuable suggestions during the completion of this work.

REFERENCES

1. A. G. Kaestner, *Geschichte der Mathematik*, volume 1, Göttingen (1796), 415.
2. Isaac Newton, *Treatise on the Method of Flexions and Infinite Series*, London, 1737.

3. D. Shanks, Improving an Approximation for pi, *Amer. Math. Monthly* **99** (1992), 263.
4. T. M. Apostol, *Mathematical Analysis*, Addison-Wesley Publishing Company, New York, 1960, 464–465.
5. S. Ramanujan, Modular equations and approximations to π, *Quart. J. Math. Oxford* **45** (1914), 350–372.

A π-less Buffon's Needle Problem

DAVID RICHESON
Dickinson College
Carlisle, PA 17013
richesod@dickinson.edu

In 1733 the French naturalist Georges Louis Leclerc, Comte de Buffon, posed and solved the following problem in geometric probability: when a needle of length L is dropped onto a wooden floor constructed from boards of width D (where $D \geq L$), what is the probability that it will lie across a seam between two boards? Buffon determined that the probability is $2L/D\pi$. His proof of the now-famous *Buffon's needle problem* appeared in print 44 years later [**5**].

The solution to this problem is straightforward, requiring only the integral of a trigonometric function, and is accessible to students in an integral calculus course (a solution without integration can be found in [**9**, §1.1]). In 1812 Laplace noticed that it is possible to approximate π by tallying repeated drops of the needle (the most remarkable and most suspect example being Mario Lazzarini's 1901 approximation of π to six decimal places after only 3408 tosses [**10, 11, 4**]). This means of estimating π is now a classic application of the Monte Carlo method.

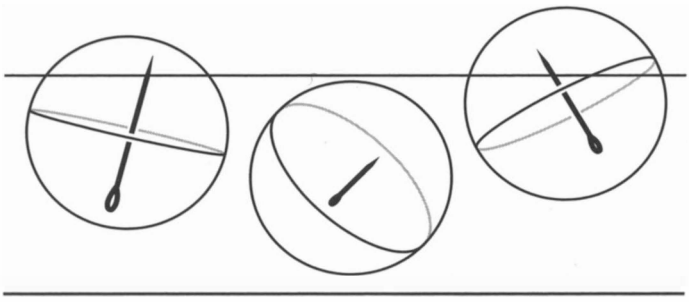


Figure 1

There have been numerous extensions and variations of the Buffon's needle problem (see, e.g. [**1, 6, 7, 8, 10, 12, 13, 14**]). In this note we propose another variation. Suppose that a needle of length L is pushed completely into a clear rubber ball of diameter L. If the ball is dropped onto a wooden floor with slats of width D ($D \geq L$), what is the probability that the needle will lie above a seam in the floor? (This is equal to the probability that a randomly-placed needle in $\mathbb{R}^3$ intersects the set of planes $\{x = kD : k \in \mathbb{Z}\}$.) We call this the *Buffon's ball problem*. We will see that unlike the classical case, π does not appear in the probability, and when $L = D$ the probability is unexpectedly simple.

The exact location of the needle depends on the location of the ball in the plane and on the orientation of the needle in the ball. Thus, it corresponds to a point in the space $\mathbb{R}^2 \times S$, where S is the sphere of diameter L. Taking advantage of symmetries we can reduce the configuration space to $X = [0, D/2] \times S$. That is, we need only know the

distance between the base of the ball and the nearest seam (we assume that the boards run east-west) and the orientation of the needle inside the ball. When we say the ball is dropped at random, we mean that the possible configurations are uniformly distributed in X.

For now, fix y, the distance between the base of the ball and the nearest seam. We wish to compute $P(y)$, the probability that the needle crosses a seam for this value of y. Let R_y denote the region of the sphere corresponding to tip locations that yield a crossing and let $A(y)$ denote the area of R_y. Then,

$$P(y) = \frac{A(y)}{\text{area of sphere}} = \frac{A(y)}{L^2\pi}.$$

When $y > L/2$ the needle cannot cross the seam, so $R_y = \emptyset$, $A(y) = 0$, and $P(y) = 0$. When $0 \leq y \leq L/2$, R_y consists of two identical spherical caps centered about the north and south poles (as in FIGURE 2)—one cap corresponds to the tip being north of the seam and the other to the eye being north of the seam.

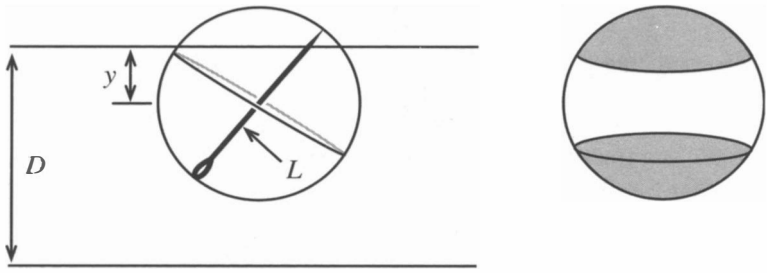


Figure 2

To compute the areas of these spherical caps we turn to Archimedes. In his work *On the Sphere and Cylinder, Book I* [**3**, Proposition 43, p. 53] he proved the remarkable theorem that the area of a spherical cap be expressed in terms of only the slant height of the cap, r (see FIGURE 3). Just like a circle in the plane, the area of the cap is πr^2.

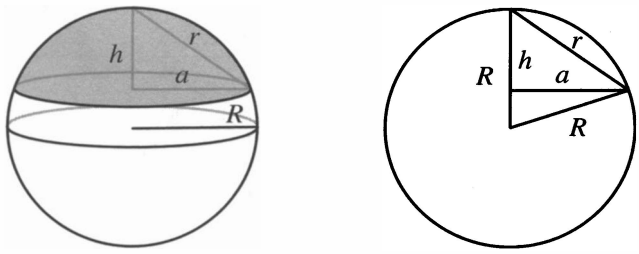


Figure 3

We would like the area of the cap in terms of the radius of the sphere, R, and the height of the cap, h. Taking a to be the radius of the base of the cap, we have

$$a^2 + h^2 = r^2 \quad \text{and} \quad (R - h)^2 + a^2 = R^2.$$

Eliminating a we obtain $r^2 = 2Rh$. Thus, the surface area of the spherical cap is $2\pi Rh$.

Since our spherical caps have height $L/2 - y$, the area of R_y is

$$A(y) = 2\left(2\pi \cdot \frac{L}{2} \cdot \left(\frac{L}{2} - y\right)\right) = L^2\pi\left(1 - \frac{2y}{L}\right).$$

Thus, for the fixed value of y $(0 \le y \le L/2)$, the probability that the needle crosses the seam is

$$P(y) = \frac{L^2\pi(1 - 2y/L)}{L^2\pi} = 1 - \frac{2y}{L}.$$

Also note that $P(y) = 0$ for $L/2 \le y \le D/2$.

Finally we are able to solve the Buffon's ball problem. Since every possibility $0 \le y \le D/2$ is equally likely, the probability that the needle will lie above a seam is

$$\frac{2}{D}\int_0^{D/2} P(y)\,dy = \frac{2}{D}\int_0^{L/2}\left(1 - \frac{2y}{L}\right)dy,$$

which, by elementary geometry (see FIGURE 4), equals $L/2D$.

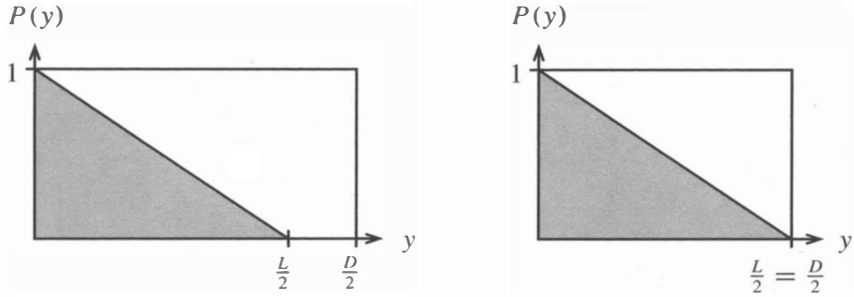


Figure 4

Notice that unlike the classical Buffon's needle problem, π makes no appearance in the probability. Moreover, as is easily seen in FIGURE 4, when $D = L$ the probability of the needle crossing a seam is the same as a coin toss!

Although we justified these surprising conclusions mathematically, it would be nice to gain an intuitive understanding of why they hold. To do so, we return to the formula for the area of a spherical cap, $A = 2\pi Rh$. Using this formula it is easy to show that when a sphere and its circumscribing cylinder are sliced by two planes perpendicular to the axis of the cylinder (see FIGURE 5) they produce slices of equal lateral surface area (this result can be found in [2]). Viewed another way, when a sphere is cut by two

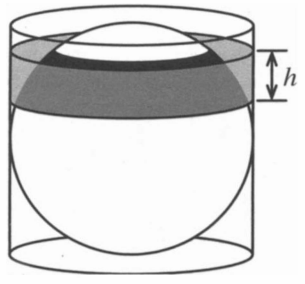


Figure 5

parallel planes, the area bounded by the planes depends only on their separation and not on their orientation.

In the context of the Buffon's ball problem this uniformity implies that the lengths of all projections of the needle onto the north-south axis are equally likely! So, the expected length of the projection is $L/2$. Since the distribution is uniform, the probability of a crossing is $(L/2)/D = L/2D$.

Surveying the literature we find that the classical Buffon's needle problem can be extended in several directions. Of particular interest are the generalizations found in Ramaley's paper **[13]**. First, if the needle is longer than the width of the boards ($L > D$), then it may cross more than one seam when it falls on the floor. In this case the expression found by Buffon, $2L/D\pi$, is the expected value for the number of line-crossings. In fact, Ramaley shows that this interpretation holds even if the needle is not straight, but is a polygonal curve of length L, or, in the limiting case, a curve of length L. That is, if a piece of string of length L is tossed on the floor then $2L/D\pi$ is the expected number of line-crossings! He called this the *Buffon's noodle problem.*

Examining Ramaley's arguments we see that they apply to our situation with no modification except the replacement of Buffon's probability with ours. So, if a needle of length L is placed in a ball of diameter L and is dropped onto a wooden floor with boards of width D, then the expected number of seams that the needle will cross is $L/2D$. Likewise, if a piece of string of length L is suspended inside this same ball so that it cannot move and it is dropped on the same floor, then $L/2D$ is the expected number of seams it will cross (FIGURE 6). Viewed in another way, this says that on average a curve of length L in $\mathbb{R}^3$ will intersect the set of planes $\{x = kD : k \in \mathbb{Z}\}$ in $L/2D$ points.

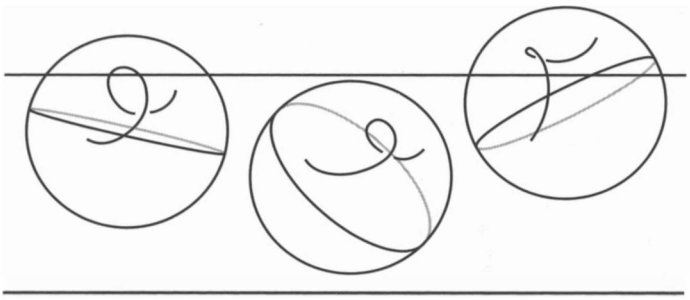


Figure 6

Acknowledgment. The author thanks the referees for their insightful comments on the first version of this note.

REFERENCES

1. Ani Adhikari and Jim Pitman. The shortest planar arc of width 1. *Amer. Math. Monthly*, **96** (1989), 309–327.
2. Tom M. Apostol and Mamikon A. Mnatsakanian. A fresh look at the method of Archimedes. *Amer. Math. Monthly*, **111** (2004), 496–508.
3. Archimedes. *The works of Archimedes: Edited in modern notation with introductory chapters by T. L. Heath with a supplement, "The method of Archimedes"*. Dover Publications Inc., New York, 1953. Reprint of the 1897 edition and the 1912 supplement.
4. Lee Badger. Lazzarini's lucky approximation of π. this MAGAZINE, **67** (1994), 83–91.
5. Georges Louis Leclerc Comte de Buffon. Essai d'arithmétique morale. In *Histoire naturelle, générale er particulière, Supplément*, volume 4, pages 46–123. 1777.
6. Duane W. DeTemple and Jack M. Robertson. Constructing Buffon curves from their distributions. *Amer. Math. Monthly*, **87** (1980), 779–784.

7. R. L. Duncan. A variation of the Buffon needle problem. this MAGAZINE, **40** (1967), 36–38.
8. H. J. Khamis. Buffon's needle problem on radial lines. this MAGAZINE, **64** (1991), 56–58.
9. Daniel A. Klain and Gian-Carlo Rota. *Introduction to geometric probability*. Lezioni Lincee. [Lincei Lectures]. Cambridge University Press, Cambridge, 1997.
10. Pierre Simon Laplace. *Théorie Analytique des Probabilités*. 1812.
11. Mario Lazzarini. Un' applicazione del calcolo della probabilità alla ricerca sperimentale di un valore approssimato di π. *Periodico di Matemtica*, **4** (1901), 140–143.
12. M. F. Neuts and P. Purdue. Buffon in the round. this MAGAZINE, **44** (1971), 81–89.
13. J. F. Ramaley. Buffon's noodle problem. *Amer. Math. Monthly*, **76** (1969), 916–918.
14. J. V. Uspensky. *Introduction to Mathematical Probability*. McGraw Hill Book Company, Inc., New York, 1937.

Conditions Equivalent to the Existence of Odd Perfect Numbers

JUDY A. HOLDENER
Kenyon College
Gambier, OH 43022
holdenerj@kenyon.edu

The abundancy index $I(n)$ of a positive integer n is defined to be the ratio $I(n) = \sigma(n)/n$, where $\sigma(n) = \sum_{d|n} d$. This index is a useful tool in determining whether a number is deficient, abundant, or perfect. In particular, n is *deficient* if $I(n) < 2$, it is *abundant* if $I(n) > 2$, and *perfect* if $I(n) = 2$. Some of the oldest open problems in mathematics relate to the abundancy of a number. Are there infinitely many perfect numbers? Does there exist an odd perfect number? Here are just two questions that were posed by the Greeks over two thousand years ago, and yet they remain unanswered today.

In recent years, this MAGAZINE has published several interesting articles examining the abundancy index of a number [**1, 3, 4**]. In [**1**], R. Laatsch provided a comprehensive summary of what is known about the abundancy index, including a proof that the image of $I(n)$ is dense in the interval $(1, \infty)$. He also posed several interesting questions, one of which was: *Is every rational number $q > 1$ the abundancy index of some integer?* In [**4**], P. A. Weiner answered Laatsch's question in the negative by providing an infinite family of rational numbers in $(1, \infty)$ which fail to be an abundancy index of any integer. Even more interesting, Weiner proved that the set of rationals in $(1, \infty)$ not in the range of $I(n)$ is actually dense in $(1, \infty)$. Finally, Weiner proved the following result:

THEOREM. (WEINER, 2000) *If $I(n) = \frac{5}{3}$ for some n, then $5n$ is an odd perfect number.*

In [**3**], R. F. Ryan then generalized this theorem of Weiner by proving the following:

THEOREM. (RYAN, 2003) *If there exists a positive integer n and an odd positive integer m such that $2m - 1$ is a prime not dividing n and*

$$I(n) = \frac{2m - 1}{m},$$

then $n(2m - 1)$ is an odd perfect number.

In Theorem 1 below, we generalize Ryan's result further by providing a condition that is actually equivalent to the existence of odd perfect numbers. The generalization

is also helpful in that it sheds some light on the theorems cited above by revealing the connection that they have to Euler's well-known characterization of odd perfect numbers. (Euler proved that if an odd perfect number exists, it must have the form $p^\alpha m^2$, where p is a prime satisfying $\gcd(p, m) = 1$ and $p \equiv \alpha \equiv 1 \pmod 4$ [**2**, p. 267].)

THEOREM 1. *There exists an odd perfect number if and only if there exist positive integers p, n, and α such that $p \equiv \alpha \equiv 1 \pmod 4$, where p is a prime not dividing n, and*

$$I(n) = \frac{2p^\alpha(p-1)}{p^{\alpha+1}-1}.$$

Proof. If N is an odd perfect number, then Euler showed that N must have the form $N = p^\alpha m^2$, where p is a prime satisfying $\gcd(p, m) = 1$ and $p \equiv \alpha \equiv 1 \pmod 4$. Hence $\sigma(N) = \sigma(p^\alpha m^2) = \sigma(p^\alpha)\sigma(m^2) = 2p^\alpha m^2$, and

$$I(m^2) = \frac{\sigma(m^2)}{m^2} = \frac{2p^\alpha}{\sigma(p^\alpha)} = \frac{2p^\alpha(p-1)}{p^{\alpha+1}-1}.$$

This proves the forward direction of the theorem.

Conversely, assume that there exists a positive integer n such that

$$I(n) = \frac{2p^\alpha(p-1)}{p^{\alpha+1}-1},$$

where $p \equiv \alpha \equiv 1 \pmod 4$ and p is a prime satisfying $p \nmid n$. Then

$$I(n \cdot p^\alpha) = I(n)I(p^\alpha) = \frac{2p^\alpha(p-1)}{p^{\alpha+1}-1} \cdot \frac{p^{\alpha+1}-1}{p^\alpha(p-1)} = 2.$$

So $n \cdot p^\alpha$ is a perfect number.

Now we claim that $n \cdot p^\alpha$ cannot be even. For if it were, it would have the Euclid-Euler form for even perfect numbers:

$$n \cdot p^\alpha = 2^{m-1}(2^m - 1)$$

where $2^m - 1$ is prime. Since $2^m - 1$ is the only odd prime factor on the right hand side, $p^\alpha = p^1 = 2^m - 1$. But $p \equiv 1 \pmod 4$ and $2^m - 1 \equiv 3 \pmod 4$ (because m must be at least 2 in order for $2^m - 1$ to be prime). This contradiction shows that $n \cdot p^\alpha$ is not even. Therefore it is an odd perfect number. ∎

In closing, we note that if $I(n) = 5/3$ then 5 cannot be a divisor of n. (See the proof of Theorem 3 in [**4**].) Hence Theorem 1 yields Weiner's result when $p = 5$ and $\alpha = 1$. If $\alpha = 1$ and $p = 4k + 1$, then

$$\frac{2p^\alpha(p-1)}{p^{\alpha+1}-1} = \frac{2p}{p+1} = \frac{2(4k+1)}{4k+2} = \frac{2(2k+1)-1}{2k+1},$$

and by setting $m = 2k + 1$, we see that Theorem 1 generalizes Ryan's result as well.

Acknowledgment. I would like to thank Paul Weiner who suggested the simpler proof of the converse to Theorem 1 appearing here. Thanks, too, to an anonymous referee who offered valuable feedback on this work and to the Department of Mathematics at the University of Colorado in Boulder (and David Grant, in particular) for the support they provided me during the writing of this paper.

REFERENCES

1. R. Laatsch, Measuring the abundance of integers, this MAGAZINE **59** (1986), 84–92.
2. K. H. Rosen, *Elementary Number Theory and its Applications*, 5th ed., Pearson Addison Wesley, Boston, 2005.
3. R. F. Ryan, A simpler dense proof regarding the abundancy index, this MAGAZINE **76** (2003), 299–301.
4. P. A. Weiner, The abundancy index, a measure of perfection, this MAGAZINE **73** (2000), 307–310.

The Associativity of the Symmetric Difference

MAJID HOSSEINI
State University of New York at New Paltz
New Paltz, NY 12561–2443
hosseinm@newpaltz.edu

The symmetric difference of two sets A and B is defined by $A\Delta B = (A \setminus B) \cup (B \setminus A)$. It is easy to verify that Δ is commutative. However, associativity of Δ is not as straightforward to establish, and usually it is given as a challenging exercise to students learning set operations (see [**1**, p. 32, exercise 15], [**3**, p. 34, exercise 2(a)], and [**2**, p. 18]).

In this note we provide a short proof of the associativity of Δ. This proof is not new. A slightly different version appears in Yousefnia [**4**]. However, the proof is not readily accessible to anyone unfamiliar with Persian.

Consider three sets A, B, and C. We define our universe to be $X = A \cup B \cup C$. For any subset U of X, define the characteristic function of U by

$$\chi_U(x) = \begin{cases} 1, & \text{if } x \in U; \\ 0, & \text{if } x \in X \setminus U. \end{cases}$$

Two subsets U and V of X are equal if and only if $\chi_U = \chi_V$. The following lemma is the key to our proof.

LEMMA. *For any two subsets U and V of X and for any $x \in X$,*

$$\chi_{U\Delta V}(x) = (\chi_U(x) - \chi_V(x))^2 \tag{1}$$

$$= \chi_U(x) + \chi_V(x) - 2\chi_U(x)\chi_V(x). \tag{2}$$

Proof. Note that both sides of (1) are equal to 1 exactly when x belongs to one of U or V, but not to both. The identity (2) follows immediately from (1) and the fact that $\chi_S^2 = \chi_S$ for any set S. ∎

PROPOSITION. *Let A, B, and C be three sets. Then*

$$(A\Delta B)\Delta C = A\Delta(B\Delta C).$$

Proof. From the Lemma we see that

$$\begin{aligned} \chi_{(A\Delta B)\Delta C} &= \chi_{A\Delta B} + \chi_C - 2\chi_{A\Delta B}\chi_C \qquad (3) \\ &= (\chi_A + \chi_B - 2\chi_A\chi_B) + \chi_C - 2(\chi_A + \chi_B - 2\chi_A\chi_B)\chi_C \\ &= \chi_A + \chi_B + \chi_C - 2\chi_A\chi_B - 2\chi_A\chi_C - 2\chi_B\chi_C + 4\chi_A\chi_B\chi_C. \end{aligned}$$

Since the last line in (3) is symmetric with respect to A, B, and C, we conclude that

$$\chi_{(A\Delta B)\Delta C} = \chi_{(B\Delta C)\Delta A}. \tag{4}$$

But Δ is commutative, and therefore (4) completes the proof of the Proposition. ■

By using modular arithmetic, we can make the above proof even shorter. Note that $U = V$ if and only if $\chi_U \equiv \chi_V \pmod 2$. Thus, the Lemma becomes

$$\chi_{U\Delta V} \equiv \chi_U + \chi_V \pmod 2.$$

The associativity of the symmetric difference now follows from the associativity of addition in modular arithmetic.

REFERENCES

1. H.B. Enderton, *Elements of set theory* (Academic Press, New York, 1977).
2. P.R. Halmos, *Naive set theory* (D. Van Nostrand, Princeton, 1960)
3. K. Kuratowski, *Introduction to Set Theory and Topology* (Pergamon Press, Oxford, 1961)
4. M. Yousefnia, A proof of associativity of symmetric difference of sets, *Roshd-e-Amuzesh-e-Riyaazi* **3** (1986), 25–26.

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

Assistant Editors: RĂZVAN GELCA, Texas Tech University; ROBERT GREGORAC, Iowa State University; GERALD HEUER, Concordia College; VANIA MASCIONI, Ball State University; BYRON WALDEN, Santa Clara University; PAUL ZEITZ, The University of San Francisco

Proposals

To be considered for publication, solutions should be received by May 1, 2007.

1756. *Proposed by Courtney H. Moen and William P. Wardlaw, U. S. Naval Academy, Annapolis, MD.*

For which nonnegative integers n do there exist nonnegative integers a and b such that $n! = 2^a(2^b - 1)$?

1757. *Proposed by Ken Ross, University of Oregon, Eugene, OR.*

Consider $M \geq 3$ equally likely properties, like birthdays, that objects, such as people, can possess. We assume that each object possesses exactly one of the M properties. For $n \leq M$, let $P(n;\ M)$ be the probability that in a random sample of n objects, all of the objects possess different properties, and let $\beta(M)$ be the least n such that $P(n; M) < \frac{1}{2}$. (The familiar birthday problem is based on the fact that $\beta(365) = 23$.) It is well known, based on estimations, that $\beta(M)$ is close to $\sqrt{2M \ln 2}$. Show that, in fact, $\beta(M)$ is one of the two integers in the interval $(\sqrt{2M \ln 2},\ \sqrt{2M \ln 2} + 2)$.

1758. *Proposed by Michael Goldenberg and Mark Kaplan, The Ingenuity Project, Baltimore Polytechnic Institute, Baltimore, MD.*

Let F_n be the nth Fibonacci number, that is, $F_0 = 0,\ F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$. Prove that

$$\prod_{n=2}^{\infty} \frac{F_{2n} + 1}{F_{2n} - 1} = 3.$$

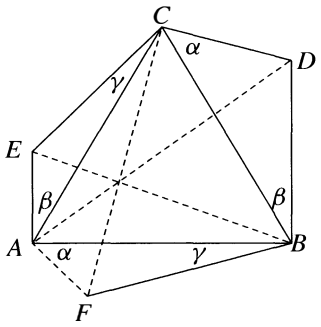
We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie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

Solutions should be written in a style appropriate for this MAGAZINE. Each solution should begin on a separate sheet.

Solutions and new proposals should be mailed to Elgin Johnston, Problems Editor, Department of Mathematics, Iowa State University, Ames IA 50011, or mailed electronically (ideally as a LaTeX file) to `ehjohnst@iastate.edu`. All communications should include the reader's name, full address, and an e-mail address and/or FAX number. Please make sure your name appears on all pages, including electronic pages.

1759. *Proposed by Larry W. Cusick and Maria Nogin, California State University, Fresno, CA.*

In the accompanying figure, $\triangle ABC$ is equilateral. In addition, $\angle FAB \cong \angle DCB$, $\angle FBA \cong \angle ECA$, and $\angle EAC \cong \angle DBC$. Prove that segments $\overline{AD}$, $\overline{BE}$, and $\overline{CF}$ are concurrent.



1760. *Proposed by Péter Ivády, Budapest Hungary.*

Prove that for $0 < a < b < \infty$,

$$\sqrt{\frac{a^2+b^2}{2}} + \sqrt{ab} - \frac{a+b}{2} > \frac{b-a}{\ln b - \ln a}.$$

Quickies

Answers to the Quickies are on page 399.

Q965. *Proposed by Götz Trenkler, University of Dortmund, Dortmund, Germany.*

Let A and B be normal $n \times n$ matrices with complex entries. Prove that AB and BA have the same rank.

Q966. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let $\{f_n\}$ be a sequence of bounded, real valued functions defined on a set $S \subseteq (-\infty\ \infty)$ and let f be a bounded real valued function defined on S. Show that if

$$\lim_{n\to\infty} (f_n(x_n) - f(x_n)) = 0 \tag{1}$$

for every sequence $\{x_n\}$ in S, then the sequence $\{f_n\}$ converges to f uniformly on S. Is the result still true if we only assume (1) for every convergent sequence $\{x_n\}$ in S that converges to an element in S?

Solutions

A perimeter inequality — December 2005

1731. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let A', B', and C' be points on sides $\overline{BC}$, $\overline{CA}$, and $\overline{AB}$, respectively, of triangle ABC, and let M be the point at which $\overline{AA'}$ intersects $\overline{B'C'}$. Prove that either $p(A'C'B) > p(A'C'M)$ or that $p(A'B'C) > p(A'B'M)$, where $p(XYZ)$ denotes the perimeter of triangle XYZ.

Solution by Fejéntaláltuka Szeged Problem Solving Group, Hungary.

We assume that none of A', B', C' coincides with a vertex of the triangle. If $\overline{B'C'} \parallel \overline{BC}$, then there is a point D on $\overline{A'C}$ with $\overline{DB'} \parallel \overline{MA'}$. Then $B'DA'M$ is a parallelogram and

$$\begin{aligned} p(A'B'M) &= p(A'B'D) \\ &= A'B' + B'D + A'D < A'B' + (B'C + CD) + A'D \\ &= p(A'B'C). \end{aligned}$$

If $\overline{B'C'} \not\parallel \overline{BC}$, then there is a point E on either $\overline{AB}$ or $\overline{AC}$ with $\overline{A'E} \parallel \overline{B'C'}$. Without loss of generality, assume that E is on $\overline{AB}$. Then there is a point F on $\overline{A'E}$ such that $\overline{C'F} \parallel \overline{MA'}$. Then from parallelogram $C'FA'M$ we have

$$\begin{aligned} p(A'C'M) = p(A'C'F) &= A'C' + C'F + FA' < A'C' + (C'E + EF) + FA' \\ &= A'C' + C'E + EA' < A'C' + C'E + (EB + BA') \\ &= A'C' + C'B + BA' = p(A'C'B). \end{aligned}$$

Also solved by Deeparnab Chakrabarty (India), Paul Weisenhorn (Germany), and the proposer. There was one partial solution.

The AM-GM region **December 2005**

1732. *Proposed by Cafè Dalat Problem Solving Group, Washington D.C.*

Let n be a positive integer and define the function $f : [0, 1]^n \longrightarrow [0, 1]^2$ by

$$f(x_1, x_2, \ldots, x_n) = \left(\frac{x_1 + x_2 + \cdots + x_n}{n}, \sqrt[n]{x_1 x_2 \cdots x_n} \right).$$

Let $I(n)$ denote the range of f in $[0, 1]^2$. Determine $\cup_{n=1}^{\infty} I(n)$.

Solution by Robert Calcaterra, University of Wisconsin-Platteville, Platteville, WI.

Let $S = \cup_{n=1}^{\infty} I(n)$. Because the arithmetic mean of a finite set of nonnegative numbers cannot be less than its geometric mean, S is a subset of $\{(x, y) : 0 \le y \le x \le 1\}$. Furthermore, if $(x_1 + \cdots + x_n)/n = 1$ with $x_k \in [0, 1]$ for each k, then each $x_k = 1$ and $f(x_1, x_2, \ldots, x_n) = (1, 1)$. Thus, $(1, y) \in S$ if and only if $y = 1$.

Now assume that $0 \le g \le a < 1$. Choose a positive integer n with $\frac{n-1}{n} > a$, let $b = \frac{na}{n-1}$, and let $F(t) = at(b + at - bt)^{n-1}$. Because $F(0) = 0$ and $F(1) = a^n$, there exists $r \in [0, 1]$ such that $F(r) = g^n$. Set $x_1 = ar$ and $x_k = b + (a - b)r$ for $k = 2, 3, \ldots, n$. Then $0 \le x_k < 1$ for all k and $f(x_1, x_2, \ldots, x_n) = (a, g)$. Consequently, $S = \{(x, y) : 0 \le y \le x < 1\} \cup \{(1, 1)\}$.

Also solved by Tshhaye Andeberhan (Eritrea), Michael Andreoli, Michel Bataille (France), Jean Bogaert (Belgium), Cal Poly Pomona Problem Solving Group, Chip Curtis, Larry W. Cusick, Knut Dale (Norway), Jim Delany, Jeffrey M. Groah, Eugene A. Herman, Tom Leong, Northwestern University Math Problem Solving Group, Gabriel T. Prăjitură, Thomas Q. Sibley, Albert Stadler (Switzerland), Marian Tetiva (Romania), and the proposers. There was one incorrect submission.

A limit from the exponential function **December 2005**

1733. *Proposed by Ovidiu Furdui, student, Western Michigan University, Kalamazoo, MI.*

Let x be a fixed real number. For positive integer n define

$$a_n = \left(1 + \frac{1}{n}\right)^{nx} \quad \text{and} \quad b_n = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}.$$

Determine

$$\lim_{n\to\infty}(a_{n+1}+a_{n+2}+\cdots+a_{2n}-b_{n+1}-b_{n+2}-\cdots-b_{2n}).$$

Solution by Michael Goldenberg and Mark Kaplan, The Ingenuity Project, Baltimore Polytechnic Institute, Baltimore, MD.

Let $c_k=(a_k-e^x)-(b_k-e^x)$. We show that $\lim_{n\to\infty}\sum_{k=n+1}^{2n}c_k=-xe^x\ln\sqrt{2}$. First observe that for $k+2>2|x|$,

$$|b_k-e^x|\le\sum_{j=0}^{\infty}\frac{|x|^{k+1+j}}{(k+1+j)!}\le\frac{|x|^{k+1}}{(k+1)!}\sum_{j=0}^{\infty}\frac{|x|^j}{(k+2)^j}$$
$$=\frac{|x|^{k+1}}{(k+1)!}\cdot\frac{1}{1-\frac{|x|}{k+2}}<\frac{2|x|^{k+1}}{(k+1)!}.$$

Thus, for fixed x, $\lim_{n\to\infty}\sum_{k=n+1}^{2n}(b_k-e^x)=0$. Next note that

$$a_k-e^x=\left(1+\frac{1}{k}\right)^{kx}-e^x=e^{kx\ln\left(1+\frac{1}{k}\right)}-e^x=e^{x\left(1-\frac{1}{2k}+O\left(\frac{1}{k^2}\right)\right)}-e^x$$
$$=e^x\left(e^{-\frac{x}{2k}+O\left(\frac{1}{k^2}\right)}-1\right)=e^x\left(-\frac{x}{2k}+O\left(\frac{1}{k^2}\right)\right).$$

Therefore,

$$\lim_{n\to\infty}\sum_{k=n+1}^{2n}(a_k-e^x)=\lim_{n\to\infty}\sum_{k=n+1}^{2n}e^x\left(-\frac{x}{2k}+O\left(\frac{1}{k^2}\right)\right)$$
$$=-\frac{1}{2}xe^x\lim_{n\to\infty}\sum_{k=n+1}^{2n}\frac{1}{k}=-xe^x\ln\sqrt{2}.$$

Also solved by Michel Bataille (France), Jean Bogaert (Belgium), Paul Bracken and N. Nadeau, Robert Calcaterra, Eugene A. Herman, Hosam M. Mahmoud, Nicholas C. Singer, Albert Stadler (Switzerland), Michael Vowe (Switzerland), and the proposer. There was one incorrect submission.

Values of a polynomial **December 2005**

1734. *Proposed by H. A. ShahAli, Tehran, Iran.*

Let α be a fixed irrational number and let P be a polynomial with integer coefficients and with $\deg(P)\ge 1$. Prove that there are infinitely many pairs (m, n) of integers such that $P(m)=\lfloor\alpha n\rfloor$.

Solution by Nicholas C. Singer, Annandale, VA.

For positive real number A define $x \bmod A$ by

$$x \bmod A := x-A\left\lfloor\frac{x}{A}\right\rfloor.$$

Note that $x \bmod A\in[0,\ A)$ for all real x. Given integer m, set

$$n=n(m)=(\operatorname{sgn}\alpha)\left(\left\lfloor\frac{P(m)}{\alpha}\right\rfloor+1\right).$$

If $P(m) \bmod |\alpha| > |\alpha| - 1$, then

$$\lfloor n\alpha \rfloor = \left\lfloor |\alpha| \left(\left\lfloor \frac{P(m)}{|\alpha|} \right\rfloor + 1 \right) \right\rfloor = \lfloor P(m) + |\alpha| - (P(m) \bmod |\alpha|) \rfloor = P(m),$$

because $0 < |\alpha| - (P(m) \bmod |\alpha|) < 1$.

For $|\alpha| < 1$ the condition $P(m) \bmod |\alpha| > 0 > |\alpha| - 1$ holds automatically.

Now assume that $|\alpha| > 1$. Weyl's theorem says that for a nonconstant polynomial $p(x)$ with real coefficients and irrational leading coefficient, the sequence $\{p(n)\}$ is uniformly distributed mod 1. Thus, if $p(x) = P(x)/|\alpha|$, then there are infinitely many m for which $1 - \frac{1}{|\alpha|} < p(m) \bmod 1 < 1$. For these m we have $P(m) \bmod |\alpha| > |\alpha| - 1$.

Also solved by Gabriel Dospinescu (France) and Marian Tetiva (Romania), G.R.A.20 Math Problems Group (Italy), Albert Stadler (Switzerland), and the proposer. There was one incorrect submission.

Zeros of a polynomial **December 2005**

1735. *Proposed by George Gilbert, Texas Christian University, Fort Worth, TX.*

Find the complex zeros of the polynomial

$$p_n(z) = \det \begin{bmatrix} -1 & z & 0 & & & \cdots & 0 \\ 1-z & -1 & z & 0 & & \cdots & 0 \\ 0 & 1-z & -1 & z & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & 1-z & -1 & z \\ 0 & 0 & \cdots & & 0 & 1-z & -1 \end{bmatrix},$$

where the matrix is an $n \times n$ tridiagonal matrix.

Solution by Knut Dale, Telemark University College, Bø, Norway.

We have $p_1(z) = -1$ and $p_2(z) = 1 - z(1 - z)$. Expanding on the first row of the matrix leads to the recurrence formula

$$p_n + p_{n-1} + z(1 - z)p_{n-2} = 0.$$

The characteristic equation for this recurrence formula has solutions $t_1 = z - 1$ and $t_2 = -z$. If $z = \frac{1}{2}$ then $t_1 = t_2 = -\frac{1}{2}$ and $p_n(\frac{1}{2}) = (n + 1)(-\frac{1}{2})^n \neq 0$. If $z \neq \frac{1}{2}$ then

$$p_n(z) = \frac{1}{2z - 1} \left((z - 1)^{n+1} - (-z)^{n+1} \right).$$

To get $p_n(z) = 0$ we must have $z - 1 = -z\, w^k$ where $w = \exp(\frac{2\pi i}{n+1})$, $k = 0, 1, \ldots, n$, that is, $z = \dfrac{1}{1 + w^k}$. Since $z \neq \frac{1}{2}$ we must have $k \neq 0$. If $n = 2m - 1$ is odd, we must also have $k \neq m$ to avoid $w^k = -1$. Hence, if n is even p_n has n different zeros, and if n is odd p_n has $n - 1$ different zeros.

Also solved by Michel Bataille (France), Jean Bogaert (Belgium), Brian Bradie, Robert Calcaterra, David Callan, Chip Curtis, Paul Deiermann, Jim Delany, Matthew Hudelson, Kim McInturff, Gabriel T. Prăjitură, Nicholas C. Singer, Albert Stadler (Switzerland), Michael Vowe (Switzerland), and the proposer. There was one incorrect submission and one incomplete submission.

Problem 1720 Revisited **April 2005**

1720. *Proposed by Stephen J. Herschkorn, HIghland Park, NJ.*

Let X be a standard normal random variable and let a be a positive number. Show that $E[X \,|\, |X - a| < t]$ is strictly decreasing in nonnegative t.

Solution by the proposer.

Note that $E[Z \,|\, |Z - a| < t] = a + E[Z - a \,|\, |Z - a| < t]$. The desired result follows because the random variable $Z - a$ satisfies the hypotheses of the following result.

PROPOSITION. Let X be an absolutely continuous random variable whose range is an interval I symmetric about 0. Assume that X has a strictly positive, piecewise continuous density function f on I such that $f(-x)/f(x)$ is strictly increasing in nonnegative x. Then $E[X \,|\, |X| < t]$ is strictly decreasing in nonnegative $t \in I$.

Proof. Let $\beta(x) = f(-x)/f(x)$ and

$$g(t) = E[X \,|\, |X| < t] = \frac{\int_{-t}^{t} x f(x)\, dx}{\int_{-t}^{t} f(x)\, dx} = \frac{\int_{0}^{t} x(1 - \beta(x)) f(x)\, dx}{\int_{0}^{t} (1 + \beta(x)) f(x)\, dx}.$$

Then $g'(t)$ exists for all but at most countably many values of t, and where it exists,

$$g'(t) = \frac{f(t)}{(P[|X| < t])^2} \int_{0}^{t} h(x, t) f(x)\, dx,$$

where

$$h(x, t) = t\,(1 - \beta(t))\,(1 + \beta(x)) - x\,(1 + \beta(t))\,(1 - \beta(x))\,.$$

It suffices to show that $h(x, t) < 0$ for $0 < x < t$. To this end, note that

$$\frac{1 - \beta(y)}{1 + \beta(y)} = \frac{2}{1 + \beta(y)} - 1$$

is decreasing in nonnegative in y because β is strictly increasing. Thus for $0 < x < t$,

$$\frac{x\,(1 - \beta(x))}{1 + \beta(x)} > \frac{x\,(1 - \beta(t))}{1 + \beta(t)} > \frac{t\,(1 - \beta(t))}{1 + \beta(t)},$$

where the latter inequality follows from that fact that β is bounded below by $\beta(0) = 1$.

This concludes the proof of the proposition and establishes the desired result. ■

I thank Thomas Struppeck for the suggestion that the monotonicity of β was the crucial hypothesis.

Note. A solution to this problem appeared in the April 2006 issue. In that solution it was noted that the requested expected value was the x-coordinate of the center of mass of the region below the graph of $y = e^{-x^2}$ and above the x-axis for x between $a - t$ and $a + t$. It was then remarked that because of the shape of the graph, this x-coordinate was a decreasing function of $t > 0$. The proposer forwarded the following example showing that the decreasing nature of this coordinate is actually a subtle point. Let f

be defined on $\mathbb{R}$ by

$$f(x) = \begin{cases} 5 & x \le -1 \\ 3 - 2x & -1 < x \le 1 \\ 9x - 8 & 1 < x \le \dfrac{4}{3} \\ 4 & x > \dfrac{4}{3}. \end{cases}$$

Let $a(t)$ be the x-coordinate of the center mass of the region under the graph of $y = f(x)$ and above the x-axis for x between $-t$ and t. Then it is easily shown that $a(\frac{4}{3}) < a(2)$. Thus, as t increases from $\frac{4}{3}$ to 2, the x-coordinate of the center of mass of the region increases inspite of the fact that we are adding more mass on the left than on the right.

Answers

Solutions to the Quickies from page 394.

A965. It is well known that

$$\operatorname{rank}(AB) = \operatorname{rank}(B) - \dim(\operatorname{Im} B \cap \operatorname{Ker} A).$$

See, for example, Matrix Theory, by Fuzhen Zhang, Springer, 1999, page 46. Using the same result we also have

$$\operatorname{rank}(BA) = \operatorname{rank}(A^*B^*) = \operatorname{rank}(B^*) - \dim(\operatorname{Im} B^* \cap \operatorname{Ker} A^*).$$

Now $\operatorname{rank}(B) = \operatorname{rank}(B^*)$, and because A and B are normal, $\operatorname{Im} B = \operatorname{Im} B^*$ and $\operatorname{Ker} A = \operatorname{Ker} A^*$. It follows that $\operatorname{rank}(AB) = \operatorname{rank}(BA)$.

A966. For each positive integer n, let $M_n = \sup_{x \in S} |f_n(x) - f(x)|$. Because f and f_n are bounded, each M_n is finite. For each n there exists an $x_n \in S$ such that

$$|f_n(x_n) - f(x_n)| > M_n - \frac{1}{n}.$$

We then have

$$0 \le M_n < |f_n(x_n) - f(x_n)| + \frac{1}{n}.$$

Because $f_n(x_n) - f(x_n) \to 0$, it follows that $M_n \to 0$. Therefore f is the uniform limit of $\{f_n\}$ on S.

The result is not true if we are only given that $f_n(x_n) - f(x_n) \to 0$ for every sequence $\{x_n\}$ in S that converges to a value in S. As an example, take $f_n(x) = x^n$ and $f(x) = 0$ on $S = (0, 1)$. If $\{x_n\}$ is any sequence in S such that $x_n \to x \in S$, then $f_n(x_n) - f(x_n) = f_n(x_n) \to 0$. However, $\{f_n\}$ does not converge uniformly to f on S.

Editor's Note. In the February 2006 issue, Proposal 1736 and Quickie 958 appeared with the proposer designation "Anonymous." Both the proposal and the quickie were submitted by Claude Bégin of Quebec, Canada.

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.

Gerstein, Mark, Reclusive genius (letter), *New York Times* (3 September 2006), `http://www.nytimes.com/2006/09/03/opinion/103math.html` . Nasar, Sylvia, and David Gruber, Manifold destiny, *New Yorker* (28 August 2006) `http://www.newyorker.com/printables/fact/060828fa_fact2` . Unraveling toughest puzzle outstanding, rate mathematicians, *People's Daily* (China) (5 June 2006), `http://english.people.com.cn/200606/05/eng20060605_271113.html` . Shing-Tung Yau, `http://www.doctoryau.com` .

You have no doubt long since heard news of the award of Fields Medals at the International Congress of Mathematicians in Madrid in August, and of Grigory Perelman's refusal to accept his award. Perelman's conduct and attitude are a refreshing twist on the stereotype of mathematician as weirdo (think of Erdős, Gödel, Nash), harking back to Newton, who was reluctant to publish because "It would perhaps increase my acquaintance, the thing which I cheify [sic] study to decline." Perelman "apparently puts deep thought and seeking knowledge ahead of personal accolades and career recognition.... Perhaps tranquil reclusion is a prerequisite for brilliant thought" (Mark Gerstein). Perelman can remain reclusive, but tranquility will surely elude him; a tempest has begun. Two Chinese mathematicians, Zhu Xiping (Zhongshang University) and Cao Huaidong (Lehigh University) claim that Perelman provided only "guidelines," not a proof, for the Poincaré Conjecture, and that they have provided a "complete proof," building on the contributions of Perelman and others. Fame and celebrity are at stake, not to mention the Clay Foundation's $1 million prize. Perelman's comment: "Apparently, Zhu did not understand [my] argument and reworked it." As a side show, we have Shing-Tung Yau, mentor to Zhu and Cao and himself a Fields Medal winner (1982), alleging that Nasar and Gruber's article defames his reputation; his and others' responses to their article are at his Website. Meanwhile, Nasar and Gruber report that "Perelman repeatedly said that he had retired from the mathematics community and no longer considered himself a professional mathematician." (Even if that claim is erroneous, I somehow doubt that he will sue the authors.) Perhaps, unlike most sports champions but like chess champion Bobby Fischer, he is quitting while he is (at least momentarily) ahead. Who said that mathematics doesn't make for lively entertainment?

Hayes, Brian, Gauss's day of reckoning, *American Scientist* (May–June 2006) 200–205, `http://www.americanscientist.org/template/AssetDetail/assetid/50686` .

Author Hayes investigates the famous tale about Gauss's ingenious way of beating the tedium of the in-school assignment of summing the integers from 1 to 100. Quips Hayes, "it's really hard to do it the hard way," because shortcuts become obvious. His conclusions are that there is no account from Gauss's lifetime (the first is in a funerary tribute by a colleague at Göttingen), we do not know how Gauss solved the problem, and we need to show that there is a place in mathematics for students who were not brilliant at age seven.

Roberts, Siobhan, *King of Infinite Space: Donald Coxeter, the Man Who Saved Geometry*, Walker & Company, 2006; xv + 386 pp, $27.95. ISBN 10: 0–8027–1499–4; ISBN-13: 978–0–8027–1499–2.

Donald Coxeter (1907–2003), "guided almost completely by a profound sense of what is beautiful" (Freeman Dyson), almost single-handedly brought geometry back from the death to which it was condemned by the Bourbaki movement and the New Math (a pedagogical "reform" movement in the 1960s that emphasized terminology and abstraction). This engaging and thoroughly-researched book (100 pp of endnotes and bibliography) relates Coxeter's life without going into great detail about his geometry; his encounters with Wittgenstein, Escher, and Buckminster Fuller, and Jeff Weeks reflect his lifelong interest in symmetry. One of the several appendices lists Coxeter's mathematical publications. Aargh! No index!—that's really unacceptable for a definitive biography.

Nahin, Paul J., *Dr. Euler's Fabulous Formula Cures Many Mathematical Ills*, Princeton University Press, 2006; xx + 380 pp, $29.95. ISBN 0–691–11822–1.

This book supplements the author's *An Imaginary Tale: The Story of* $\sqrt{-1}$ (1998) with what he considers the "sexy part" of complex analysis. (To the author: Saying doesn't make it so, false advertising does not help the cause, and surely mathematics should arouse other passions. But I like the book.) The author asks for two years of calculus, differential equations, and matrix algebra and probability, leaving "more than a few otherwise educated readers out in the cold." Since on p. 10 the reader contemplates the infinite product expansion for the sine function, the author is definitely right about the prerequisites, and the "more than a few" would likely to freeze to death before finishing the subsequent 300 pp. The book is not just about $e^{\pi i} + 1 = 0$ (which one of our math majors has tattooed on his forearm!); it makes strong connections that should inspire mathematics students, and may rescue complex numbers from the oblivion that the high school and college curricula consign them to—a cure much to be wished for. Also, where else would you be reminded that Jean-Luc Picard of *Star Trek* finds relaxation in the 24th century trying to resolve Fermat's Last Theorem (will Wiles prove to be wrong after all???)?

Daniel, James W., and Leslie Jane Federer Vaaler, *Mathematical Interest Theory*, Pearson Prentice Hall, 2007; xv + 496 pp, $113.33. ISBN 0–13–147285–2.

"Mathematics of Finance" fell out of the college curriculum years ago but is making a partial comeback in courses in quantitative literacy. Meanwhile, the first few actuarial exams have altered emphasis away from mathematics and toward economics, reflecting the fact that insurance companies earn far more from investments than from premiums. But an understanding of interest calculations is still essential for an actuary, important in business, and useful in personal finance. New textbooks on the subject are rare; Stephen G. Kellison's *Theory of Interest* (2nd ed., 1991) has long dominated the field. This new book includes an introduction to the use of the Texas Instruments BA II Plus calculator; the usual material on interest and return, annuities, loans, and bonds; and further material on stocks, arbitrage, and interest rate sensitivity. Answers are given to all exercises, and each chapter also features problems whose answer requires writing a paragraph or essay.

Stewart, Ian, *Letters to a Young Mathematician*, Basic Books, 2006; xii + 210 pp, $22.95. ISBN 0–465–08231–9.

Stewart takes off from Hardy's *A Mathematician's Apology* (1940), offering updated attitudes about the nature of the mathematical enterprise, in the form of letters to an imaginary aspiring mathematician as she advances from high school to assistant professor. The book will also interest parents and friends of potential mathematicians, with its letters answering questions such as: Why do math? What is math? Hasn't it all been done? Can't computers do it all? Why are proofs necessary? How do mathematicians see the world differently? What's the best way to learn mathematics? ("Read around your subject. Do not read only the assigned text.") The last few letters address how to advance in the academic profession.

NEWS AND LETTERS

Acknowledgments

In addition to our Associate Editors, the following referees have assisted the MAGAZINE *during the past year. We thank them for their time and care.*

Aboufadel, Edwad F., *Grand Valley State University, Allendale, MI*

Adams, Colin, *Williams College, Williamstown, MA*

Apostol, Tom, *Caltech, Pasadena, CA*

Appleby, Glenn D., *Santa Clara University, Santa Clara, CA*

Ash, J. Marshall, *De Paul University, Chicago, IL*

Axtell, Michael C., *Wabash, Crawfordsville, IN*

Baeth, Nicholas R., *Central Missouri State University, Warrensburg, MO*

Barksdale, James B. Jr., *Western Kentucky University, Bowling Green, KY*

Bauer, Craig P., *York College of PA, York, PA*

Beauregard, Raymond A., *University of Rhode Island, Kingston, RI*

Beezer, Robert A., *University of Puget Sound, Tacoma, WA*

Benjamin Arthur T., *Harvey Mudd College, Claremont, CA*

Bennett, Curtis D., *Loyola Marymount University, Los Angeles, CA*

Berndt, Bruce, *University of Illinois at Urbana-Champaign, Urbana, IL*

Bhutani, Kiran R., *The Catholic University of America, Washington, DC*

Blair, Steve, *Grand Valley State University, Allendale, MI*

Blecksmith, Richard, *Northern Illinois University, Dekalb, IL*

Bogomolny, Alex, *East Brunswick, NJ*

Bonan-Hamada, Catherine, *Mesa State College, Grand Junction, CO*

Borwein, Jonathan, *Dalhousie University, Halifax, NS Canada*

Bradley, Robert, *Adelphia University, Garden City, NJ*

Bressoud, David M., *Macalester College, Saint Paul, MN*

Bullington, Grady, *University of Wisconsin, Oshkosh, Oshkosh, WI*

Calkin, Neil J., *Clemson University, Clemson, SC*

Callan, David, *University of Wisconsin, Madison, WI*

Canada, Daniel, *Spokane, WA*

Case, Jeremy, *Taylor University, Upland, IN*

Chen, Hang, *Central Missouri State University, Warrensburg, MO*

Craft, David, *Muskingum College, New Concord, OH*

Curran, Stephen J., *University of Pittsburgh at Johnstown, Johnstown, PA*

Davis, Philip J., *Brown University, Providence, RI*

De Angelis, Valerio, *Xavier University of Louisiana, New Orleans, LA*

DeTemple, Duane, *Washington State University, Pullman, WA*

Dickinson, William C., *Grand Valley State University, Allendale, MI*

Dodge, Clayton, *Orono, ME*

Dunne, Edward, *American Mathematical Society, Providence, RI*

Edwards, Christopher, *University of Wisconsin, Oshkosh, Oshkosh, WI*

Eenigenburg, Paul, *Western Michigan University, Kalamazoo, MI*

Emert, John Wesley, *Ball State University, Muncie, IN*

Ensley, Douglas, E., *Shippensburg University of Pennsylvania, Shippensburg, PA*

Epp, Susanna, *DePaul University, Chicago, IL*

Eroh, Linda, *University of Wisconsin, Oshkosh, Oshkosh, WI*

Evans, Anthony, *Wright State University, Dayton, OH*

Feil, Todd, *Denison University, Grandville, OH*

Feroe, John, *Vassar College, Poughkeepsie, NY*

Fisher, J. Chris, *University of Regina, Regina, Saskatchewan, Canada*

Fraser, Craig, *University of Toronto, Toronto, ON, Canada*

Fredricks, Gregory, *Lewis and Clark University, Portland, OR*

Fung, Maria, *Western Oregon University, Monmouth, OR*

Gallian, Joseph A., *University of Minnesota Duluth, Duluth, MN*

Goddard, Wayne D., *Clemson University, Clemson, SC*

Graham, Ronald L., *La Jolla, CA*

Grossman, Jerrold W., *Oakland University, Rochester, MI*

Grünbaum, Branko, *University of Washington, Seattle, WA*

Guy, Richard K., *University of Calgary, Calgary, AB, Canada*

Hamburger, Peter, *Indiana University-Purdue, Fort Wayne, IN*

Hayes, Leslie, *St. Joseph University, Philadelphia, PA*

Henle, James, *Smith College, Northampton, MA*

Hodge, Jonathan, *Grand Valley State University, Allendale, MI*

Hoehn, Larry, *Austin Peay State University, Clarksville, TN*

Hoffoss, Diane, *University of San Diego, CA*

Howard, Fred, *Wake Forest University, Winston-Salem, NC*

Hull, Thomas, *Merrimack College, North Andover, MA*

Isaksen, Daniel C., *Wayne State University, Detroit, MI*

Jepsen, Charles H., *Grinnell College, Grinnell, IA*

Johnson, Brenda, *Union College, Schenectady, NY*

Kallaher, Michael, *Pullman, WA*

Kalman, Dan, *American University, Washington, DC*

Kemp, Daniel C., *South Dakota State University, Brookings, SD*

Kerckhove, Michael G., *University of Richmond, Richmond, VA*

Kleiner, Israel, *York University, Toronto, ON, Canada*

Knapp, Michael P., *Loyola College of MD., Baltimore, MD*

Kung, Sidney, *Cupertino, CA*

Kwong, Harris, *State University of New York at Fredonia, Fredonia, NY*

Lang, Robert J., *Alamo, CA*

Larson, Loren, *Saint Olaf College, Northfield, MN*

Lautzenheiser, Roger, *Rose Hulman Inst. of Technology, Terre Haute, IN*

Ledyaev, Yuri, *Western Michigan University, Kalamazoo, MI*

Lo Bello, Anthony, *Allegheny College, Meadville, PA*

Loud, Warren S., *Minneapolis, MN*

McCleary, John H., *Vassar College, Pouhkeepsie, NY*

McLaughlin, James G., *West Chester University of Pennsylvania, West Chester, PA*

Merrill, Kathy D., *Mancos, CO*

Meyerson, Mark D., *US Naval Academy, Annapolis, MD*

Michael, T. S., *US Naval Academy, Annapolis, MD*

Miller, Steven, *Brown University, Providence, RI*

Miller, Zevi, *Miami University, Oxford, OH*

Molinek, Donna K., *Davidson College, Davidson, NC*

Monson, Barry R., *University of Brunswick, Fredericton, NB, Canada*

Montgomery, Hugh L., *University of Michigan, Ann Arbor, MI*

Moretti, Christopher, *Southeastern Oklahoma State University, Durant, OK*

Morics, Steven W., *Redlands, CA*

Moses, Peter, *Moparmatic Company, Astwood Bank, UK*

Nelson, Donald J., *Western Michigan University, Kalamazoo, MI*

Nowakowski, Richard, *Dalhousie University, Halifax, NS, Canada*

Oakley, Patricia A., *Goshen College, Goshen, IN*

O'Leary, Michael, *College of DuPage, Glen Ellyn, IL*

O'Loughlin, Daniel J., *The College of St. Catherine, St. Paul, MN*

Olsovsky, Gregor, M., *Behrend College, Erie, PA*

Otero, Daniel E., *Xavier University, Cincinnati, OH*

Pence, Dennis, *Western Michigan University, Kalamazoo, MI*

Perkel, Manley, *Wright State University, Dayton, OH*

Perline, Ronald K., *Drexel University, Philadelphia, PA*

Index to Volume 79

AUTHORS

TITLES

PROBLEMS

The letters P, Q, and S refer to Proposals, Quickies, and Solutions, respectively; page numbers appear in parentheses. For example, Q959(151) refers to Quickie 959, which appears on page 151.

Deutsch, Emeric, P1739(67)

Díaz-Barrero, José Luis, P1738(67)

Doucette, Robert, S1714(71)

Edgar, Gerald A., P1745(151)

Fejéntalálltuka Szeged Problem Solving Group, S1731(395)

George, John C., S1753(309)

Goldenberg, Michael, and Mark Kaplan, P1744(151), P1758(393), S1733(396)

G.R.A. 20 Problems Group, S1718(316)

Herman, Eugene, S1725(223), S1730(314)

Herschkorn, Stephen, P1746(214), P1747(214), S1720(398)

Hillar, Christopher J., P1750(219)

Ivády, Péter, P1760(394)

Jaroma, John H., Q957(68)

Kaplan, Mark, and Michael Goldenberg P1744(151), P1758(393), S1733(396)

Kimberling, Clark, Q959(151)

Klamkin, Murray, Q962(219)

Kwong, Harris, S1719(154)

Lang, David P., P1743(150)

Langford, Jeffrey and Luz M. DeAlba, P1742(150)

Lee, Jung-Jin, Q964(310)

Leong, Tom, S1728(312)

Lindstrom, Peter W., S1724(222)

Moen, Courtney H. and William P. Wardlaw, P1756(393)

Moser, William, S1718(315)

Nogin, Maria, and Larry W. Cusick, P1759(394)

Northwestern Math Problem Solving Group, S1726(310)

Rosentrater, C. Ray, S1715(72)

Ross, Ken, P1757(393)

ShahAli, H. A., Q963(310)

Singer, Nicholas C, S1734(396)

Spivey, Michael Z., P1737(67)

Sternitzky, John and Robert Calcaterra, P1742(309)

Trenkler, Götz, Q965(394)

Wardlaw, William P., and Courtney H. Moen, P1756(393)

Zerger, Tom, S1717(152)

Zhou, Li, S1720(155), S1724(222)

Thanks too to Steve Dunbar for refereeing many problems.

CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, DC 20036