Vol. 90, No. 2, April 2017

MAA

MATHEMATICS MAGAZINE

- The dodecahedron and icosahedron are Rupert
- Tipping weebles and eggs
- Chasing lights in Lights Out
- Polyominoes and errant pixels for digital cameras

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.

Submissions of articles are required via the *Mathematics Magazine*'s Editorial Manager System. The name(s) of the author(s) should not appear in the file. Initial submissions in pdf or LaTeX form can be sent to the editor at www.editorialmanager.com/mathmag/.

The Editorial Manager System will cue the author for all required information concerning the paper. Questions concerning submission of papers can be addressed to the editor at mathmag@maa.org. Authors who use LaTeX are urged to use the Magazine article template. However, a LaTeX file that uses a generic article class with no custom formatting is acceptable. The template and the Guidelines for Authors can be downloaded from www.maa.org/pubs/mathmag.

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, NW, Washington, DC 20036 and Lancaster, PA, in the months of February, April, June, October, and December.

Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Address advertising correspondence to

MAA Advertising
1529 Eighteenth St. NW
Washington, DC 20036
Phone: (202) 319-8461
E-mail: advertising@maa.org

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

Change of address, missing issue inquiries, and other subscription correspondence can be sent to:

The MAA Customer Service Center
P.O. Box 91112
Washington, DC 20090-1112
(800) 331-1622
(301) 617-7800
maaservice@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, NW, Washington, DC 20036-1385.

Printed in the United States of America.

COVER IMAGE

Qunitic © 2017 David A. Reimann (*Albion College*). Used by permission.
Contour lines of quintic polynomial $f(z) = z^5 – 5z + 12$ are depicted for $|z| \leq 3$. The white lines show the magnitude of f at {2, 4, 8, 16, 32, 64, 128}. The black lines show the functions phase in multiples of $\pi/4$. The paper by Lunsford was the inspiration for the cover art.

Vol. 90, No. 2, April 2017

MATHEMATICS MAGAZINE

LETTER FROM THE EDITOR

Can a cube pass through another cube of the same size? No, this is not the beginning of a special April Fools' Day article. Other polyhedra can pass through themselves, too. Whether or not such polyhedra exist is known as Rupert's problem. The April issue begins with an article by Richard Jerrard, John Wetzel, and Liping Yuan that shows that the dodecahedron and icosahedron satisfy the Rupert property.

As a child of the 1980s, I remember the advertising slogan "Weebles wobble, but they don't fall down!" And, although I never had a Weeble growing up, I certainly remember the ad campaign. Subhranil De examines the mathematics of Weebles and other roly poly toys and explains why these toys right themselves when tilted.

In English, we read from left to right and compute decimals that way, too. Indeed, divide 4175 by 17; if you used long division, then you were working from left to right. Is it possible to work from right to left to determine the decimal expansion of a rational number? No, this article is not an April Fools' Day joke either. Surprisingly, the answer to this question is "yes." Amitabha Tripathi describes how to do so in the third article of the issue.

Other articles in the issue include Matt Lunsford's foray into Galois theory over prime finite fields, in which he seeks a condition for the existence of irreducible radical extensions. Also, Paul Levrie applies Euler's product formula for the gamma function to consider the asymptotic behavior of binomial coefficients to show the interval of convergence of the Maclaurin series for $f(x) = (1 + x)^{\alpha}$. C. David Leach revisits the Lights Out puzzle and explains a process known as light chasing and how a look-up table can be constructed so that light chasing can be used to solve (when possible) the Lights Out puzzle.

In the penultimate article, Daniel Heath and Robert Rydberg view defects in a digital image as polyominoes. They use the polyominoes to count the number of multipixel defects to solve an engineering problem related to manufacturing a digital camera. Romeo Mestrovic provides a congruence modulo p^3, p a prime, that is reminiscent of a similar congruence by Morley from 1895.

In between the articles are three proofs without words. Ángel Plaza visually proves a result involving the partial column sums in Pascal's triangle, while Tom Edgar offers a proof without words that shows that triangular numbers and other figurate numbers satisfy a recursive equation. Roger Nelsen uses pictures to prove Diophantus of Alexandria's sum of squares identity.

There is no crossword puzzle in this issue, but there is a puzzle by Lai Van Duc Thinh. The Pinemi puzzle is reminiscent of Mine Sweeper with an additivity constraint. There are also two poems in the issue. One is by Kay Shultz and the other is by Del Corey. As usual, there are Problems and Reviews. At the end of the Problems section, Adrian Chun Pong Chu generalizes the solution to a past Quickie. The issue concludes with the solutions to the 77th William Lowell Putnam Competition.

Let me sign off by thanking Paul Stockmeyer for serving as an associate editor on the editorial board of MATHEMATICS MAGAZINE since 2007.

Michael A. Jones, Editor

ARTICLES

Platonic Passages

RICHARD P. JERRARD[1]
Department of Mathematics
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

JOHN E. WETZEL[2]
1601 Lakeside Drive Unit A
Champaign, IL 61821
jewetz@comcast.net

LIPING YUAN
Hebei Normal University
Shijiazhuang 050016 P. R. China
lpyuan@hebtu.edu.cn

According to J. Wallis [**11**, pp. 470–71], more than three hundred years ago Prince Rupert of the Rhine[3] (1619–1682) won a wager that a hole could be cut through a cube large enough to permit another cube of the same size to slide through. In 1950, D. J. E. Schrek [**7**] published a detailed proof of this somewhat surprising fact together with a careful review of its history. C. J. Scriba [**8**] showed in 1968 that the regular tetrahedron and octahedron have this same property: each can transit through a suitable tunnel in another of the same size and type.

We show here that the remaining two platonic solids, the dodecahedron and icosahedron, also have this property (as announced in [**3**]).

Many convex bodies in R^3 share this Rupert property, but not all. It is easy to give examples of convex bodies that do not have the Rupert property, for example, the unit ball and the equilateral drum (a circular cylinder of unit diameter and height closed on each end by disks). But we know of no convex polyhedron that does not have the Rupert property. With a certain hesitancy, we even suggest that perhaps every convex polyhedron in R^3 has the Rupert property. In any case, an example of a convex polyhedron lacking the Rupert property would be of considerable interest.

Preliminaries

By a *convex body* in R^3 we mean a compact, convex set with nonempty interior. There seems to be little ambiguity in what is meant by the somewhat imprecise language, "a hole can be cut" in a convex body $\mathfrak{K}$; by a "hole" is meant simply a "straight tunnel."

Let π_n be a plane with unit normal vector n, and let $\varpi_n : R^3 \to \pi_n$ be the projection map of R^3 onto π_n. Let γ be a simple closed curve that lies in the plane π_n, and let I_γ be the domain in π_n interior to γ.

Math. Mag. **90** (2017) 87–98. doi:10.4169/math.mag.90.2.87.
MSC: Primary 51M16, Secondary 51M04; 52C15.

[1]Deceased July 23, 2014.

[2]Corresponding author.

[3]Count Palatine of the Rhine and Duke of Bavaria, son of Frederick V, the Winter King, Elector Palatine, and king of Bohemia, and Elizabeth, daughter of James I of England.

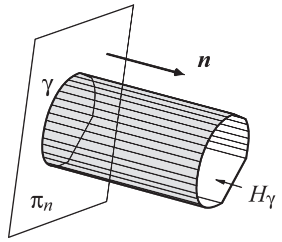


Figure 1 A right cylinder.

The *right cylinder* C_γ with *directrix* γ and *direction* n (Figure 1) is the set

$$C_\gamma = \{y + tn \in R^3:\ y \in \gamma,\ -\infty < t < \infty\},$$

and the *tunnel* H_γ defined by the right cylinder C_γ is just its interior region:

$$H_\gamma = \{y + tn \in R^3:\ y \in I_\gamma,\ -\infty < t < \infty\}.$$

Since $\mathfrak{K}$ is a convex body, its projection $\varpi_n(\mathfrak{K})$ on a plane π_n is a convex set with nonempty interior, and it follows that the boundary of the projection is a simple closed convex curve γ. The passage of $\mathfrak{K}$ through the tunnel H_γ determined by the right cylinder with directrix γ and direction n is described completely by

$$\mathfrak{K}_t = tn + \mathfrak{K} \subset H_\gamma, \quad -\infty < t < \infty.$$

Note that H_γ is an open set in R^3; so we demand in particular that $\mathfrak{K}_t$ not touch the bounding cylinder C_γ during the transit.

Now the following fundamental fact is virtually obvious.

Theorem 1. *Let $\mathfrak{K}$ be a convex body in R^3. If there are planes π_m and π_n so that the projection $P_i = \varpi_n(\mathfrak{K})$ of $\mathfrak{K}$ onto the plane π_n fits in the interior of the projection $P_o = \varpi_m(\mathfrak{K})$ of $\mathfrak{K}$ onto the plane π_m, then $\mathfrak{K}$ can be passed through the tunnel H_γ whose direction is m and whose directrix γ is the boundary of the projection P_o.*

We call P_i and P_o the *inner* and *outer* projections of $\mathfrak{K}$, respectively.

When the conditions of this theorem are met, we say that the convex body $\mathfrak{K}$ has the *Rupert property,* or is *Rupert*.

Nieuwland constants If a convex body $\mathfrak{K}$ has the Rupert property, a natural question to ask (and asked by Rupert in the case of the cube) is how large a body $\mathfrak{K}'$ similar to $\mathfrak{K}$ can be passed through a hole in $\mathfrak{K}$, i.e., for how large a positive scalar ν can the convex body $\nu\mathfrak{K}$ be passed through a suitable tunnel in $\mathfrak{K}$? We call this *Nieuwland's question* after P. Nieuwland (1764–1794), who answered this question for the cube. (Nieuwland's results for the cube were published posthumously by Swinden [**10**, pp. 512–513, 608–610].) Define the Nieuwland constant $\nu(\mathfrak{K})$ of a convex body $\mathfrak{K}$ by

$$\nu(\mathfrak{K}) = \sup\{\nu > 0 : \text{there is a tunnel in } \mathfrak{K} \text{ through which } \nu\mathfrak{K} \text{ can pass}\}.$$

Thus $\mathfrak{K}$ has the Rupert property if and only if $\nu(\mathfrak{K}) \geq 1$. Nieuwland showed that $\nu(\mathfrak{K}) = \frac{3}{4}\sqrt{2}$ if $\mathfrak{K}$ is a cube, so a cube of any edge $e < \frac{3}{4}\sqrt{2}$ can be passed through a suitable tunnel in a unit cube, but no cube of edge $e > \frac{3}{4}\sqrt{2}$ can be so passed. Determining the Nieuwland constant for a convex body $\mathfrak{K}$ is generally difficult, in part because there may be a multiplicity of tunnels to consider.

Table 1 collects the known estimates of the Nieuwland constants for the tetrahedron, cube, and octahedron, and it includes the new estimates for the dodecahedron and icosahedron given by Jerrard and described in the last section.

TABLE 1: Nieuwland constant estimates.

Platonic Solid	Nieuwland Estimate
Tetrahedron $\mathfrak{T}$	$\nu(\mathfrak{T}) \geq \frac{2}{5}\sqrt{3}(\sqrt{6}-1) > 1.004\,235$
Cube $\mathfrak{C}$	$\nu(\mathfrak{C}) = \frac{3}{4}\sqrt{2} \geq 1.060\,660$
Octahedron $\mathfrak{O}$	$\nu(\mathfrak{O}) \geq \frac{3}{4}\sqrt{2} \geq 1.060\,660$
Dodecahedron $\mathfrak{D}$	$\nu(\mathfrak{D}) \geq \frac{171}{170} > 1.005\,882$
Icosahedron $\mathfrak{I}$	$\nu(\mathfrak{I}) \geq 1108/1098 > 1.009\,107$

Platonic solids, I

According to Theorem 1, to show that a convex body $\mathfrak{K}$ has the Rupert property, we must exhibit the two projections $P_o(\mathfrak{K})$ and $P_i(\mathfrak{K})$ so that $P_i(\mathfrak{K})$ fits in the (relative) interior of $P_o(\mathfrak{K})$. The tetrahedron, cube, and octahedron can easily be handled geometrically, but the dodecahedron and icosahedron are a bit more difficult. We begin with the tetrahedron.

Tetrahedron Let $\mathfrak{T} = ABCD$ be a regular tetrahedron with unit edge labeled with the equilateral triangle BCD as base and apex A, and take the outer projection P_o to be an equilateral triangle CDE of unit side, placed as shown in Figure 2a with E in the plane π through the line CD perpendicular to the plane of the base of $\mathfrak{T}$. The orthogonal projection of $\mathfrak{T}$ into π nearly fits into the interior of P_o, the difficulty being at the vertices C and D. But clearly a small rotation of $\mathfrak{T}$ about the altitude EM

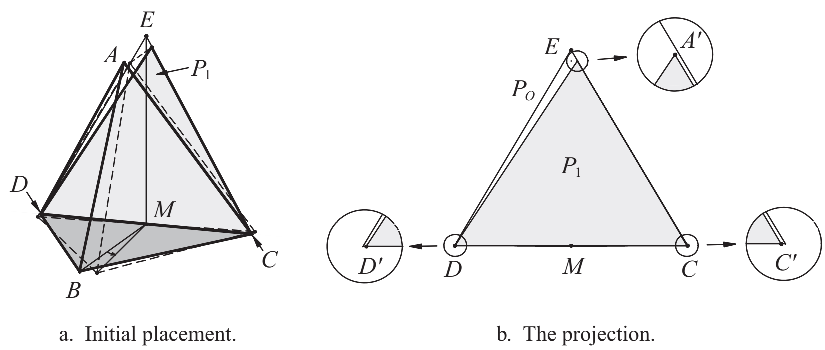


Figure 2 The tetrahedron $\mathfrak{T}$.

of triangle CED moves the projections C' and D' of C and D on π into the edge CD of triangle CED keeping the projection A' of A inside P_o. It follows that $\mathfrak{T}$ has the Rupert property, because a small upward translation of the rotated $\mathfrak{T}$ moves the projections A', B', C', and D' of all four vertices inside P_o.

To estimate the Nieuwland constant $\nu(\mathfrak{T})$ we chose the angle ϑ of rotation so that the projected segment $A'C'$ lies inside P_o and parallel to the side EC (Figure 2b). One can show that

$$\vartheta = 60^\circ - \arcsin \tfrac{1}{3}\sqrt{6} \approx 5.264\,389^\circ. \tag{1}$$

Finally, since $A'E > C'C = D'D$, $CD = 1$, and $C'D' = \cos\vartheta$, the least upper bound of the ratio by which T can be expanded and still pass through a similarly situated tunnel in T is

$$\frac{CD}{C'D'} = \sec\vartheta = \frac{2}{5}\sqrt{3}(\sqrt{6}-1) > 1.004\,235, \tag{2}$$

where the surd expression follows from (1). It seems likely that $\lambda_{\mathfrak{T}} = \frac{2}{5}\sqrt{3}(\sqrt{6}-1)$, but in any case, $\nu(\mathfrak{T}) > 1.004\,235$.

Remark. A similar argument shows that every tetrahedron (not just the regular tetrahedron $\mathfrak{T}$) is Rupert. (Let P_o be the face with the greatest area, so that the projection of the fourth vertex into the plane of P_o lies in P_o.)

Cube The fact that the cube has the Rupert property is very well known and has become a staple of recreational mathematics. A careful and detailed development was given in 1950 by Schrek [**7**] (see [**4**]); here we settle for a drawing that shows the essential details. Let $\mathfrak{C}$ be the unit cube with opposite vertices A and B. The two points that lie on adjacent edges of $\mathfrak{C}$ at distance $3/4$ from A and the two points on the two parallel edges at distance $3/4$ from B (as pictured in Figure 3) are coplanar, and the plane on which they lie meets the cube in a square of side $\frac{3}{4}\sqrt{2} \approx 1.06 > 1$,

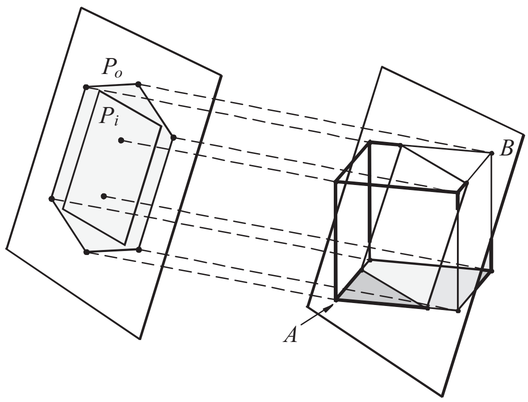


Figure 3 The cube $\mathfrak{C}$.

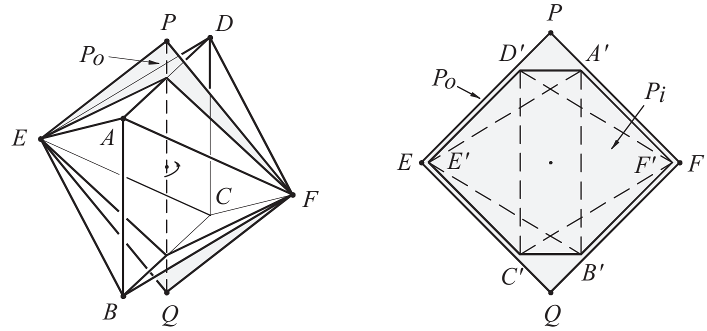
a square that is, in fact, the largest square that fits in the cube. Take P_o to be the projection of $\mathfrak{C}$ on that plane (pictured on a parallel plane in Figure 3 for clarity), and take the inner projection P_i to be the unit square. Since the inner projection P_i fits in the interior of the outer projection P_o, the cube has the Rupert property. Scaling the cube so that its largest square just touches the edges of P_o shows that $\nu(\mathfrak{C}) = \frac{3}{4}\sqrt{2}$ (for details, see [**7**]).

Remarks. Every box (i.e., every rectangular parallelepiped) is Rupert (see [**4**]). In fact, similar methods show that every parallelepiped is Rupert.

Determining the Nieuwland constant $\nu(\mathfrak{K})$ of the cube involves finding the largest square that fits in a unit cube. In his "Mathematical Games" column in the November 1966 issue of *Scientific American,* Martin Gardner asked for the largest cube that fits in a tesseract (a 4-cube) of unit edge (see [**1**, 172–73]). The question was answered in 1996, when Kay Shultz showed that the edge of the largest such cube is the square root of the smaller of the two real roots of the polynomial $4x^4 - 28x^3 - 7x^2 + 16x + 16$, approximately 1.0074348. (See the end of this article for a poem by Kay Schulz about solving this problem.)

More generally, what is the edge $f(m,n)$ of the largest m-cube that can fit in an n-cube of unit edge? Most of what is currently known about $f(m,n)$ is summarized by G. Huber in his preface to his reprinting of Shultz's 1996 notes, [**9**].

Octahedron The unit octahedron, $\mathfrak{O} = EABCDF$ in Figure 4a, is formed by eight equilateral faces formed into two congruent four-sided pyramids sharing a square base.



a. Projections. b. P_i fits in P_o.

Figure 4 The octahedron $\mathfrak{O}$.

Its projection in the direction $\overrightarrow{EF}$ is a unit square; let P_o be the unit square $PEQF$ positioned as pictured, with one diagonal EF and the other diagonal passing through the midpoints of the sides AD and CD of the medial square $ABCD$. It is clear that a small rotation of $\mathfrak{O}$ about the axis PQ (leaving the square P_o in place) moves the projections E' and F' into the square while leaving the projections A', B', C', and D' of the vertices A, B, C, and D inside the square; in other words, the projection P_i of the rotated octahedron $\mathfrak{O}$ lies in the interior of P_o (Figure 4b). This establishes the claim that the octahedron has the Rupert property.

To bound the Nieuwland constant $\nu(\mathfrak{O})$, we choose the angle of rotation so that the projection $E'C'$ on P_o of the rotated edge EC is parallel to the edge EQ of the square. One can see that the appropriate angle ϑ of rotation is given by

$$\vartheta = \arccos \frac{1}{\sqrt{3}} - \arccos \sqrt{\frac{2}{3}} \approx 19.5^\circ. \tag{3}$$

Then the projections of the rotated edges BF, FA, and ED are all parallel to sides of the unit square, and the projection P_i of (the rotated octahedron) $\mathfrak{O}$ is as pictured in Figure 4b. A short calculation from (3) shows that

$$\cos \vartheta = \frac{2\sqrt{2}}{3},$$

and it follows that

$$E'F' = \sqrt{2} \cos \vartheta = \frac{4}{3}.$$

Consequently,

$$\nu(\mathfrak{O}) \geq \frac{EF}{E'F'} = \frac{3\sqrt{2}}{4} > 1.060\,660,$$

as shown in Table 1. It seems likely that the equality holds.

Remark. Similarly, the cuboctahedron (the solid formed by clipping each vertex of a cube by a plane through the midpoint of the three adjacent edges) is Rupert.

Platonic solids, II

We employ visual geometric reasoning to handle the dodecahedron, and then, for variety, we use coordinate methods to deal with the icosahedron. Many of these results were first worked out by Jerrard prior to 2005, and some of what follows has been assembled from notes he prepared more than a decade ago.

Dodecahedron Let O be the center of symmetry of the dodecahedron $\mathfrak{D}$, and label the 20 vertices of $\mathfrak{D}$ as illustrated in Figure 5. The vertices and edges visible "from above" are drawn bold, and those not visible are drawn in gray. Note that the vertices labeled k and $21 - k$ are symmetric with respect to the center O. Let P be the center of the face 1-2-3-4-5 and M the midpoint of the edge 7-12.

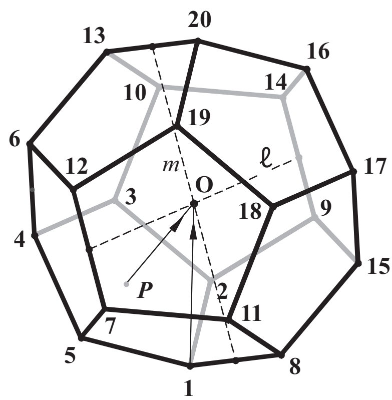


Figure 5 Labeled dodecahedron $\mathfrak{D}$.

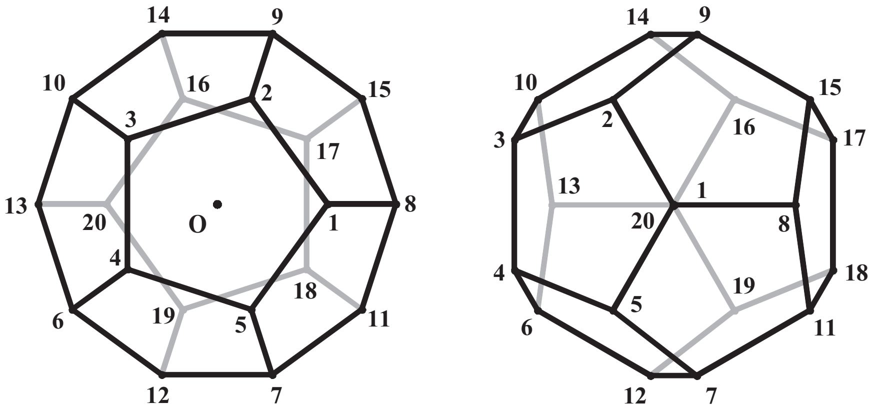
Let P_o be the projection of $\mathfrak{D}$ onto a faceplane, viz., in the direction of $\overrightarrow{PO}$, pictured in Figure 6a. Its boundary is a regular decagon.

Let P_1 be the projection of $\mathfrak{D}$ from a vertex, viz., in the direction of $\overrightarrow{1O}$, pictured in Figure 6b. Its boundary is an irregular dodecagon whose opposite edges are parallel.

In Figure 7a, the projection of the vertex labeled k of $\mathfrak{D}$ retains the label k in the projection P_1 but is underscored in the projection P_o.

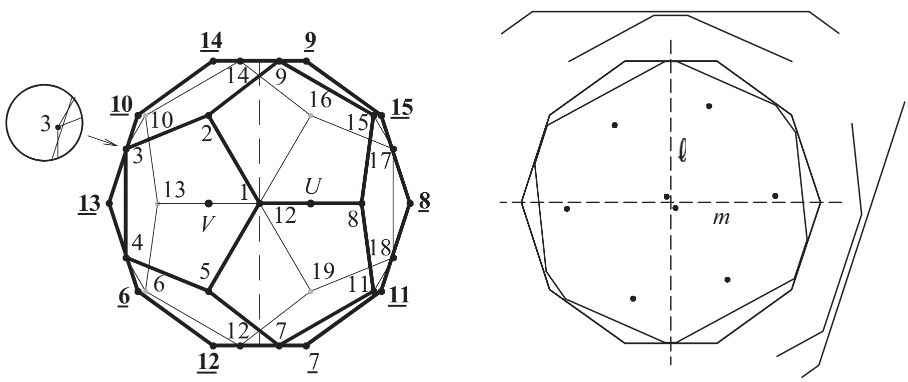
Placed as pictured with its center of symmetry at the center of P_o, the projection P_1 very nearly fits in P_o. The edge 14-9 of $\mathfrak{D}$ projects into the edge $\underline{14}$-$\underline{9}$ of the decagon boundary of P_o, the opposite edge 12-7 projects into the edge $\underline{12}$-$\underline{7}$, and the four vertices 3, 6, 17, and 18 project to points that lie just outside of P_o, as illustrated in the detail.

The dodecahedron can be rotated a little so that its projection P_i fits in the interior of P_o. We first make a small rotation of $\mathfrak{D}$ about the axis ℓ that joins the midpoints of the edges 14-9 and 12-7, moving the projections of both the edges 3-4 and 17-18 into the interior of P_o. This small rotation leaves the projected edges 14-9 and 12-7 on the edges $\underline{14}$-$\underline{9}$ and $\underline{12}$-$\underline{7}$. The axis m that joins the midpoints of the edges 8-1 and 13-20 (most easily seen in Figure 5), is parallel to and midway between the edges 7-12 and 9-14. It follows that a second small rotation of the rotated dodecahedron $\mathfrak{D}$ about m moves both the rotated edges 9-14 and 7-12 into the interior of P_o. Although this rotation moves the images of the four vertices 3, 4, 17, and 18 toward the boundary decagon of P_o, if the rotation is sufficiently small it leaves the projection of all of the vertices of the rotated dodecahedron (the 12 vertices of the interior irregular dodecagon



a. P_o. b. P_1.

Figure 6 Projections P_o and P_1 of $\mathfrak{D}$.



a. P_1 on P_o. b. P_i on P_o.

Figure 7 Fitting.

and the 8 marked points) in the interior of P_o (Figure 7b, in which P_i is centrally symmetric and concentric with P_o). So the projection P_i of the rotated dodecahedron lies in the interior of P_o.

Consequently $\mathfrak{D}$ is Rupert.

We find this intuitive geometric reasoning convincing. Jerrard introduced coordinates, determined suitable rotation angles (1.55° about m and 6.45° about ℓ), and verified numerically that the 20 vertices of P_i actually lie in the interior of P_o. (In Figure 7b the vertices are the 12 vertices of the projected dodecagon and the 8 dots. The fit is quite tight.) We omit the numerical details.

In 1776, Euler showed by an elementary geometric argument that the product of two rotations about nonparallel axes in R^3 is again a rotation. A contemporary discussion can be found, for example, in [**6**]. Consequently it is possible to rotate the dodecahedron $\mathfrak{D}$ about a suitable single axis in such a way that the rotated dodecahedron will pass through a suitable tunnel in a second dodecahedron of the same size.

Jerrard also determined the shortest distance from the inner projection P_i to the boundary of the outer projection P_o and showed that the inner projection can be expanded by about 1 part in 170, i.e., by a factor of at least $171/170 = 1.005\,882$, and still fit in P_o. This is the bound for $\nu(\mathfrak{D})$ included in Table 1.

Icosahedron Position the icosahedron $\mathfrak{I}$ with its center at the origin O, and label the 20 vertices as shown in Figure 8, with k and $13 - k$ symmetric with respect to O. The edges and vertices that are visible "from above" are drawn bold, and those hidden are in gray.

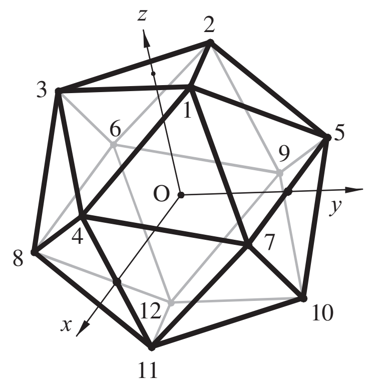


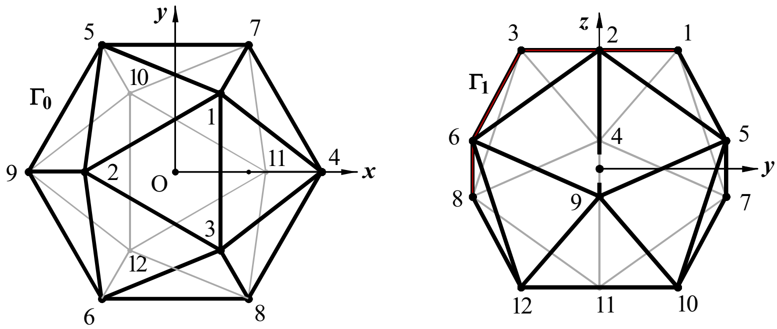
Figure 8 Labeled icosahedron $\mathfrak{I}$.

We choose an orthogonal coordinate system as follows: the origin is O, the positive x-axis passes through the midpoint of the edge 4-11, the positive y-axis passes through the midpoint of the edge 5-7, and the positive z-axis passes through the center of the face 1-2-3. The projection of $\mathfrak{I}$ into the xy-plane is a regular hexagon, and we chose the unit distance so that this hexagon has edge of length 2. The coordinates of the vertices are given in Table 2.

TABLE 2: Vertex coordinates of $\mathfrak{I}$.

Vertices	Coordinates	Values
1 and -12	$\left(1+\frac{1}{\sqrt{5}},\ \sqrt{3}(1+\frac{1}{\sqrt{5}}),\ 2(1+\frac{2}{\sqrt{5}})\right)$	(1.4472, 2.5066, 3.7889)
2 and -11	$\left(-2(1+\frac{1}{\sqrt{5}}),\ 0,\ 2(1+\frac{2}{\sqrt{5}})\right)$	(−2.8944, 0, 3.7889)
3 and -10	$\left(1+\frac{1}{\sqrt{5}},\ -\sqrt{3}(1+\frac{1}{\sqrt{5}}),\ 2(1+\frac{2}{\sqrt{5}})\right)$	(1.4472, −2.5066, 3.7889)
4 and -9	$\left(2(1+\frac{3}{\sqrt{5}}),\ 0,\ \frac{2}{\sqrt{5}}\right)$	(4.6833, 0, 0.8944)
5 and −8	$\left(-(1+\frac{3}{\sqrt{5}}),\ \sqrt{3}(1+\frac{3}{\sqrt{5}}),\ \frac{2}{\sqrt{5}}\right)$	(−2.3416, 4.0558, 0.8944)
6 and −7	$\left(-(1+\frac{3}{\sqrt{5}}),\ -\sqrt{3}(1+\frac{3}{\sqrt{5}}),\ \frac{2}{\sqrt{5}}\right)$	(−2.3416, −4.0558, 0.8944)

Let P_o be the projection of $\mathfrak{I}$ onto the xy-plane, shown in Figure 9a. Its boundary Γ_o is a regular hexagon. To avoid confusion, we retain the same names for the projected vertices, so, for example, vertex 1 in this figure is the projection of vertex 1 into the xy-plane in Figure 8. The (x, y) coordinates of the points in this figure are obtained by setting the z-coordinate to zero in Table 2.



a. P_o, in the xy-plane. b. P_1, in the yz-plane.

Figure 9 Projections P_o and P_1.

Let P_1 be the projection of $\mathfrak{I}$ onto the yz-plane, shown in Figure 9b. Its boundary Γ_1 is an irregular octagon whose opposite sides are parallel. Since this projection is in the direction of the median from vertex 2 to the edge 1-3 in the face triangle 1-2-3 of $\mathfrak{I}$, it has been called the "face normal" projection (see "Regular icosahedron," *Wikipedia, the Free Encyclopedia* (accessed June 2015)). In this drawing edge 2-9 has been broken to show the origin and the edge 4-9, which lie behind it. The (y, z) coordinates of the points in this figure are obtained by setting the x-coordinate to zero in Table 2.

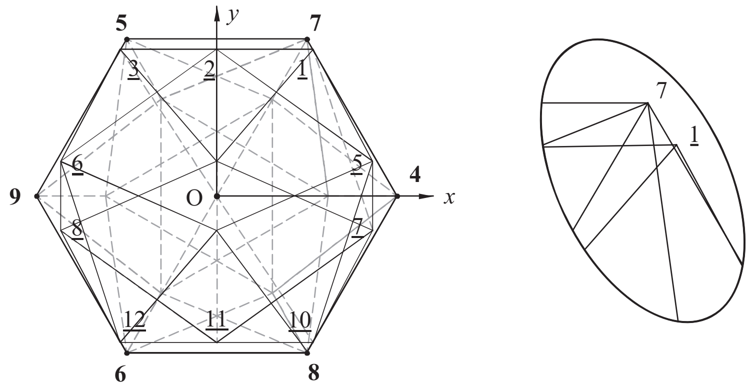
Placed with its center at the center of P_o, the projection P_1 very nearly but not quite fits in P_o (Figure 10a, in which the axes from P_1 have been replaced by those of P_o). In this figure the vertices on the boundary of P_o are named as in Figure 9a, but to avoid confusion the vertices of the projection P_1 are underscored. Figure 10b shows an enlarged detail of the fit near $\underline{1}$. The situation at $\underline{3}$, $\underline{10}$, and $\underline{12}$ is symmetrically the same.

Note that when P_1 is placed on P_o, the y-axis lands on the x-axis of P_o (Figure 10a). If $\mathfrak{I}$ is rotated by an angle φ about the y-axis, both the line segments $\underline{1}$-$\underline{3}$ and $\underline{10}$-$\underline{12}$ move into the interior of P_o, the segments $\underline{6}$-$\underline{8}$ and $\underline{5}$-$\underline{7}$ remain in the interior, but since the y-axis is parallel to the faceplane $\underline{1}$-$\underline{2}$-$\underline{3}$, the projections of the vertices $\underline{2}$ and $\underline{11}$ move closer to the boundary segments 5-7 and 6-8, respectively.

TABLE 3: Coordinates of projected vertices.

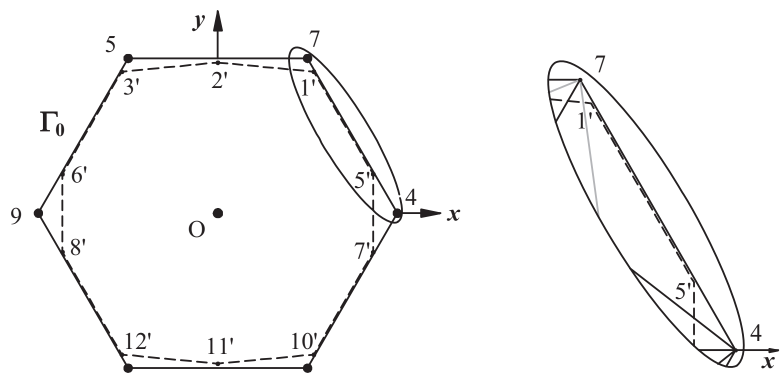
Vertices	(x, y) coordinates in P_o.	$\varphi = 3.1289°$
$1'$ and $-12'$	$\left(\sqrt{3}(1+\frac{1}{\sqrt{5}}),\ (2+\frac{4}{\sqrt{5}})\cos\varphi-(1+\frac{1}{\sqrt{5}})\sin\varphi\right)$	(2.5066, 3.7042)
$2'$ and $-11'$	$\left(0,\ (2+\frac{4}{\sqrt{5}})\cos\varphi+(2+\frac{2}{\sqrt{5}})\sin\varphi\right)$	(0.0, 3.9412)
$3'$ and $-10'$	$\left(-\sqrt{3}(1+\frac{1}{\sqrt{5}}),\ (2+\frac{4}{\sqrt{5}})\cos\varphi-(1+\frac{1}{\sqrt{5}})\sin\varphi\right)$	(−2.5066, 3.7042)
$4'$ and $-9'$	$\left(0,\ \frac{2}{\sqrt{5}}\cos\varphi-(2+\frac{6}{\sqrt{5}})\sin\varphi\right)$	(0.0, 0.6375)
$5'$ and $-8'$	$\left(\sqrt{3}(1+\frac{3}{\sqrt{5}}),\ \frac{2}{\sqrt{5}}\cos\varphi+(1+\frac{3}{\sqrt{5}})\sin\varphi\right)$	(4.0558, 1.0209)
$6'$ and $-7'$	$\left(-\sqrt{3}(1+\frac{3}{\sqrt{5}}),\ \frac{2}{\sqrt{5}}\cos\varphi+(1+\frac{3}{\sqrt{5}})\sin\varphi\right)$	(−4.0558, 1.0209)

Because $\underline{3}$ and $\underline{1}$ are so much closer to the boundary segments than is $\underline{2}$, it seems clear that if φ is sufficiently small, all the projected vertices will be inside the boundary



a. P_1 on P_o. b. Detail.

Figure 10 P_1 nearly fits in P_o.



a. P_1 on P_o. b. Detail.

Figure 11 P_i fits inside P_o.

of P_o, as desired. Let $\mathfrak{I}$ be the rotated icosahedron, and denote the projection into P_o of the vertex k by k'. Table 3 shows the coordinates of the projected vertices k' in terms of φ, found from the familiar formulas for the rotation through the angle φ about the origin in the xz-plane. The coordinates are presented in the (x, y) axes in P_o, and to the same scale as the vertices given in Table 2.

Choosing φ so that the projection of the edge $\underline{1}$-$\underline{5}$ is inside and parallel to the edge 7-4 minimizes the maximum distance between these two segments, and a calculation shows that this is accomplished for $\varphi = 3.1289^\circ$ (and similarly for $\underline{7}$-$\underline{10}$, $\underline{3}$-$\underline{6}$, and $\underline{12}$-$\underline{8}$), while leaving the image of vertices $\underline{2}$ and $\underline{11}$ inside P_o (Figure 11a).

The interior of P_o is described by the system of inequalities (4).

$$\begin{cases} -2\sqrt{3}(1+\frac{3}{\sqrt{5}}) - \sqrt{3}x < y < 2\sqrt{3}(1+\frac{3}{\sqrt{5}}) - \sqrt{3}x \\ -\sqrt{3}(1+\frac{3}{\sqrt{5}}) < y < \sqrt{3}(1+\frac{3}{\sqrt{5}}) \\ -2\sqrt{3}(1+\frac{3}{\sqrt{5}}) - \sqrt{3}x < y < 2\sqrt{3}(1+\frac{3}{\sqrt{5}}) + \sqrt{3}x \end{cases} \tag{4}$$

Verifying that these coordinates satisfy the constraints (4) establishes that the vertices of $\mathfrak{I}_i$ are in the interior of P_0, and this completes the coordinate proof that the icosahedron $\mathfrak{I}$ is Rupert. The tunnel cut in $\mathfrak{I}$ leaves but a thin shell behind, so making a physical model showing the interpenetration is likely to be difficult.

By computing the minimum distance between the boundary of P_i and P_o, one can see that the inner projection P_i can be enlarged by at least 1 part in 109.8, so the Nieuwland constant $\nu(\mathfrak{I})$ is greater than $1108/1098 > 1.009\,107\,47$, as shown in Table 1. We omit further details.

Remarks. There are interesting analogous questions in R^n that have not been studied. Which convex bodies in R^n can be passed through a hole in another?

The mathematics literature includes many "passage" problems related to Rupert's in which more general motions are permitted to accomplish the transit. In 1920 Zindler [**13**] described an affine cube that can pass through a circular ring of radius smaller than that of its smallest circumscribed cylinder, and this and related notions have been investigated by many authors since. See, for example, Zamfirescu [**12**] and the references included therein.

Acknowledgment We are indebted to Prof. Yuan's student Ying Chai for her interest in and contributions to these results, in particular for the numerical calculations for the icosahedron. We thank Dr. Daniel Lichtblau of Wolfram Research for his assistance with Mathematica. We are also indebted to several members of the DesignCAD User Forum, who helped with the preparation of some of the more difficult drawings. We want especially to thank Kevin Chapman (aka Droll Troll), IMSI/Designs, Novato, CA, for his insightful help and his patience. Finally, we express our gratitude to Prof. Richard Bishop and other participants in the UIUC Geometric Potpourri Seminar for their interest in this problem and their helpful suggestions.

REFERENCES

1. M. Gardner, *The Colossal Book of Mathematics*. W. W. Norton, New York, 2001.
2. R. P. Jerrard, J. E. Wetzel, Multipurpose stoppers are Rupert, *College. Math. J.* **39** (2008) 90–94.
3. ———, Platonic passages, 07T-51-2, *Abstracts Amer. Math. Soc.* **28** (2007) 351.
4. ———, Prince Rupert's rectangles, *Amer. Math. Monthly* **111** (2004) 22–31.
5. ———, Problem 11291, *Amer. Math. Monthly* **114** (2007) 451.
6. B. Palais, R. Palais, S. Rodi, A disorienting look at Euler's theorem on the axis of a rotation, *Amer. Math. Monthly* **116** (2009) 892–909, doi:10.4169/000298909X477014.
7. D. J. E. Schrek, Prince Rupert's problem and its extension by Pieter Nieuwland, *Scripta Math.* **16** (1950) 73–80, 261–267.
8. C. J. Scriba, Das problem des Prinzen Ruprecht von der Pfalz, *Praxis Math.* **10** (1968) 241–246.
9. K. R. Shultz, Largest m-cube in an n-cube: Partial solution, with a preface by G. Huber, KITP, NSF-ITP-13-142 (2013).
10. J. H. van Swinden, *Grondbeginsels der Meetkunde* (Elements of Geometry). Pieter den Hengst and Son, Amsterdam, 1816.
11. J. Wallis, *De Algebra Tractatus; Historicus & Practicus*, 1685, in Opera Mathematica, vol. 2, Oxoniae: E Theatro Sheldoniano, 1693.
12. T. Zamfirescu, Pushing convex and other bodies through rings and holes, *An. Univ. Vest Timiş. Ser. Mat.-Inform* **XLVIII** no. 3 (2010) 00–10.
13. K. Zindler, Über konvexe Gebilde, *Monatsh. Math. Phys.* 30 (1920) 87–102, 10.1007/BF01699908.

Summary. It is well known that a hole can be cut in a cube large enough to permit a second cube of equal size to pass through, a result attributed to Prince Rupert of the Rhine by J. Wallis more than three centuries ago. C. Scriba showed nearly 50 years ago that the tetrahedron and the octahedron have this same property. Somewhat surprisingly, the remaining two platonic solids, the dodecahedron and the icosahedron, also have this property: each can be passed through a suitable tunnel in another of the same size and kind. We supply the details.

RICHARD JERRARD (MR Author ID: 94525) received his bachelor's and master's degrees in mechanical engineering from the University of Wisconsin in 1950 and, after military service, his mathematics Ph.D. from the University of Michigan in 1957. He joined the mathematics faculty at the University of Illinois in 1958, spent three years at Cambridge University and the University of Warwick, and retired in 1995. Early in his career he

worked in applied mathematics, and later he became interested in topology and geometry. He died in 2014 after a long period of decline.

JOHN WETZEL (MR Author ID: 182185) did his undergraduate work at Purdue University, received his Ph.D. in mathematics from Stanford University in 1964, and spent his academic career (apart from sabbaticals) at the University of Illinois from which he retired in 1999. Always interested in classical geometry, he has studied the ways in which one shape fits in another—questions he regards as "fitting problems for retirement."

LIPING YUAN (MR Author ID: 689149) received her bachelor's, master's, and doctor's degrees in mathematics from Hebei Normal University (P.R. China) in 1993, 2000, and 2003, her second mathematics Ph.D. from Dortmund University (Germany) in 2006. She joined the mathematics faculty at Hebei Normal University in 1995. She is working in the area of discrete and combinatorial geometry.

Consider a cube for a minute,
And imagine the largest square in it.
Then if you're a math whiz,
Tell me how big it is;
It's tricky to even begin it!

Now let us move up one dimension:
Find the cube of the largest extension
That fit's (neatly packed)
Into one tesseract,
And, boy, will you have stress and tension!

Martin Gardner proposed this last question,
And I solved it, at no one's suggestion.
It took 15 years
Of blood, sweat and tears,
And gave me severe indigestion.

My proof fills up 100 pages;
Till I solved it, it stumped all the sages.
It was recently checked
And pronounced quite correct,
But it hasn't augmented my wages.

–Kay R. Pechenick DeVicci Shultz

Weebles Only Wobble But Eggs Fall Down

SUBHRANIL DE
School of Natural Sciences Indiana University Southeast
New Albany, IN 47150

A Weeble is a once-popular, typically egg-shaped toy that tends to stand upright on the less pointed end, similar to ones made from plastic Easter egg shells shown in Figure 1. They belong to the broader category of roly-poly toys that have the same basic property: they right themselves when tipped from the erect position. Roly-poly toys would often represent an animal or even a person, like the bopping bag (see Figure 2) that was popular during the 1950s before Weebles hit the market.

Figure 1 Weeble-like toys made from plastic Easter egg shells weighted at the bottom on the inside.

As the 1970s advertising slogan said, "Weebles wobble, but they don't fall down." This is because any roly-poly toy is at stable equilibrium in the upright position. As a contrast, for an actual egg the upright position is one of unstable equilibrium. For any solid of revolution with a convex vertex where the center of mass is located on its symmetry axis, an upright position of standing over a horizontal plane is an equilibrium position. From a physical perspective, the equilibrium is called stable if the object returns to the equilibrium position when released from a slight tilt. On the other hand, if the object deviates away further from equilibrium when released from a slight tilt, the equilibrium is unstable. Interestingly, a fairly straightforward principle involving the object's contour around the vertex and the position of its center of mass governs if the equilibrium will be stable or unstable. The 'contour' refers to the planar curve which when rotated about the symmetry axis produces the 'solid of revolution.' For example, rotating an ellipse about its semimajor axis produces a 'prolate spheroid,' while rotating it about its semi-minor axis produces an 'oblate spheroid.'

Math. Mag. **90** (2017) 99–107. doi:10.4169/math.mag.90.2.99.
MSC: Primary 70C20.

Figure 2 'Bozo the clown' bopping bag. (Courtesy: www.retroplanet.com)

The principle mentioned above dictates that *if at the equilibrium position, the vertical height of the center of mass from the horizontal plane is smaller than or greater than the radius of curvature of the solid's contour at the vertex, the equilibrium will be stable or unstable, respectively* [**1, 2, 3**]. At this juncture, the notion of curvature is worth a brief review, since it plays a big role in the subject matter of this article. Let us refer to any planar curve, like the one shown in Figure 3.

For the curve shown, the straight line PC is normal to the curve at the point P, while QC is normal to the curve at another point Q close to P. Let the angle between CP and CQ be $\Delta\theta$ and the arc length along the curve between P and Q be Δs. Now let us consider the situation where P is kept fixed and Q is brought increasingly closer to P. The radius of curvature r of the curve at point P is defined as the value of the distance $\overline{PC}$ in the limit as the arc length $\widehat{PQ}$ goes to 0. This also implies the following relation:

$$r = \lim_{\Delta\theta \to 0} \left(\frac{\Delta s}{\Delta\theta} \right) \equiv \frac{ds}{d\theta}.$$

The position of the intersection point C in the same limit is the center of curvature of the curve at P, and the reciprocal of the radius of curvature, namely the quantity $1/r$ is called the curvature at P (the 'extrinsic curvature,' to be specific). In simplest terms, the radius of curvature is the radius of the circle that approximates the curve around the point of interest.

For the egg-shaped object under consideration, let us assume a simple model consisting of a hemispherical surface at the bottom joined with a prolate hemispheroid at the top (Figure 4). In the case of a roly-poly toy, the center of mass (shown as the

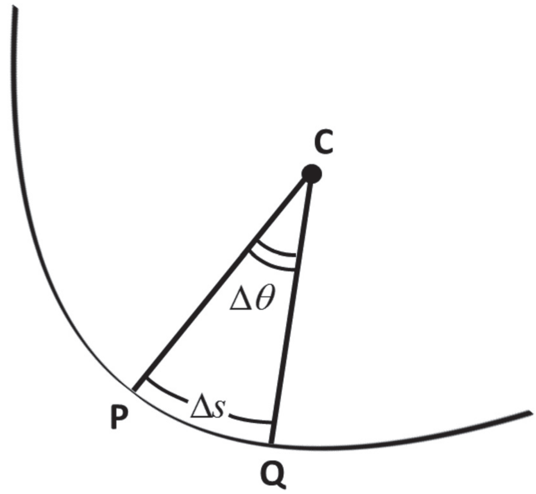


Figure 3 The straight line PC is normal to the curve at point P, and the straight line QC normal at point Q. The center of curvature C of the curve at P is the position of the intersection point C in the limit as the arc length $\widehat{PQ}$ goes to zero. The corresponding limiting value of the length $\overline{PC}$ is the radius of curvature of the curve at P.

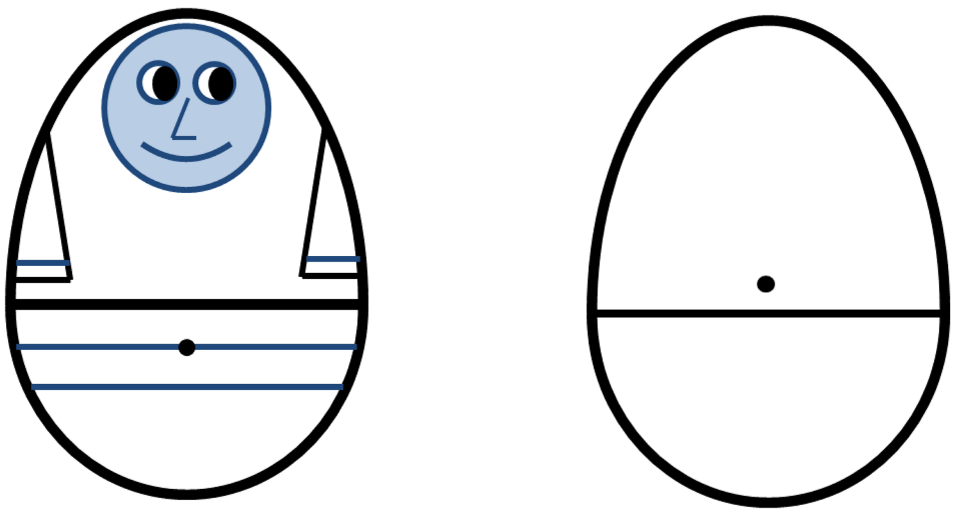


Figure 4 A roly-poly toy and an egg shown in their equilibrium positions standing on the less pointed end. For the roly-poly toy (left), the equilibrium is stable. For the egg (right) though, the equilibrium is unstable. The center of masses are given by the black dots.

small black dot) is below the equatorial plane of the hemisphere, while for the egg it is above. Consequently, at the upright position, the height of the center of mass for the roly-poly toy is smaller than the radius of the hemispherical bottom, while for the egg it is the other way around. This is sufficient to govern the nature of the equilibrium of each object in this situation!

However, even if the bounding surface around the vertex of the solid is not perfectly spherical, the aforementioned principle still holds. But why does this simple principle work? Classic textbooks of mechanics and statics contain several different derivations [**1, 2, 3**]. In this article, we will provide a contrasting analysis which not only confirms the above principle, but also brings us closer to addressing the case when the equilibrium is critical, i.e., when the height of the center of mass is exactly equal to the radius of curvature. A transparent treatment of that, so far in literature, seems to have remained obscure. Is there another straightforward higher-order principle that lets us infer the nature of the equilibrium in that critical case as well?

Analysis

Figure 5 shows the side view of the contour of the rigid body (placed over a horizontal plane) tilted at an angle θ relative to the equilibrium position. First, let us go over the fundamental mathematical rules governed by mechanics that define the two kinds of mechanical equilibrium. The fact that $\theta = 0$ (i.e., the axis being vertical) is an equilibrium position implies that the vertical height h of the center of mass (measured from the plane) of the body as a function of the tilt angle θ is at a local extremum at $\theta = 0$. If the equilibrium is stable, then h is at a local minimum (following from the fact that the gravitational potential energy of the object must be a minimum), whereas h is at a local maximum (since the potential energy must be a maximum) if the equilibrium is unstable. Hence, assuming h to be an analytic function of θ within some domain (i.e., derivatives of all orders of the function exist at all points within the said domain) around $\theta = 0$, the condition $\left(\frac{dh}{d\theta}\right)_{\theta=0} = 0$ holds. In addition, if $\left(\frac{d^2h}{d\theta^2}\right)_{\theta=0}$ is nonzero, a positive value of the second derivative implies a stable equilibrium, while a negative value implies an unstable equilibrium.

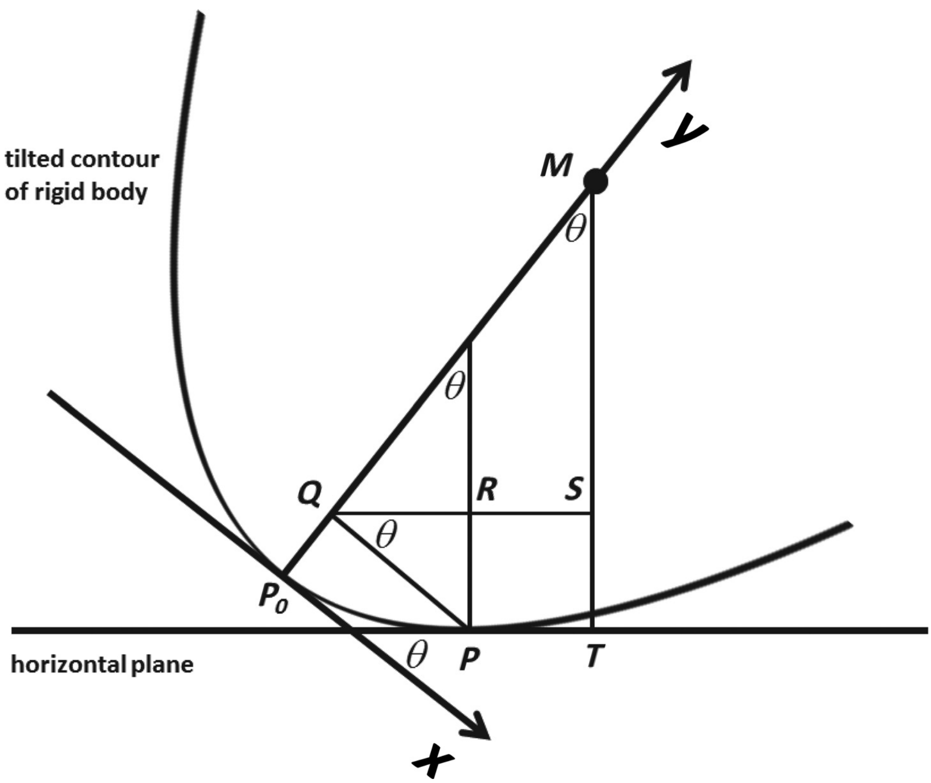


Figure 5 Side view of the rigid body contour tilted at an angle θ with respect to the equilibrium position.

For a tilted solid, we define an xy-coordinate system in which the y-axis is the symmetry axis of the solid; see Figure 5. We assume that within some finite domain around $x = 0$ the contour of the solid is described by some function $y = f(x)$ that is analytic. This function is even (i.e., $f(x) = f(-x)$) as well, since the y-axis is an axis of symmetry. The point of contact P of the tilted body over the horizontal plane has coordinates (x, y). M is the position of the center of mass on the y-axis, the length $\overline{MP_0}$ being equal to h_0, the height of the center of mass at the equilibrium position. The length $\overline{MT}$ is equal to the present height h of the center of mass.

From the diagram it is evident that

$$\overline{MT} = \overline{MS} + \overline{ST}. \tag{1}$$

Now, from the right triangle MQS we can write

$$\overline{MS} = \overline{MQ}\cos\theta \text{ and } \overline{MQ} = \overline{MP_0} - \overline{QP_0}. \tag{2}$$

But, at the same time, $\overline{MP_0} = h_0$ and $\overline{QP_0} = y$. So, by (2), we can write:

$$\overline{MS} = (h_0 - y)\cos\theta. \tag{3}$$

Also, since $STPR$ is a rectangle,

$$\overline{ST} = \overline{RP}, \tag{4}$$

and from the right triangle QRP we can write

$$\overline{RP} = \overline{QP}\sin\theta. \tag{5}$$

Recognizing that $\overline{QP} = x$, we can combine (4) and (5) to write

$$\overline{ST} = x\sin\theta. \tag{6}$$

Finally, combining (1), (3), and (6) we arrive at

$$h = x\sin\theta + (h_0 - y)\cos\theta, \tag{7}$$

which is the expression for the height of the center of mass of the solid when tilted.

Because $\frac{dy}{dx} = \tan\theta$, $r = \frac{ds}{d\theta}$ at point P, and

$$\frac{dx}{ds} = \cos\theta \text{ and } \frac{dy}{ds} = \sin\theta, \tag{8}$$

then

$$\frac{dx}{d\theta} = r\cos\theta \text{ and } \frac{dy}{d\theta} = r\sin\theta \tag{9}$$

Now, differentiating (7) with respect to θ, we obtain

$$\frac{dh}{d\theta} = \frac{dx}{d\theta}\sin\theta + x\cos\theta - (h_0 - y)\sin\theta - \frac{dy}{d\theta}\cos\theta. \tag{10}$$

We can simplify (10) by using the equations in (9) to yield

$$\frac{dh}{d\theta} = x\cos\theta - (h_0 - y)\sin\theta.$$

Differentiating this equation with respect to θ, we obtain

$$\frac{d^2h}{d\theta^2} = \frac{dx}{d\theta}\cos\theta - x\sin\theta - (h_0 - y)\cos\theta + \frac{dy}{d\theta}\sin\theta.$$

Using (9) to simplify this equation, and then using the identity $\sin^2\theta + \cos^2\theta = 1$, we get

$$\frac{d^2h}{d\theta^2} = r - x\sin\theta - (h_0 - y)\cos\theta,$$

which, because of (7), implies that

$$\frac{d^2h}{d\theta^2} = r - h. \tag{11}$$

Because $(dh/d\theta)_{\theta=0} = 0$, the condition that $\left(d^2h/d\theta^2\right)_{\theta=0} > 0$ or $\left(d^2h/d\theta^2\right)_{\theta=0} < 0$ implies that at $\theta = 0$ the quantity h is a local minimum or a maximum, respectively. Equation (11) confirms that $\left(d^2h/d\theta^2\right)_{\theta=0} > 0$ or $\left(d^2h/d\theta^2\right)_{\theta=0} < 0$ if $r_0 > h_0$ or $r_0 < h_0$, respectively, where r_0 is the value of r at $\theta = 0$. Hence, at $\theta = 0$, h is a local minimum or a maximum depending on if $r_0 > h_0$ or $r_0 < h_0$, respectively.

In other words, we have confirmed the relationship between the height of the center of mass and the type of equilibrium. This is formally stated below as a theorem.

Theorem 1. *Consider the equilibrium position of a solid of revolution resting on its vertex over a horizontal plane. At this position, if the vertical height of the center of mass from the plane is smaller than or larger than the radius of curvature of the solid's contour at its vertex, the equilibrium will be stable or unstable, respectively.*

If $r_0 = h_0$, then $\left(d^2h/d\theta^2\right)_{\theta=0} = 0$, and the equilibrium is then critical. However, equation (11) helps us infer the nature of the equilibrium even in the critical case.

First, recall both h and r are even functions of θ, which follows from the fact that $y = f(x)$ is an even function of x. Hence, all odd-order derivatives of h and r with respect to θ are odd functions of θ, and will vanish at $\theta = 0$. Therefore, for each of the quantities h and r, in order to determine the nature of its extremum at $\theta = 0$, we must examine its lowest even-order derivative that is nonzero at $\theta = 0$.

Differentiating $\frac{d^2h}{d\theta^2} = r - h$ two more times with respect to θ, we obtain

$$\frac{d^4h}{d\theta^4} = \frac{d^2r}{d\theta^2} - \frac{d^2h}{d\theta^2}.$$

The above relation tells us that if $\left(d^2h/d\theta^2\right)_{\theta=0} = 0$, then $\left(d^4h/d\theta^4\right)_{\theta=0} = \left(d^2r/d\theta^2\right)_{\theta=0}$. Hence, in that case $\left(d^4h/d\theta^4\right)_{\theta=0}$ will have the same sign as $\left(d^2r/d\theta^2\right)_{\theta=0}$, assuming the latter is nonzero. Hence, at $\theta = 0$, h will be a local minimum or a maximum depending on if r is a local minimum or a maximum, respectively.

What makes this logical scheme intriguing is that beyond this point it can be continued to any even order of derivatives as needed. Hence, we have arrived at a higher-order principle governing the nature of the equilibrium.

Theorem 2. *Consider the equilibrium position of a solid of revolution resting on its vertex over a horizontal plane. In case the equilibrium is critical due to the fact that the vertical height of the center of mass from the plane is equal to the radius of curvature of the solid's contour at its vertex, the equilibrium will be stable or unstable if at the vertex the radius of curvature itself is a local minimum or a maximum, respectively.*

Examples of applications

The simple yet powerful *second-order principle* (Theorem 1) stated above has been used in numerous examples before, including the different examples in references

[**1, 2, 3, 4**], as well as our own model of the roly-poly toy and the egg (Figure 4) in the present article.

The use of the *higher-order principle* (Theorem 2) when necessary is deemed convenient by the fact that in many situations it is obvious from the shape of the solid's contour whether it is the 'least curved' or the 'most curved' at the vertex. For example, in the case when the bounding surface around the vertex is part of a paraboloid or a prolate hemispheroid, it is obviously most curved at the vertex (Figures 6(a) and 6(b)). Hence the radius of curvature is at a local minimum at the equilibrium contact point, and using Theorem 2 we can readily infer that at the critical point the equilibrium must be stable, corroborating the findings in [**4**] achieved through direct calculations. On the other hand, if the bounding surface around the vertex is, for example, part of an oblate hemispheroid (Figure 6(c)), it is least curved at the vertex, and hence at the critical point the equilibrium must be unstable.

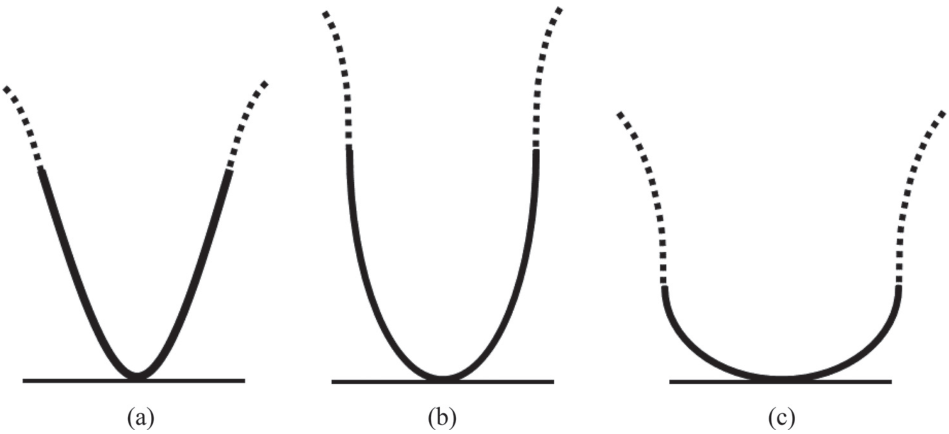


Figure 6 The contours of three solids of revolution resting on their vertices over a horizontal plane. Around the vertex, the bounding surface is a portion of (a) a paraboloid, (b) a prolate hemispheroid, and (c) an oblate hemispheroid, respectively.

In the special case where all derivatives of r vanish at $\theta = 0$, that implies that r is a constant within some domain around $\theta = 0$, and in that case at the critical point the rigid body will truly be at *neutral equilibrium*. This is not surprising since r being a constant physically implies that the portion of the bounding surface around the vertex is truly a part of a sphere, like our model of the roly-poly toy and the egg in Figure 2 where we took the bottom part of the bounding surface to be a hemisphere. Hence, in case the center of mass is exactly on the equatorial plane within the egg-like structure, the resulting critical equilibrium will be neutral, implying that in case the structure is tilted by an angle relative to the upright position and released, it will just stay there. This behavior is reminiscent of that of a uniform sphere over a horizontal plane.

A doubly critical roly-poly toy

Utilizing the two theorems stated above, we can come up with the mathematical description of a roly-poly toy that captures the salient features of both. The bounding surface of this roly-poly toy is achieved by joining a prolate hemispheroid and an oblate hemispheroid with mutually reciprocal aspect ratios as shown in Figure 7.

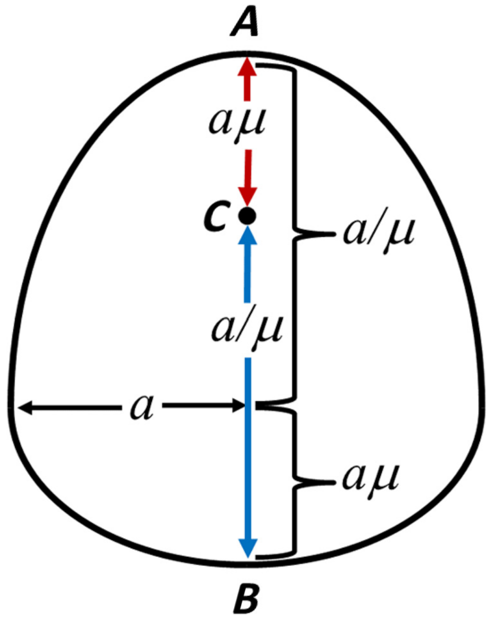


Figure 7 The contour of a doubly critical roly-poly toy with a value of μ chosen to be $\mu = 2/3$. The lower part is the oblate hemispheroid with its vertex at point B, while the upper part is a prolate hemispheroid with its vertex at point A. Point C is the point where the center of curvature of both vertices happens to be located.

The oblate hemispheroid has a semimajor axis of length a and a semiminor axis of length $a\mu$, where μ is a proper fraction. Correspondingly, the prolate hemispheroid has a semiminor axis of length a and a semimajor axis of length a/μ. Hence, the two vertices of the roly-poly toy, namely the points A and B, belong to the prolate and the oblate parts, respectively. At this point we recall that the radius of curvature on an ellipse at the vertex of a semiaxis c is given by a^2/c, where a is the other semiaxis. This makes the location of the point C on the axis of symmetry AB, taken to be at a distance $a\mu$ from A (and consequently at a distance a/μ from B), to be the center of curvature at both A and B.

With the above-mentioned dimensions in place, we now consider what happens for different positions of the center of mass of the roly-poly toy on the symmetry axis AB. If it is located between the points A and C (shown in red), according to Theorem 1, the upright position on A becomes a stable equilibrium while the upright position on B becomes an unstable equilibrium. If the location of the center of mass is between the points B and C (shown in blue) instead, the nature of the equilibrium flips for both vertices. Finally, if the center of mass is located exactly at the point C, the equilibrium on both A and B will be critical. However, in that situation, using Theorem 2 we can infer that the critical equilibrium on A will actually be stable, while the one on B will be unstable! Figure 8 shows a plot of h vs θ for the doubly critical case, where $\theta = 0$ corresponds to the roly-poly toy standing on end B, and $\theta = 180^\circ$ corresponds to the roly-poly toy standing on end A. The plot looks flat around both values of θ, since at both points the value of $\frac{d^2h}{d\theta^2}$ is zero. Nevertheless, at $\theta = 0$ the value of $\frac{d^4h}{d\theta^4}$ is negative, implying that h is a maximum, making it an unstable equilibrium. At the same time, at $\theta = 180^\circ$ the value of $\frac{d^4h}{d\theta^4}$ is positive so that h is a minimum, making it stable equilibrium.

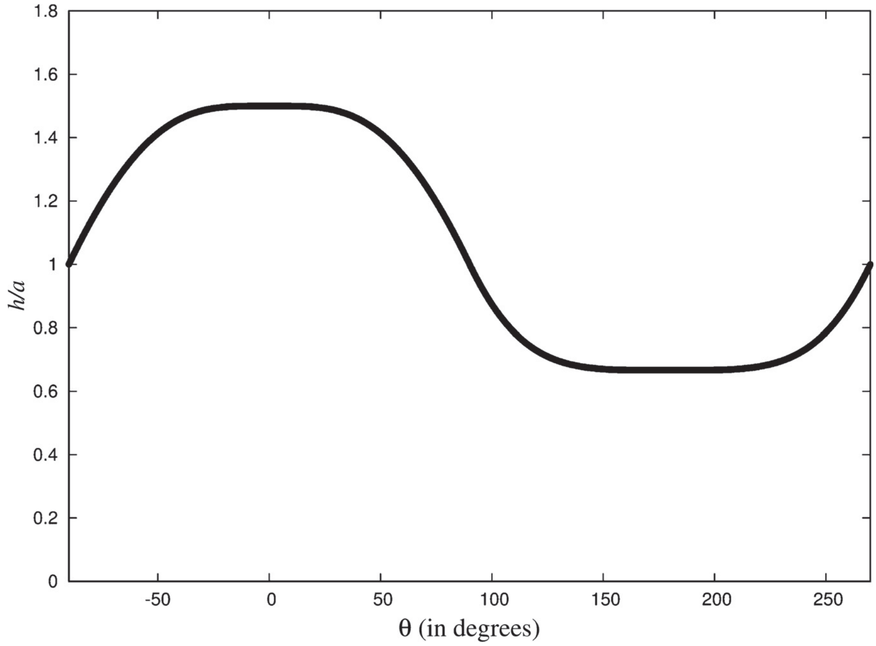


Figure 8 The plot of h vs θ for the roly-poly toy shown in Figure 7, when the center of mass is located exactly at point C. Both $\theta = 0$ and $\theta = 180^\circ$ are points of critical equilibrium, although at $\theta = 0$ the roly-poly toy is unstable, while at $\theta = 180^\circ$ it is stable.

REFERENCES

1. A. J. Du Bois, *The Mechanics of Engineering*, Vol. 1. First ed. John Wiley & Sons, New York, 1905. 312.
2. J. Greaves, *A Treatise on Elementary Statics.* Macmillan, New York, 1886. 212.
3. S. L. Loney, *An Elementary Treatise on Statics.* Cambridge Univ. Press, Cambridge, 1944. 167.
4. S. De, On static equilibrium of a hemispheroid, *Math. Gaz.* **98** (2014) 541.

Summary. Why does a roly-poly toy tend to stand upright on its less pointed end while an egg does not? The answer lies in a simple principle that governs the nature of the mechanical equilibrium of a solid of revolution. A roly-poly toy is at a stable equilibrium on its vertex while an egg is at an unstable equilibrium. The roly-poly toy's center of mass is vertically below the center of curvature of its vertex, while for the egg it is the other way around. Although this simple but powerful principle is known, in this article we present a contrasting analysis that not only verifies this principle, but also provides the answer to another interesting question: Is there another simple principle that governs the nature of the equilibrium in the critical situation where the center of curvature of the vertex happens to coincide with the center of mass?

SUBHRANIL DE (MR Author ID: 1179343) earned a Ph.D. in physics from the University of Rochester in 2003. He has a theoretical background in statistical physics and fluid mechanics. He also takes a sincere interest in pedagogical scholarship of physics and mathematics. He has been teaching physics at Indiana University Southeast since 2008. Among his interests outside the realm of physics, he enjoys lyrical poetry and classical music, loves to travel, and looks forward to throwing a Frisbee on a sunny day.

An Alternate Method to Compute the Decimal Expansion of Rational Numbers

AMITABHA TRIPATHI
Indian Institute of Technology Delhi
Hauz Khas, New Delhi – 110016, India
atripath@maths.iitd.ac.in

Every middle school student knows how to use long division to find the decimal expansion of a rational number. Doing so results in writing the digits of the decimal expansion from left to right. Is there a way to write the decimal expansion from right to left? The purpose of this note is to describe a procedure to write such a decimal expansion. While a high school student in 1976, I was aware that this procedure applied to rational numbers whose denominators were primes ending in 9. It wasn't until 2006 that I was surprised to discover that the procedure could be extended to all denominators.

Since we are interested in the decimal expansion of rational numbers, we may suppose that a/n is a rational number, in reduced form. To implement our procedure, we follow the two steps listed below.

- **Step 1.** If $\gcd(n, 10) > 1$, write a/n in terms of b/m with $\gcd(m, 10) = 1$. If $\gcd(n, 10) = 1$, then define $a = b$ and $m = n$.

As an example of Step 1, consider the decimal expansion of $r = 7/760$ so that $r = 7/(2^3 \cdot 5 \cdot 19)$. From

$$r = \frac{1}{10^3}\,\frac{5^2 \cdot 7}{19} = \frac{1}{10^3}\left(9 + \frac{4}{19}\right), \tag{1}$$

we see that the decimal expansion of r can be determined from that of $4/19$. In this example, $n = 2^3 \cdot 5 \cdot 19$ and $m = 19$, and we have linked the decimal expansion of a rational number a/n with $\gcd(n, 10) > 1$ to the decimal expansion of b/m with $\gcd(m, 10) = 1$. So when $\gcd(n, 10) > 1$, Step 1 requires us to express a/n as $b/(m \cdot 10^\gamma)$ for an appropriate choice of γ. If $\gcd(a, n) = 1$, we let $a = b$ and $m = n$, and move to Step 2.

- **Step 2.** The decimal expansion of b/m, $\gcd(m, 10) = 1$ is always a purely recurring decimal. To write down the digits from right to left, determine the last digit, a positive integer c by which to multiply each digit to get the subsequent digit, and the number of digits.

Let us illustrate Step 2 by continuing with the example in Step 1 – the decimal expansion of $4/19$. It is well known that the decimal expansion will be purely recurring. To write down the digits from right to left in a sequence, we determine (i) the starting digit (the last digit in the expansion), (ii) the positive integer c (we call this the multiplying factor), and (iii) the number of digits in the recurring part. Theorem 1 shows that we must start the sequence with 4, use the multiplying factor 2, and stop after 18 digits. Therefore, we begin with 4, then multiply 4 by 2 (the multiplying factor

Math. Mag. **90** (2017) 108–113. doi:10.4169/math.mag.90.2.108.
MSC: Primary 11A07, Secondary 11A99

for this example) to get 8 as the second digit (from right to left!). To proceed further, we multiply 8 by 2 to get 16; this gives 6 as the third digit and a carry-over 1.

To get the fourth digit from the third, we multiply 6 by 2 and add any carry-over that may have accrued. This gives 13; so 3 is the fourth digit and we again carry-over 1. A careful and patient application of this procedure results in the following sequence; the carry-overs appear in bold face above the corresponding the digits.

$$4, 8, 6^{(\mathbf{1})}, 3^{(\mathbf{1})}, 7, 4^{(\mathbf{1})}, 9, 8^{(\mathbf{1})}, 7^{(\mathbf{1})}, 5^{(\mathbf{1})}, 1^{(\mathbf{1})}, 3, 6, 2^{(\mathbf{1})}, 5, 0^{(\mathbf{1})}, 1, 2.$$

Thus

$$\frac{4}{19} = 0.\overline{210526315789473684}. \tag{2}$$

Finally, substituting (2) in (1):

$$\frac{7}{2^3 \cdot 5 \cdot 19} = \frac{1}{10^3}\left(9 + \frac{4}{19}\right) = 0.009\,\overline{210526315789473684}.$$

Suppose $a/n \in \mathbb{Q}$, with $1 \le a < n$, $\gcd(a, n) = 1$, $n = 2^\alpha 5^\beta m$, and $\gcd(m, 10) = 1$. Then the decimal expansion of a/n terminates if $m = 1$ and recurs if $m > 1$. In the latter case, if ℓ denotes the number of digits in the recurring part and γ the number of digits in the nonrecurring part, ℓ equals the order of 10 modulo m and $\gamma = \max\{\alpha, \beta\}$. The statement ℓ equals the order of 10 modulo m means that ℓ is the least positive integer for which m divides $10^\ell - 1 = \underbrace{9 \cdots 9}_{\ell \text{ times}}$; we denote this by $\ell = \operatorname{ord}_m 10$. Given a positive integer m, we know that $\operatorname{ord}_m b$ exists if and only if $\gcd(b, m) = 1$, or in other words, when b is an element of the multiplicative group of units in the ring $\mathbb{Z}_m$. A proof of the statement that $\ell = \operatorname{ord}_m 10$ can be found in [**1**, pp. 109–111], but we give a short and self-contained proof nevertheless.

The decimal expansion of a/n terminates if and only $10^e a/n$ is an integer, which is true if and only if n divides 10^e, so that $m = 1$. Assume now that $n = 2^\alpha 5^\beta m$, with $\gcd(m, 10) = 1$ and $m > 1$. If we reduce each term of the infinite sequence $\{a10^i\}_{i \ge 1}$ mod n, by the *pigeonhole principle*, two terms must be equal. If we choose the first such pair, say $a10^s$ and $a10^{s+\ell}$, then $10^s \equiv 10^{s+\ell} \mod n$, since $\gcd(a, n) = 1$. From $\gcd(m, 10) = 1$ it follows that $10^\ell \equiv 1 \mod m$. Since this implies $a10^{s+k} \equiv a10^{s+\ell+k}$ mod n for each $k \ge 0$, the sequence is eventually recurring and the length of the period equals $\ell = \operatorname{ord}_m 10$. From $10^s \equiv 10^{s+\ell} \mod n$ we see that both 2^α and 5^β must divide 10^s since each is coprime to $10^\ell - 1$. The smallest such s equals $\gamma = \max\{\alpha, \beta\}$, and this represents the length of the nonrecurring part.

We may suppose a/n is a positive rational number, in reduced form and less than 1, for the purpose of decimal expansion. Let $n = 2^\alpha 5^\beta m$, where $\gcd(m, 10) = 1$, and set $\gamma = \max\{\alpha, \beta\}$. Since

$$10^\gamma \frac{a}{n} = 2^{\gamma-\alpha} 5^{\gamma-\beta} \frac{a}{m} = \frac{b}{m},$$

the decimal expansion of a/n is the same as that of b/m but with the decimal shifted γ places to the left. The general case is therefore easily reduced to the case where $\gcd(n, 10) = 1$. For example, to obtain the decimal expansion of $7/(2^3 \cdot 5 \cdot 19)$, we have already noted that

$$\frac{7}{2^3 \cdot 5 \cdot 19} = \frac{1}{10^3} \frac{5^2 \cdot 7}{19} = \frac{1}{10^3}\left(9 + \frac{4}{19}\right).$$

We now consider rational numbers with denominator coprime to 10. The decimal expansion of such rational numbers is purely recurring, and the digits when sequenced from right to left satisfy a recurrence equation.

Theorem 1. *Let a, n be positive integers, with $1 \le a < n$ and $\gcd(a, n) = \gcd(n, 10) = 1$. Then*

$$\frac{a}{n} = 0.\overline{a_1\, a_2\, a_3\, \cdots\, a_\ell},$$

where $\ell = \mathrm{ord}_n\, 10$, and the finite sequence $\{a_k\}_{k=1}^{\ell}$ satisfies the recurrence

$$a_k \equiv \frac{10c-1}{10} a_{k+1} + \frac{10c-1}{10^2} a_{k+2} + \cdots + \frac{10c-1}{10^{\ell-k-1}} a_{\ell-1} + \frac{an^\star}{10^{\ell-k}} \pmod{10}$$

for $1 \le k \le \ell - 1$, with the initial condition

$$a_\ell \equiv an^\star \pmod{10},$$

where $c \equiv 10^{-1} \mod n$, $1 \le c \le n-1$ and $n^\star \equiv -n^{-1} \mod 10$, $1 \le n^\star \le 9$.

Proof. Write $a/n = 0.\overline{a_1\, a_2\, a_3\, \cdots\, a_\ell}$; we know that $\ell = \mathrm{ord}_n\, 10$. Let A_k denote the k-digit number $a_1 a_2 \cdots a_k$ for $k \ge 1$. Since $A_k = \lfloor a 10^k / n \rfloor$, we have

$$a 10^k = n A_k + r_k, \tag{3}$$

with $1 \le r_k < n$. Since a/n is the sum of an infinite geometric series with first term $A_\ell / 10^\ell$ and common ratio $1/10^\ell$, we have

$$n A_\ell = a(10^\ell - 1). \tag{4}$$

Thus $n a_\ell \equiv n A_\ell \equiv -a \pmod{10}$, and $a_\ell \equiv -an^{-1} \equiv an^\star \pmod{10}$.

Define c such that

$$10c \equiv 1 \pmod{n},$$

with $1 \le c \le n-1$. Since $n \cdot \frac{10c-1}{n} \equiv -1 \pmod{10}$ and both $\frac{10c-1}{n}$ and $n^\star$ lie between 1 and 9, it follows that

$$10c - 1 = nn^\star. \tag{5}$$

From (3), for $k > 1$,

$$r_{k-1} \equiv a 10^{k-1} \equiv ac 10^k \equiv c r_k \pmod{n}.$$

In particular, since $r_\ell = a$ by reducing (3) mod n, we have $r_{\ell-1} \equiv ac \pmod{n}$.

From (3), we have

$$10 r_{k-1} - r_k = 10(a 10^{k-1} - n A_{k-1}) - (a 10^k - n A_k) = n(A_k - 10 A_{k-1}) = n a_k \tag{6}$$

valid for $2 \le k \le \ell$.

For each k, $1 \le k \le \ell - 1$, let us denote the $(\ell - k)$-digit number with digits $a_{k+1} a_{k+2} \cdots a_\ell$ by A'_k. Thus

$$A_\ell = 10^{\ell-k} A_k + A'_k,$$

so that by (3) and (4),

$$\begin{aligned} n A'_k + a &= n(A_\ell - 10^{\ell-k} A_k) + a = a 10^\ell - n 10^{\ell-k} A_k = 10^{\ell-k}(a 10^k - n A_k) \\ &= 10^{\ell-k} r_k. \end{aligned} \tag{7}$$

From (5), (6), and (7),

$$\frac{(10c-1)A'_k + an^\star}{10^{\ell-k}} = \frac{n^\star(nA'_k + a)}{10^{\ell-k}} = n^\star r_k \equiv a_k \mod 10. \quad (8)$$

To complete the proof, replace A'_k by $a_{k+1}10^{\ell-k-1} + a_{k+2}10^{\ell-k-2} + \cdots + a_\ell$ in (8). ■

Remark. Let n be a positive integer such that $\gcd(n, 10) = 1$. Then

$$n^\star \equiv \begin{cases} -n \bmod 10 & \text{if } n \equiv \pm 1 \pmod{10}, \\ n \bmod 10 & \text{if } n \equiv \pm 3 \pmod{10}, \end{cases}$$

and

$$c \equiv \begin{cases} \dfrac{9n+1}{10} & \text{if } n \equiv 1 \pmod{10}, \\ \dfrac{3n+1}{10} & \text{if } n \equiv 3 \pmod{10}, \\ \dfrac{7n+1}{10} & \text{if } n \equiv 7 \pmod{10}, \\ \dfrac{n+1}{10} & \text{if } n \equiv 9 \pmod{10}. \end{cases}$$

Proof. We observe that $\gcd(n, 10) = 1$ is equivalent to stating that $n \equiv \pm 1$ or ± 3 modulo 10. Since $n^2 \equiv 1 \pmod{10}$ when $n \equiv \pm 1 \pmod{10}$ and $n^2 \equiv -1 \pmod{10}$ when $n \equiv \pm 3 \pmod{10}$, we have $n^{-1} \equiv n \mod 10$ in the first case and $n^{-1} \equiv -n \pmod{10}$ in the second case. This leads to the results on $n^\star$.

Recall that c is the unique integer satisfying $10c \equiv 1 \pmod{n}$ and $1 \leq c \leq n-1$. It is easy to verify that the case-by-case values of c given above satisfy both conditions and are integers. This completes the proof. ■

Let us revisit Step 2 of the procedure to determine the decimal expansion of 4/19 in light of Theorem 1. Theorem 1 applies only to cases where n ends in 1, 3, 7, or 9, and so Step 2 applies to only those cases. For our example, $n^\star = 1$ and $c = 2$ by Remark. Thus $a_\ell \equiv an^\star \equiv a \pmod{10}$, and so $a_\ell = 4$ is the first term in the sequence. The number of terms in the sequence is given by $\ell = \text{ord}_n 10$, as remarked earlier. Recall that ℓ is the least positive integer for which $10^\ell - 1 = \underbrace{9\cdots9}_{\ell \text{ times}}$ is a multiple of n. There is no simple way to compute ℓ, but because 10 is an element of the multiplicative group of units in the ring $\mathbb{Z}_n$ we know that $\ell \mid \phi(n)$.

The following reformulation of the recurrence in Theorem 1 leads to a computationally interesting consequence:

$$a_k \equiv ca_{k+1} + \left(\frac{ca_{k+2} - a_{k+1}}{10} + \cdots + \frac{ca_\ell - a_{\ell-1}}{10^{\ell-k-1}} + \frac{an^\star - a_\ell}{10^{\ell-k}} \right) \pmod{10} \quad (9)$$

for $1 \leq k \leq \ell - 2$, with

$$a_\ell \equiv an^\star \mod 10, \quad a_{\ell-1} \equiv ca_\ell + \frac{an^\star - a_\ell}{10} \pmod{10}.$$

For each k, $1 \leq k \leq \ell - 2$, let us denote by λ_{k+1} the expression within brackets in (9), and let $\lambda_\ell = (an^\star - a_\ell)/10$. For $1 \leq k \leq \ell - 1$, set

$$ca_{k+1} + \lambda_{k+1} = 10\mu_k + a_k.$$

Then

$$\mu_k = \frac{ca_{k+1} - a_k}{10} + \frac{ca_{k+2} - a_{k+1}}{10^2} + \cdots + \frac{ca_\ell - a_{\ell-1}}{10^{\ell-k}} + \frac{an^\star - a_\ell}{10^{\ell-k+1}} = \lambda_k,$$

so that

$$ca_{k+1} + \lambda_{k+1} = 10\lambda_k + a_k$$

for $1 \leq k \leq \ell - 2$.

We summarize the procedure for the computation of the sequence of digits $a_\ell, a_{\ell-1}, \ldots, a_1$ that appear in the decimal expansion of a/n when $\gcd(n, 10) = 1$ in our concluding remark.

Remark. Let a, n be positive integers, with $1 \leq a < n$ and $\gcd(a, n) = \gcd(n, 10) = 1$. Then $a/n = 0.\overline{a_1\, a_2\, \cdots\, a_\ell}$, where $\ell = \mathrm{ord}_n\, 10$. Define two positive integers $n^\star$ and c as follows:

$$10 \mid (nn^\star + 1), \quad c = \frac{nn^\star + 1}{10},$$

where $1 \leq n^\star \leq 9$.
The sequence $\{a_\ell, a_{\ell-1}, \ldots, a_1\}$ and the sequence $\{\lambda_\ell, \lambda_{\ell-1}, \ldots, \lambda_1\}$ are interlinked. The first sequence consists of integers in [0, 9] and satisfies (9):

$$a_k \equiv ca_{k+1} + \lambda_{k+1} \mod 10 \tag{10}$$

for $1 \leq k \leq \ell - 1$, with

$$a_\ell \equiv an^\star \mod 10, \quad \lambda_\ell = \frac{an^\star - a_\ell}{10}. \tag{11}$$

The second sequence consists of nonnegative integers and is given by the expression within brackets in (9):

$$\lambda_{k+1} = \frac{ca_{k+2} - a_{k+1}}{10} + \cdots + \frac{ca_\ell - a_{\ell-1}}{10^{\ell-k-1}} + \frac{an^\star - a_\ell}{10^{\ell-k}} \tag{12}$$

for $1 \leq k \leq \ell - 1$, with $\lambda_\ell = (an^\star - a_\ell)/10$. The two sequences are connected via the formulae:

$$\begin{aligned}
an^\star &= 10\lambda_\ell + a_\ell,\\
ca_\ell + \lambda_\ell &= 10\lambda_{\ell-1} + a_{\ell-1},\\
ca_{\ell-1} + \lambda_{\ell-1} &= 10\lambda_{\ell-2} + a_{\ell-2},\\
&\vdots\\
ca_{k+1} + \lambda_{k+1} &= 10\lambda_k + a_k,\\
&\vdots\\
ca_2 + \lambda_2 &= 10\lambda_1 + a_1.
\end{aligned}$$

When $\gcd(n, 10) = 1$, the digits in the decimal expansion of a/n are obtained recursively by computing the sequence of digits is $a_\ell, a_{\ell-1}, \ldots, a_1$. The rightmost digit a_ℓ is $an^\star \bmod 10$, the multiplying factor c equals $(nn^\star + 1)/10$, and the rest of the sequence $a_{\ell-1}, \ldots, a_1$ is computed via the sequence $\lambda_\ell, \lambda_{\ell-1}, \ldots, \lambda_1$ of carry-overs.

We close this article by applying the result in Remark to a special example. Repunits are numbers like $1, 11, 111, \ldots$ that contain only the digit 1; the ℓ-digit repunit equals $(10^\ell - 1)/9$. Repdigits are numbers like $d, dd, dd, \ldots$ that contain only the digit d; the ℓ-digit repdigit containing only the digit d equals $(10^\ell - 1)d/9$. The decimal expansion of a/n, $\gcd(a, n) = 1$, when $n = 10^\ell - 1$ has a particularly simple form. We may assume, as before, that $1 \le a < 10^\ell - 1$. With $a = d_1 10^{\ell-1} + d_2 10^{\ell-2} + \cdots + d_\ell$, and assuming $\gcd(a, n) = 1$, we use Remark to show that $a/n = \overline{d_1\, d_2\, d_3\, \cdots\, d_\ell}$. Note that if a has k digits and $k < \ell$, then $d_i = 0$ for $1 \le i \le \ell - k$.

We know that the decimal expansion of a reduced rational number a/n is purely recurring, and that the length of the recurring part equals $\mathrm{ord}_n 10$. It is easy to see directly from the definition of order that $\mathrm{ord}_n 10 = \ell$ when $n = 10^\ell - 1$. Thus a/n is of the form $\overline{a_1\, a_2\, a_3\, \cdots\, a_\ell}$, and we must show that $a_i = d_i$ for $1 \le i \le \ell$.

From Remark we get $n^\star = 1$ and $c = (n + 1)/10 = 10^{\ell-1}$. Hence $a_\ell = d_\ell$ and $\lambda_\ell = A_{\ell-1}$ by (11). Suppose $a_i = d_i$ for $i \in \{k + 1, \ldots, \ell\}$. Then from (10), (12) and the induction hypothesis, we get

$$\begin{aligned} a_k &\equiv -\frac{d_{k+1}}{10} - \cdots - \frac{d_{\ell-1}}{10^{\ell-k-1}} + \frac{a - d_\ell}{10^{\ell-k}} \pmod{10} \\ &= \frac{a - (d_{k+1}10^{\ell-k-1} + \cdots + d_{\ell-1}10 + d_\ell)}{10^{\ell-k}} \pmod{10} \\ &\equiv d_k \pmod{10}. \end{aligned}$$

This proves our claim by induction.

All this can be seen more directly by observing that the sum of the infinite geometric series with both first term and common ratio equal to $1/10^\ell$ is $1/(10^\ell - 1)$. So

$$\frac{a}{10^\ell - 1} = \frac{a}{10^\ell} + \frac{a}{10^{2\ell}} + \frac{a}{10^{3\ell}} + \cdots = \overline{a_1\, a_2\, a_3\, \cdots\, a_\ell},$$

where $a_1 a_2 a_3 \cdots a_\ell$ represents the number a. As before, if a has k digits and $k < \ell$, then $a_i = 0$ for $1 \le i \le \ell - k$. This immediately leads to the decimal expansion of a/n when $1 \le a < n$ and $\gcd(a, n) = \gcd(n, 10) = 1$. We know that $n \mid (10^\ell - 1)$ for some positive integer ℓ; write $nd = 10^\ell - 1$. This gives $a/n = ad/(10^\ell - 1)$ with $1 \le ad < 10^\ell - 1$, and we can immediately write down the decimal expansion of a/n.

Acknowledgment The author wishes to thank the two anonymous referees and the Editor for their comments that have greatly improved the exposition of this article.

REFERENCE

1. G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers.* Fourth ed. Clarendon Press, Oxford Univ. Press, New York, 1959.

Summary. Rational numbers, except those with denominators of the form $2^a 5^b$, have a recurring decimal expansion. It is usual to write these digits from left to right. We give a procedure to write the decimal digits from the opposite end—right to left.

AMITABHA TRIPATHI (MR Author ID: 262251) enjoys lecturing on mathematical topics that are not routine, in particular mathematical problem solving. Outside of mathematics, he enjoys playing table tennis, working on crosswords and sudoku, listening to music, and when given an opportunity, demonstrating his computational abilities.

Capelli–Rédei Theorem, Solvable Quintics, and Finite Fields

MATT D. LUNSFORD
Union University
Jackson, TN 38305
mlunsford@uu.edu

The question of solvability by radicals of the general quintic polynomial remained a significant open problem in mathematics from the time of the publication of Cardano's *Ars Magna* (1545) until the early decades of the 19th century. Recall that a polynomial $f(x)$ is said to be *solvable by radicals* if it is possible to express all of the roots of the polynomial equation $f(x) = 0$ using only known quantities, field operations, and the extraction of roots. Building upon the work of Lagrange, Neils Henrik Abel proved in 1824 that the general quintic polynomial is not solvable by radicals. Soon afterward, Evariste Galois, in a paper published posthumously in 1846, gave a group-theoretic criterion for determining which specific polynomials are solvable by radicals. While the general quintic polynomial is not solvable by radicals, one can use the criterion of Galois to show that many quintics are solvable.

In 1991, D. S. Dummit [**1**] published an explicit formula in radicals for the roots of solvable quintics with rational coefficients. Furthermore, this formula holds for solvable quintics with coefficients taken from any field of characteristic not equal to 2 or 5. To demonstrate, consider the solvable quintic polynomial $f(x) = x^5 - 5x + 12$. Dummit's formula [**1**, pp. 399–400] yields the five roots as

$$
\begin{aligned}
x_1 &= (r_1 + r_2 + r_3 + r_4)/5\\
x_2 &= \left(\zeta^4 r_1 + \zeta^3 r_2 + \zeta^2 r_3 + \zeta r_4\right)/5\\
x_3 &= \left(\zeta^3 r_1 + \zeta r_2 + \zeta^4 r_3 + \zeta^2 r_4\right)/5\\
x_4 &= \left(\zeta^2 r_1 + \zeta^4 r_2 + \zeta r_3 + \zeta^3 r_4\right)/5\\
x_5 &= \left(\zeta r_1 + \zeta^2 r_2 + \zeta^3 r_3 + \zeta^4 r_4\right)/5
\end{aligned}
$$

where ζ is a designated primitive fifth root of unity, and (with fifth roots chosen appropriately)

$$r_1 = \sqrt[5]{-3125 - 1250\sqrt{5} - \frac{750}{2}\sqrt{100 + 20\sqrt{5}} - \frac{375}{2}\sqrt{100 - 20\sqrt{5}}},$$

$$r_2 = \sqrt[5]{-3125 + 1250\sqrt{5} + \frac{375}{2}\sqrt{100 + 20\sqrt{5}} - \frac{750}{2}\sqrt{100 - 20\sqrt{5}}},$$

$$r_3 = \sqrt[5]{-3125 + 1250\sqrt{5} - \frac{375}{2}\sqrt{100 + 20\sqrt{5}} + \frac{750}{2}\sqrt{100 - 20\sqrt{5}}},$$

$$r_4 = \sqrt[5]{-3125 - 1250\sqrt{5} + \frac{750}{2}\sqrt{100 + 20\sqrt{5}} + \frac{375}{2}\sqrt{100 - 20\sqrt{5}}}.$$

Math. Mag. **90** (2017) 114–116. doi:10.4169/math.mag.90.2.114.
MSC: Primary 12C05

One might ask if these radical expressions r_1, r_2, r_3, r_4 that yield the five roots belong to the smallest extension field of Q containing the roots? In the comparable case of an irreducible cubic equation with rational coefficients having three real roots, the answer is surprisingly no. This famous result, known as the *casus irreducibilis*, states that to express in radicals the real roots of the irreducible cubic with rational coefficients requires elements from the field of complex numbers.

In a previous note in THIS MAGAZINE [**3**], it was shown that a situation analogous to the classic *casus irreducibilis* exists in the context of irreducible cubic polynomials over finite fields if it is required that solvability by radicals means solvability by irreducible radicals. Here we wish to extend this analogy to solvable irreducible quintics over finite fields. Before beginning, we review the necessary essentials.

Let F be a field. A polynomial $f(x) \in F[x]$ is *irreducible* over F if it cannot be written as the product of two polynomials in $F[x]$ of lower degree. A field E is called a *splitting field* of the polynomial $f(x) \in F[x]$ if $F \subseteq E$, $f(x)$ splits into linear factors over E, and E is generated by F and the roots of $f(x)$. A field $K = F(b)$ is an *irreducible radical extension* of F if it is a radical extension (i.e., $b^m = a \in F$) and the polynomial $x^m - a$ is irreducible over F. Similarly, a chain of fields $F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t$ is an *irreducible radical tower* over F if each successive field extension is an irreducible radical extension of the preceding field. Finally, an element is expressible in terms of *irreducible radicals* if it lies in some irreducible radical tower.

We now restrict our attention to finite fields. Let F_p denote the prime finite field of characteristic prime p and let F_p^* denote the multiplicative cyclic group of order $p-1$ consisting of the nonzero elements of F_p. Also, let $GF(p^n)$ denote the Galois field with p^n elements, which is a radical extension of degree n over F_p. Finally, $f(x) \in F_p[x]$ will be an irreducible polynomial of degree n over F_p and $E = GF(p^n)$ will denote the splitting field of $f(x)$ over F_p.

Of particular interest is the case when $E = GF(p^n)$ is an irreducible radical extension of F_p. To this end, recall the following theorem that establishes the existence of irreducible binomial polynomials over finite fields.

Theorem. (Capelli, Rédei) [**2**, p. 115], *Let $m \geqslant 2$ be an integer and let $a \in F_p^*$. The binomial polynomial $x^m - a$ is irreducible over F_p if and only if the following two conditions are satisfied:*

i) *each prime factor t of m divides the order of a, denoted by $o(a)$, in F_p^*, but t does not divide $\frac{p-1}{o(a)}$, and*
ii) $p \equiv 1 \pmod 4$ *if* $m \equiv 0 \pmod 4$.

To demonstrate, consider the case of $m = 4$ and $p = 5$, so that ii) above is satisfied. Then, the binomial polynomial $x^4 - 2$ is irreducible over F_5 since in F_5^*, $o(2) = 4$ and 2 divides $o(a)$ but 2 does not divide $\frac{p-1}{o(a)} = \frac{4}{4} = 1$. However, if we change from $p = 5$ to $p = 7$, then ii) above is no longer satisfied and $x^4 - 2$ is reducible over F_7. In fact, it factors as $x^4 - 2 = (x+2)(x+5)\left(x^2+4\right)$.

We now seek a condition establishing the existence of irreducible radical extensions over prime finite fields.

Theorem. *Let $m \geqslant 2$ be an integer. $GF(p^m)$ is an irreducible radical extension of F_p if and only if the following two conditions are satisfied:*

i) $p \equiv 1 \pmod t$ *for each prime factor t of m, and*
ii) $p \equiv 1 \pmod 4$ *if* $m \equiv 0 \pmod 4$.

Proof. Let t be a prime divisor of the degree m. Suppose $GF(p^m)$ is an irreducible radical extension of F_p. Then there exists $a \in F_p^*$ such that $x^m - a$ is irreducible over F_p. From the Capelli–Rédei theorem, t divides o(a). By Lagrange's theorem, $o(a)$ must divide the order of the cyclic group F_p^*, which has $p - 1$ elements. It follows that t divides $p - 1$. Thus, $p \equiv 1 \pmod{t}$. Finally, if $m \equiv 0 \pmod{4}$ then $p \equiv 1 \pmod{4}$ by ii) of Capelli–Rédei. Conversely, suppose that i) and ii) above hold. Choose $a \in F_p^*$ such that $o(a) = p - 1$, which is possible since F_p^* is a cyclic group. Then t divides $o(a)$ but does not divide $\frac{p-1}{o(a)} = 1$. Therefore, $x^m - a$ is irreducible over F_p by the Capelli–Rédei theorem. Hence, $GF(p^m)$ is an irreducible radical extension of F_p. ■

Two corollaries follow immediately from the theorem.

Corollary. *Let m be a prime. $GF(p^m)$ is an irreducible radical extension of F_p if and only if $p \equiv 1 \pmod{m}$.*

Corollary. *Let $m = 2^k,\ k \geqslant 2$. $GF(p^m)$ is an irreducible radical extension of F_p if and only if $p \equiv 1 \pmod{4}$.*

The two corollaries give a simple divisibility condition to determine whether or not the roots of an irreducible polynomial $f(x) \in F_p[x]$ of degree m are expressible in terms of irreducible radicals present in the splitting field $E = GF(p^m)$. In particular, the roots can be expressed in terms of irreducible radicals in E exactly when $p \equiv 1 \pmod{3}$ for irreducible cubics, $p \equiv 1 \pmod{4}$ for irreducible quartics, and $p \equiv 1 \pmod{5}$ for solvable irreducible quintics.

Returning to the earlier example, we note that $f(x) = x^5 - 5x + 12$ is irreducible over both F_7 and F_{11}. If $p = 7$, then the roots of this polynomial cannot be expressed in terms of irreducible radicals in $GF(7^5)$. In particular, a square root of 5 and a primitive fifth root of unity are not present in the splitting field $GF(7^5)$. These elements, however, do lie in the extension field $GF(7^{20})$. If, on the other hand, $p = 11$, then $GF(11^5)$ is an irreducible radical extension of F_{11} that contains the required elements. Using Dummit's formula, the roots of $x^5 - 5x + 12 = 0$ are expressible in terms of irreducible radicals in $GF(11^5)$, where $\zeta = 3$ is the designated primitive fifth root of unity.

REFERENCES

1. D. S. Dummit, Solving solvable quintics, *Math Comput* **57** no. 195 (1991) 387–401.
2. R. Lidl, H. Neiderreiter, *Introduction to Finite Fields and their Applications*. Cambridge Univ. Press, Cambridge, 1986.
3. M. D. Lunsford, Cardano, *casus irreducibilis*, and finite fields, *Math Mag* **87** no. 5 (2014) 377–380.

Summary. A situation analogous to the classic *casus irreducibilis* exists in the context of irreducible cubic polynomials over finite fields if it is required that solvability by radicals means solvability by irreducible radicals. In this paper we extend this analogy to irreducible quintics over finite fields by use of the Capelli–Rédei theorem.

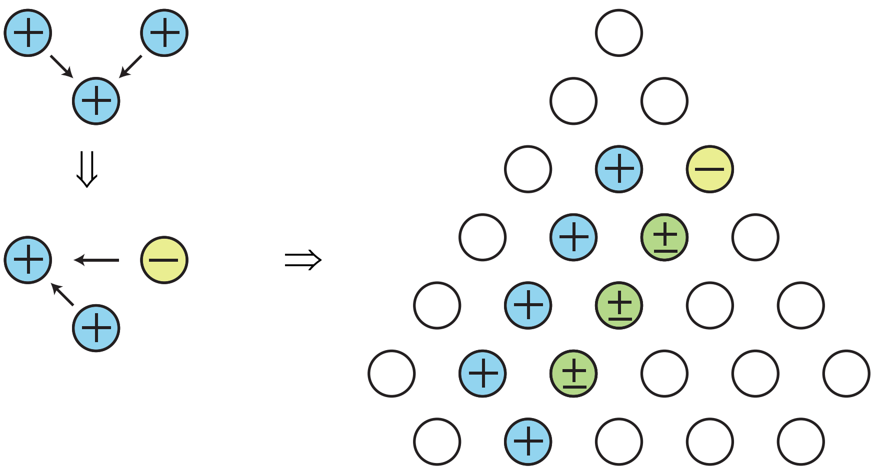
MATT D. LUNSFORD (MR Author ID: 367743) is Professor of Mathematics at Union University in Jackson, TN, where he has been a faculty member since 1993. He holds a doctorate in mathematics from Tulane University. His current research interests include classical Galois theory and finite fields. He and his wife Deanna have three children: Cara, Thomas, and Emma, and one son-in-law: Brennan.

Proof Without Words: Partial Column Sums in Pascal's Triangle

ÁNGEL PLAZA
Universidad de Las Palmas de Gran Canaria
Las Palmas, Canaria, Spain
angel.plaza@ulpgc.es

Theorem. For any integers $j \le m \le n$: $\sum_{k=m}^{n} \binom{k}{j} = \binom{n+1}{j+1} - \binom{m}{j+1}$.

Proof.



$$\binom{k}{j} = \binom{k+1}{j+1} - \binom{k}{j+1} \Rightarrow \sum_{k=m}^{n} \binom{k}{j} = \binom{n+1}{j+1} - \binom{m}{j+1}.$$

■

Exercise. Show that for any integers $j \ge 0$ and $m \le n$:

$$\sum_{k=m}^{n} \binom{k+j}{k} = \binom{n+j+1}{n} - \binom{m+j}{m-1}.$$

Note: Setting $j = m$ and $\binom{m}{m+1} = 0$ in the theorem yields one form of the hockey stick identity, while setting $m = 0$ and $\binom{m+j}{-1} = 0$ in the exercise yields another form of the hockey stick identity. See [**1, 2**].

Math. Mag. **90** (2017) 117–118. doi:10.4169/math.mag.90.2.117. © Mathematical Association of America
MSC: Primary 05A10.

REFERENCES

1. P. Hilton, J. Pedersen, Looking into Pascal's triangle: Combinatorics, arithmetic, and geometry, *Math. Mag.* **60** no. 5 (1987) 305–316.
2. S. Mehri, J. Pedersen, The hockey stick theorems in Pascal and trinomial triangles, arXiv:1404.5106v1. [math.HO] 21 Apr. 2014.

Summary. Based on the binomial property $\binom{k+1}{j+1} = \binom{k}{j} + \binom{k}{j+1}$, written as $\binom{k}{j} = \binom{k+1}{j+1} - \binom{k}{j+1}$, the sum of consecutive column entries of Pascal's triangle is written as a difference of two binomial coefficients in the next column, which generalizes the so-called hockey stick identities.

ÁNGEL PLAZA (MR Author ID: 350023) received his masters degree from Universidad Complutense de Madrid in 1984 and his Ph.D. from Universidad de Las Palmas de Gran Canaria in 1993, where he is a full professor in applied mathematics. He is interested in mesh generation and refinement, combinatorics and visualization support in teaching and learning mathematics.

PINEMI PUZZLE

			8		5		4	4	
	16			6		7	7		3
			12	11	7	8		8	4
6	9	8							
	4		10			9	9		8
4		7		8	10		12	11	
	7		6		9				9
6		5	6				11		
	7		6		10		8		
4		5				7		6	3

How to play. Place one jamb (|), two jambs (||), or three jambs (|||) in each empty cell. The numbers indicate how many jambs there are in the surrounding cells—including diagonally adjacent cells. Each row and each column has 10 jambs. Note that no jambs can be placed in any cell that contains a number.

The solution is on page 141.

—contributed by Lai Van Duc Thinh,
Vietnam; fibona2cis@gmail.com

The Asymptotic Behavior of the Binomial Coefficients

PAUL LEVRIE
University of Antwerp, Belgium
KU Leuven, Belgium
paul.levrie@cs.kuleuven.be

Each year, when teaching my first-year calculus class, I get stuck when I want to write down the interval of convergence of the binomial series, the Maclaurin series for the function $f(x) = (1+x)^\alpha$ with α a real number. That the series in question,

$$(1+x)^\alpha = 1 + \sum_{n=1}^{\infty} \binom{\alpha}{n} x^n, \tag{1}$$

converges for all x with $|x| < 1$, is easy to prove using the ratio test. But what happens at the endpoints of the interval of convergence? The answer may be found in some calculus books [**1**], as shown in Table 1. In the last case of Table 1, we normally have $[-1, 1]$ except when α is an integer and the corresponding series terminates.

TABLE 1: Intervals of convergence of the binomial series for $(1+x)^\alpha$.

Case	Value of α	Interval of convergence
a	$\alpha \le -1$	$(-1, 1)$
b	$-1 < \alpha < 0$	$(-1, 1]$
c	$0 \le \alpha$	$[-1, 1]$ or $\mathbb{R}$

Did you ever wonder where these 3 cases come from? Note that using Abel's theorem [**2**] the case $\alpha < 0$ and $x = -1$ can be dealt with. The answer for the other cases lies in the asymptotic behavior of the binomial coefficients, where

$$\left|\binom{\alpha}{n}\right| \sim C \cdot \frac{1}{n^{1+\alpha}} \quad \text{as} \quad n \to \infty. \tag{2}$$

The key to the proof of (2) is the gamma function, and more specifically Euler's product formula for the gamma function:

$$\Gamma(s) = \lim_{n\to\infty} \frac{n!n^s}{s(s+1)(s+2)\cdots(s+n-1)(s+n)} \quad (s \neq 0, -1, -2, \ldots). \tag{3}$$

In this note, we will give a rigorous proof of this formula, and we'll use it to determine the asymptotic behavior of the binomial coefficients. As an application we'll prove the three cases of convergence of the binomial series at the points $x = \pm 1$.

Math. Mag. **90** (2017) 119–123. doi:10.4169/math.mag.90.2.119. © Mathematical Association of America
MSC: Primary 40A05

Euler's product formula for the gamma function

Formula (3) is not as well known as the usual definition of the gamma function, in which

$$\Gamma(s) = \int_0^{+\infty} t^{s-1} e^{-t} dt \quad (s > 0). \tag{4}$$

For negative values of $s \neq -1, -2, \ldots$, the gamma function is then defined using the functional equation

$$\Gamma(s+1) = s \cdot \Gamma(s). \tag{5}$$

We will first prove (3) using (4) and (5). Note that for s a positive integer, $\Gamma(s) = (s-1)!$.

Theorem 1. *For all $s \neq 0, -1, -2, \ldots,$ we have*

$$\Gamma(s) = \lim_{n \to \infty} \frac{n! n^s}{s(s+1)(s+2)\cdots(s+n)}. \tag{6}$$

The proof uses the following lemma.

Lemma. For $n = 1, 2, 3, \ldots$, the function F_n with $F_n(t) = e^{-t} - \left(1 - \frac{t}{n}\right)^n$ satisfies

$$0 \leq F_n(t) \leq \frac{e^{-1}}{n}$$

for all t in $[0, n]$.

Proof of the lemma (see also [**3**]). For $n = 1$ we use the fact that $F_1'(t) > 0$ for $t > 0$, with $F_1(0) = 0$ and $F_1(1) = e^{-1}$. For $n > 1$ the first inequality follows immediately from

$$0 \leq 1 - u \leq e^{-u} \quad \text{for } u \in [0, 1]$$

by replacing u by $\frac{t}{n}$ and raising the result to the nth power.

The second inequality is a consequence of the continuity of F_n. Since $F_n(0) = 0$, $F_n(n) = e^{-n}$, and $F_n'(n) = -e^{-n}$, the function F_n reaches a maximum value at some point t_0 between 0 and n, with $F_n(t) \leq F_n(t_0)$ for all $t \in [0, n]$. Since F_n is differentiable, we have that $F_n'(t_0) = 0$. Hence,

$$e^{-t_0} - \left(1 - \frac{t_0}{n}\right)^{n-1} = 0.$$

We use this to rewrite $F_n(t_0)$ so that

$$F_n(t_0) = e^{-t_0} - e^{-t_0} \cdot \left(1 - \frac{t_0}{n}\right) = \frac{t_0 e^{-t_0}}{n}.$$

Since $H(t) = te^{-t}$ reaches its maximum value for $t = 1$, we have that

$$F_n(t) \leq F_n(t_0) \leq \frac{e^{-1}}{n} \quad \text{for } t \in [0, n].$$

This proves the lemma. ∎

Proof of Theorem 1. The case that s is a positive integer is easy: The limit at the right-hand side of (6) reduces to $(s-1)!$.

The rest of the proof is in two parts. In the first part we prove the theorem for values of s with $0 < s < 1$. We start by writing the improper integral as a limit. It is a consequence of the lemma that

$$\int_0^{+\infty} t^{s-1}e^{-t}dt = \lim_{n\to\infty}\int_0^n t^{s-1}\Big(1-\frac{t}{n}\Big)^n dt. \tag{7}$$

Indeed, we can rewrite the inequality of the lemma in this way:

$$0 \le t^{s-1}e^{-t} - t^{s-1}\Big(1-\frac{t}{n}\Big)^n \le \frac{e^{-1}}{n}t^{s-1} \text{ for } t \in (0, n].$$

We then integrate between 0 and n to yield

$$0 \le \int_0^n \left(t^{s-1}e^{-t} - t^{s-1}\Big(1-\frac{t}{n}\Big)^n\right)dt \le \frac{e^{-1}}{n}\frac{n^s}{s}.$$

Note that the improper integral in the middle converges by the comparison test. Taking the limit as $n \to \infty$ proves (7), since the right-hand side of this last inequality has limit 0 (we have assumed that $s < 1$).

Now we write the integral at the right-hand side of (7) as a product. This is done by repeatedly using integration by parts:

$$\begin{aligned}
\int_0^n t^{s-1}\Big(1-\frac{t}{n}\Big)^n dt &= \frac{1}{s}\left(\left[t^s\Big(1-\frac{t}{n}\Big)^n\right]_0^n + \frac{n}{n}\int_0^n t^s\Big(1-\frac{t}{n}\Big)^{n-1}dt\right)\\
&= \frac{1}{s}\frac{n}{n}\int_0^n t^s\Big(1-\frac{t}{n}\Big)^{n-1}dt\\
&= \frac{1}{s}\frac{n}{n}\frac{1}{s+1}\frac{n-1}{n}\int_0^n t^{s+1}\Big(1-\frac{t}{n}\Big)^{n-2}dt = \cdots\\
&= \frac{1}{s}\frac{n}{n}\frac{1}{s+1}\frac{n-1}{n}\cdots\frac{1}{s+n-1}\frac{1}{n}\int_0^n t^{s+n-1}\Big(1-\frac{t}{n}\Big)^0 dt\\
&= \frac{1}{s}\frac{n}{n}\frac{1}{s+1}\frac{n-1}{n}\cdots\frac{1}{s+n-1}\frac{1}{n}\frac{n^{s+n}}{s+n}.
\end{aligned}$$

Hence we find that

$$\int_0^n t^{s-1}\Big(1-\frac{t}{n}\Big)^n dt = \frac{n!n^s}{s(s+1)(s+2)\cdots(s+n)}.$$

Taking the limit as $n \to \infty$ proves the case $0 < s < 1$.

In the second part we extend this result to other values of s. This is done using the functional equation for the gamma function. From

$$\Gamma(s) = \lim_{n\to\infty}\frac{n!n^s}{s(s+1)(s+2)\cdots(s+n)} = \frac{\Gamma(s+1)}{s} \text{ for } 0 < s < 1,$$

we deduce that

$$\Gamma(s+1) = \lim_{n\to\infty} \frac{n!n^s}{(s+1)(s+2)\cdots(s+n)}$$
$$= \lim_{n\to\infty} \frac{n!n^{s+1}}{(s+1)(s+2)\cdots(s+n)(s+1+n)}.$$

Hence (3) is true for $1 < s < 2$, and we can repeat this process for all other intervals for s. ∎

Asymptotic behavior of the binomial coefficients

Now we're ready to prove our main result.

Theorem 2. *For* $\alpha \neq 0, 1, 2, \ldots$, *we have*

$$\binom{\alpha}{n} \sim \frac{(-1)^n}{\Gamma(-\alpha)n^{1+\alpha}}, \qquad n \to \infty.$$

Proof. Comparing the definition of the binomial coefficient,

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)}{n!},$$

with the right-hand side of (3), it is easy to see that this product formula is just what we need. We can rewrite Euler's formula so that

$$\Gamma(s) = \lim_{n\to\infty} \frac{n!n^{s-1}}{s(s+1)(s+2)\cdots(s+n-1)} = \lim_{n\to\infty} \frac{(-1)^n n^{s-1}}{\binom{-s}{n}}$$

($s \neq 0, -1, -2, \ldots$) and by replacing $-s$ by α we get

$$\lim_{n\to\infty} (-1)^n \binom{\alpha}{n} \Gamma(-\alpha) n^{1+\alpha} = 1$$

which implies that

$$\binom{\alpha}{n} \sim \frac{(-1)^n}{\Gamma(-\alpha)n^{1+\alpha}}, \qquad n \to \infty$$

for $\alpha \neq 0, 1, 2, \ldots$.

Convergence of the binomial series

Using Theorem 2 it's a nice exercise on convergence tests for series (see [**2**]) to explain the convergence behavior of the binomial series (1) at the endpoints of the convergence interval.

We will use the fact that the sequence of binomial coefficients $\binom{\alpha}{n}$ is ultimately alternating and decreasing.

Case a. $\alpha \leq -1$

Both series, for $x = -1$ as well as for $x = 1$, diverge as a consequence of the *divergence test*, since, with C as in (2), $\lim_{n\to\infty} C/n^{1+\alpha} \neq 0$.

Case b. $-1 < \alpha < 0$

The series with $x = -1$ diverges. From some n on, the terms of the series will all be positive. Divergence follows from the divergence of the p-series $\sum_{n=0}^{\infty} 1/n^{1+\alpha}$ using the *limit comparison test*.

For $x = 1$ the series converges. The series is ultimately alternating, with monotonically decreasing terms going to zero. Convergence follows from the *alternating series test*.

Case c. $0 \leq \alpha$ (we assume that α is not an integer)

The series with $x = -1$ converges. From some n on, the terms of the series will all be positive. Convergence follows from the convergence of the p-series $\sum_{n=0}^{\infty} 1/n^{1+\alpha}$ using the *limit comparison test*.

For $x = 1$ the series converges by the *absolute convergence test*.

Acknowledgment The author would like to thank the anonymous referee for his interesting comments.

REFERENCES

1. D. M. Bressoud, *A Radical Approach to Real Analysis*. The Mathematical Association of America, Washington, DC, 2007.
2. K. Knopp, *Theory and Application of Infinite Series*. Dover Publications, New York, 1990.
3. P. Levrie, W. Daems, Evaluating the probability integral using Wallis's product formula for π, *Amer. Math. Monthly* **116** (2009) 538–541.

Summary. In this note we give a rigorous proof by elementary means of the asymptotic behavior of the binomial coefficients. As an application we look at the convergence of the binomial series at the endpoints of the convergence interval.

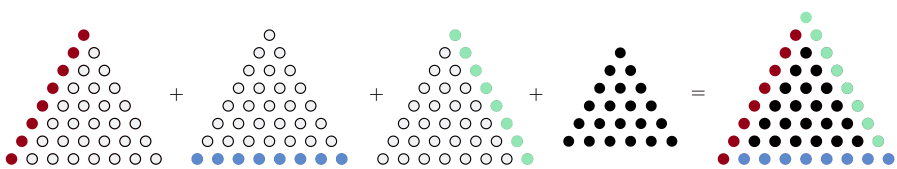
PAUL LEVRIE (MR Author ID: 230199) obtained a doctorate in mathematics from the Catholic University of Leuven, Belgium in 1987 in the field of numerical analysis. Since then he has been busy teaching and trying to find easy proofs of well-known mathematical results. He coauthors a Dutch blog entitled Wiskunde is Sexy (Mathematics is Sexy), in an attempt to make mathematics more popular in Flanders.

Proof Without Words: A Recursion for Triangular Numbers and More

TOM EDGAR
Pacific Lutheran University
Tacoma, WA
edgartj@plu.edu

Theorem. *The triangular numbers,* $t_n := 1 + 2 + 3 + \cdots + n$*, satisfy the recursion* $t_n = 3t_{n-1} - 3t_{n-2} + t_{n-3}$.

Proof.



■

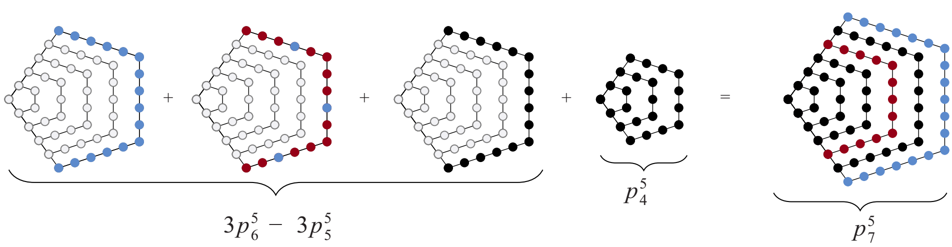
If a_n and b_n are sequences that both satisfy the same recurrence, any sequence $s_n := ca_n + db_n$ satisfies the recurrence as well; so, the theorem provides a wordless proof for each of the following three statements (since $z_n = 1$ satisfies the recurrence).

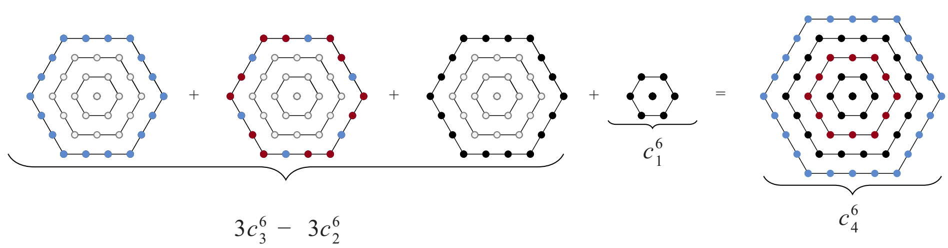
Corollary 1. *The* k*-gonal numbers, given by either* $p_n^k = \sum_{i=0}^{n-1} ((k-2) \cdot i + 1)$ *or* $p_n^k = (k-3)t_{n-1} + t_n$*, satisfy the recursion* $p_n^k = 3p_{n-1}^k - 3p_{n-2}^k + p_{n-3}^k$.

Corollary 2. *The centered* k*-polygonal numbers, given by* $c_n^k = kt_{n-1} + 1$*, satisfy the recursion* $c_n^k = 3c_{n-1}^k - 3c_{n-2}^k + c_{n-3}^k$.

Corollary 3. *Any quadratic sequence* $f_n = an^2 + bn + c$ *satisfies the recursion* $f_n = 3f_{n-1} - 3f_{n-2} + f_{n-3}$ *since* $n^2 = t_n + t_{n-1}$ *and* $n = t_n - t_{n-1}$.

Corollaries 1 and 2 can be visualized directly using diagrams analogous to the following for the pentagonal and centered hexagonal numbers.





Math. Mag. **90** (2017) 124–125. doi:10.4169/math.mag.90.2.124. © Mathematical Association of America
MSC: Primary 11B37

Summary. We visually demonstrate a recurrence satisfied by the triangular numbers and hence all quadratic sequences including all types of polygonal numbers.

TOM EDGAR (MR Author ID: 821633) is an associate professor of mathematics at Pacific Lutheran University in Tacoma, Washington.

A New Theory of Strings

Yes, no one knows if atoms are real
but they sure seem like they exist.
After all, what are we made of
beyond skin and bones and cells
and maybe some other things like love.

So, a name for something tiny
might as well be pieces of string,
Which reminds me of a shiny harp!
Yes, that's the kind of thing
it could be, because when
the harpist plucks, watch the strings.

They seem to vibrate with music,
yes invisible sounds that penetrate
thru our ears to our very souls,
which, of course, no one can see,
but we know exist, in reality,
or theory?

— Del Corey, Emeritus
Macomb Community College
delmcorey@yahoo.com

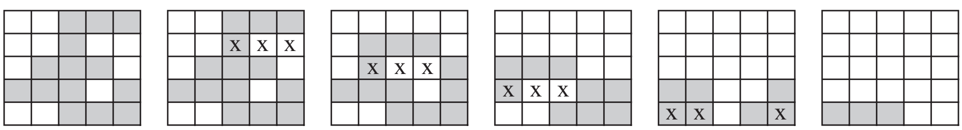
Chasing the Lights in Lights Out

C. DAVID LEACH
University of West Georgia
Carrollton, GA 30118
cleach@westga.edu

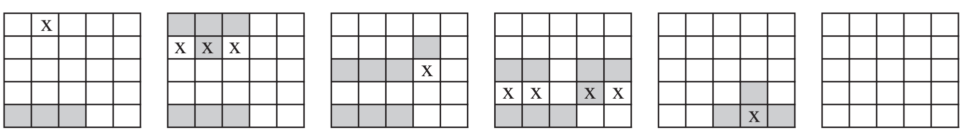
Lights Out is a game played on a 5×5 grid of light-up buttons. It was marketed in the 1990s as a handheld electronic game by Tiger Electronics®, and versions and variations of it are available online. Each button can be in one of two states: on or off. The neighbors of a button include the buttons directly above, below, right, and left of the button. When a button is pressed, its state and the state of its neighbors are toggled. At the start of the game, the lights are lit in a seemingly random configuration; the goal is to turn all of the lights off.

Determining the optimal solution—one requiring the fewest button presses—for a given configuration can be done using Gauss–Jordan elimination. However, the matrices involved are rather large and do not lend themselves well to repeated play of the game.

Let's play a game using a very simple strategy: starting with the second row, press every button in that row that is immediately beneath a light in row 1 that is on. After completing this task, all lights in row 1 are off. Now do the same with row 3, pressing those buttons beneath the lights in row 2 that are now lit. Continue row by row until you get to the bottom of the board. The diagram below shows our progress through a game, with the initial configuration shown on the leftmost board. White cells are off, shaded cells are on, and X represents a button being pressed.



At this point our simple strategy seems to break down, since it does not provide any means for turning out the lights in the bottom row. With no clear way to proceed, suppose that we "randomly" press the second button in the first row. This turns on a few lights, so we repeat the process of clearing the rows from top to bottom; however, this time when we finish, all the lights are off.



The strategy we employed here is popularly known as *light chasing*. Light chasing consists of three stages. Stage 1, shown in the first diagram above, is a what we term a *chase*. At the end of stage 1, the only lights that remain on are those in the bottom row. Stage 2 involves a lookup using Table 1, which can be found on the web at a number of places, including [**14**]. If we let 0 and 1 represent the off and on states, respectively, the bottom row of our board gives the configuration 11100. The lookup table tells us to press only the second button in the top row. If the lights in the bottom row are not

Math. Mag. **90** (2017) 126–133. doi:10.4169/math.mag.90.2.126.
MSC: Primary 91A60, Secondary 15A24.

in one of the configurations in the lookup table, then the game is unsolvable. Stage 3 is another chase, only at the end of stage 3 all the lights are off.

Light chasing requires practically no computation, but it does require that we have the lookup table at our disposal. Generating the lookup table takes a bit of computational effort, but once it is done, it can be used repeatedly.

TABLE 1: Lookup table for Lights Out.

Lights Lit in Bottom Row	Buttons to Press in Top Row
00111	00010
01010	10010
01101	10000
10110	00001
10001	11000
11011	00100
11100	01000

History

The mathematics behind Lights Out has been studied by many different authors, and several results have been discovered and published more than once. This has led to confusion as to who made certain discoveries first.* Much of the reason for this confusion is because different authors approached the problem from different perspectives and used different terminology to describe the game. The name *Lights Out* was introduced by Tiger Electronics in 1995, and many articles, including [**1**] have used that name. Other names that have been used to describe the same problem include the *switch-setting problem* [**5, 4**] the *lamp-lighting problem* [**10**], and the *nine tails problem* (for a 3×3 board). Some authors approached the game as a 2-dimensional cellular automaton, and used the names σ*-game* and σ*-automata* [**2, 7, 13**]. The earliest reference to a Lights Out-type problem comes from [**11**], in which Lovász credits Tibor Gallai for solving the *all-ones problem*, a related problem in graph theory, in the early 1970s.

The concept of light chasing has been examined for the on-off game in [**2, 5, 4, 13**]. The multicolor case, in which the lights cycle through a sequence of colors before turing off, was examined in [**7**]. However, the term *light chasing*, while popular on internet discussions of Lights Out, does not appear in any of those articles.

The main purpose of this paper is to examine light chasing, with a particular emphasis on generating lookup tables. We will begin by generating Table 1, then generate lookup tables for variations of the game with grids of other sizes and lights that cycle through multiple colors. The methods we use can be implemented using the free software SageMath [**12**] or another computer algebra system.

Modeling the game

Since a light has two states—on and off—we can represent its state with an element from $\mathbb{Z}_2$. Pressing a button adds 1 to its state and its neighbors' states modulo 2. This

*The author would like to thank the anonymous referees for providing much of the historical information contained in this section.

simple observation has two useful consequences: since addition is commutative, the order in which the buttons are pressed is unimportant; since addition in $\mathbb{Z}_2$ is an order 2 operation, no button needs to be pressed more than once. At the beginning of a game, the lights have some initial configuration. There are 2^{25} possible configurations, but only 1/4 of them are solvable [**1**].

Before we analyze light-chasing, let us briefly outline the usual method for solving the game with linear algebra. For a more detailed description, see [**1**]. We label the 25 buttons from the upper left to the lower right as $b_1, \ldots, b_{25}$. A configuration of the game board is described by the vector $\vec{b} = [b_1, b_2, \ldots, b_{25}]$, where $b_i = 1$ if light i is on and 0 if off. We define the 25×25 matrix M by $m_{i,j} = 1$ if pressing button b_i affects light b_j and 0 otherwise. Solving $M\vec{x} = \vec{b}$ for $\vec{x}$ gives the set of buttons to press to turn all the lights off.

Let A be the the tridiagonal 5×5 submatrix in the upper left corner of M. The entries of A describe which buttons in row i affect which lights in row i. The matrix A is repeated along the diagonal—once for each row on the game board—and is used extensively in the analysis of light chasing.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1	1				1																			
2	1	1	1				1																		
3		1	1	1				1																	
4			1	1	1				1																
5				1	1					1															
6	1					1	1				1														
7		1				1	1	1				1													
8			1				1	1	1				1												
9				1				1	1	1				1											
10					1				1	1					1										
11						1					1	1				1									
12							1				1	1	1				1								
13								1				1	1	1				1							
14									1				1	1	1				1						
15										1				1	1					1					
16											1					1	1				1				
17												1				1	1	1				1			
18													1				1	1	1				1		
19														1				1	1	1				1	
20															1				1	1					1
21																1					1	1			
22																	1				1	1	1		
23																		1				1	1	1	
24																			1				1	1	1
25																				1				1	1

$= M$

Light chasing

Before we attempt to solve the game using light chasing, let's examine the effect that a single chase would have on the rows of the game board if we were to begin with all the lights off and we press an arbitrary set of buttons in the top row. For each row i we define two vectors: let $\vec{p}_i = [p_{i1}\ p_{i2}\ p_{i3}\ p_{i4}\ p_{i5}]$, where p_{ij} is the number of times that the jth button in row i will be pressed during the chase; let $\vec{s}_i = [s_{i1}\ s_{i2}\ s_{i3}\ s_{i4}\ s_{i5}]$, where s_{ij} is the state of the jth light in row i after row i is pressed, but before row $i+1$ is pressed during the chase. Although $\vec{p}_i$ and $\vec{s}_i$ refer to rows of the game board, in all of our calculations, they will be used as column vectors. Once we choose an arbitrary $\vec{p}_1$, the remaining $\vec{p}_i$ and all the $\vec{s}_i$ vectors are completely determined. The value of $\vec{s}_i$ is determined only by the buttons pressed in rows i and $i-1$, and is given by the following equation from [**4**]:

$$\vec{s}_i = A\vec{p}_i + \vec{p}_{i-1}. \tag{1}$$

We adopt the convention that $\vec{p}_0 = \vec{0}$, since there are no buttons to be pressed above row 1. To turn off the lights in row $i - 1$, we must have that

$$\vec{p}_i = -\vec{s}_{i-1} \tag{2}$$

for $i \geq 2$. Since our calculations are being done over $\mathbb{Z}_2$, we could dispense with negatives and rewrite equation (2) as $\vec{p}_i = \vec{s}_{i-1}$; however, we will retain the negative here and in subsequent equations in order to generalize to $\mathbb{Z}_p$ for $p > 2$ later. Starting with $\vec{p}_0 = \vec{0}$ and arbitrary $\vec{p}_1$, by alternately applying equations (1) and (2), we can write

$$\vec{s}_i = S_i\vec{p}_1, \tag{3}$$

where S_i is a 5×5 matrix. We can check that for $1 \leq i \leq 5$, the S_i are given by the polynomials $S_1 = A$, $S_2 = -A^2 + I_5$, $S_3 = A^3 - 2A$, $S_4 = 3A^2 - A^4 - I_5$, and

$$S_5 = A^5 - 4A^3 + 3A = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}. \tag{4}$$

Since row 5 is the last row, the vector $S_5\vec{p}_1$ tells the overall effect that the chase has on the last row. More precisely, the chase has the overall effect of adding the vector $S_5\vec{p}_1$ to the last row. The other rows all get zeroed out by the end of the chase.

In our computation of S_5, we use a recursive process that requires computing S_i for all $i \leq 5$, but it's possible to compute any value of S_i directly. Goldwasser et al. [**5**] were the first to show that S_i is given by the $(k + 1)$st Fibonacci polynomial, reduced modulo 2 and evaluated at A. Later we will examine this connection in more detail. For now, let's consider an actual game.

Suppose we have already completed the first chase and the only lights that remain on are in row 5. Let $\vec{s}_b$ (b for bottom) be the state vector of row 5 at this stage of the game. Note that $\vec{s}_b$ is the state of row 5 immediately after the *first* chase is completed, while $\vec{s}_5$ is the state of row 5 after the *second* chase is completed. It is our goal to determine the value of $\vec{p}_1$ that will result in $\vec{s}_5 = \vec{0}$ for the given value of $\vec{s}_b$. Thus we need to solve the equation $S_5\vec{p}_1 + \vec{s}_b = \vec{0}$ for $\vec{p}_1$, which we rearrange to get

$$S_5\vec{p}_1 = -\vec{s}_b. \tag{5}$$

Since $\text{rank}(S_5) = 3$ and $\mathbb{Z}_2$ has only two possible values for each coefficient, the column space of S_5 contains exactly eight vectors—precisely the seven bottom-row vectors listed in the lookup table along with the zero vector.

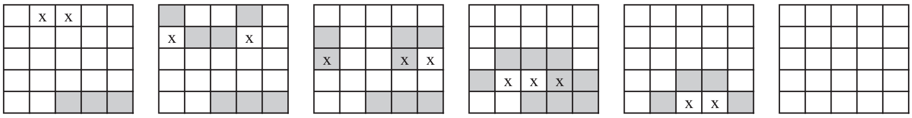
Note that the top row given in the lookup table is not the only solution possible. For example, suppose that the bottom row is $\vec{s}_b = [0\,0\,1\,1\,1]$, and let $\vec{v}_i$ be the ith column of S_5. Solving

$$c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + c_4\vec{v}_4 + c_5\vec{v}_5 = [0\,0\,1\,1\,1] \tag{6}$$

gives the general solution

$$\vec{p}_1 = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + c_5 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \tag{7}$$

for any choice of c_4 and c_5. If we let both be zero, we get $\vec{p}_1 = [0\,1\,1\,0\,0]$, and the game proceeds as follows:



If we choose $c_4 = 1$ and $c_5 = 0$, we get $\vec{p}_1 = [0\,0\,0\,1\,0]$, as given in the lookup table.

Variations of the game

Grids with five columns and *m* rows. First, let's keep the number of columns at 5, and let m be the number of rows. As before we can alternately apply equations (1) and (2) to find matrix S_m for any value of m that we like and proceed exactly as above. However, if m is large, this becomes tedious. It would be preferable to find a closed-form formula for S_m. One way to compute S_m is by using the Fibonacci polynomials.

The Fibonacci polynomials are defined by $F_1(x) = 1$, $F_2(x) = x$, and $F_n(x) = xF_{n-1}(x) + F_{n-2}(x)$ for $n \geq 3$, and are given by the well-known formula

$$F_n(x) = \sum_{i=0}^{\lfloor (n-1)/2 \rfloor} \binom{n-i-1}{i} x^{n-2i-1}. \tag{8}$$

Several authors [**2, 5, 13**] have shown that S_m, when reduced modulo 2, is equivalent to $F_{m+1}(A) \pmod 2$. Making the necessary substitutions into equation (8) gives

$$S_m = F_{m+1}(A) = \sum_{i=0}^{\lfloor m/2 \rfloor} \binom{m-i}{i} A^{m-2i} \pmod 2. \tag{9}$$

Another way to compute S_m is by combining equations (1) and (2) to get the second-order recurrence relation

$$\vec{p}_{i+1} = -A\vec{p}_i - \vec{p}_{i-1} \tag{10}$$

with given initial conditions $\vec{p}_0 = \vec{0}$ and $\vec{p}_1$. In order to solve this for $\vec{p}_i$, we rewrite the recurrence as the first-order recurrence

$$\begin{bmatrix} \vec{p}_{i+1} \\ \vec{p}_i \end{bmatrix} = \begin{bmatrix} -A & -I \\ I & \vec{0} \end{bmatrix} \begin{bmatrix} \vec{p}_i \\ \vec{p}_{i-1} \end{bmatrix}. \tag{11}$$

Let L denote the square matrix in recurrence (11), and note that L is a 10×10 partitioned matrix. Repeatedly applying recurrence (11) amounts to repeated multiplication, giving the following general formulas for $\vec{p}_i$ and $\vec{s}_i$:

$$\begin{bmatrix} \vec{p}_{i+1} \\ \vec{p}_i \end{bmatrix} = L^i \begin{bmatrix} \vec{p}_1 \\ \vec{p}_0 \end{bmatrix} \qquad \begin{bmatrix} \vec{s}_i \\ \vec{s}_{i-1} \end{bmatrix} = -L^i \begin{bmatrix} \vec{p}_1 \\ \vec{p}_0 \end{bmatrix}. \tag{12}$$

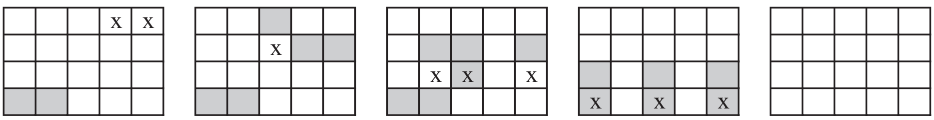
From equation (12) we can extract the equation $\vec{s}_i = S_i\vec{p}_1$, where S_i is the upper left quadrant of the matrix $-L^i$.

Example: Suppose $m = 4$ and $\vec{s}_b = [1\,1\,0\,0\,0]$. We need to solve the equation $S_4\vec{p}_1 = [1\,1\,0\,0\,0]$. To obtain S_4, we take the upper left quadrant of $-\begin{bmatrix} -A & -I \\ I & 0 \end{bmatrix}^4$, which is

$$S_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{13}$$

This particular matrix is quite nice for a few reasons: Since $\operatorname{rank}(S_4) = 5$, all configurations are solvable; since the back diagonal contains all ones, $S_4\vec{x}$ reverses the vector $\vec{x}$, making for a very easy-to-remember strategy.

Solving $S_4\vec{p}_1 = [1\,1\,0\,0\,0]$ gives $\vec{p}_1 = [0\,0\,0\,1\,1]$, and the game proceeds as follows:



Grids with other than 5 columns. Changing the number of columns affects the sizes of the matrices and vectors involved, but the analysis of the game is otherwise unchanged. If we let n denote the number of columns, then A is the $n \times n$ matrix defined by $a_{ij} = 1$ if $|i - j| \leq 1$ and 0 otherwise, and vectors $\vec{p}_i$ and $\vec{s}_i$ now have n entries instead of 5. With these modifications, we can proceed exactly as before.

Lights with more than two states. Another common variation on the lights out game is to have lights that take on more than two states. The states are sometimes represented with different colors, but we will consider them to be the numbers in $\mathbb{Z}_p = \{0, \ldots, p - 1\}$, with 0 being considered off. We will restrict ourselves to the case where p is a prime number so that each element of $\mathbb{Z}_p$ has a multiplicative inverse, making $\mathbb{Z}_p$ a finite field. This restriction is not necessary, but it simplifies the tasks of solving $S_m\vec{p}_1 = -\vec{s}_b$ and finding $\operatorname{rank}(S_m)$ on a computer algebra system, since S_m can be easily row-reduced.

A button's state value increases by 1 (mod p) whenever it or one of its neighbors is pressed. The 0 state is considered off. As in the $\mathbb{Z}_2$ case, the order in which the buttons are pressed is unimportant. In optimal play, no button would need to be pressed more than $p - 1$ times; with light chasing no button needs to be pressed more than $p - 1$ times during each chase. We can use equation (12) to compute S_m, by proceeding as before but performing all arithmetic in the field $\mathbb{Z}_p$.

If we want a closed-form expression for S_m, we can't use equation (9) for $p > 2$; however, we can modify it to get the following closed-form formula for S_m:

$$S_m = \sum_{i=0}^{\lfloor m/2 \rfloor} (-1)^{n-i+1} \binom{m-i}{i} A^{m-2i}. \tag{14}$$

This formula accounts for the sign changes that occur during the recursive calculation of S_m, whereas equation (9) simply discards them since sign changes have no effect modulo 2. This is essentially the method shown in [**7**] involving Chebyshev polynomials.

Let's examine the 5×5 grid over $\mathbb{Z}_3$. (This game is playable on Tiger Electronics' Lights Out 2000, and on the web at many places including [**8**] and [**9**]). The matrix A

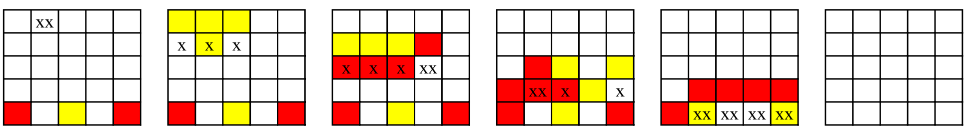
looks identical to the A from before, but the entries are now from $\mathbb{Z}_3$. Using equation (9) or equation (14) to compute S_5 over $\mathbb{Z}_3$ gives

$$S_5 = \begin{bmatrix} 2 & 1 & 1 & 1 & 2 \\ 1 & 0 & 2 & 0 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 0 & 2 & 0 & 1 \\ 2 & 1 & 1 & 1 & 2 \end{bmatrix}.$$

Example: Suppose that the bottom row is $\vec{s}_b = [1\,0\,2\,0\,1]$. We wish to add $-\vec{s}_5 = [2\,0\,1\,0\,2]$ to the bottom row, so we solve $S_5\vec{p}_1 = [2\,0\,1\,0\,2]$ for $\vec{p}_1$. We get the general solution

$$\vec{p}_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + c_5 \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{15}$$

where c_3, c_4, and c_5 are any scalars from $\mathbb{Z}_3$. If we choose them all to be zero, we get $\vec{p}_1 = [0\,2\,0\,0\,0]$. Letting red and yellow represent states 1 and 2, respectively, the game proceeds as follows (2 X's in a cell means the button is pressed twice):



Now let's create a lookup table for this game. The column space of S_5 consists of all solvable bottom-row configurations; since $\text{rank}(S_5) = 2$, there are nine such vectors. Each vector has the form $c_1\vec{v}_1 + c_2\vec{v}_2$, where $\vec{v}_1 = [1\,0\,2\,0\,1]$ and $\vec{v}_2 = [0\,1\,0\,1\,0]$ form a basis of the column space.

Using all possible values of c_1 and c_2 (9 pairs total), we get the nine solvable bottom-row configurations. Each is a value of $\vec{s}_b$ for which $S_m\vec{p}_1 = -\vec{s}_b$ has a solution. For each of these, we solve for $\vec{p}_1$, which is listed in the table as the Top Row vector to be pressed. We get the table shown below.

TABLE 2: Lookup table for 5×5 over $\mathbb{Z}_3$.

c_1	c_2	Bottom Row Configuration	Top Row Vector to Press
0	0	00000	00000
0	1	01010	22000
0	2	02020	11000
1	0	10201	02000
1	1	11211	21000
1	2	12221	10000
2	0	20102	01000
2	1	21112	20000
2	2	22122	12000

Now that we have the tools, we leave the reader with a few exercises. With the help of a computer algebra system, create lookup tables for other grid sizes and prime numbers of colors. You might first try these particular values, and check your tables by playing online at a site such as [**9**].

1. The 4×4 game over $\mathbb{Z}_2$.
2. The 4×4 game over $\mathbb{Z}_5$.
3. The 7×7 game over $\mathbb{Z}_3$.

REFERENCES

1. M. Anderson, T. Feil, Turning Lights Out with liner algebra, *Math. Mag* **71** no. 4 (1998) 300–303.
2. R. Barua, S. Ramakrishnan, σ-game, σ^+-game, and two-dimensional cellular automata, *Theoret. Comput. Sci.* **154** no. 2 (1996) 349–366.
3. J. Goldwasser, W. Klostermeyer, Maximization versions of "Lights Out" games in grids and graphs, *Congr. Numer.* **126** (1997) 99–111.
4. J. Goldwasser, W. Klostermeyer, G. Trapp, Characterizing switch-setting problems, *Linear Multilinear Algebra* **43** (1997) 121–135.
5. J. Goldwasser, W. Klostermeyer, G. Trapp, C.-Q. Zhang, *Setting Switches on a Grid, Technical Report TR-95-20*, Dept. of Statistics and Computer Science, West Virginia University, 1995.
6. J. Goldwasser, W. Klostermeyer, H. Ware, Fibonacci polynomials and parity domination in grid graphs, *Graphs Combin.* **18** (2002) 271–283.
7. M. Hunziker, A. Machiavelli, J. Park, Chebyshev polynomials over finite fields and reversibility of sigma-automata on square grids, *Theoret. Comput. Sci.* **320** (2004) 465–483.
8. S. Jaap, *Jaap's Puzzle Page*, http://www.jaapsch.net/puzzles/lights.htm.
9. N. Kato, *A Lights Out Puzzle with Solver*, http://www.ueda.info.waseda.ac.jp/~n-kato/lightsout/index.html.
10. J. Sjstrand, Note on the lamp lighting problem, *Adv. Appl. Math.* **27** (2001) 357–366.
11. L. Lovsz, *Combinatorial Problems and Exercises*, Chap. 5, problem 17c. North-Holland Publishing, Amsterdam, 1979.
12. SageMath, the Sage Mathematics Software System, The Sage Developers, http://www.sagemath.org.
13. K. Sutner, σ-automata and Chebyshev polynomials, *Theoret. Comput. Sci.* **230** (2000) 49–73. (Author claims this was written in 1996; see, for example, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.27.1593).
14. Wikipedia contributors, Lights Out, *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/wiki/Lights_Out_(game).

Summary. Lights Out is a game played on a grid of light-up buttons. At the start of the game, the lights are lit in a seemingly random configuration. A button changes states whenever it or one of its neighbors is pressed. The goal is to turn all of the lights out. Light chasing is a simple strategy for winning Lights Out by using a lookup table. In this article we examine how to construct these lookup tables for several variations of the game.

DAVID LEACH (MR Author ID: 684867) received his Ph.D. from Auburn University in 2002 and has been a member of the mathematics faculty at the University of West Georgia since then. His primary research areas are graph theory and combinatorics. He enjoys mixing mathematics with his interests in music, electronics, toys, and games, as evidenced by the present article on Lights Out. He and his wife Lori live in a house with several cats.

Proof Without Words: Diophantus of Alexandria's Sum of Squares Identity

ROGER B. NELSEN
Lewis & Clark College
Portland, OR
nelsen@lclark.edu

Theorem [**1**]. *If two positive integers are each sums of two squares, then their product is a sum of two squares in two different ways, i.e.*,

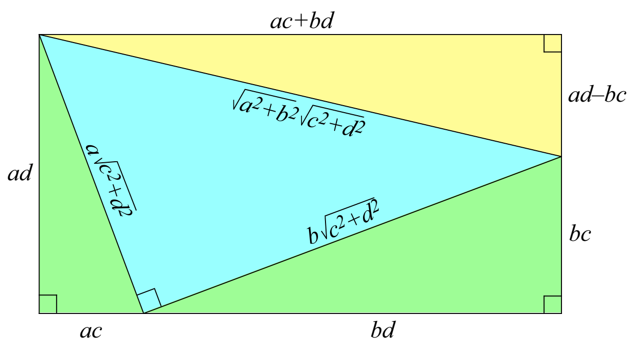
$$\left(a^2+b^2\right)\left(c^2+d^2\right)=(ac+bd)^2+(ad-bc)^2 \tag{1}$$

and

$$\left(a^2+b^2\right)\left(c^2+d^2\right)=(ad+bc)^2+(ac-bd)^2 \tag{2}$$

Example: $65=13\cdot 5=\left(3^2+2^2\right)\left(2^2+1^2\right)=8^2+1^2=7^2+4^2$.

Proof. (of (1) when $ad>bc$, the case $ad<bc$ is similar).



$$\left(\sqrt{a^2+b^2}\sqrt{c^2+d^2}\right)^2=(ac+bd)^2+(ad-bc)^2 \qquad \blacksquare$$

Exchanging c and d in the figure yields a proof of (2).

REFERENCE

1. T. L. Heath, *A History of Greek Mathematics* Vol. II, Oxford Univ. Press, 1921. 481–482.

Summary. We wordlessly prove $\left(a^2+b^2\right)\left(c^2+d^2\right)=(ac\pm bd)^2+(ad\mp bc)^2$ using the Pythagorean theorem.

ROGER NELSEN (MR Author ID 237909) is a professor emeritus at Lewis & Clark College, where he taught mathematics and statistics for 40 years.

Math. Mag. **90** (2017) 134. doi:10.4169/math.mag.90.2.134.
MSC: Primary 11E25

On Polyominoes and Digital Cameras

DANIEL J. HEATH
Pacific Lutheran University
Tacoma, WA 98447
heathdj@plu.edu

R. ROBERT RYDBERG, III
Schweitzer Engineering Labporatoris
Pullman, WA 99163
rob_rydberg@selinc.com

Here's an interesting (and mostly true) story. We were working on a brand-new CMOS (complimentary metal oxide semiconductor) imager, a rectangular array of photosensitive silicon sensors that convert light to electrons, for a semiconductor manufacturer. We used polyominoes to analyze trends and patterns of the sensors in the imager.

CMOS imagers are found in cellular phones, tablets, or most any device that records digital images. This new CMOS imager was larger and faster than any imager we'd made before. In particular, it was a 16 megapixel array that captured data up to 60 frames per second. We had a customer who wanted to use this imager in a high-speed digital camera and had just received our first prototypes. As part of acceptance testing, the customer told us they would allow for 16,000 single-pixel defects. The catch was that they would not allow for any defects involving two or more adjacent pixels.

This was a problem for us because we knew, from our manufacturing data, that the smaller imagers that we produce have multipixel defects. We knew it was impossible to manufacture a CMOS imager of the size they wanted that has up to 16,000 single-pixel defects and no dual-pixel defects. The challenge for us was that the client did not believe us. They insisted that if we could not manufacture such an imager, then they would find another manufacturer.

Our manager was concerned that we could lose a good client due to a technical misunderstanding. Our manager challenged us to prove theoretically that it was impossible to manufacture with 16,000 single-pixel defects and no multipixel defects. To add further stress, we had 36 hours to arrive at a theoretical solution since our manager was leaving in a few hours to board a plane to explain this technical misunderstanding to them in person. We had to solve the problem and prepare a few slides to be the core of their presentation to the customer.

We very quickly realized that this was a probability problem that involved counting polyomino configurations. We chose to take two approaches to it; if they matched, then we knew we had calculated correctly. Once we had the count correct for polyominoes representing single-, dual-, or triple-pixel defects, we could determine the probability and expected number of each kind of defect. In this article, we discuss only one of our methods in which we show essentially that if we expect x defective pixels in a single CMOS imager, then we should expect y multipixel defects. We intentionally do not discuss tolerances or the root cause of any particular defect since our objective is only to convince our client that multipixel defects will exist in a CMOS imager with their allowable number of single-pixel defects.

Math. Mag. **90** (2017) 135–141. doi:10.4169/math.mag.90.2.135.
MSC: Primary 05B99.

Mathematical introduction

The client wished to buy an imager, to use in a new digital video camera, that had 16 million pixels (in a 5000×3200 array). They were willing to accept up to 16,000 single-pixel defects, and no multipixel defects.

When we refer to a multipixel defect, we mean a defect in which two or more (cardinally or diagonally) adjacent pixels are defective (and by cardinally adjacent, we mean adjacent in the north-south-east-west sense). In the graph induced by the lattice structure of an array of pixels, including horizontal, vertical, and diagonal edges, a *k-pixel defect* is a connected k-vertex maximal subgraph of defective pixels. For example, each of the defective pixels (in black) in Figure 1 is part of a six-pixel defect. We note the similarity to *polyominoes*, about which much has been written. Essentially, a k-pixel defect is a generalization of a k-square polyomino in which pixels (squares) may be diagonally adjacent as well as cardinally adjacent.

```
o o o o o o
o • o o o o
o o • o • o
o o • • o o
o o o o • o
o o o o o o
```

Figure 1 A 6 pixel defect

In our analysis, we shall make the same assumptions as the engineering team. First, defects are uniformly distributed, and whether or not a given pixel is defective is not influenced by defects in other pixels, that is, the probability of a given pixel being defective is independent of that of any other pixel; see [**1**]. The chemical processes for silicon manufacturing are stochastic in nature, but dependent errors can be introduced by the mechanical processes. For example, a piece of dust in the clean room could cause dependent adjacent defects. In fact, mechanical sources of dependent defects are more likely to result in adjacent defects, so our assumption of independence should result in an underestimate when we use it to calculate the expected number of multipixel defects.

Second, we will assume that every pixel is an interior pixel, that is, adjacent to eight other pixels. This assumption is, of course, false. However, with a rectangular 5000×3200 array of pixels, there are $4998 \times 3198 = 15983604$ interior pixels and 16396 border pixels, so approximately one-tenth of one percent of the pixels are on the border. With such a small percentage we can effectively ignore the border without drastically affecting the results. Thus we imagine the camera as consisting of an infinite rectangular array, or alternately, as a rectangular array with opposite edges identified (i.e., a torus).

Our team believed that we could keep the number of defects down to about 5000, a considerable improvement on the client's allowance for 16000. We chose to keep this information to ourselves, however, since 5001 defects would be too many if 5000 or fewer were allowed, while it would be well within parameters if the original 16000 or fewer were allowed.

We shall use P to represent the maximum allowed probability that a given pixel is defective, and $Q = 1 - P$. With an allowance for a maximum of $\tau = 16000$ defects, $P = \frac{16000}{16000000} = \frac{1}{1000}$, and $Q = \frac{999}{1000}$. If the allowance is $\tau = 5000$ defects, we obtain $P = \frac{1}{3200}$, $Q = \frac{3199}{3200}$. In our calculations, we will simply use P and Q and substitute the appropriate values of P and Q at the end of the calculation. We look separately at each pixel, making strong use of independence. In effect, we are looking at each pixel in a set of adjacent pixels as an independent trial.

2	1	2
1	x	1
2	1	2

Figure 2 A 1 pixel defect

The probability that a given pixel lies on a one-pixel defect, that is, the pixel is defective but is surrounded by (eight) defect-free pixels, is $P \cdot Q^8$. This is pictured in Figure 2, where "x" represents the given (defective) pixel, and each number (either "1" or "2", which we shall refer to later) represents a defect-free pixel.

Using our allowance of $\tau = 5000$ defects, this probability is about 0.000311720. Multiplying this times the number of pixels should produce the expected number of one-pixel defects: 4987.51. Since the number of defects expected is 5000, but only 4987.51 are accounted for by counting the expected number of one-pixel defects, we see that the expected number of pixels that are part of multipixel defects is 12.49.

Here, the number of individual defects we are assuming is far below the client's allowed number, yet we still expect to find several multipixel defects. If we use the client's allowance of $\tau = 16000$, the situation is much worse: We can expect about 128 pixels in multipixel defects!

At this point, our team was already convinced that it would be impossible to manufacture the camera to the client's specifications. However, since we had some time remaining before our 36-hour deadline and to further back up our claims, we chose to count some specific types of multipixel defects.

Multiple defects

Figure 2 already begins the process of counting two-pixel defects. There are two possible configurations of two-pixel defects, each appears with four different orientations. The two configurations are: those in which the second defective pixel is cardinally adjacent (marked with a "1" in Figure 2) and those in which it is diagonally adjacent (marked with a "2").

In Figure 3, we have shown the two different types of two-pixel defects. The numbered pixels surrounding the two defective pixels (marked with "x" for the pixel of focus and "•" for the other) will be used later for determining the possible three-pixel defects.

2	1	2
3	•	3
4	x	4
5	6	5

A

	9	8	7
10	3	•	8
5	x	3	9
11	5	10	

B

Figure 3 Two different types of two-pixel defects

Each of the two types has four different orientations and two defective pixels. The first type is surrounded by 10 defect-free pixels and the second by 12, so the probability of any specific pixel being part of a two-pixel defect is

$$4 \cdot P^2 \cdot \left[Q^{10} + Q^{12}\right] \approx 7.78569 \times 10^{-7},$$

assuming $\tau = 5000$. Multiplying this times the number of pixels, we obtain the expected number of pixels in two-pixel defects: 12.46. (And using $\tau = 16000$ we find about 127 pixels in two-pixel defects.)

To count expected three-pixel defects is considerably more involved. We refer back to Figure 3A and 3B, where we have used "x" to mark the pixel of focus, "•" to mark the second defective pixel, and some number to mark the third. Since the third defective pixel must be adjacent to either or both of the first two, it must be in one of the numbered positions shown. There are 11 different configurations of three-pixel defects when we ignore orientation. Each possible choice of "x•3," for example, is the same configuration up to some action of D_4, the group of symmetries of the square. (This means that it might be rotated or reflected [or both], but otherwise is the same configuration.) Each configuration has distinct orbits when acted on by some subgroup of D_4 and thus can appear in one, two, four, or eight different orientations; see [**4**], [**7**].

For example, the arrangement shown in Figure 4 in the upper right-hand corner actually represents eight possible configurations. Each of the eight can be obtained by reflecting the shown arrangement about a horizontal mirror (two arrangements) and/or rotating by some multiple of 90° (four configurations).

Clearly, we must take into consideration a host of possible arrangements, as shown in Figure 4, each appearing in the array with the largest subgroup of D_4 that takes it to distinct orbits, with the numbers representing defect-free pixels (but which we shall use later to count four-pixel defects). Each of these arrangements can be obtained by overlaying two two-pixel defects on one another in some way. Further, we know that we have all possible configurations since we have listed all possible combinations in Figure 3.

Z_2	Z_4	D_4
2 3 4 3 2 1 • x • 1 2 3 4 3 2	5 6 7 8 9 1 x • • 10 5 6 7 8 9	11 12 13 14 2 x • 8 15 22 21 20 • 16 19 18 17
23 24 25 24 • 26 27 25 26 x 26 25 27 26 • 24 25 24 23	23 28 29 28 x 30 31 29 30 • 32 33 31 32 • 34 33 34 35	24 36 37 46 x 20 38 39 45 44 • • 40 43 42 32 41
	29 37 47 37 29 45 • 48 • 45 51 50 x 50 51 27 49 27	13 7 20 30 48 • • 44 52 x 53 54 26 3 50
	12 6 21 52 • 53 21 4 x • 6 55 4 52 12	11 26 55 49 5 • x 3 56 22 50 52 • 46 57 36 28
		58 43 31 51 54 • 38 59 56 x 7 • 60 25 57 47 62 61

Figure 4 Table of three-pixel defects

Thus, there are 11 different types of three-pixel defects up to orientation, or a total of 60 if we take different orientations as distinct types.

We distill Figure 4 into the table below by replacing the arrangements of pixels with the number of defect-free pixels surrounding the three defective pixels and replacing the subgroups with the number of distinct orbits:

2	4	8
12	12	14
16	16	14
–	15	12
–	12	14
–	–	15

Thus, the probability of any specific pixel being part of a three-pixel defect is

$$\begin{aligned} 2 \cdot P^3 \cdot Q^{12} + 4 \cdot P^3 \cdot Q^{12} + 8 \cdot P^3 \cdot Q^{14} &+ \\ 2 \cdot P^3 \cdot Q^{16} + 4 \cdot P^3 \cdot Q^{16} + 8 \cdot P^3 \cdot Q^{14} &+ \\ 4 \cdot P^3 \cdot Q^{15} + 8 \cdot P^3 \cdot Q^{12} &+ \\ 4 \cdot P^3 \cdot Q^{12} + 8 \cdot P^3 \cdot Q^{14} &+ \\ 8 \cdot P^3 \cdot Q^{15}&, \end{aligned}$$

or about 1.82317×10^{-9} if we use $\tau = 5000$. Multiplied times the number of pixels, we obtain 0.0292; this is the expected number of pixels in three-pixel defects if there are 5000 total defects. On the other hand, using $\tau = 16000$, we'd expect to find the number of pixels in three-pixel defects to be about 1.

In Figure 4 we have already begun the process of enumerating all possible four-pixel defects. Each such defect arises by placing one more defective pixel adjacent to some three-pixel defect and thus in one of the positions numbered 1 through 62. It is tedious but elementary to check that each four-pixel defect with the same number is equivalent, so, for example, each possible choice of "x••3" is the same configuration up to some action of D_4. Further, it is straightforward to check that each configuration has eight orientations except 1, 4, 10, 14, 23, 27, 35, 42, 47, 48, 49, 53, 55, and 58, each of which has only four orientations.

Thus, there are 62 types of four-pixel defects up to orientation, or 440 if we take different orientations to be distinct types.

The number of types of n-pixel defects known so far can be summed up in the following table. We use a_n to refer to the number of n-pixel defects with symmetries removed and b_n to refer to the total number of n-pixel defects.

n	1	2	3	4
a_n	1	2	11	62
b_n	1	8	60	440

Obviously, this process of enumerating each possible n-pixel defect by hand cannot continue indefinitely, leaving open questions. Can the sequences a_n and b_n be given in closed form? A search of [**5**] shows that $b_n = \{1, 8, 60, 440, \ldots\}$ is the "number of rooted two-dimensional polyominoes with n octagonal cells, with no symmetries removed" (sequence A094169), so we can expect the number of five-pixel defects, not counting orientation, to be 3230, meaning that the number with symmetries removed is at least 404. For six-pixel defects, we obtain 23688 (and thus at least 2961 with symmetries removed). We doubt anyone really wants to count those by hand.

The table below shows the expected number of pixels in 1-, 2-, 3- and (4+)-pixel defects, given the allowances of $\tau = 16000$ and $\tau = 5000$ individual defects.

	16,000	5,000
1	15872.447	4987.514
2	126.599	12.457
3	0.947	0.0292
4+	0.007	< 0.0001

Conclusion?

At this point our "story" is complete since the client accepted our explanation of why a camera matching their specifications would be impossible for us (or any manufacturer) to construct.

However, it is fascinating to note that we have not "solved" the problem in general. How many different arrangements of five-pixel defects are there? How about n-pixel defects? Asymptotically, this is equivalent to counting polyominoes with no symmetries removed, divided by 8, a problem with a large literature; see [**2**], [**3**]. Of course, as n increases, the probability of encountering the boundary likewise increases and nullifies our argument that we don't really need to count the boundary. What happens if we eliminate our time-saving assumption that our lattice is a rectangular torus?

What if the client had only been concerned with cardinal adjacency? This would have made our calculations much simpler since the number of "pointed polyominoes with n cells" is known to at least $n = 18$; see [**5**], sequence A126202. Thus, we know, for example, that there are 46 different five-pixel "cardinal" defects, not counting orientation (it is more difficult to show that there are 319 counting orientation).

What if the pixels were arranged in an isometric (versus rectangular) grid? Then each pixel would be surrounded by six (versus eight [adjacent] or four [cardinally adjacent]) others. There would be only one type of two-pixel defect, and it would have distinct orbits when acted on by $\mathbb{Z}_6$, a subgroup of D_6. Repeating the arguments in the last section, we can show that there are five types of three-pixel defects up to orientation (33 when considering orientation) and 18 types of four-pixel defects (182 when considering orientation). Using [**10**] to count possible pointed "pentahexes" up to orientation, we find 90 five-pixel defects up to orientation; we haven't bothered to count the distinct orientations. Hence, in this case, the associated integer sequences are $\{1, 1, 5, 18, 90, \ldots\}$ up to orientation and $\{1, 6, 33, 182, \ldots\}$ with orientations counted as distinct. The former of these doesn't appear in the Online Encyclopedia of Integer Sequences as of this writing. Although a sequence beginning with the same four numbers exists for the latter, there is no reason to believe that it is the same sequence other than the first four terms. Hence, we have no clue how to count polyhexes for large n, though [**9**] might provide some starting point.

Leaving the application behind and merely looking at the abstract problem is also fascinating. For example, what if our array is n-dimensional instead of merely two-dimensional? In n dimensions, we note that there are n different types of two-pixel defects, and the number of ways that each can appear is governed by a subgroup of symmetries of an n-dimensional (hyper-)cube.

REFERENCES

1. J. Fridrich, *Sensor Defects in Digital Image Forensic*. Springer, 2013.
2. S. Golomb, *Polyominoes*. Charles Scribners' Sons, 1965.

3. A. J. Guttman, *Polygons, Polyominoes and Polycubes,* Springer Lecture Notes in Physics 775.
4. G. Pólya, Kombinatorische Anzahlbestimmungen f ur Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68** no. 1 (1937).
5. *The Online Encyclopedia of Integer Sequences,* oeis.org.
6. D. H. Redelmeier, Counting polyominoes: Yet another attack, *Discrete Math.* **36** no. 2 (1981) 191–203.
7. J. H. Redfield, The Theory of Group-Reduced Distributions, *Amer. J. Math.* **49** no. 3 (1927) 433–455.
8. G. C. Rhoads, Planar tilings by polyominoes, polyhexes, and polyiamonds, *J. Comput. Appl. Math.* **174** no. 2 (2005) 329–353.
9. M. Voege, A. J. Guttmann, I. Jensen, On the number of benzenoid hydrocarbons, *J. Chem. Inf. Comp. Sci.* **42** no. 3 (2002) 456–466.
10. Wikipedia: *Pentahexes,* `http://en.wikipedia.org/wiki/Polyhex_(mathematics)`

Summary. We give an example of how mathematics is used in engineering by discussing whether a high-resolution CMOS imager, used in digital cameras, can be manufactured according to a client's specifications. We relate defects in a rectangular array of pixels to polyominos in order to count how many multipixel, or adjacent, defects are expected for a given number of single-pixel, or isolated, defects in a pixel array. This is done to quantify the expected number of multipixel defects in a given pixel array.

DANIEL J. HEATH (MR Author ID: 612809) is an associate professor and the chair of mathematics at Pacific Lutheran University. In his spare time, he plays jazz steelpan, practices paper engineering, and goes mushroom hunting with his son.

R. ROBERT RYDBERG, III (MR Author ID: 1185712) earned his Ph.D. in electrical and computer engineering at Washington State University and presently works as a software engineer at Schweitzer Engineering Laboratories, Inc. developing power systems control equipment. He spends some of his free time teaching Boy Scouts leadership, teamwork, and logistics while pretending it's about outdoor skills and high adventure.

SOLUTION TO PINEMI PUZZLE

<table>
<tr><td>||</td><td>||</td><td>||</td><td>8</td><td>|</td><td>5</td><td>||</td><td>4</td><td>3</td><td>|</td></tr>
<tr><td>|</td><td>16</td><td>||</td><td>|||</td><td>6</td><td>||</td><td>7</td><td>7</td><td>||</td><td>3</td></tr>
<tr><td>||</td><td>|||</td><td>||</td><td>12</td><td>11</td><td>7</td><td>8</td><td>|||</td><td>8</td><td>4</td></tr>
<tr><td>6</td><td>9</td><td>8</td><td>||</td><td>|||</td><td>|</td><td>|</td><td>|</td><td>|</td><td>|</td></tr>
<tr><td>|</td><td>4</td><td>|</td><td>10</td><td>||</td><td>|||</td><td>9</td><td>9</td><td>|||</td><td>8</td></tr>
<tr><td>4</td><td>||</td><td>7</td><td>||</td><td>8</td><td>10</td><td>|||</td><td>12</td><td>11</td><td>|||</td></tr>
<tr><td>|</td><td>7</td><td>||</td><td>6</td><td>|</td><td>9</td><td>|</td><td>|||</td><td>||</td><td>9</td></tr>
<tr><td>6</td><td>||</td><td>5</td><td>6</td><td>|</td><td>||</td><td>|</td><td>11</td><td>|</td><td>|||</td></tr>
<tr><td>|||</td><td>7</td><td>|</td><td>7</td><td>|</td><td>10</td><td>||</td><td>8</td><td>|</td><td>||</td></tr>
<tr><td>4</td><td>|</td><td>5</td><td>|||</td><td>|</td><td>||</td><td>7</td><td>|||</td><td>6</td><td>3</td></tr>
</table>

A Morley-Like Congruence Arising From Morley's Congruence

ROMEO MEŠTROVIĆ
University of Montenegro
Dobrota 36, 85330 Kotor, Montenegro
romeo@ac.me

The main result and its proof

As early as 1895, with the help of De Moivre's formula, Frank Morley [**4**] proved that for any prime $p \geq 5$,

$$\binom{p-1}{(p-1)/2} \equiv (-1)^{(p-1)/2} 4^{p-1} \pmod{p^3}. \tag{1}$$

Morley's proof of the congruence (1) made an ingenious use of integration of trigonometric sums. Subsequently, two alternate proofs that used the properties of Bernoulli numbers were given in 1913 by N. Nielsen [**5**, p. 353] and in 1938 by E. Lehmer [**3**, p. 360]. In 2012, C. Aebi and G. Cairns [**1**] gave an elementary proof of Morley's congruence (1). Notice also that in 1953 L. Carlitz [**2**, the congruence (3.1)] extended Morley's congruence (1) modulo p^4 for any prime $p \geq 5$.

In this note, we prove the following Morley-like congruence.

Theorem 1. *Let p be any odd prime. Then*

$$\binom{3(p-1)/2}{(p-1)/2} \equiv -\frac{8p(3p+1)}{4^p} \pmod{p^3}. \tag{2}$$

Proof. Notice that for any positive integer n we have

$$\binom{2n}{n} = \frac{\prod_{i=0}^{n-1}(2n-i)}{n!} = \frac{\prod_{i=1}^{n}((2n+1)-i)}{n!},$$

and

$$\binom{3n}{n} = \frac{\prod_{i=0}^{n-1}(3n-i)}{n!} = \frac{\prod_{i=1}^{n}(2n+i)}{n!} = \frac{2n+1}{3n+1} \cdot \frac{\prod_{i=1}^{n}((2n+1)+i)}{n!}.$$

Multiplying the above two equalities and taking $2n+1 = p$, i.e., $n = (p-1)/2$, we find that

Math. Mag. **90** (2017) 142–143. doi:10.4169/math.mag.90.2.142.
MSC: Primary 11A07, Secondary 11B65

$$\begin{aligned}\binom{p-1}{(p-1)/2}\binom{3(p-1)/2}{(p-1)/2} &= \frac{2p}{3p-1}\cdot\frac{\prod_{i=1}^{(p-1)/2}(p^2-i^2)}{\left(\left(\frac{p-1}{2}\right)!\right)^2}\\ &\equiv \frac{2p}{3p-1}\cdot\frac{\prod_{i=1}^{(p-1)/2}(-i^2)}{\left(\left(\frac{p-1}{2}\right)!\right)^2} \pmod{p^3}\\ &= \frac{2p}{3p-1}\cdot\frac{(-1)^{(p-1)/2}\left(\left(\frac{p-1}{2}\right)!\right)^2}{\left(\left(\frac{p-1}{2}\right)!\right)^2}\\ &= \frac{2(-1)^{(p-1)/2}p}{3p-1}\\ &\equiv -2(-1)^{(p-1)/2}p(3p+1) \pmod{p^3}.\end{aligned} \tag{3}$$

Substituting the congruence (1) into (3), we immediately obtain the congruence (2) for any prime $p \geq 5$. Finally, a direct calculation shows that (2) is also satisfied for $p = 3$.

REFERENCES

1. C. Aebi, G. Cairns, Morley's other miracle, *Math. Mag.* **85** no. 3 (2012) 205–217.
2. L. Carlitz, A theorem of Glaisher, *Canad. J. Math.* **5** (1953) 306–316.
3. E. Lehmer, On congruences involving Bernoulli numbers and the quotients of Fermat and Wilson, *Ann. of Math.* **39** (1938) 350–360.
4. F. Morley, Note on the congruence $2^{4n} \equiv (-1)^n(2n)!/(n!)^2$, where $2n+1$ is a prime, *Ann. of Math.* **9** (1895) 168–170.
5. N. Nielsen, Recherches sur les nombres de Bernoulli, *Danske Vidensk. Selsk. Skr.* (7) **10** (1913) 285–366.

Summary. In this note, we prove a congruence by applying the famous Morley congruence to a simple binomial identity.

ROMEO MEŠTROVIĆ (MR Author ID: 616301) received his B.S. from the Department of Mathematics, Faculty of Science at University of Zagreb and his M.S. and Ph.D. from Faculty of Natural Sciences and Mathematics at University of Montenegro, where he is currently a full professor of mathematics and applied mathematics at the Maritime Faculty Kotor. His diverse research interests include complex and functional analysis, combinatorial and elementary number theory, and applications of queuing theory in modeling port and terminals. He has more than 50 mostly self-published papers in peer-reviewed journals. During the last ten years, he was seven times the team leader of the Montenegro teams at Balkan Mathematical Olympiads and Junior Balkan Mathematical Olympiads.

PROBLEMS

Proposals

To be considered for publication, solutions should be received by September 1, 2017.

2016. *Proposed by Proposed by Rick Mabry, LSU Shreveport, Shreveport, LA.*

Let $\mathcal{P} = A_1A_2A_3A_4A_5A_6A_7A_8$ be a regular octagon. Let $A_0 = A_8$, $A_9 = A_1$, and $A_{10} = A_2$. For $j = 1, 2, \dots, 8$, let M_j be the midpoint of $\overline{A_jA_{j+1}}$, and let A'_j be the point of intersection of the segments $\overline{A_jM_{j+2}}$ and $\overline{A_{j-1}M_{j+1}}$. Prove that the inner octagon $\mathcal{P}' = A'_1A'_2A'_3A'_4A'_5A'_6A'_7A'_8$ has one-third the area of $\mathcal{P}$.

2017. *Proposed by Kim Sung Soo, Hanyang University, Ansan, Korea.*

Consider the following modification of the classical game of Nim. Initially, there are one or more heaps, each consisting of one or more stones. Players Alice and Bob take turns (Alice plays first). A player's move consists of choosing one or more heaps and removing exactly one stone from each of them. The player who takes the last stone loses. Determine all initial states for which Alice has a winning strategy.

2018. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

A *derivative function* is a function $f : \mathbb{R} \to \mathbb{R}$ such that f is the derivative of some differentiable function on $\mathbb{R}$.

(i) If f is a derivative function and $c \in \mathbb{R}$, show that the set $\{x \in \mathbb{R} \mid f(x) \geq c\}$ is a countable intersection of open subsets of $\mathbb{R}$.

(ii) Find a derivative function f and $c \in \mathbb{R}$ such that $\{x \in \mathbb{R} \mid f(x) \geq c\}$ is not a closed subset of $\mathbb{R}$.

Math. Mag. **90** (2017) 144–150. doi:10.4169/math.mag.90.2.144.

We invite readers to submit original problems appealing to students and teachers of advanced undergraduate mathematics. Proposals must always be accompanied by a solution and any relevant bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. Submitted problems should not be under consideration for publication elsewhere.

Proposals and solutions should be written in a style appropriate for this MAGAZINE.

Authors of proposals and solutions should send their contributions using the Magazine's submissions system hosted at http://mathematicsmagazine.submittable.com. *More detailed instructions are available there. We encourage submissions in PDF format, ideally accompanied by LaTeX source. General inquiries to the editors should be sent to* mathmagproblems@maa.org.

2019. *Proposed by Greg Oman, University of Colorado, Colorado Springs, CO.*

Recall that a (strict) linear ordering of a set X is any binary relation $\prec$ on X such that:

- $x \prec y$ and $y \prec z$ implies $x \prec z$ for all $x, y, z \in X$ (transitivity), and
- exactly one of the three possibilities $x \prec y$, $y \prec x$, $x = y$ holds for all $x, y \in X$ (trichotomy).

A strict linear ordering $\prec$ of an additive abelian group $(G, +)$ is said to be *translation-invariant* when, for all $x, y, z \in G$, if $x \prec y$, then $x + z \prec y + z$. Consider the group $(\mathbb{Z} \times \mathbb{Z}, +)$ of pairs of integers under the operation of coordinate-wise addition $(a, b) + (c, d) = (a + c, b + d)$. Prove or disprove: There exist infinitely many distinct translation-invariant linear orderings of $(\mathbb{Z} \times \mathbb{Z}, +)$.

2020. *Proposed by Julien Sorel, PNI, Piatra Neamt, Romania.*

Find all natural numbers n such that the integral $I_n := \int_0^1 x^n \arctan x \, dx$ is a rational number.

Quickies

Q1069. *Proposed by Proposed by George Stoica, Saint John, New Brunswick, Canada.*

Let $f : [0, \infty) \to [0, \infty)$ be a nonincreasing function such that $f(x) \to 0$ as $x \to \infty$. Show that there exists a continuous function function $g : [0, \infty) \to (0, \infty)$ such that

$$\int_0^\infty g(x)dx = \infty, \quad \text{but} \quad \int_0^\infty f(x)g(x)dx < \infty.$$

Q1070. *Proposed by Lokman Gökçe, Turkey.*

On the sides of the regular pentagon $ABCDE$, construct equilateral triangles $\triangle BCP$ and $\triangle DEQ$ so that the point lies P inside, and the point Q outside, of $ABCDE$. Find the ratio PQ/AB.

Solutions

Sums of powers of Pythagorean triples **February 2016**

1986. *Proposed by Spiros P. Andriopoulos, Third High School of Amaliada, Eleia, Greece.*

Let (a, b, c) be a Pythagorean triple, i.e., a, b, c are positive integers such that $c^2 = a^2 + b^2$, and let $n \geq 0$. Prove that

$$\frac{a^{2n+1} + b^{2n+1} + c^{2n+1}}{a + b + c}$$

is an integer.

Solution by Timothy Woodcock, Stonehill College, MA.
First we show that $a+b+c$ divides ab. Let $s=(a+b+c)/2$. Since squaring an integer preserves its parity and $c^2=a^2+b^2$, it follows that s is an integer. Furthermore,

$$ab=\frac{1}{2}[(a+b)^2-(a^2+b^2)]=\frac{1}{2}[(a+b)^2-c^2]=\frac{1}{2}(a+b+c)(a+b-c)=2s(s-c).$$

Hence, ab is a multiple of $2s=a+b+c$.

Now we prove that $t_n=a^{2n+1}+b^{2n+1}+c^{2n+1}$ is a multiple of $a+b+c$ by induction on n. The assertion trivially holds for $n=0$ since $t_0=a+b+c$. Inductively assume the property holds for some $n\geq 0$. We have

$$\begin{aligned} t_{n+1} &= a^{2n+3}+b^{2n+3}+c^{2n+3}=a^2a^{2n+1}+b^2b^{2n+1}+(a^2+b^2)c^{2n+1} \\ &= (a^2+b^2)t_n-ab(ab^{2n}+a^{2n}b). \end{aligned}$$

Since both ab and t_n (by the inductive hypothesis) are divisible by $a+b+c$, so is t_{n+1}, completing the inductive proof.

Also solved by Adnan Ali (India), Arkady Alt, Michel Bataille (France), Dionne Bailey & Elsie Campbell & Charles Diminnie, Brian Beasley, James Brawner, Paul Budney, Robert Calcaterra, Robin Chapman (UK), Hongwei Chen, L. Cíosoǵ, Ross Dempsey & Amit Gupta, Joseph DiMuro, Robert Doucette, Habib Y. Far, Dmitry Fleischman, Natacha Fontes-Merz, Michael Goldberg & Mark Kaplan, John Hawkins & David Stone, Nam Gu Heo (Korea), Eugene Herman, Brian Hogan, Tom Jager, Sojung Kang (Korea), Harris Kwong, Peter McPolin (UK), Jerry Minkus, Missouri State University Problem Solving Group, Omarjee Moubinool (France), Northwestern University Math Problem Solving Group, Stanley E. Payne, Ángel Plaza (Spain), Gary Raduns, Mohammad Riazi-Kermani, Joel Schlosberg, Allen Schwenk, Michael Vowe (Switzerland), Edward White, and the proposer. There was 1 incomplete or incorrect solution.

Matrices with integral entries and given determinant **February 2016**

1987. *Proposed by Valeriy Karachik and Leonid Menikhes, South Ural State University, Chelyabinsk, Russia.*

Let n, k, d be integers such that $n\geq 2$. Is there an $n\times n$ matrix $A=(a_{ij})$ with $\det A=d$ whose entries are integers $a_{ij}\geq k$?

Solution by L. Cíosoǵ, Fresno, CA.
Yes, such a matrix exists. If $n=2$, the matrix

$$\begin{pmatrix} d+k & d+k+1 \\ k & k+1 \end{pmatrix} \quad \text{if } d\geq 0,$$

$$\begin{pmatrix} k & k+1 \\ -d+k & -d+k+1 \end{pmatrix} \quad \text{if } d<0,$$

satisfies the requirements. Next, given any 2×2 matrix

$$B=\begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

satisfying the requirements of the problem for $n=2$, we construct an $n\times n$ matrix A

for each $n \geq 2$ as follows. Start with the $n \times n$ matrix

$$C = \begin{pmatrix} x & y & 0 & \cdots & 0 \\ z & w & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

whose lower right-hand block is the $(n-2) \times (n-2)$ identity matrix. Note that $\det C = \det B = d$. Adding the second column of C to the third and following columns, and subsequently adding the second row of the resulting matrix third and following rows, we obtain the matrix

$$A = \begin{pmatrix} x & y & y & \cdots & y \\ z & w & w & \cdots & w \\ z & w & w+1 & \cdots & w \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z & w & w & \cdots & w+1 \end{pmatrix}.$$

Since the row and column operations applied to C preserve its determinant, the matrix A clearly satisfies the requirements of the problem.

Also solved by Robert Calcaterra, Robin Chapman (UK), Joseph DiMuro, Natacha Fontes-Merz, Eugene Herman, Miguel Lerma, Mark McKinzie, Mohammad Riazi-Kermani, Joel Schlosberg, John H. Smith, and the proposer.

Integers that are divisor-sum composite **February 2016**

1988. *Proposed by Lenny Jones and Alicia Lamarche, Shippensburg University, Shippensburg, PA.*

Call a positive integer n *divisor-sum composite (DSC)* if the sum of two or more (distinct) divisors of n is always composite. Let $(a_1, a_2, \ldots, a_k)$ be a k-tuple of positive integers for some $k \geq 1$. Prove that there exist infinitely many DSC numbers n of the form

$$n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$$

for suitable distinct primes $p_1, p_2, \ldots, p_k$.

Solution by Brian D. Beasley, Presbyterian College, Clinton, SC.

Given $(a_1, a_2, \ldots, a_k)$, let $M = (a_1 + 1)(a_2 + 1) \ldots (a_k + 1)$ (the number of positive divisors of n). By Dirichlet's theorem, there are infinitely many primes q with $q \equiv 1 \pmod{M!}$. Choose any k such distinct primes $p_1, p_2, \ldots, p_k$, and let $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$. We show that every n of this form is divisor-sum composite.

Any positive divisor d of n factors as a product of (zero or more) factors p_1, p_2, $\ldots$, p_k, perhaps with repetitions; therefore, $d \equiv 1 \pmod{M!}$. It follows that any sum s of $l \geq 2$ distinct divisors $d_1, \ldots, d_l$ of n satisfies $s > l$ and $s \equiv 1 + \cdots + 1 = l \pmod{M!}$. Since n has M distinct positive divisors, we have $l \leq M$, so l divides $M!$. Therefore, $s \equiv l \equiv 0 \pmod{l}$, hence l divides s and s is composite. We conclude that n is divisor-sum composite.

Also solved by Adnan Ali (India), Michel Bataille (France), Robert Calcaterra, L. Cíosoǵ, Robin Chapman (UK), Robert Doucette, Dmitry Fleischman, Tom Jager, David Stone, and the proposer.

A binomial sum with harmonic weights **February 2016**

1989. *Proposed by Michel Bataille, Rouen, France.*

Let n be a positive integer and let a be a positive real number. Define

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \quad \text{and} \quad S_n(a) = \sum_{k=1}^{n} \binom{n}{k} a^k H_k.$$

Evaluate

$$\lim_{n\to\infty} \left(\frac{S_n(a)}{(a+1)^n} - \ln n \right).$$

Editor's Note. Prof. Donald Knuth pointed out that the sum in the statement of the problem is discussed in Theorem 1.2.7A of his book. (Knuth, Donald E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms,* Addison-Wesley. 1st ed., 1968; 3rd ed., 1997.)

Solution by Robin Chapman, University of Exeter, UK.
The value of the limit is

$$\gamma - \ln\left(\frac{a+1}{a}\right),$$

where $\gamma = \lim_{n\to\infty}(H_n - \ln n) = 0.577\ldots$ is the Euler–Mascheroni constant. To prove this, note that

$$H_n = \sum_{k=1}^{n} \int_0^1 x^{k-1} dx = \int_0^1 \sum_{k=1}^{n} x^{k-1} dx = \int_0^1 \frac{1-x^n}{1-x}\, dx \tag{1}$$

for all positive integers n, so we have

$$\begin{aligned} S_n(a) &= \int_0^1 \sum_{k=0}^{n} \binom{n}{k} a^k \cdot \frac{1-x^k}{1-x} dx = \int_0^1 \frac{(a+1)^n - (ax+1)^n}{1-x}\, dx \\ &= (a+1)^n \int_0^1 \left(1 - \left(\frac{ax+1}{a+1}\right)^n\right) \frac{dx}{1-x}, \end{aligned}$$

by the binomial theorem. Substituting $y = (ax+1)/(a+1)$, we have

$$\frac{S_n(a)}{(a+1)^n} = \int_{1/(a+1)}^{1} \frac{1-y^n}{1-y}\, dy. \tag{2}$$

Combining equations (1) and (2) we have

$$H_n - \frac{S_n(a)}{(a+1)^n} = \int_0^{1/(a+1)} \frac{1-y^n}{1-y}\, dy \to \int_0^{1/(a+1)} \frac{dy}{1-y} = \ln\left(\frac{a+1}{a}\right) \quad \text{as } n\to\infty,$$

since $y^n \to 0$ uniformly on $[0, 1/(a+1)]$ (because $a > 0$). Therefore,

$$\frac{S_n(a)}{(a+1)^n} - \ln n = (H_n - \ln n) - \left(H_n - \frac{S_n(a)}{(a+1)^n}\right) \to \gamma - \ln\left(\frac{a+1}{a}\right) \quad \text{as } n\to\infty.$$

Also solved by Robert A. Agnew, Khristo Boyadzhiev, Robert Calcaterra, Hongwei Chen, Eugene A. Herman, Peter McPolin (UK), Omarjee Moubinool (France), and the proposer. There was one incomplete or incorrect solution.

A rational quadratic inequality in three variables February 2016

1990. *Proposed by Nermin Hodžić (student), University of Tuzla, Tuzla, Bosnia and Herzegovina and Salem Malikić (student), Simon Fraser University, Burnaby, BC, Canada.*

Let a, b, c be nonnegative real numbers such that

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} = 2. \tag{3}$$

Prove that

$$\frac{1}{2} \leq \frac{ab+bc+ca}{a^2+b^2+c^2} \leq \frac{7}{11}. \tag{4}$$

Solution by Kee-Wai Lau, Hong Kong, China.
Let $s_1 = a+b+c$, $s_2 = ab+bc+ca$, $s_3 = abc$. We have

$$F(s_1, s_2, s_3) = s_1^2 s_2^2 - 4s_1^3 s_3 + 18 s_1 s_2 s_3 - 4s_2^3 - 27 s_3^2 = (a-b)^2(b-c)^2(c-a)^2 \geq 0. \tag{*}$$

It follows from equation (3) that at most one of the nonnegative numbers a, b, c is zero; in particular, $s_1 = a+b+c > 0$. Since equation (3) and inequality (4) are both homogeneous in a, b, c, we assume $s_1 = 1$ without loss of generality. Equation (3) simplifies to $s_3 = (4s_2 - 1)/5$. Regard $F(s_1, s_2, s_3) = F(1, s_2, (4s_2-1)/5)$ as a function $f(s_2)$ of the variable s_2 only. Straightforward algebra from equation (*) gives

$$0 \leq f(s_2) = -\frac{1}{25}(s_2+1)(4s_2-1)(25s_2-7),$$

hence $1/4 \leq s_2 \leq 7/25$, inasmuch as s_2 is nonnegative. Thus, we have $1/2 \leq s_2/(1-2s_2) \leq 7/11$. This proves inequality (4) since $a^2+b^2+c^2 = (a+b+c)^2 - 2(ab+bc+ca) = 1 - 2s_2$.

Also solved by Michel Bataille (France), James Brawner, Hongwei Chen, John Christopher, Eugene Herman, Michael Vowe (Switzerland), and the proposer. There were 5 incomplete or incorrect solutions.

Answers

Solutions to the Quickies from page 145.

A1069. Let $h : [0, \infty) \to (0, \infty)$ be any strictly decreasing differentiable function such that $h(x) \geq f(x)$ and $h(x) \to 0$ as $x \to \infty$. We claim that $g = -h'/h$ satisfies the requirements. Since h is strictly decreasing and positive, $g(x)$ is positive on $[0, \infty)$. We have

$$\int_0^\infty g(x)dx = -\int_0^\infty \frac{h'(x)}{h(x)}dx = \lim_{x\to\infty} [\ln(h(0)) - \ln h(x)] = \infty$$

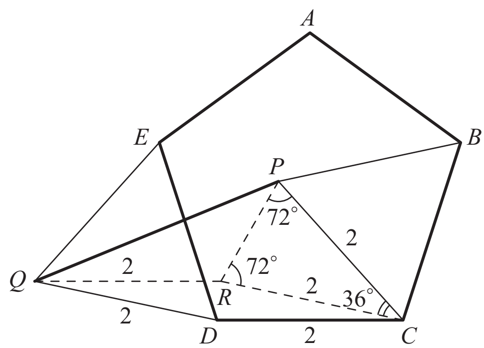
since $h(x) \to 0$ as $x \to \infty$, whereas

$$\int_0^\infty f(x)g(x)dx \le \int_0^\infty h(x)g(x)dx = -\int_0^\infty h'(x)dx$$
$$= \lim_{x\to\infty} [h(0) - h(x)] = h(0) < \infty.$$

A1070. Without loss of generality, we may assume that the pentagon and equilateral triangles all have side length 2. Rotate P by 36° about C (away from B) to obtain a point R (see figure below). Clearly $\angle RCD = 108^\circ - 60^\circ - 36^\circ = 12^\circ$ and $\angle QDC = 108^\circ + 60^\circ = 168^\circ$ are supplementary angles, so $CDQR$ must be a side-2 rhombus. In the isosceles triangle $\triangle PCR$, we have $PR = 2\,CR \sin 18^\circ = \sqrt{5} - 1$. We also have $\angle PRQ = 360^\circ - (168^\circ + 72^\circ) = 120^\circ$, so the law of cosines in triangle $\triangle QPR$ gives

$$PQ^2 = 2^2 + (\sqrt{5} - 1)^2 - 2 \cdot 2 \cdot (\sqrt{5} - 1) \cdot (-\tfrac{1}{2}) = 8.$$

Therefore, $PQ/AB = \sqrt{8}/2 = \sqrt{2}$.



When a Sum of Powers Equals a Power

In Quickies 1065 [1], a solution in positive integers for the equation $a^2 + b^7 + c^{13} + d^{14} = e^{15}$ is given. We here generalize this result.

Theorem. *Let $r_1, \ldots, r_n, s$ be positive integers such that each r_i is coprime to s. Then the equation $x_1^{r_1} + \cdots + x_n^{r_n} = y^s$ has a solution in positive $x_1, \ldots, x_n, y$.*

Proof. Since $t := \mathrm{lcm}\{r_1, \ldots, r_n\}$ is coprime to s, there exist positive integers a and b such that $at + 1 = bs$ by Bézout's identity, hence

$$(n^{at/r_1})^{r_1} + \cdots + (n^{at/r_n})^{r_n} = n \cdot n^{at} = (n^b)^s. \quad \blacksquare$$

REFERENCE

1. A. Chu, Quickies 1065, *Math. Mag.*, **89** (2016) 379, 385; http://www.jstor.org/stable/10.4169/math.mag.89.5.385.

—Submitted by Adrian Chun Pong Chu
Chinese University of Hong Kong

http://dx.doi.org/10.4169/math.mag.90.2.150
MSC: Primary 11D41

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles, books, and other materials 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.

O'Neil, Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishers, 2016; x+260 pp, $26. ISBN 978-0-553-41881-1

________, Weapons of math destruction, *Discover* (October 2016) 50–55.

________, Weapons of math destruction, 12:15 min. video, https://www.youtube.com/watch?v=gdCJYsKlX_Y.

The computer science major at my institution, in the need to offer a full roster of "content" courses, no longer has room for a course about social and ethical dimensions of computing. Cathy O'Neil points out the ethicsless "dark side" of the much-touted "Big Data" world that our graduates will function in: Humans implement judgments made by algorithms generated from mathematical models based on black-box "machine learning" and enormous amounts of data; because of the authority of mathematics, the certitude of computers, and the myriad collated details collected about individuals, the verdicts of so-called "evidence-based" algorithms are often beyond dispute or appeal. For examples, consider computer-based systems used to assess credit worthiness, what books to discard from libraries, the risk of recidivism of prisoners up for parole, and evaluation of teachers by "added value," plus the now-ubiquitous automatic résumé readers for job applicants. (Do you realize that your credit score can matter more than your driving record in determining your auto insurance premium?) And of course Big Data, in the form of "microtargeting," can be used to determine how to change the minds of individualized pockets of voters. The ill effects of judgments based on these systems include built-in prejudices, including racism and discrimination against the poor. Such "collateral damage" is brought about "because data scientists ... don't dwell on those errors. Their feedback is money, which is also their incentive." I would like to think that mathematics and mathematicians should not have to bear the brunt of the blame—why not the data scientists? or the managers in charge? or (ultimately) lack of consideration and human compassion? But just as it was our former students who in the early 2000s helped bring about the Great Recession, it is likely to be our future former students who "improve," certify the efficiency of, and anoint the algorithms in question.

Beineke, Jennifer, and Jason Rosenhouse (eds.), *The Mathematics of Various Entertaining Subjects: Research in Recreational Math*, Princeton University Press, 2016; xv+272 pp, $75. ISBN 978-0-691-16403-8.

The chapters in this book represent some of the work presented at the MOVES Conference in 2013 (MOVES being an acronym for Mathematics of Various Entertaining Subjects). There are authoritative essays on one-move puzzles (in squash racquets, when should you "set two" under English scoring?), solving the Tower of Hanoi with random moves, detecting false coins, Heartless poker, tic-tac-toe on affine planes, and a dozen others. Abstracts of talks at the 2015 conference are at http://momath.org/wp-content/uploads/2015/07/MOVES_Abstracts-2015-version721.pdf. The 2017 conference will be held Aug. 6–8 at the National Museum of Mathematics in New York.

Math. Mag. **90** (2017) 151–152. doi:10.4169/math.mag.90.2.151.

Fernandez, Oscar, *The Calculus of Happiness: How a Mathematical Approach to Life Adds Up to Health, Wealth, and Love*, Princeton University Press, 2017; xiii+155 pp, $24.95. ISBN 978-0-691-16863-0.

This book suggests that we can use mathematics to improve our lives in the dimensions of health, wealth, and love. "Research-backed" equations are given to determine how many calories you should eat, how to improve cholesterol and lose weight, and how many years of life you forego at a specific level of waist-to-height ratio. Also touted are "math-backed" strategies to increase take-home pay, beat Wall Street, guarantee no cheating among couples, and make joint decisions. The mathematics developed and applied are linear functions (resting metabolic rate), polynomials (years of life lost), piecewise-linear functions (tax rates), logarithms (time to pay off a loan), standard deviations (investment risk), stable matchings (partners), and dynamical systems (relationships). Each chapter ends with "Mathematical Takeaways," "Nonmathematical Takeaways," and "A Few Practical Tips," and the calculational details are postponed to an appendix. Much as I am enthused about popularizing the practicality of mathematics, such as in Jordan Ellenberg's *How Not to Be Wrong: The Power of Mathematical Thinking* (2014), I am very ill at ease with a mathematician prescribing diet, investment, and dating on the basis of the "authority" of mathematics. In the preface, the author suggests consulting relevant experts before changing your life based on the book—advice that I would prefer to see repeated in boldface in every collection of "practical tips," together with further disclaimers.

Devlin, Keith, *Finding Fibonacci: The Quest to Rediscover the Forgotten Mathematical Genius Who Changed the World*, Princeton University Press, 2017; vi+250 pp, $29.95. ISBN 978-0-691-17486-0.

This volume recounts the incidents leading to author Devlin's *The Man of Numbers: Fibonacci's Arithmetic Revolution* (2011). Like that book, this one celebrates Fibonacci's influence on commerce of his day, particularly through a "textbook" in Italian that could have been written only by him. Devlin admires Fibonacci's role as expositor of not only Hindu–Arabic numerals and their calculus but also problems that relate to modern financial instruments.

Elwes, Richard, The top 10 mathematical achievements of the last 5ish years, maybe, https://richardelwes.co.uk/2015/06/18/the-top-10-mathematical-achievements-of-the-last-5ish-years-maybe/.

How would you answer a nonmathematical colleague who asked you what is new in mathematics? This brief webpage, from early 2016, cites recent mathematical achievements, such as the proof of the weak Goldbach conjecture, progress on bounded gaps between primes, and a finite formula for partition numbers, together with links to sources. And what about in 2016? Solution to the cap set problem, the densest packings in dimensions 8 to 24, a bounded envy-free cake-cutting algorithm, ...

Singmaster, David, The utility of recreational mathematics, *The UMAP Journal of Undergraduate Mathematics and Its Applications* 37 (4) (2016) 339–380.

For many years, David Singmaster has researched the history of recreational mathematics in all its manifestations, much of it collected in his *Sources in Recreational Mathematics: An Annotated Bibliography* (8th preliminary edition (2004), http://www.puzzlemuseum.com/singma/singma-index.htm). This article, based on talks over the years, delves into the nature of recreational mathematics and why it is useful: It can lead to serious mathematics, it has turned up ideas of genuine utility, it has great pedagogic utility, and it is very useful to the historian of mathematics. Singmaster offers numerous examples, from games to patents, from Sanskrit poetry to chain codes, from Rubik's cubes to Penrose pieces, all illustrated with full-color photographs and reproductions of ancient sources. (Full disclosure: I am the editor of the publication in which this article appeared, for which I solicited the article.)

NEWS AND LETTERS

77th Annual William Lowell Putnam Mathematical Competition

Editor's Note: Additional solutions will be printed in the *Monthly* later in the year.

PROBLEMS

A1. Find the smallest positive integer j such that for every polynomial $p(x)$ with integer coefficients and for every integer k, the integer

$$p^{(j)}(k) = \left. \frac{d^j}{dx^j} p(x) \right|_{x=k}$$

(the jth derivative of $p(x)$ at k) is divisible by 2016.

A2. Given a positive integer n, let $M(n)$ be the largest integer m such that

$$\binom{m}{n-1} > \binom{m-1}{n}.$$

Evaluate

$$\lim_{n\to\infty} \frac{M(n)}{n}.$$

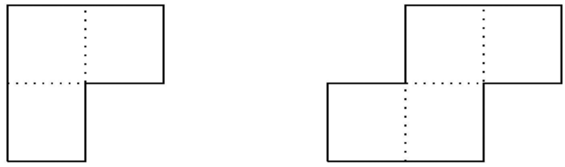
A3. Suppose that f is a function from $\mathbb{R}$ to $\mathbb{R}$ such that

$$f(x) + f\left(1 - \frac{1}{x}\right) = \arctan x$$

for all real $x \neq 0$. (As usual, $y = \arctan x$ means $-\pi/2 < y < \pi/2$ and $\tan y = x$.) Find

$$\int_0^1 f(x)\, dx.$$

A4. Consider a $(2m-1) \times (2n-1)$ rectangular region, where m and n are integers such that $m, n \geq 4$. This region is to be tiled using tiles of the two types shown:



(The dotted lines divide the tiles into 1×1 squares.) The tiles may be rotated and reflected, as long as their sides are parallel to the sides of the rectangular region. They must all fit within the region, and they must cover it completely without overlapping.

Math. Mag. **90** (2017) 153–164. doi:10.4169/math.mag.90.2.153.

What is the minimum number of tiles required to tile the region?

A5. Suppose that G is a finite group generated by the two elements g and h, where the order of g is odd. Show that every element of G can be written in the form

$$g^{m_1}h^{n_1}g^{m_2}h^{n_2}\cdots g^{m_r}h^{n_r}$$

with $1 \le r \le |G|$ and $m_1, n_1, m_2, n_2, \ldots, m_r, n_r \in \{1, -1\}$. (Here $|G|$ is the number of elements of G.)

A6. Find the smallest constant C such that for every real polynomial $P(x)$ of degree 3 that has a root in the interval [0, 1],

$$\int_0^1 |P(x)|\,dx \le C \max_{x\in[0,1]} |P(x)|.$$

B1. Let $x_0, x_1, x_2, \ldots$ be the sequence such that $x_0 = 1$ and for $n \ge 0$,

$$x_{n+1} = \ln(e^{x_n} - x_n)$$

(as usual, the function ln is the natural logarithm). Show that the infinite series

$$x_0 + x_1 + x_2 + \cdots$$

converges and find its sum.

B2. Define a positive integer n to be *squarish* if either n is itself a perfect square or the distance from n to the nearest perfect square is a perfect square. For example, 2016 is squarish, because the nearest perfect square to 2016 is $45^2 = 2025$ and $2025 - 2016 = 9$ is a perfect square. (Of the positive integers between 1 and 10, only 6 and 7 are not squarish.)

For a positive integer N, let $S(N)$ be the number of squarish integers between 1 and N, inclusive. Find positive constants α and β such that

$$\lim_{N\to\infty} \frac{S(N)}{N^\alpha} = \beta,$$

or show that no such constants exist.

B3. Suppose that S is a finite set of points in the plane such that the area of triangle $\triangle ABC$ is at most 1 whenever A, B, and C are in S. Show that there exists a triangle of area 4 that (together with its interior) covers the set S.

B4. Let A be a $2n \times 2n$ matrix, with entries chosen independently at random. Every entry is chosen to be 0 or 1, each with probability $1/2$. Find the expected value of $\det(A - A^t)$ (as a function of n), where A^t is the transpose of A.

B5. Find all functions f from the interval $(1, \infty)$ to $(1, \infty)$ with the following property:

if $x, y \in (1, \infty)$ and $x^2 \le y \le x^3$, then $f(x)^2 \le f(y) \le f(x)^3$.

B6. Evaluate

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{n=0}^{\infty} \frac{1}{k\,2^n + 1}.$$

SOLUTIONS

Solution to A1. (Based on a student paper.) We must have $j > 7$, because for $p(x) = x^7$ and $j \le 7$, $p^{(j)}(1) = 5040/(7-j)!$ is not divisible by 2016. For $j = 8$

and any polynomial $p(x)$ with integer coefficients, every term of $p^{(8)}(x)$ has a coefficient that is divisible by the product of eight consecutive integers. [Specifically, the coefficient of x^k in $p^{(8)}(x)$ is divisible by $(k+8)(k+7)\cdots(k+1)$.] Of any eight consecutive integers, at least one is divisible by 7, at least two are divisible by 3, and two are divisible by 4, of which one is divisible by 8. Thus their product is divisible by $7 \cdot 3 \cdot 3 \cdot 4 \cdot 8 = 2016$. It follows that for $j = 8$ and any polynomial $p(x)$ with integer coefficients, $p^{(j)}(k)$ is divisible by 2016 for every integer k, and so $j = 8$ is the smallest positive integer with the desired property.

Solution to A2. We first rewrite the condition $\binom{m}{n-1} > \binom{m-1}{n}$ as

$$\begin{aligned} 0 &< \binom{m}{n-1} - \binom{m-1}{n} \\ &= \frac{m!}{(n-1)!(m-n+1)!} - \frac{(m-1)!}{n!(m-n-1)!} \\ &= \frac{(m-1)!}{n!(m-n+1)!}\,[mn - (m-n+1)(m-n)], \end{aligned}$$

which shows that it is equivalent to

$$mn - (m-n+1)(m-n) > 0, \quad \text{that is,} \quad m^2 - (3n-1)m + (n^2-n) < 0.$$

This condition is satisfied if and only if $r_1 < m < r_2$, where

$$r_{1,2} = \frac{3n - 1 \pm \sqrt{(3n-1)^2 - 4(n^2-n)}}{2} = \frac{3n-1 \pm \sqrt{5n^2 - 2n + 1}}{2}$$

are the roots of the quadratic equation $x^2 - (3n-1)x + (n^2 - n) = 0$. For sufficiently large n, there is certainly an integer m strictly between r_1 and r_2, and so the integer $M(n)$ is then given by

$$M(n) = \left\lceil \frac{3n - 1 + \sqrt{5n^2 - 2n + 1}}{2} \right\rceil - 1,$$

where $\lceil a \rceil$ indicates the least integer $\geq a$. We then have

$$M(n) = \frac{3n - 1 + \sqrt{5n^2 - 2n + 1}}{2} - \varepsilon(n)$$

with $0 < \varepsilon(n) \leq 1$, and it follows that

$$\begin{aligned} \lim_{n\to\infty} \frac{M(n)}{n} &= \lim_{n\to\infty} \frac{3n - 1 + \sqrt{5n^2 - 2n + 1}}{2n} - \lim_{n\to\infty} \frac{\varepsilon(n)}{n} \\ &= \lim_{n\to\infty} \frac{3 - \frac{1}{n} + \sqrt{5 - \frac{2}{n} + \frac{1}{n^2}}}{2} - 0 \\ &= \frac{3 + \sqrt{5}}{2}. \end{aligned}$$

Solution to A3. Let $\phi(x) = 1 - \frac{1}{x}$. Note that for $x \neq 0, 1$,

$$\phi(\phi(x)) = \frac{1}{1-x} \quad \text{and} \quad \phi(\phi(\phi(x))) = x.$$

Therefore, from the given equation $f(x) + f(\phi(x)) = \arctan x$, if we repeatedly replace x by $\phi(x)$ we get $f(\phi(x)) + f(\phi(\phi(x))) = \arctan \phi(x)$ and then $f(\phi(\phi(x))) + f(x) = \arctan \phi(\phi(x))$. Thus, for $x \neq 0, 1$, we have

$$\begin{aligned} f(x) &= \frac{1}{2}\big(\arctan x - \arctan \phi(x) + \arctan \phi(\phi(x))\big) \\ &= \frac{1}{2}\big(\arctan x + \arctan(-\phi(x)) + \arctan \phi(\phi(x))\big). \end{aligned}$$

This function is continuous on the open interval (0, 1), and from standard properties of arctan, it has finite limit $\frac{3\pi}{8}$ as $x \to 0^+$ and as $x \to 1^-$; thus, we may assume that f is continuous on the entire interval [0, 1]. Meanwhile, we can replace x by $1 - x$ in the last equation to get

$$\begin{aligned} f(1-x) &= \frac{1}{2}\big(\arctan(1-x) + \arctan(-\phi(1-x)) + \arctan \phi(\phi(1-x))\big) \\ &= \frac{1}{2}\left(\arctan(1-x) + \arctan(-\phi(1-x)) + \arctan \frac{1}{x}\right). \end{aligned}$$

Now note that $\phi(x)\phi(1-x) = 1$, and so if we add the expressions for $f(x)$ and $f(1-x)$ and rearrange, we get an expression involving three sums of arctangents of reciprocals:

$$f(x) + f(1-x) = \frac{1}{2}\left(\arctan x + \arctan \frac{1}{x} + \arctan a + \arctan \frac{1}{a} + \arctan b + \arctan \frac{1}{b}\right),$$

where $a = -\phi(x)$ and $b = 1 - x$. For x in the open interval (0, 1), the numbers x, a, b are all positive, so each pair of arctangents of reciprocals sums to $\frac{\pi}{2}$. Therefore,

$$f(x) + f(1-x) = \frac{1}{2}\left(\frac{\pi}{2} + \frac{\pi}{2} + \frac{\pi}{2}\right) = \frac{3\pi}{4}.$$

Finally,

$$\int_0^1 f(x)\,dx = \int_0^1 f(1-x)\,dx$$

by the substitution $u = 1 - x$, so

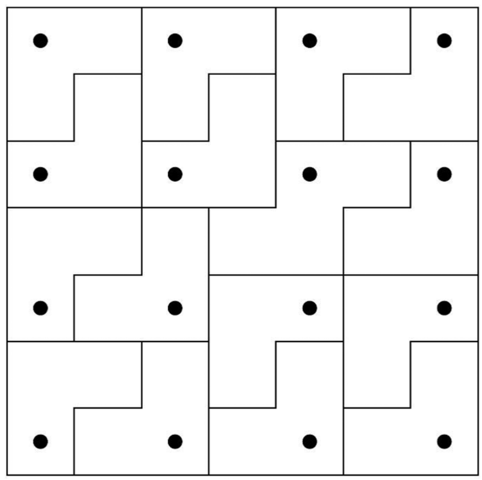
$$\begin{aligned} \int_0^1 f(x)\,dx &= \frac{1}{2}\left(\int_0^1 f(x)\,dx + \int_0^1 f(1-x)\,dx\right) \\ &= \frac{1}{2}\int_0^1 (f(x) + f(1-x))\,dx = \frac{1}{2}\int_0^1 \frac{3\pi}{4}\,dx = \frac{3\pi}{8}. \end{aligned}$$

Comment. Although $\phi(x)$ and $\phi(1-x)$ are reciprocals, they are negative and thus their arctangents sum to $-\frac{\pi}{2}$.

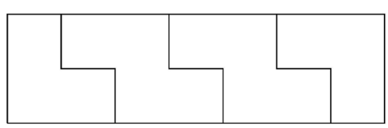
Solution to A4. The region can be tiled with mn tiles, but not with fewer. To show this, start by dividing the region into 1×1 squares; there are $(2m-1)(2n-1)$

such squares, and they can be labeled in a natural way by ordered pairs of integers (i, j) with $1 \le i \le 2m - 1$, $1 \le j \le 2n - 1$. Consider the squares with labels (k, l) for which k and l are both odd. (In the figure below showing an example for $m = n = 4$, these squares are indicated by dots.) There are mn such "dotted" squares, and any of the given tiles can cover at most one of them, so at least mn tiles are necessary.

To show that mn tiles are sufficient, we give an inductive construction of a tiling with mn tiles. For the base case, we have $m = n = 4$, so we need to tile a 7×7 region with 16 tiles. Because each "L"-shaped tile has area 3 and each "N"-shaped tile has area 4, we must use 15 "L"-shaped tiles and just one "N"-shaped tile. Here is an example of such a tiling:



In order to tile a $7 \times (2n - 1)$ region for any $n > 4$, we can start with the tiling above and attach $n - 4$ tilings of vertical 7×2 strips to the right edge of the tiled 7×7 square region. Each such strip can be tiled with four tiles as shown (rotated through 90 degrees to save space) below, for a total of $16 + 4(n - 4) = 4n$ tiles, as required.



Finally, to tile a $(2m - 1) \times (2n - 1)$ rectangular region for any $m > 4$, we can start with a tiling of a $7 \times (2n - 1)$ region as described above and attach $m - 4$ tilings of horizontal $2 \times (2n - 1)$ strips to the bottom edge of that region. Each such strip can be tiled with n tiles in a similar way to the 2×7 rectangle shown above (using $n - 2$ "N"-shaped tiles instead of two, along with two "L"-shaped tiles at the end of the strip), for a total of $4n + (m - 4)n = mn$ tiles, as required.

Solution to A5. (Based on a student paper.) Let $x = gh$, $y = g^{-1}h$, and consider the subgroup $H = \langle x, y \rangle$ of G generated by x and y. We have $xy^{-1} = g^2 \in H$; because g has odd order, we have $g^{2k-1} = e$ for some positive integer k, so $g = g^{2k} = (g^2)^k \in H$ and $h = g^{-1}x \in H$, so $H = G$. Also, because G is finite, the inverses of x and y can be written as powers of x and y with positive exponents. It follows that every element of G can be written in the form $t_1 t_2 \cdots t_s$ where each t_i is either x or y. For a given element $z \in G$, consider such an expression $z = t_1 t_2 \cdots t_r$ for which $r > 0$ is as small as possible; we will show that $r \le |G|$, which will prove the desired statement (in fact, a somewhat stronger one, because while the exponents m_i from the problem statement can be either 1 or -1, the exponents n_i will all be 1). Suppose that, instead, $r > |G|$, and consider the "partial products"

$$t_1, t_1 t_2, t_1 t_2 t_3, \ldots, t_1 t_2 \cdots t_r = z.$$

Because these are all elements of G, by the Pigeonhole Principle at least two of them would be equal. That is, for some $1 \leq i < j \leq r$ we would have

$$t_1 t_2 \cdots t_i = t_1 t_2 \cdots t_j.$$

But this could be used to find the shorter expression

$$z = t_1 t_2 \cdots t_i \, t_{j+1} t_{j+2} \cdots t_r$$

for z, contradicting the minimality of r.

Solution to A6. We will show that $C = \frac{5}{6}$. Let $M = \max_{x \in [0,1]} |P(x)|$. Because $P(x)$ has degree 3, we have $M \neq 0$, and we can replace $P(x)$ by $P(x)/M$ without affecting the problem, so we may (and will) assume that $M = 1$.

We first consider the case that $P(0) = 0$ and that this is the only root of $P(x)$ in the interval [0, 1]. Without loss of generality, we may then assume that $P(x) > 0$ for all $x \in (0, 1]$. Because $M = 1$, there must be at least one $\alpha \in (0, 1]$ with $P(\alpha) = 1$; on the other hand, if there is such an α with $\alpha < 1$, then $P(x)$ has a local maximum there. Because $P(x)$ has degree 3, it can have only one local maximum, so there are three subcases, as follows.

Subcase 1: $P(0) = 0$, $0 < P(x) < 1$ for all $x \in (0, 1]$ except $x = \alpha$ and $x = 1$, $P(\alpha) = P(1) = 1$, for some $\alpha \in (0, 1)$. In this subcase, $P(x) - 1$ has a double root at α and a root at 1, and using $P(0) = 0$, we see that

$$P(x) - 1 = (x - 1)\left(\frac{x}{\alpha} - 1\right)^2, \quad \text{so} \quad P(x) = \frac{1}{\alpha^2}(x^3 - x^2) - \frac{2}{\alpha}(x^2 - x) + x.$$

By direct calculation,

$$\int_0^1 P(x)\,dx = -\frac{1}{12\alpha^2} + \frac{1}{3\alpha} + \frac{1}{2} = -\frac{1}{12}\beta^2 + \frac{1}{3}\beta + \frac{1}{2},$$

where $\beta = \frac{1}{\alpha}$. The integral has a maximum value of $\frac{5}{6}$, attained for $\alpha = \frac{1}{2}, \beta = 2$. In particular, we have now shown that no smaller constant C is possible.

Subcase 2: $P(0) = 0$, $0 < P(x) < 1$ for all $x \in (0, 1]$ except $x = \alpha$, $P(\alpha) = 1$, for some $\alpha \in (0, 1)$. In this subcase, $P(x) - 1$ has a double root at α, value -1 at 0, and a negative value at 1, from which we see that

$$P(x) = 1 + (cx - 1)\left(\frac{x}{\alpha} - 1\right)^2$$

for some c with $c < 1$. Thus we have

$$\int_0^1 P(x)\,dx < \int_0^1 1 + (x - 1)\left(\frac{x}{\alpha} - 1\right)^2 dx \leq \frac{5}{6}$$

by our calculation from the previous subcase.

Subcase 3: $P(0) = 0$, $0 < P(x) < 1$ for all $x \in (0, 1)$, $P(1) = 1$. Note that if $k \geq 0$ is a constant such that $P(x) + kx(1 - x)^2 < 1$ for all $x \in (0, 1)$, then we can replace

$P(x)$ by the new polynomial $P(x) + kx(1-x)^2$ without affecting the conditions, and this will increase the integral. With this motivation, let

$$k = \inf_{x \in (0,1)} \frac{1 - P(x)}{x(1-x)^2}; \text{ note that } k \geq 0.$$

Because k is the infimum of a continuous function on an open interval, it is either achieved for some $x = \alpha \in (0, 1)$, or approached as x approaches one of the end points of the interval. If it is achieved, then the polynomial

$$Q(x) = P(x) + kx(1-x)^2$$

satisfies the conditions of subcase 1 (in particular, its degree must be 3, because $Q(x) - 1$ has a double root at α and a root at 1) and we have

$$\int_0^1 P(x)\,dx \leq \int_0^1 Q(x)\,dx \leq \frac{5}{6}$$

by our earlier work. If the infimum is not achieved, we must have

$$k = \lim_{x \uparrow 1} \frac{1 - P(x)}{x(1-x)^2},$$

and so the polynomial $1 - P(x)$ must be divisible by $(1-x)^2$. Combining this with $P(0) = 0$, we see that $P(x)$ can be written in the form

$$P(x) = 1 + (ax - 1)(1-x)^2$$

for some constant a; because $P(x)$ does not take on the value 1 on the open interval $(0, 1)$, we have $a \leq 1$, else $P(x) = 1$ for $x = \frac{1}{a}$. We then have the estimate

$$\int_0^1 P(x)\,dx \leq \int_0^1 1 + (x-1)(1-x)^2\,dx = \int_0^1 1 - (1-x)^3\,dx = \frac{3}{4} < \frac{5}{6}.$$

We now proceed to the general case, in which $P(x)$ can have anywhere from one to three roots in the interval $[0, 1]$; we still have $M = 1$. Partition the interval into subintervals

$$[r_i, r_{i+1}] \quad \text{with} \quad 0 = r_0 < r_1 < \cdots < r_{s+1} = 1$$

such that each root of $P(x)$ occurs among $r_0, \ldots, r_{s+1}$ and exactly one end point of each subinterval $[r_i, r_{i+1}]$ is a root of $P(x)$. (In the special case considered so far, there was only one subinterval.) For $i = 0, 1, \ldots, s$, define $P_i(x) = P(r_i + (r_{i+1} - r_i)x)$. Then $P_i(x)$ is a polynomial of degree 3 which has a root at exactly one of the end points of the interval $[0, 1]$ and no other roots in $[0, 1]$. We have

$$\int_0^1 |P(x)|\,dx = \sum_{i=0}^{s} \int_{r_i}^{r_{i+1}} |P(x)|\,dx = \sum_{i=0}^{s} (r_{i+1} - r_i) \int_0^1 |P_i(u)| du$$

by using the substitution $x = r_i + (r_{i+1} - r_i)u$ on the ith subinterval. For each i, we can replace $P_i(u)$ by $-P_i(u)$ if necessary to get $0 \leq P_i(u)$ for all $u \in [0, 1]$, and then replace $P_i(u)$ by $P_i(1-u)$ if necessary to get $P_i(0) = 0$. Then for each i we have a polynomial that satisfies the conditions of our special case, except that the maximum value of $|P_i(u)|$ on [0, 1] may not be 1. However, the values of $|P_i(u)|$ on [0, 1] are among the values of $|P(x)|$ on [0, 1], so they are at most 1. We therefore have

$$\int_0^1 |P(x)|\,dx = \sum_{i=0}^{s}(r_{i+1} - r_i)\int_0^1 |P_i(u)|du \le \sum_{i=0}^{s}(r_{i+1} - r_i)\frac{5}{6} = \frac{5}{6}(r_{s+1} - r_0) = \frac{5}{6}.$$

Solution to B1. From the given equation $x_{n+1} = \ln(e^{x_n} - x_n)$ we see that $e^{x_{n+1}} = e^{x_n} - x_n$, so $x_n = e^{x_n} - e^{x_{n+1}}$. Thus the partial sums of the given series are telescoping:

$$x_0 + x_1 + \cdots + x_{N-1} = e^{x_0} - e^{x_1} + e^{x_1} - e^{x_2} + \cdots + e^{x_{N-1}} - e^{x_N} = e - e^{x_N}.$$

So to show the series converges, we should investigate the existence of $L = \lim_{N\to\infty} x_N$. Let $f(x) = \ln(e^x - x)$, so that $x_{n+1} = f(x_n)$. Note that by the well-known inequality $e^x \ge 1 + x$, for any $x > 0$ we have $1 \le e^x - x < e^x$ and so

$$0 = \ln(1) \le f(x) < \ln(e^x) = x.$$

Therefore, the sequence (x_n) is decreasing and bounded below (by 0), so its limit L exists; it follows that the given series converges and has sum $e - e^L$. Finally, taking the limit of both sides of $x_{n+1} = f(x_n)$, by the continuity of f we get $L = f(L) = \ln(e^L - L)$, so $e^L = e^L - L$ and $L = 0$. So the desired sum is $e - 1$.

Solution to B2. We'll show that $\alpha = \frac{3}{4}$, $\beta = \frac{4}{3}$. Consider a perfect square k^2, where k is a positive integer. The integers for which k^2 is the nearest perfect square are the integers from $k^2 - k + 1$ to $k^2 + k$, inclusive, and so the squarish integers for which k^2 is the nearest perfect square are the integers $k^2 - m^2$ with $0 < m^2 \le k - 1$, the integers $k^2 + m^2$ with $0 < m^2 \le k$, and k^2 itself. In particular, the number of squarish integers for which k^2 is the nearest perfect square is $\lfloor\sqrt{k-1}\rfloor + \lfloor\sqrt{k}\rfloor + 1$, where $\lfloor a\rfloor$ indicates the greatest integer $\le a$. Then if the integer N is of the form $N = n^2 + n$, we have

$$\begin{aligned} S(N) &= \sum_{k=1}^{n}\left(\left\lfloor\sqrt{k-1}\right\rfloor + \left\lfloor\sqrt{k}\right\rfloor + 1\right) \le \sum_{k=1}^{n}\left(2\left\lfloor\sqrt{k}\right\rfloor + 1\right) \\ &\le 2\left(\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}\right) + n \\ &< 2\int_0^{n+1}\sqrt{x}\,dx + n \\ &= \frac{4}{3}(n+1)^{3/2} + n, \end{aligned}$$

while also

$$\begin{aligned} S(N) &= \sum_{k=1}^{n}\left(\left\lfloor\sqrt{k-1}\right\rfloor + \left\lfloor\sqrt{k}\right\rfloor + 1\right) \ge \sum_{k=1}^{n} 2\left\lfloor\sqrt{k}\right\rfloor \\ &\ge 2\left(\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}\right) - 2n \\ &> 2\int_0^{n}\sqrt{x}\,dx - 2n \\ &= \frac{4}{3}n^{3/2} - 2n. \end{aligned}$$

Now for any positive integer N, we have

$$(n-1)^2 + (n-1) = n^2 - n \le N \le n^2 + n$$

for either $n = \lfloor \sqrt{N} \rfloor$ or $n = \lfloor \sqrt{N} \rfloor + 1$, and then

$$\frac{4}{3}\left(\sqrt{N}-2\right)^{3/2} - 2\sqrt{N} < \frac{4}{3}(n-1)^{3/2} - 2(n-1) < S(n^2-n) \le S(N)$$

and

$$S(N) \le S(n^2+n) < \frac{4}{3}(n+1)^{3/2} + n < \frac{4}{3}(\sqrt{N}+2)^{3/2} + \sqrt{N} + 1$$

by the estimates above. It follows by the "squeeze theorem" that

$$\lim_{N\to\infty} \frac{S(N)}{N^{3/4}} = \frac{4}{3},$$

and we are done.

Solution to B3. If all the points in S are on the same line, there is no difficulty. Otherwise, choose points A, B, C in S such that $\triangle ABC$ has maximal area. Let l_A be the line through A parallel to the opposite side BC; similarly, let l_B, l_C be the lines through B, C parallel to AC, AB, respectively. Note that these three lines form a triangle PQR such that A is the midpoint of QR, B is the midpoint of RP, and C is the midpoint of PQ. The region enclosed by $\triangle PQR$ is divided by the sides of $\triangle ABC$ into four congruent triangular regions of equal area, so area$(\triangle PQR) = 4 \cdot$ area$(\triangle ABC)$ is at most 4; we will show that $\triangle PQR$ covers the set S. If necessary, we can then enlarge $\triangle PQR$ to get a triangle of area 4 that covers S.

Let D be a point in S. Then area$(\triangle ABD) \le$ area$(\triangle ABC)$, so the altitude from D to AB is at most as long as the altitude from C to AB, so either D lies on the line l_C or D is on the same side of l_C as AB. Similarly, D is in the same closed half-plane bounded by l_B as CA, and in the same closed half-plane bounded by l_A as BC. It follows that D is on or inside $\triangle PQR$, completing the proof.

Solution to B4. Let the i, j entry of A be the random variable $X_{i,j}$ and the i, j entry of $A - A^t$ be the random variable $Y_{i,j} = X_{i,j} - X_{j,i}$. Note that the expectation of $Y_{i,j}$ is $\mathbb{E}(Y_{i,j}) = \mathbb{E}(X_{i,j}) - \mathbb{E}(X_{j,i}) = \frac{1}{2} - \frac{1}{2} = 0$ for all i and j. Also, $Y_{j,j} = 0$ for all j, while for $i \neq j$, $Y_{i,j}$ takes on the values $1, 0, -1$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, respectively, so $Y_{i,j}^2$ takes on the values $1, 0$ each with probability $\frac{1}{2}$, so $\mathbb{E}(Y_{i,j}^2) = \frac{1}{2}$. Meanwhile, $Y_{i,j}$ and $Y_{k,l}$ are independent random variables unless $(i, j) = (k, l)$ or $(i, j) = (l, k)$.

Turning to the determinant, its expectation is given by

$$\begin{aligned}\mathbb{E}(\det(A - A^t)) &= \mathbb{E}\left(\sum_{\sigma \in S_{2n}} \operatorname{sgn}(\sigma) \prod_{i=1}^{2n} Y_{i,\sigma(i)}\right) \\ &= \sum_{\sigma \in S_{2n}} \operatorname{sgn}(\sigma)\, \mathbb{E}\left(\prod_{i=1}^{2n} Y_{i,\sigma(i)}\right).\end{aligned}$$

Now if σ is a permutation that fixes any $j \in \{1, 2, \ldots, 2n\}$, then $\prod_{i=1}^{2n} Y_{i,\sigma(i)}$ is zero because of the factor $Y_{j,j}$. On the other hand, if σ is a permutation with $\sigma^2(j) \neq j$ for some j, then for that j, $Y_{j,\sigma(j)}$ and $\prod_{i \neq j} Y_{i,\sigma(i)}$ are independent, and so

$$\mathbb{E}\left(\prod_{i=1}^{2n} Y_{i,\sigma(i)}\right) = \mathbb{E}(Y_{j,\sigma(j)})\mathbb{E}\left(\prod_{i\neq j} Y_{i,\sigma(i)}\right)$$
$$= 0 \cdot \mathbb{E}\left(\prod_{i\neq j} Y_{i,\sigma(i)}\right) = 0.$$

So the only permutations that don't contribute 0 to the expectation of the determinant are the permutations that are products of n disjoint 2-cycles. For such a permutation σ, $\operatorname{sgn}(\sigma) = (-1)^n$, and $\mathbb{E}\left(\prod_{i=1}^{2n} Y_{i,\sigma(i)}\right)$ is the product of the expectations of n independent random variables of the form $Y_{j,\sigma(j)}Y_{\sigma(j),j} = -Y^2_{j,\sigma(j)}$; we have seen that those separate expectations are $-\frac{1}{2}$, so we get

$$\operatorname{sgn}(\sigma)\,\mathbb{E}\left(\prod_{i=1}^{2n} Y_{i,\sigma(i)}\right) = (-1)^n\left(-\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n$$

for such a permutation. Meanwhile, the number of such permutations is

$$(2n-1)(2n-3)\cdots 3\cdot 1 = \frac{(2n)!}{2n(2n-2)\cdots 2} = \frac{(2n)!}{2^n\cdot n!},$$

leading to the final answer $\left(\frac{1}{2}\right)^n \cdot \frac{(2n)!}{2^n\cdot n!} = \frac{(2n)!}{4^n\cdot n!}$.

Solution to B5. (Based on a student paper.) It is easy to check that for any positive real number a, the function $f\colon (1,\infty)\to(1,\infty)$ defined by $f(x) = x^a$ for all $x\in(1,\infty)$ has the property stated in the problem. We now show that these are the only such functions, using the following lemma.

Lemma. Let α, β be fixed positive real numbers whose ratio is irrational. Then, given any constant $\varepsilon > 0$, there exist integers $m_1, n_1 > 0$ such that $0 < m_1\alpha - n_1\beta < \varepsilon$, and also integers $m_2, n_2 > 0$ such that $0 < n_2\beta - m_2\alpha < \varepsilon$. As a result, *any* open interval contains a number of the form $m\alpha - n\beta$ (as well as a number of the form $n\beta - m\alpha$) with m, n positive integers.

To prove the lemma, choose N such that $N > \beta/\varepsilon$, and divide the interval $[0,\beta]$ into N subintervals of equal size. Consider the numbers $k\alpha \pmod{\beta}$ in that interval, as k varies from 1 to $N+1$. By the Pigeonhole Principle, two of these numbers must lie in the same subinterval, so we have $k_1 < k_2$ such that $k_1\alpha - j_1\beta$, $k_2\alpha - j_2\beta$ are in the same subinterval for some nonnegative integers j_1, j_2; note that $j_1 \le j_2$ because $k_1\alpha < k_2\alpha$. Then for $m = k_2 - k_1 > 0$, $n = j_2 - j_1 \ge 0$,

$$0 < |m\alpha - n\beta| \le \beta/N < \varepsilon.$$

By replacing ε in this argument by a smaller positive constant that is less than α, we can guarantee that $n > 0$. We can now take either
$m_1 = m, n_1 = n$ or $n_2 = n, m_2 = m$, depending on the sign of $m\alpha - n\beta$. If we could only find (say) m_1, n_1 satisfying the condition but not m_2, n_2, we could, by making ε smaller, find m_1', n_1' with $m_1' > m_1$ and $0 < m_1'\alpha - n_1'\beta < m_1\alpha - n_1\beta$ (and consequently $n_1' > n_1$). But then we could take $m_2 = m_1' - m_1$, $n_2 = n_1' - n_1$, and we would then have $m_2, n_2 > 0$, $0 < n_2\beta - m_2\alpha < \varepsilon$ after all. This completes the proof of the lemma.

Given a function f with the desired property, define a function $g\colon (0,\infty) \to (0,\infty)$ by the condition $g(\ln x) = \ln(f(x))$, or equivalently $g(x) = \ln(f(e^x))$.

Note that if $x, y \in (0, \infty)$ and $2x \le y \le 3x$, then $(e^x)^2 \le e^y \le (e^x)^3$, so $(f(e^x))^2 \le f(e^y) \le (f(e^x))^3$ and taking logarithms, $2\,g(x) \le g(y) \le 3\,g(x)$. Because this condition on g is linear, it will be enough to find the functions $g : (0, \infty) \to (0, \infty)$ for which both **a)** $g(1) = 1$ and **b)** $2x \le y \le 3x$ implies $2\,g(x) \le g(y) \le 3\,g(x)$.

Given such a function g, we repeat the transformation above by defining a function $h\colon (-\infty, \infty) \to (-\infty, \infty)$ by $h(\ln x) = \ln(g(x))$, or $h(x) = \ln(g(e^x))$. Then we have $h(0) = 0$, and $x + \ln 2 \le y \le x + \ln 3$ implies $h(x) + \ln 2 \le h(y) \le h(x) + \ln 3$.

We set $\alpha = \ln 2,\ \beta = \ln 3$, so our last implication states that

$$x + \alpha \le y \le x + \beta \quad \text{implies} \quad h(x) + \alpha \le h(y) \le h(x) + \beta.$$

Then for $y = x + \alpha$, we have $h(x) + \alpha \le h(y) = h(x + \alpha)$, and for $y = x + \beta$, we have $h(x + \beta) = h(y) \le h(x) + \beta$. Starting from $h(0) = 0$ and iterating the first of these inequalities m times and then the second inequality n times, we obtain

$$m\alpha - n\beta \le h(m\alpha - n\beta) \quad \text{for all integers } m, n \ge 0.$$

Therefore, for any integers $m \ge 1, n \ge 0$ and any element y of the closed interval $[m\alpha - n\beta, m\alpha - n\beta + \beta - \alpha]$, because we have $x + \alpha \le y \le x + \beta$ for $x = (m-1)\alpha - n\beta$, we have $h(x) + \alpha \le h(y)$ and thus

$$h(y) \ge h(x) + \alpha = h((m-1)\alpha - n\beta) + \alpha \ge (m-1)\alpha - n\beta + \alpha = m\alpha - n\beta.$$

By the lemma, for any fixed real number x_0 and any $\varepsilon > 0$ there exist integers $m \ge 1$, $n \ge 0$ such that $x_0 > m\alpha - n\beta > x_0 - \varepsilon$. Then provided ε is small enough (specifically, $\varepsilon \le \beta - \alpha$), we have $x_0 \in [m\alpha - n\beta, m\alpha - n\beta + \beta - \alpha]$ and so $h(x_0) \ge m\alpha - n\beta > x_0 - \varepsilon$ by the estimate above. Letting $\varepsilon \to 0$, we see that $h(x_0) \ge x_0$.

Similarly, we have $h(x) - \alpha \ge h(x - \alpha),\ h(x - \beta) \ge h(x) - \beta$, and iterating these inequalities we get

$$h(n\beta - m\alpha) \le n\beta - m\alpha \quad \text{for all integers } m, n \ge 0.$$

Then for any integers $m \ge 0, n \ge 1$ and any element z of the closed interval $[n\beta - m\alpha - (\beta - \alpha), n\beta - m\alpha]$, we have
$h(z) \le h((n-1)\beta - m\alpha) + \beta \le n\beta - m\alpha$. By the lemma, for our fixed x_0 and small enough $\varepsilon > 0$, we can arrange for x_0 to be in that closed interval and for $x_0 < n\beta - m\alpha < x_0 + \varepsilon$, and so $h(x_0) \le x_0 + \varepsilon$. Letting $\varepsilon \to 0$, we see that $h(x_0) \le x_0$, and so, finally, $h(x_0) = x_0$.

We have now shown that $h(x) = x$ for all x, so $g(x) = x$. Removing the condition $g(1) = 1$, we see that the possible functions g are given by $g(x) = ax$, where a is a positive constant, and therefore the possible functions f are given by $f(x) = x^a$, completing the proof.

Solution to B6. (Based on a student paper.) The series is absolutely convergent, because

$$\frac{1}{k}\sum_{n=0}^{\infty}\frac{1}{k\,2^n + 1} < \frac{1}{k}\sum_{n=0}^{\infty}\frac{1}{k\,2^n} = \frac{1}{k}\frac{2}{k} = \frac{2}{k^2}$$

and $\sum_{k=1}^{\infty}\frac{2}{k^2}$ converges. Let

$$f(k) = \frac{(-1)^{k-1}}{k}\sum_{n=0}^{\infty}\frac{1}{k\,2^n + 1}, \quad \text{so the desired sum is} \quad \sum_{k=1}^{\infty} f(k).$$

Each value of k can be written uniquely in the form $k = q\,2^j$ where q is odd and $j \geq 0$ is the number of factors 2 in k, so we can rewrite the sum as

$$\sum_{k=1}^{\infty} f(k) = \sum_{q \text{ odd}} \sum_{j=0}^{\infty} f(q\,2^j).$$

Now note that for any odd q, we have

$$\begin{aligned}
\sum_{j=0}^{\infty} f(q\,2^j) &= f(q) + \sum_{j=1}^{\infty} f(q\,2^j) \\
&= \frac{1}{q}\sum_{n=0}^{\infty} \frac{1}{q\,2^n + 1} + \sum_{j=1}^{\infty} \frac{-1}{q\,2^j} \sum_{n=0}^{\infty} \frac{1}{q\,2^{n+j} + 1} \\
&= \frac{1}{q}\sum_{n=0}^{\infty} \frac{1}{q\,2^n + 1} + \sum_{j=1}^{\infty} \frac{-1}{q\,2^j} \sum_{n=j}^{\infty} \frac{1}{q\,2^{n} + 1} \\
&= \frac{1}{q}\sum_{n=0}^{\infty} \frac{1}{q\,2^n + 1} - \sum_{n=1}^{\infty} \sum_{j=1}^{n} \frac{1}{q\,2^j(q\,2^n + 1)} \\
&= \frac{1}{q}\sum_{n=0}^{\infty} \frac{1}{q\,2^n + 1} - \frac{1}{q}\sum_{n=0}^{\infty} \Big(1 - \frac{1}{2^n}\Big) \frac{1}{q\,2^n + 1} \\
&= \sum_{n=0}^{\infty} \frac{1}{q\,2^n(q\,2^n + 1)}.
\end{aligned}$$

Therefore, the desired sum equals

$$\sum_{q \text{ odd}} \sum_{n=0}^{\infty} \frac{1}{q\,2^n(q\,2^n + 1)}.$$

However, as seen earlier, as q runs through all odd integers and n runs through all nonnegative integers, $q\,2^n$ runs through all positive integers, and so this last sum is simply the standard telescoping series

$$\sum_{m=1}^{\infty} \frac{1}{m(m+1)} = \sum_{m=1}^{\infty} \left(\frac{1}{m} - \frac{1}{m+1}\right) = \lim_{M\to\infty} \left(1 - \frac{1}{M+1}\right) = 1.$$

MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth St., NW
Washington, DC 20036

CONTENTS

ARTICLES

PROBLEMS AND SOLUTIONS

REVIEWS

NEWS AND LETTERS