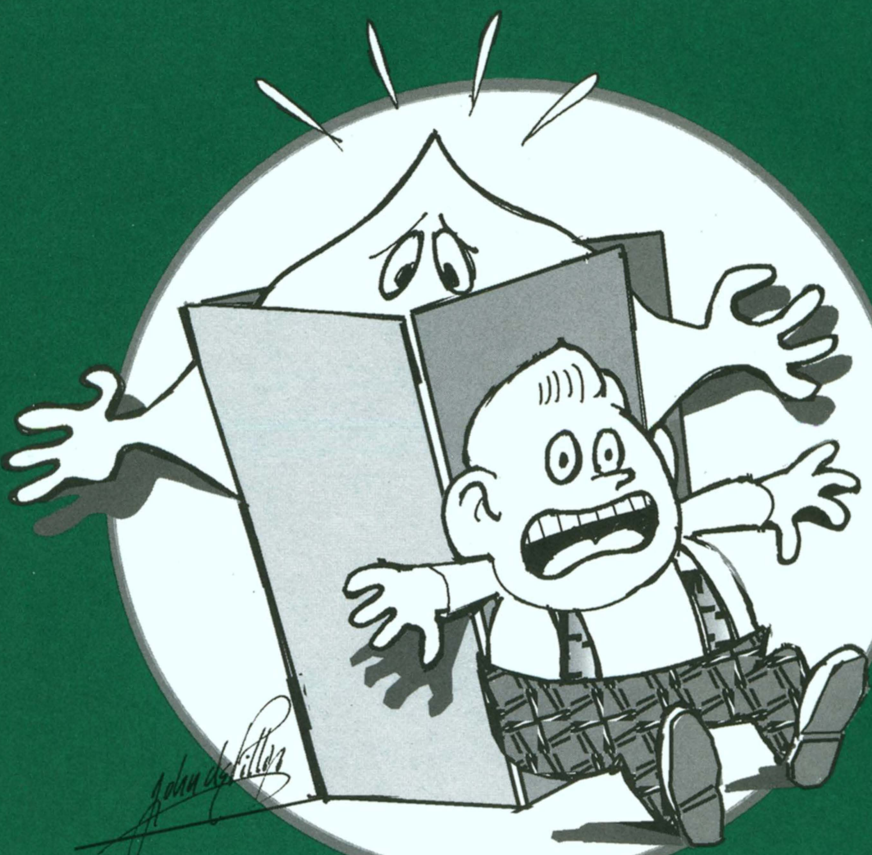
Vol. 80, No. 2, April 2007

MATHEMATICS MAGAZINE



Mom! There's an Astroid in My Closet!

- Minimum Area Venn Diagrams Whose Curves are Polyominoes
- Primes and Probability: The Hawkins Random Sieve

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

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

The **cover image** was drawn by John de Pillis, a mathematician at the University of California, Riverside, and MAA author of *777 Mathematical Conversation Starters*. It depicts our young mathematician confining the astroid in the closet. But he eventually coaxed the astroid out of the closet and into the article in this issue.

AUTHORS

Frank Ruskey is a Professor of Computer Science at the University of Victoria where he has been a faculty member since 1978. His research interests are in combinatorial algorithms and combinatorial mathematics. He owes his interest in Venn diagrams to a talk by Anthony Edwards and the stimulating writings of Branko Grünbaum. Together with his student Mark Weston he maintains an online Survey of Venn diagrams at the Electronic Journal of Combinatorics.

Stirling Chow is a Ph.D. student in the Computer Science Department at the University of Victoria. His studies are supported by a Canada Graduate Scholarship. His research interests are in combinatorial algorithms and in data visualization. His thesis is about the drawing of area-proportional Venn diagrams.

In their spare time both authors enjoy fishing in the beautiful waters around Victoria, B. C.

Richard Brazier received his BA from Bath University in the UK and his Masters and Ph.D. degrees in Applied mathematics from University of Arizona in Tucson. His interests include his family, seismology, gardening, home remodeling, and philately.

Eugene Boman received his BA from Reed College in 1984 and his MA in 1986 and Ph.D. in 1993 from the University of Connecticut. He has been at Penn State since 1996, first at the DuBois campus and more recently at the Harrisburg campus. He was minding his own business in his office one day when a freshman calculus student, Derek Seiple by name, came in and posed the following problem: How much less area does a bifold door need to open and close than a normal door? Seiple said he had been thinking about the problem since it had occurred to him as he was preparing for school a couple of years prior. Boman made a couple of suggestions and sent Seiple on his way expecting that would be the last he heard of the matter. A couple of days later Seiple returned with a partial solution. Boman made a few more suggestions; Seiple returned a few days later having made more progress. This pattern continued until the original problem had been solved and then generalized a couple of different ways. This article is the result.

Derek Seiple is currently finishing his undergraduate degree in mathematics. His Bachelor of Science Degree will be awarded by The Pennsylvania State University in May of 2007. Derek's future plans include the pursuit of a Ph.D. in mathematics.

John Lorch was educated at Palmer High School, the University of Colorado at Colorado Springs, and Oklahoma State University. When he is not fiddling with Hawkins primes or caught in the throes of some other brief mathematical addiction, John enjoys spending time with his family (Crystal, Carolyn, and Rob), reading Victorian tales of terror and the supernatural, and playing guitar in a pop cover band.

Giray Ökten is an Associate Professor of Mathematics at Florida State University. He has also taught at Ball State University, University of Alaska Fairbanks, and Claremont Graduate University where he received his Ph.D. in 1997. His research interests include Monte Carlo methods, computational finance, and in general, applied probability. In his free time, he plays with his children, watches movies with his wife, reads, and dreams.

Vol. 80, No. 2, April 2007

MATHEMATICS MAGAZINE

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

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

Advertising correspondence should be addressed to

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

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

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

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

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

Printed in the United States of America

ARTICLES

Minimum Area Venn Diagrams Whose Curves Are Polyominoes

STIRLING CHOW
Department of Computer Science
University of Victoria
Victoria, B.C., Canada V8W 3P6
schow@cs.uvic.ca

FRANK RUSKEY
Department of Computer Science
University of Victoria
Victoria, B.C., Canada V8W 3P6
ruskey@cs.uvic.ca

While working at the Berlin Academy, the renowned Swiss mathematician Leonard Euler was asked to tutor Frederick the Great's niece, the Princess of Anhalt-Dessau, in all matters of natural science and philosophy [**17**]. Euler's tutelage of the princess continued from 1760 to 1762 and culminated in the publishing of the popular and widely-translated "Letters to a German Princess" [**9**]. In the letters, Euler eloquently wrote about diverse topics ranging from why the sky was blue to free will and determinism.

In his lesson on categorical propositions and syllogisms, Euler used diagrams comprised of overlapping circles; these diagrams became known as Eulerian circles, or simply Euler diagrams. In an Euler diagram, a proposition's classes are represented as circles whose overlap depends on the relationship established by the proposition. For example, the propositions

All arachnids are bugs
Some bugs are cannibals

can be represented by FIGURE 1.

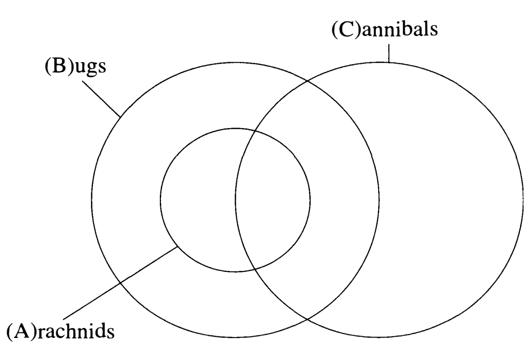


Figure 1 An example of an Euler diagram

In 1880, a Cambridge priest and mathematician named John Venn published a paper studying special instances of Euler diagrams in which the classes overlap in all possible ways [**27**]; although originally applied to logic reasoning, these Venn diagrams are now commonly used to teach students about set theory. For example, the Venn diagram in FIGURE 2 shows all the ways in which three sets can intersect. The primary difference between Venn and Euler diagrams is how they represent empty sets (e.g., the set of arachnids which are *not* bugs in the example of FIGURE 1). In an Euler diagram, regions representing empty sets are omitted, while in Venn diagrams they are included but denoted by shading.

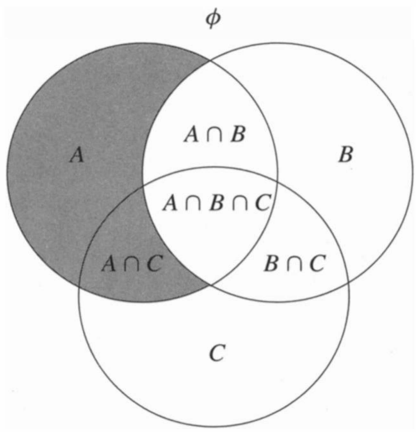


Figure 2 A Venn diagram that represents the Euler diagram in FIGURE 1 by shading the missing regions

Informally, an n-Venn diagram is a set of n simple, closed curves that subdivide the plane into 2^n connected regions with each region uniquely mapping to a subset of the n curves consisting of those curves which enclose it. The regions are usually referred to by their enclosing curves. For example, the 3-Venn diagram in FIGURE 2 with curves $\{A, B, C\}$ has regions $\{\emptyset, A, B, C, AB, AC, BC, ABC\}$. If an n-Venn diagram's curves are equivalent to each other modulo translations, rotations, and reflections, then the diagram is referred to as a *congruent* n-Venn diagram.

In recent years, there has been renewed interest in studying the combinatorial and geometric properties of Venn diagrams [**7, 24**]. Of paramount importance is how to draw a Venn diagram for a given number of sets. John Venn proposed an iterative algorithm in his original Venn paper [**27**]; unfortunately, the resulting drawings lacked an aesthetic appeal. In 1989, Anthony Edwards developed an elegant method for drawing n-Venn diagrams that produced highly symmetric drawings [**6**]. FIGURE 3 shows a comparison of 5-Venn diagrams drawn using Venn's and Edwards' algorithms.

An interesting problem popularized by Grünbaum [**13, 14, 15, 16**] is to consider which Venn diagrams can be drawn using specific shapes. FIGURE 2 shows a 3-Venn diagram comprised of circles; a natural question to ask is if such a diagram exists for four sets. It turns out the answer is no. First observe that three circles intersect to form at most eight regions. The addition of the fourth circle can intersect each of the others at most twice, and this maximum of six intersection points partitions the fourth circle into six arcs forming at most six new regions for a maximum of $8 + 6 = 14$ regions, not the 16 we require [**24**]. FIGURE 4 shows examples of Venn diagrams drawn using ellipses [**13**] and triangles [**3**]. The diagram in FIGURE 4(a) is special because it is

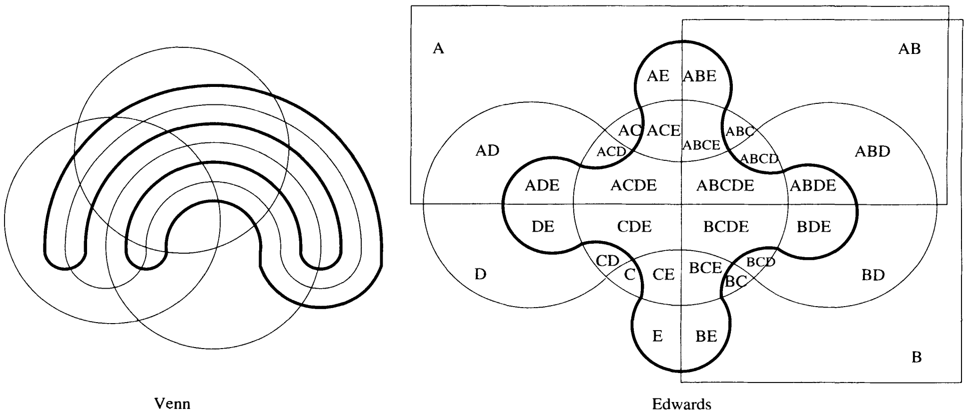


Figure 3 A 5-Venn diagram drawn iteratively using Venn's and Edwards' algorithms; the last curve drawn is highlighted.

an example of a *symmetric* Venn diagram; that is, a diagram with n-fold rotational symmetry and (necessarily) congruent curves. Symmetric Venn diagrams exist if and only if n is prime **[11]**.

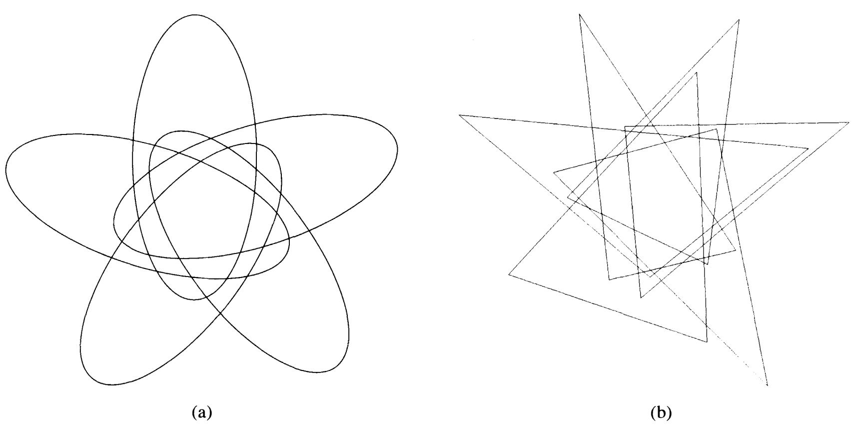


Figure 4 (a) A symmetric 5-Venn diagram using ellipses and (b) a 6-Venn diagram using triangles

On his "Math Recreations" web site **[26]**, Mark Thompson proposed the novel problem of finding *Venn polyominoes* (from now on referred to as *n-polyVenns*); these are Venn diagrams whose curves are the outlines of polyominoes. Polyominoes, or n-ominoes, are a generalization of dominoes (2-ominoes) whereby shapes are formed by gluing together n unit squares. One can also think of a polyomino as being the result of cutting a shape from a piece of graph paper where the cuts are made along the lines. Thompson found examples of congruent n-polyVenns for $n = 2, 3, 4$, and using a computer search, we found a congruent 5-polyVenn (see FIGURE 5).

In the remainder of this paper, our focus is on minimizing the total area of the drawing (relative to a scaling factor). We present some examples that minimize area according to various additional constraints. At present, these examples do not generalize, and so we develop an algorithm that comes close to minimizing the area. The algorithm is simple and utilizes symmetric chain decompositions of the Boolean lattice. We also

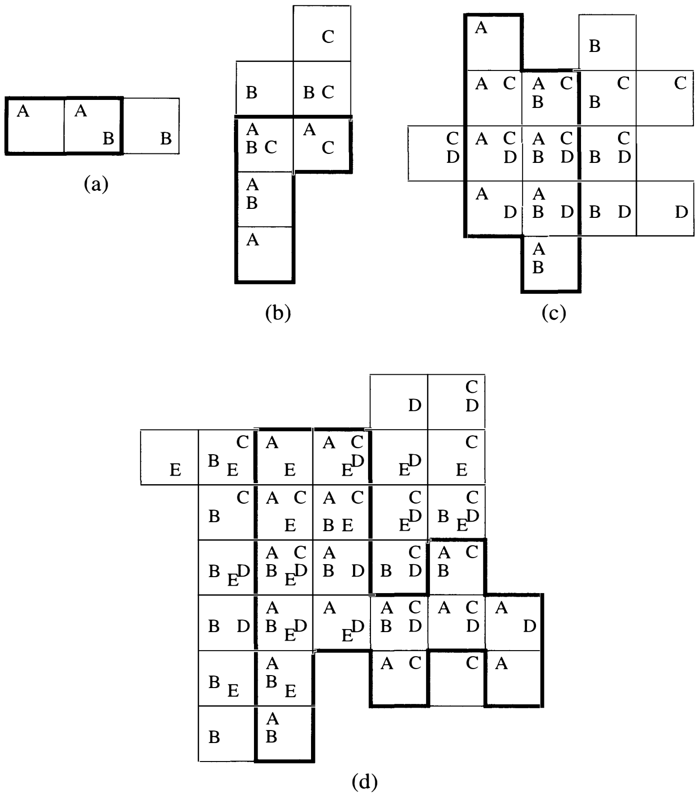


Figure 5 (a)–(c) Mark Thompson's congruent n-polyVenns for $n = 2, 3, 4$ and (d) the authors' congruent 5-polyVenn; in each case, curve A is highlighted.

provide asymptotic results that relate the area required by the algorithm's diagrams to the theoretical minimum area. We conclude by presenting some open problems related to Venn polyominoes and other shape-constrained Venn diagrams.

Polyominoes. A *polyomino* is an edge-connected set of unit squares, called *cells*, embedded in the integer lattice. Two cells are adjacent if, and only if, they share a common edge. Edge-connected means that every pair of cells is connected by a path through adjacent cells. Polyominoes are often classified by area and referred to as n-ominoes when they contain n cells. For example, the games of dominoes and Tetris are played with 2-ominoes and 4-ominoes (tetrominoes), respectively (see FIGURE 6).

Polyominoes have been extensively studied and have a wide-range of applications in mathematics and the physical sciences [**10, 19**]. The problem of counting n-ominoes has garnered considerable interest [**18, 21, 23**], and although counts up to 56-ominoes are known (see sequence A001168 [**25**]), the problem of finding a formula for the number remains open.

Several other subclasses of polyominoes have been defined. *Free* polyominoes treat polyominoes that are translations, rotations, or reflections of each other to be equivalent, whereas *fixed* polyominoes only consider translations as being equivalent. For example, FIGURE 6 shows the 19 equivalence classes of fixed tetrominoes and 5 equivalence classes (a, b, c, d, and e), of free tetrominoes.

If every column (row) of a polyomino is a contiguous strip of cells, then the polyomino is called *column-convex* (*row-convex*). A *convex* polyomino is one that is both column and row convex (see FIGURE 7). No closed-form formula is known for the

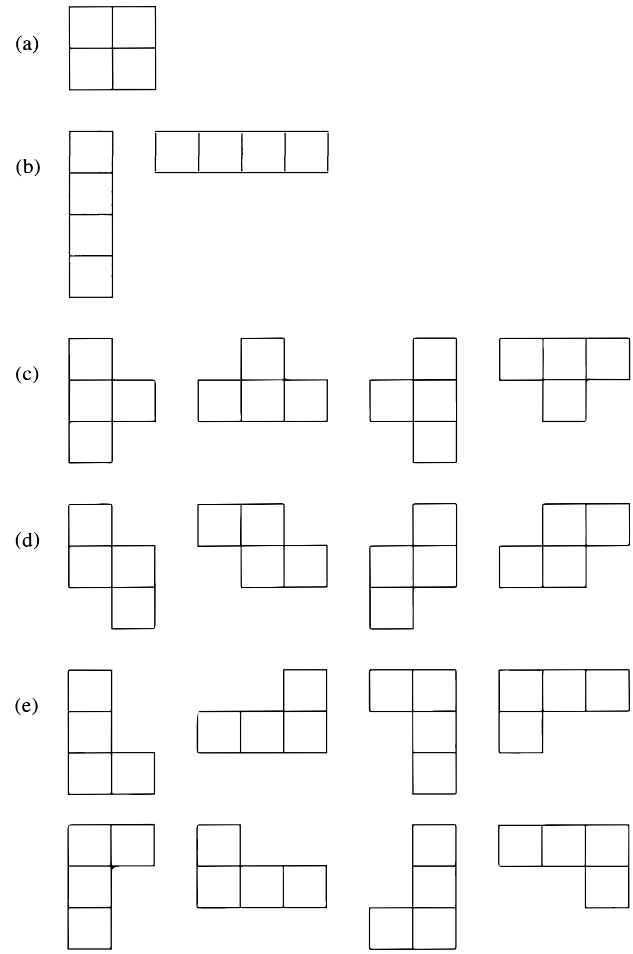


Figure 6 All possible 4-ominoes (tetrominoes)

number, $a(n)$, of fixed column-convex n-ominoes; however, Pólya [**22**] derived the recurrence relation $a(n) = 5a(n-1) - 7a(n-2) + 4a(n-3)$ with $a(1) = 1$, $a(2) = 2$, $a(3) = 6$, and $a(4) = 19$. This recurrence relation has the rational generating function

$$g(x) = \frac{x(1-x)^3}{1-5x+7x^2-4x^3}$$

(see sequence $A001169$ [**25**]).

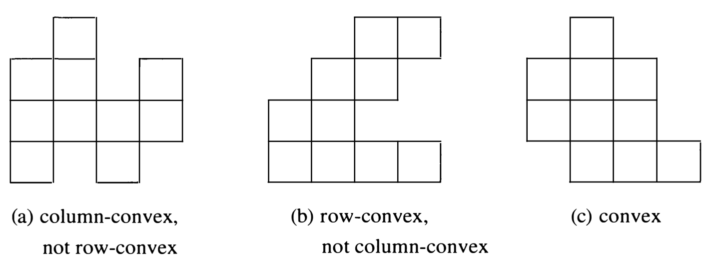


Figure 7 10-ominoes that exhibit different convexivity properties

Minimum area n-polyVenns. An n-polyVenn is a Venn diagram comprised of n curves, each of which is the perimeter of some polyomino. In particular, each polyomino must be free of holes in order for the perimeter to be a simple, closed curve, and when placed on top of another polyomino, may not partially cover any of the bottom polyomino's cells (i.e., the corners of the curves must have unit coordinates).

Referring to the examples in FIGURE 5, we see that an n-polyVenn can be drawn by tracing the curves on the lines of a piece of graph paper; in the (combinatorial) graph drawing community, this is referred to as an *orthogonal grid drawing* [2]. In fact, any orthogonal grid drawing of a Venn diagram will produce curves that are the perimeters of polyominoes. Since each bounded region must contain at least one cell and there is exactly one unbounded region, the minimum area for such a diagram is $2^n - 1$ cells. In addition, since each curve encloses 2^{n-1} regions, it must be the perimeter of at least a 2^{n-1}-omino. This leads us to the following definition of a minimum area n-polyVenn:

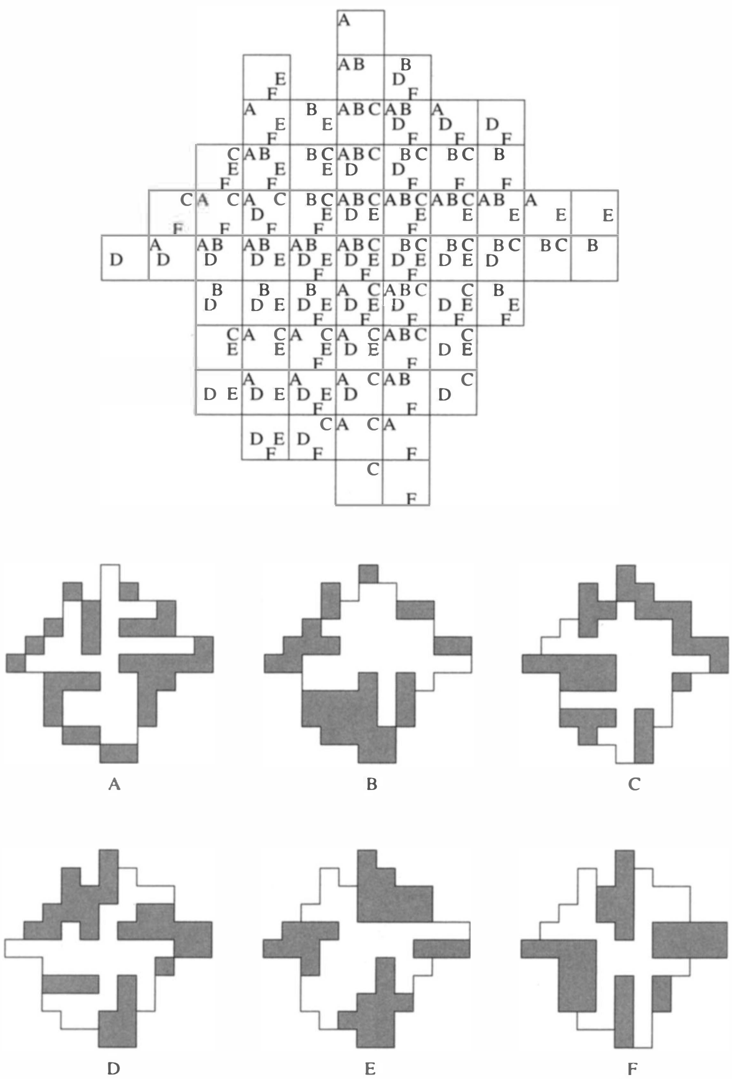


Figure 8 A minimum area 6-polyVenn

DEFINITION. A *minimum area n-polyVenn* is an orthogonal unit-grid drawing of a Venn diagram with area $2^n - 1$.

By necessity, each curve of a minimum area n-polyVenn has area 2^{n-1}. All the Venn diagrams in FIGURE 5 are minimum area congruent n-polyVenns. By trial-and-error, we have also found minimum area non-congruent n-polyVenns for $n = 6, 7$ (see Figs. 8, 9). It is unknown if minimum area n-polyVenns exist for $n \geq 8$, although we suspect there is an upper limit due to the rigid constraints of orthogonal grid drawings.

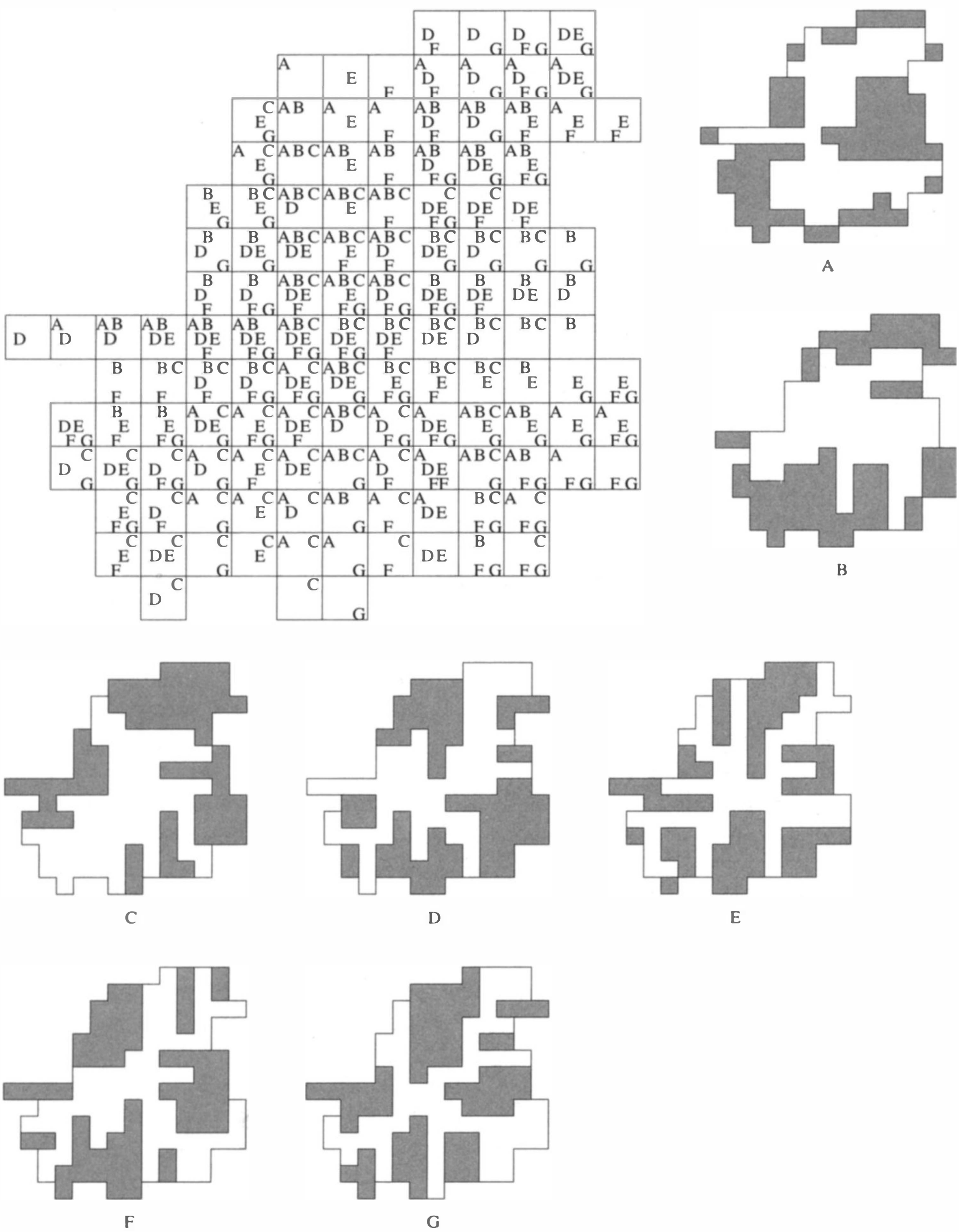


Figure 9 A minimum area 7-polyVenn

Orthogonal grid drawings of Venn diagrams were first studied by Eloff and van Zijl [**8**]; they developed a heuristic algorithm based on a greedy incremental approach. An optimization step in the algorithm attempted to reduce the overall area of the diagram, but there was no upper bound. In addition, their algorithm produced polyominoes with holes, so the resulting diagrams would not be considered Venn diagrams in the formal sense (because the sets were not represented by simple, closed curves).

In the following sections, we present algorithms for approximating minimum area n-polyVenns. The first algorithm is trivial and produces n-polyVenns with less than $3/2$ times the minimum area. The second algorithm improves upon the first by using symmetric chain decompositions of the Boolean lattice and produces n-polyVenns whose areas are asymptotically minimum (i.e., the ratio of total cells to required cells tends to one as n increases).

There is another definition of area based on the $w \times h$ bounding box that contains an n-polyVenn; such a box must also have at least one cell to represent the empty set. For example, the n-polyVenns in FIGURE 5 are contained by 4×1, 2×5, 5×5, and 7×7 bounding boxes, respectively. Since an n-polyVenn must be comprised of at least $2^n - 1$ cells, a bounding box must have area at least 2^n. This leads us to the following definition of a minimum bounding box n-polyVenn:

DEFINITION. A *minimum bounding box n-polyVenn* is an orthogonal unit-grid drawing of a Venn diagram that is enclosed by a $2^s \times 2^t$ rectangle where $s + t = n$.

Of the congruent n-polyVenns in FIGURE 5, only (a) is a minimum bounding box n-polyVenn. FIGURE 10 shows some examples of minimum bounding box non-congruent n-polyVenns.

At present, we leave minimum bounding box n-polyVenns and focus the rest of this paper on minimum area n-polyVenns.

A 3/2-APPROX algorithm. This algorithm is best explained by way of an example. Suppose we wish to draw a 5-polyVenn with the curves $\{A, B, C, D, E\}$. We begin by drawing a 1×14 rectangle and labelling it as region $ABCDE$; in other words, the curves are 1×14 rectangles stacked on top of each other. We now place 30 cells around the perimeter of $ABCDE$ and uniquely label them with the 30 remaining non-empty regions; the result is shown in FIGURE 11. After adding the perimeter cells, each curve becomes a polyomino formed by the original 1×14 rectangle with "bumps" wherever the curve encloses a perimeter cell.

In the general case, this algorithm will produce an n-Venn polyomino beginning with a $1 \times (2^{n-1} - 2)$ rectangle that has a perimeter of $2^n - 2$ (for the 2^n regions less the empty and full sets). The resulting diagrams have an area of $2^n + 2^{n-1} - 4$ which is less than $3/2$ times the minimum area of $2^n - 1$.

An asymptotically optimal algorithm. The previous algorithm can be significantly improved by noting that not all regions need to be placed adjacent to the initial rectangle; instead, if region X is a subset of region Y, then X can be placed directly above or below Y (depending on if Y is above or below the initial rectangle), and the curves will remain as polyomino perimeters. This chaining of regions can continue as long as the subset property is maintained. FIGURE 12 shows an example of 5-polyVenn that chains regions as much as possible. Note also that the resulting polyominoes are column-convex.

When regions are chained, a smaller perimeter is needed for the initial rectangle, and so the total area of the diagram is reduced. A smaller area diagram is created by

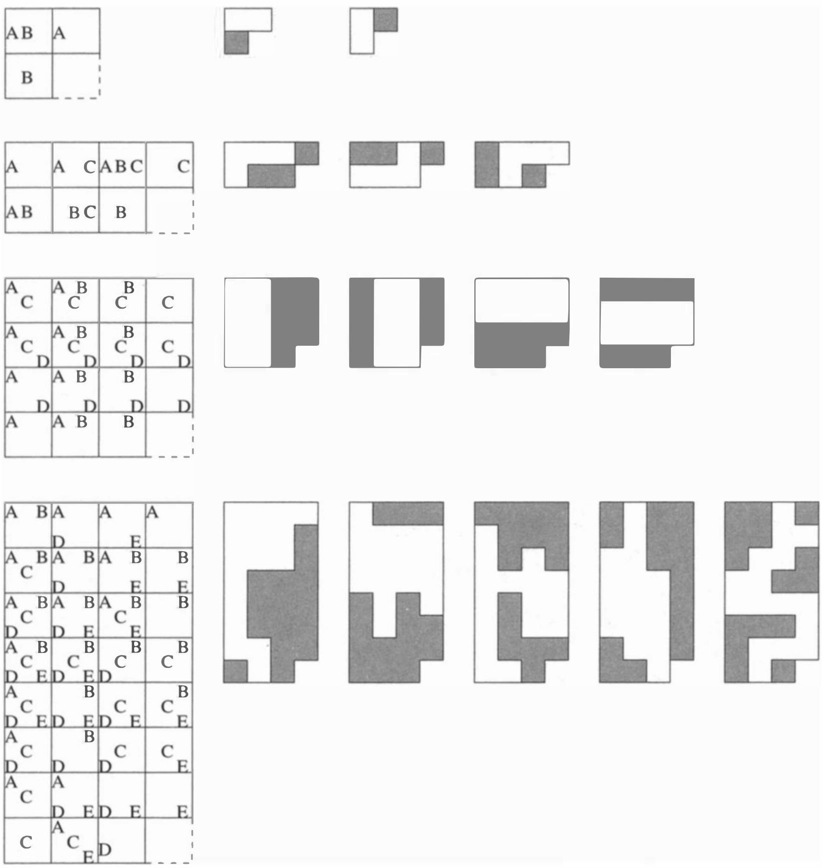


Figure 10 Minimum bounding box n-polyVenns for $2 \leq n \leq 5$

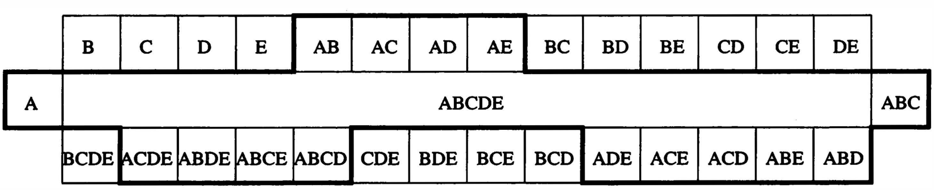


Figure 11 A naïve approximation for a minimum area 5-polyVenn; curve A is highlighted

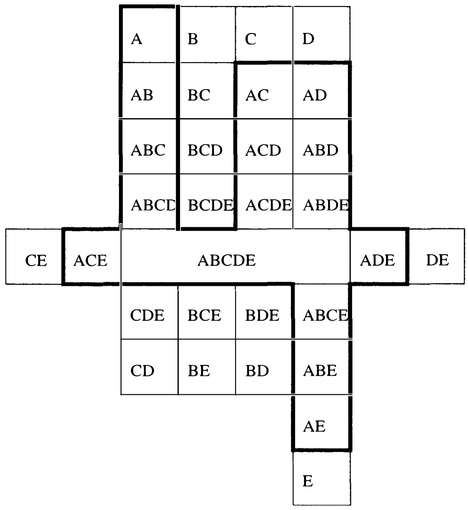


Figure 12 An approximation for a minimum area 5-polyVenn using column-convex polyominoes and symmetric chains; curve A is highlighted

minimizing the number of chains, so the question arises as to the best way to decompose the regions into chains; for this question, we need to use a result from the theory of partially ordered sets.

Given a set S with powerset $\mathcal{P}(S)$, we define the partially ordered set (poset) $\mathcal{L}(S)$ with elements $\mathcal{P}(S)$ ordered by inclusion. Since $\mathcal{L}(S)$ is closed under union, intersection, and complement, it is a Boolean lattice. FIGURE 13(a) shows an example of $\mathcal{L}(\{A, B, C, D\})$.

Let $|S| = n$. A symmetric chain decomposition (SCD) of $\mathcal{L}(S)$ is a partition of S into $\binom{n}{\lfloor n/2 \rfloor}$ symmetric chains. Each symmetric chain is a sequence of subsets $x_1, x_2, \ldots, x_t$ with the following properties:

$$x_i \subset x_{i+1} \text{ for all } 1 \leq i < t, \tag{1}$$

$$|x_i| = n - |x_{t-i+1}| \text{ for all } 1 \leq i \leq \lceil t/2 \rceil. \tag{2}$$

Symmetric chain decompositions form an essential ingredient of the recent proof of Griggs, Killian and Savage [**11**] that symmetric Venn diagrams exist if and only if the number of curves is prime.

Several algorithms exist for decomposing $\mathcal{L}(S)$ into symmetric chains; we describe two of these algorithms below. The first, due to de Bruijn, van Ebbenhorst Tengbergen, and Kruyswijk [**5**] is called the *Christmas tree pattern* by Knuth [**20**]. It is an inductive construction that creates a set T_n of $\binom{n}{\lfloor n/2 \rfloor}$ chains. Initially $T_1 = \{\emptyset \subset \{1\}\}$. To obtain T_n from T_{n-1}, take each chain $x_1 \subset x_2 \subset \cdots \subset x_t$ in T_n and replace it with the two chains $x_2 \subset \cdots \subset x_t$ and $x_1 \subset x_1 \cup \{n\} \subset x_2 \cup \{n\} \subset \cdots \subset x_t \cup \{n\}$ in T_{n+1} if $t > 1$. If $t = 1$ the first chain is empty and is ignored.

A second method, due to Aigner [**1**], can be described as a greedy lexicographic algorithm. It is efficient and easy-to-implement, and is the method that we used in creating the example diagrams. Let $m(x, y)$ be the smallest element in a set x that is not in the set y, where $m(x, y) = -\infty$ if $x \subset y$. We say that x is *lexicographically smaller* than y if $m(x, y) < m(y, x)$. In Aigner's algorithm, the following process is repeated until every element of $\mathcal{L}(\{1, 2, \ldots, n\})$ is contained in some chain. For

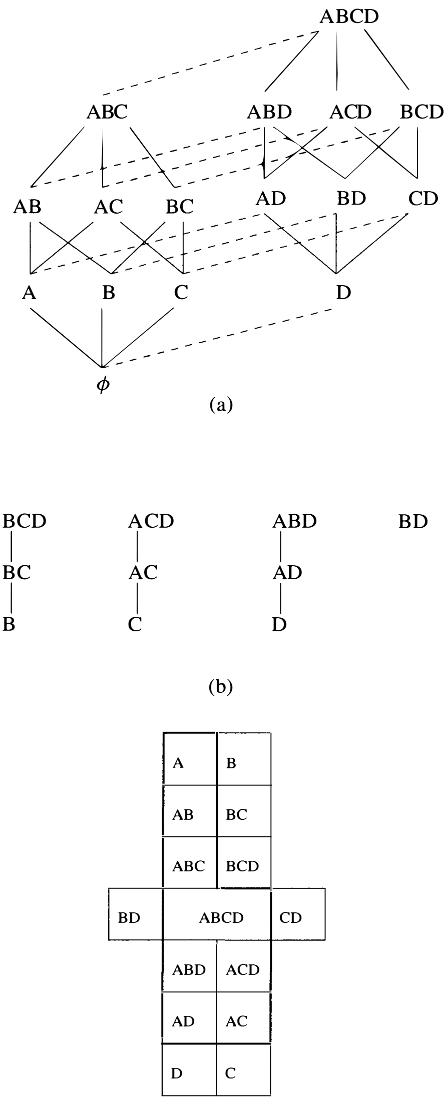


Figure 13 (a) A Hasse diagram of the poset $\mathcal{L}(\{A, B, C, D\})$, (b) one of its symmetric chain decompositions, and (c) the resulting 4-polyVenn

$k = 0, 1, 2, \ldots, n$, denote by $R(k)$ the set of subsets of $\{1, 2, \ldots, n\}$ with size k that are not yet in any chain. Let j be the smallest value for which $R(j)$ is non-empty and let x be the lexicographically smallest set in $R(j)$. The set x becomes the smallest set in a new chain $x = x_1 \subset x_2 \subset \cdots \subset x_t$. The successive elements of this chain are obtained by taking $x_{i+1} \in R(i+1)$ to be the lexicographically smallest set that contains x_i. It is by no means obvious that this algorithm is correct, but indeed it is!

Because of their subset property (1), the symmetric chains can be directly used to layout the regions of an n-polyVenn. FIGURE 13(b) shows the SCD of $\mathcal{L}(\{A, B, C, D\})$ that is produced by Aigner's algorithm, and FIGURE 13(c) shows the resulting 4-polyVenn. The 5-polyVenn in FIGURE 12 was also produced from Aigner's SCD of $\mathcal{L}(\{A, B, C, D, E\})$.

In the general case, this algorithm will produce an n-polyVenn beginning with a $1 \times (\binom{n}{\lfloor n/2 \rfloor} - 2)/2$ rectangle that has a perimeter of $\binom{n}{\lfloor n/2 \rfloor}$. The resulting diagrams have an area of $(\binom{n}{\lfloor n/2 \rfloor} - 2)/2 + 2^n - 2$. The lower bound

$$\binom{2n}{n} < \frac{2^{2n}}{\sqrt{\pi}(n^2 + n/2 + 3/32)^{1/4}}$$

[**12**] can be used to show that the algorithm produces diagrams whose area is $1 + O(1/\sqrt{n})$ times the minimum area of $2^n - 1$; therefore, as n increases, the approximation gets asymptotically close to optimal.

Open problems and final remarks. To close the paper, we list some open problems that are inspired by the examples in this paper. With the exception of the congruent n-polyVenns, the examples in this paper were constructed by hand, and it is very likely that relatively naïve programs will be able to extend them. Such extension would be interesting, but even more interesting would be general results that apply for arbitrary numbers of curves.

1. Are there congruent n-polyVenns for $n \geq 6$? FIGURE 5 shows that they exist for $n = 2, 3, 4, 5$.
2. Is there a 5-polyVenn whose curves are convex polyominoes? (The curves in FIGURE 5(d) are not both row-convex and column-convex polyominoes.)
3. Are there minimum bounding box n-polyVenns for $n \geq 6$? FIGURE 10 shows that they exist for $n = 2, 3, 4, 5$.
4. Are there minimum area n-polyVenns for $n \geq 8$? FIGURE 9 shows one for $n = 7$.
5. One problem for which we have not attempted solutions is the construction of n-polyVenns that fill a $w \times h$ box, where $wh = 2^n - 1$. Of course, a necessary condition is that $2^n - 1$ not be a Mersenne prime. For example, is there a 4-polyVenn that fits in a 3×5 rectangle or a 6-polyVenn that fits in a 7×9 or 3×21 rectangle?

Author's Note: Since submitting the original manuscript of this paper, Bette Bultena has discovered a 6-polyVenn with an 8×8 bounding box (see problem 3 above). FIGURE 8 has also been used to represent the results of experiments in plant genetics [**4**].

REFERENCES

1. Martin Aigner, Lexicographic matching in Boolean algebras, *Journal of Combinatorial Theory, Series B* **14** (1973) 187–194.
2. Guisseppi Battista, Peter Eades, Roberto Tomassia, and Ionnis Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice-Hall, 1999.
3. Jeremy Carroll, 6-Venn Triangle Problem, *Personal Homepage*, `http://www.hpl.hp.com/personal/jjc/index.html`.
4. Sandra Casimiro, Rogério Tenreiro, and António A. Monteiro, Identification of pathogenesis-related genes in the crucifer downy mildew oomycete Hyaloperonospora parasitica by differential display analysis of distinct phenotypic interactions with Brassica oleracea, *Journal of Microbiological Methods*, submitted.
5. N.G. de Bruijn, C. van Ebbenhorst Tengbergen, and D. Kruyswijk, On the set of divisors of a number, *Nieuw Archief voor Wiskunde* **2** (1951) 191–193.
6. Anthony W. F. Edwards, Venn diagrams for many sets, *New Scientist* **7** (1989) 51–56.
7. Anthony W. F. Edwards, *Cogwheels of the Mind: The Story of Venn Diagrams*, The Johns Hopkins University Press, 2004.
8. Jacques Eloff and Lynette van Zijl, An incremental construction algorithm for Venn diagrams, *SAICSIT 2000*, Cape Town, South Africa, November 2000.

9. Leonard Euler, *Lettres à une Princesse d'Allemagne sur Divers Sujets de Physique et de Philosophie*, Imperial Academy of Sciences, St. Petersburg, 1768–1772.
10. Solomon Golomb, *Polyominoes: puzzles, patterns, problems, and packings*, Princeton University Press, 1994.
11. Jerrold Griggs, Charles E. Killian, and Carla D. Savage, Venn diagrams and symmetric chain decompositions in the Boolean lattice, *Electronic Journal of Combinatorics* **11** (2004).
12. E. Grosswald, Solution of advanced problem 6019, *American Math. Monthly* **84** (1977) 63–65.
13. Branko Grünbaum, Venn diagrams and Independent Families of Sets, this MAGAZINE **48** (1975) 12–23.
14. Branko Grünbaum, Venn diagrams I, *Geombinatorics* (Vol. I) **4** (1992) 5–12.
15. Branko Grünbaum, Venn diagrams II, *Geombinatorics* (Vol. II) **2** (1992) 25–32.
16. Branko Grünbaum and Peter Winkler, A Venn diagram of 5 triangles, this MAGAZINE **55** (1982) 311.
17. Ioan James, *Remarkable Mathematicians: From Euler to von Neumann*, Cambridge University Press, 2002.
18. Iwan Jensen and Anthony J. Guttmann, Statistics of lattice animals (polyominoes) and polygons, *Journal of Physics, Series A* **33** (2000) L257–L263.
19. David A. Klarner, Some results concerning polyominoes, *Fibonacci Quarterly* **3** (1965) 9–20.
20. Donald E. Knuth, *The Art of Computer Programming*, pre-fascicle 4A (a draft of Section 7.2.1.6: generating all trees), Addison-Wesley, 2004.
21. Stephan Mertens, Counting lattice animals: a parallel attack, *Journal of Statistical Physics* **66** (1992) 669–678.
22. George Pólya, On the number of certain lattice polygons, *Journal of Combinatorial Theory* **6** (1969) 102–105.
23. D.H. Redelmeier, Counting polyominoes: yet another attack, *Discrete Mathematics* **36** (1981) 191–203.
24. Frank Ruskey and Mark Weston, A survey of Venn diagrams, *Electronic Journal of Combinatorics* **4** (1997) DS #5 (updated 2001, 2005).
25. Neil J.A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`.
26. Mark Thompson, Venn polyominoes, *Math Recreations*, `http://www.flash.net/~markthom/html/venn_polyominos.html`.
27. John Venn, On the diagrammatic and mechanical representation of propositions and reasonsings, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **9** (1880) 1–18.

Erratum

Stanley Stephens, Anderson University, Anderson IN 46012 has reported an error in the June 2006 issue of this MAGAZINE in the article Dropping Lowest Grades, *by Daniel Kane and Jonathan Kane*, pages 181–189. In the description of the Bisection Algorithm appraoch to finding q_{best} near the bottom of page 187 the numbers q_{high} and q_{low} were inadvertently interchanged. It should read:

"If $F(q_{\text{middle}}) < 0$, we reset q_{high} to q_{middle}. Otherwise we reset q_{low} to q_{middle}."

In addition, the authors report that they have learned of two related references since the article appeared:

David Eppstein and Daniel S. Hirschberg, Choosing subsets with maximum weighted average, *J. Algorithms* **24** (1997) 177–193.
Robert W. McGrail and Tracey Baldwin McGrail, A grading dilemma or the abyss between sorting and the knapsack problem, *Journal of Computer Sciences in Colleges* **19** (2004) 97–107.

Mom! There's an Astroid in My Closet!

DEREK SEIPLE
Penn State University
DuBois campus
DuBois, PA 15801

EUGENE BOMAN
Penn State University
Harrisburg campus
Middletown, PA 17057

RICHARD BRAZIER
Penn State University
DuBois campus
DuBois, PA 15801

The knowledge of the world is only to be acquired in the world, and not in a closet.

—Lord Chesterfield, from Letters to His Son (1694–1773)

Introduction

All children know that there are mysteries, sometimes frightening mysteries, hidden in closets. Adults often brush this aside as the result of an overactive imagination. But perhaps we should take a second look. Perhaps it is *our* imaginations that are *under*active. If you have a closet (or any doorway) covered with a bifold door there is an astroid lurking just inside and the only way you can get to it is to coax it carefully with a little bit of calculus. If your door has more than one fold there are even more interesting objects waiting to be discovered.

This investigation began when one of the authors (Seiple) was standing at his closet wondering how much floor space was needed to accommodate the opening and closing of the bifold door mounted on it. He was supposed to be getting dressed for school, but he was in high school at the time so perhaps he can be forgiven. When he arrived at college he described the problem to Boman and Brazier who encouraged him to investigate the problem using the calculus tools he was learning at the time. This article is the result of his investigations.

Notice that if a closet of width r has a door mounted as in FIGURE 1 (we will call this the standard mounting) then opening (or closing) the door requires $\frac{\pi r^2}{4}$ square feet of floor space be kept clear of obstacles. Adults *might* be able to do this but it can be an onerous task for a teenager.

This would seem to be the end of the story except that a survey of your closets will quickly convince you that the standard mounting is actually relatively rare on closet doors.

Our (admittedly unscientific) survey of all of the closets we have easy access to convinces us that most of the closet doors in the United States which do not use the standard mounting use a bifold mounting which we discusss next.

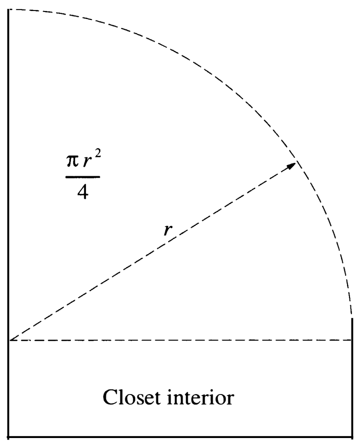


Figure 1 The floor space required for a standard door is the full quarter-circle. A bifold door requires substantially less, but how much less?

Bifold doors

Many closet doors do not have ample room for a standard mounting, therefore to save floor space closet doors are often bifold doors as shown in FIGURE 2. That is, the door is broken and hinged in the middle so that each panel is $r/2$ in length. This allows the door to be mounted in a manner similar to the standard mounting we described earlier except that only the left panel sweeps out a quarter-circle. The inner edge of the left panel of the door is fixed at the point A and the outer edge of the right panel is allowed to slide along the track.

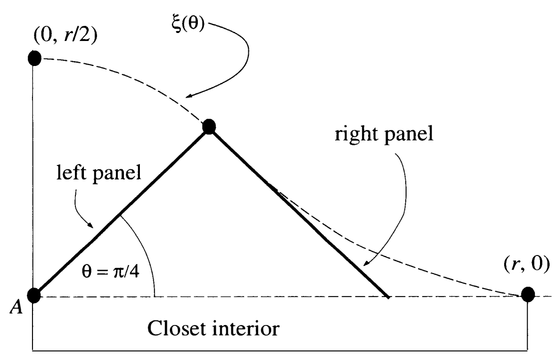


Figure 2 This figure shows the view from above a bifold door as the door closes. It first sweeps out the area under the circular arc of radius $r/2$, but when θ reaches $\pi/4$ the nature of the curve changes. Notice that "bifold" is a misnomer. There is only one fold.

To begin, consider a bifold door starting in the fully open position and closing to the right as in FIGURE 2. The floor space required to close the door is enclosed by the curve we'll call $\xi(\theta)$ and the x and y axes.

It is clear that ξ has two distinct components. The first is simply the circular arc swept out by the left panel of the door as θ proceeds from $\pi/2$ to $\pi/4$. However when $\theta = \pi/4$ the nature of ξ changes. At this point the right panel of the door is tangent to the circular arc. Thus the entire quarter-circle swept out by the left panel of the door is enclosed in the area which has already been swept out. Moreover as the door continues to close, more area outside the quarter-circle continues to be accumulated.

We seek a parameterization of the outer envelope of this area.

To that end, assume that the door is opening as in FIGURE 3 and that $0 \leq \theta \leq \pi/4$. Notice the change here. For the development we are about to present, it is easier to think of the door as opening rather than closing. In either case $\xi(\theta)$ is unchanged.

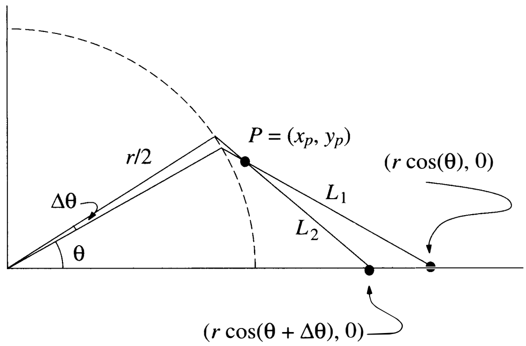


Figure 3 When θ is incremented by $\Delta\theta$ the original position of the right panel of the door and its new position will intersect. The point of intersection at P gives an approximate parameterization of the curve $\xi(\theta)$. As $\Delta\theta \to 0$ this becomes exact.

We increment θ by $\Delta\theta$ and consider the position of the right panel of the door at θ and $\theta + \Delta\theta$. If we can find the coordinates of the point P we have an approximate parameterization of the curve ξ. If $P = (x_P, y_P)$ for a fixed $\Delta\theta$ then it is clear that

$$\xi(\theta) = \lim_{\Delta\theta \to 0} (x_P, y_P)$$

is the parameterization we seek.

Let L_1 be the line of the right panel at θ and let L_2 be the line at $\theta + \Delta\theta$ and observe that the slopes of L_1 and L_2 are $-\tan\theta$ and $-\tan(\theta + \Delta\theta)$, respectively. Thus the equation of L_1 is

$$y = -\tan\theta(x - r\cos\theta) \tag{1}$$

and the equation of L_2 is

$$y = -\tan(\theta + \Delta\theta)(x - r\cos(\theta + \Delta\theta). \tag{2}$$

Combining equations 1 and 2 we get

$$x_P = r\left(\frac{\sin(\theta + \Delta\theta) - \sin\theta}{\tan(\theta + \Delta\theta) - \tan(\theta)}\right).$$

Putting this back into either L_1 or L_2 gives

$$y_P = -r\tan\theta\left(\frac{\sin(\theta + \Delta\theta) - \sin\theta}{\tan(\theta + \Delta\theta) - \tan(\theta)} - \cos\theta\right).$$

Taking the limit as $\Delta\theta \to 0$ we get

$$x = \lim_{\Delta\theta \to 0} r \left(\frac{\sin(\theta + \Delta\theta) - \sin\theta}{\tan(\theta + \Delta\theta) - \tan(\theta)} \right)$$

$$= r \lim_{\Delta\theta \to 0} \left(\frac{\frac{\sin(\theta+\Delta\theta)-\sin\theta}{\Delta\theta}}{\frac{\tan(\theta+\Delta\theta)-\tan(\theta)}{\Delta\theta}} \right).$$

Observe that the numerator and denominator of the formula above are just the derivatives of $\sin\theta$ and $\tan\theta$ respectively. Thus

$$x(\theta) = r\frac{\cos\theta}{\sec^2\theta} = r\cos^3\theta.$$

Similarly

$$y(\theta) = r\sin^3\theta.$$

Thus a parameterization for the outer envelope of the floor space used by a bifold door is given by:

$$\xi(\theta) = \begin{cases} \begin{pmatrix} r\cos^3\theta \\ r\sin^3\theta \end{pmatrix} & \text{if } 0 \le \theta \le \pi/4 \\ \begin{pmatrix} r/2\cos\theta \\ r/2\sin\theta \end{pmatrix} & \text{if } \pi/4 \le \theta \le \pi/2 \end{cases} \tag{3}$$

Notice that letting θ move from 0 to $\pi/2$ opens the door while letting θ move from $\pi/2$ to 0 closes it. To compute the area of the floor space required we need to ensure that we integrate from left to right. The area of the floor space is then given by

$$\begin{aligned} \int_{\theta=\pi/2}^{\theta=0} y(\theta)\,\mathrm{d}x &= \int_{\theta=\pi/4}^{\theta=0} r\sin^3(\theta)\frac{\mathrm{d}x}{\mathrm{d}\theta}\,\mathrm{d}\theta + \int_{\theta=\pi/2}^{\theta=\pi/4} r/2\sin\theta\frac{\mathrm{d}x}{\mathrm{d}\theta}\,\mathrm{d}\theta \\ &= 3r^2\int_0^{\pi/4}\sin^4\theta\cos^2\theta\,\mathrm{d}\theta + r^2/4\int_{\pi/4}^{\pi/2}\sin^2\theta\,\mathrm{d}\theta \\ &= \frac{5\pi r^2}{64}. \end{aligned}$$

So our initial question is resolved. If a closet r feet wide is covered by a bifold door $\frac{5\pi r^2}{64}$ square feet of floor space is required to accommodate the door. If the same closet is closed with an ordinary door then $\pi r^2/4$ square feet are needed—a savings of nearly 70%.

Adding door panels

It is clear that adding 2, 3, 4, or n folds will reduce the floor space required even further. FIGURE 4 shows the situation with 2 folds. If the doors are hinged so that the angles denoted by θ in FIGURE 4 are always equal then the problem can be approached in the same manner as before as we now show.

As before we perturb θ by $\Delta\theta$ and consider the point of intersection of the rightmost panels in FIGURE 4. In that case the equation of L_1 is (again):

$$y = -\tan\theta(x - r\cos\theta)$$

and the equation of L_2 is (again):

$$y = -\tan(\theta + \Delta\theta)(x - r\cos(\theta + \Delta\theta)).$$

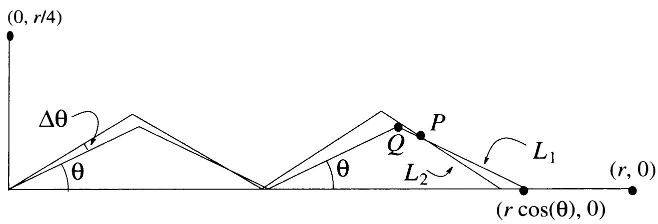


Figure 4 Each door panel has length $r/4$.

Since these are *exactly* the same equations we found in the previous section it follows that the parameterization we seek is (again) the astroid:

$$\begin{pmatrix} r\cos^3\theta \\ r\sin^3\theta \end{pmatrix}.$$

Indeed it should be clear from the above that adding more hinges has no effect on the astroidal portion of the curve. The very same astroid appears regardless of the number of folds in the door as long as all of the panels are hinged so that they make the same angle with the front of the closet (the angle θ in FIGURE 4). This assumption is critical. If the angles are allowed to differ the problem becomes considerably more complex.

Recall however that the curve $\xi(\theta)$ from the previous section had two components. The other portion was the circular arc traced out by the point corresponding to Q in FIGURE 4. To find the corresponding portion for the current curve, which we'll denote by $\xi_2(\theta)$, we need to parameterize the coordinates of the point Q.

Referring again to FIGURE 4 it is clear that

$$Q(\theta) = \begin{pmatrix} 3r/4\cos\theta \\ r/4\sin\theta \end{pmatrix}$$

and that the transition between the components of the curve occurs when $P(\theta) = Q(\theta)$, or when $\theta = \pi/6$. Thus when we have two folds in our door the curve $\xi_2(\theta)$ is

$$\xi_2(\theta) = \begin{cases} \begin{pmatrix} r\cos^3\theta \\ r\sin^3\theta \end{pmatrix}, & \text{if } 0 \le \theta \le \pi/6 \\ \begin{pmatrix} 3r/4\cos\theta \\ r/4\sin\theta \end{pmatrix}, & \text{if } \pi/6 \le \theta \le \pi/2 \end{cases}$$

and in the general case, with n folds, the curve is

$$\xi_n(\theta) = \begin{cases} \begin{pmatrix} r\cos^3\theta \\ r\sin^3\theta \end{pmatrix}, & \text{if } 0 \le \theta \le \cos^{-1}\left(\frac{2n-1}{2n}\right) \\ \begin{pmatrix} \frac{(2n-1)r}{2n}\cos\theta \\ \frac{r}{2n}\sin\theta \end{pmatrix}, & \text{if } \cos^{-1}\left(\frac{2n-1}{2n}\right) \le \theta \le \pi/2 \end{cases}.$$

It seems curious that the same curves, an ellipse and an astroid, appear regardless of how many panels we split our door into. Indeed, the same curves appear in the two-fold (four panel) case even if the panels are of two distinct sizes.

By now the calculation is very familiar so we will not belabor it. Consider the arrangement depicted in FIGURE 5. Again we have two folds (four panels) but they are no longer the same length and we have normalized the sum of the lengths of the panels to 1. If we perturb the angle θ by $\Delta\theta$ and find the intersection point P (not shown in the figure) between the original location of the rightmost panel and its perturbed location we find that the equations of L_1 and L_2 are again precisely the same as in our first problem. Thus the astroid emerges exactly as before. Moreover it is easy to show that a parameterization of the point Q in the figure is:

$$Q(\theta) = \begin{pmatrix} (1/2+\alpha)\cos\theta \\ (1/2-\alpha)\sin\theta \end{pmatrix}$$

which is again an ellipse.

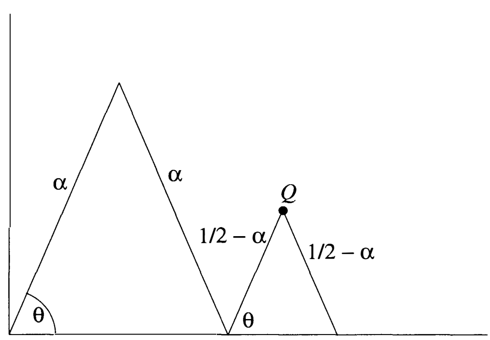


Figure 5 A bifold door with different length panels also generates an astroid and ellipse.

It seems very odd that the same curves keep emerging no matter how we try to generalize the problem.

Wiles' light switch

Andrew Wiles [**5**] has likened mathematics research to walking into an unlighted room. At first all is dark. As you fumble around you begin to get a sense of the location of the objects in the room and the relationships between them. Eventually, if you are lucky, you find the light switch and flip it. Then you see all of the structures and the relationships between them that you were already familiar with as well as new ones that you were only dimly aware of or may not have known at all.

In this section we will flip the light switch for this problem. It turns out that the relationship between ellipses and astroids, which has been our common theme, has been known since antiquity. Archimedes used it to create a mechanical device, known as the "Trammel of Archimedes," for drawing ellipses (see [**1, 2, 3**]).

Consider the following alternate construction of the astroid [**4, 6**]. We begin with vertical line segment whose endpoints are at (0, 1) and the origin (see FIGURE 6). Keeping the length of the segment constant we move the endpoints vertically toward the origin and horizontally toward (1, 0), respectively. The outer envelope of the region thus constructed is the astroid.

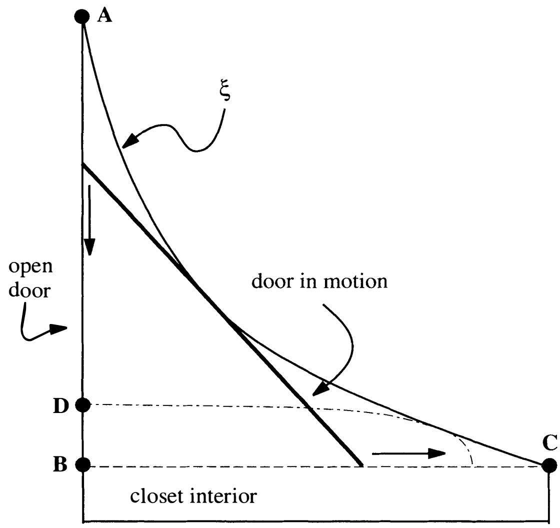


Figure 6 As the door closes points A and B move toward points B and C, respectively.

Rather, it is one quarter of the classical astroid of antiquity. This is the portion we have seen so far. If we continue in the same fashion—moving the left end of our line segment vertically to $(0, -1)$ and the right end horizontally back to the origin—we will generate the same curve reflected about the x-axis. Continuing in the same vein until our line segment has returned to its original position gives the full astroid of antiquity as it was known to Archimedes. This is shown in FIGURE 7.

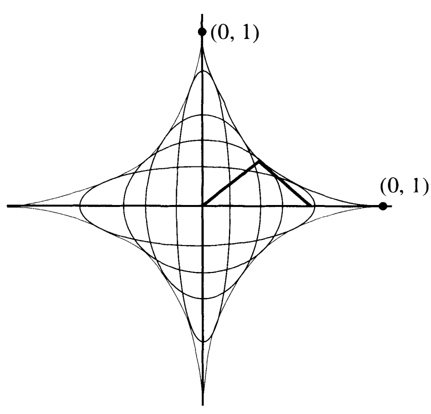


Figure 7 One way to define the astroid is as the outer envelope of a particular set of ellipses as shown here. A bifold door is shown schematically in the first quadrant.

FIGURE 7 also displays the astroid as the outer envelope of a particular set of ellipses. If we fix a point D on our original line segment (see FIGURE 6) between (0, 1) and the origin and follow it as we trace out the astroid then it is easy to show that the path it follows is an ellipse.

It seems that as our folding doors close the peak of the rightmost fold (the point Q seen in figures 4 and 5) traces out one of these ellipses (the elliptical portion of the curve $\xi_n(\theta)$) until it touches the astroid. At that point the rightmost panel of the door is tangent to both the ellipse and the astroid and our $\xi_n(\theta)$ switches modes and begins to follow the astroid.

You can't have a light without a dark to stick it in.

—Arlo Guthrie (1947–)

REFERENCES

1. http://personal.atl.bellsouth.net/e/d/edwin222/ellipsedevice.htm.
2. Xah Lee. http://www.xahlee.org/specialplanecurves_dir/astroid_dir/astroid.html.
3. Xah Lee. http://www.xahlee.org/specialplanecurves_dir/ellipse_dir/ellipse.html.
4. E. H. Lockwood, *A Book of Curves*, Cambridge University Press, 1963.
5. Andrew Wiles, The Proof, Nova, Adventures in Science, 1997, Produced by John Lynch, A co-production of BBC TV and WGBH Boston, ISBN: 1-57807-072-4.
6. Robert C. Yates, *Curves and Their Properties*, National Council of Teachers of Mathematics, 1952.

Primes and Probability: The Hawkins Random Sieve

JOHN LORCH
Ball State University
Muncie, IN 47306
jlorch@bsu.edu

GIRAY ÖKTEN
Florida State University
Tallahassee, FL 32306
okten@math.fsu.edu

While prime numbers are the fundamental building blocks of the integers, understanding how they are spread within the integers has turned out to be hard work. For example, the Prime Number Theorem stood as a conjecture for nearly a hundred years, and anyone who bags the Riemann Hypothesis first will be a million dollars richer. Modern cryptography assumes that the primes will retain their secrets for some time to come.

In the presence of hard problems, it is tempting to employ models. A good model should provide an approximation of reality which is simple enough to understand, yet accurate enough to be useful. While there are several models for the primes, in this paper we tell the story of a beautiful and compelling probabilistic model known as the *Hawkins primes*. First introduced by David Hawkins in this MAGAZINE [**11**], the model is based on a simple variation of the sieve of Eratosthenes. Over the past fifty years, the Hawkins model has been used to predict the truth, in the strongest probabilistic sense, of results (both established and conjectured) concerning the distribution of the prime numbers, including the Twin Primes Conjecture and the Riemann Hypothesis. Also, the model (or generalizations thereof) has potential to shed light on interesting integer sequences other than the primes.

The Hawkins model: Fishing for primes

In addition to computing the circumference of the earth, Eratosthenes of Cyrene (276–194 BCE) taught us to sieve primes. To employ his sieve, recall that we start with all natural numbers two and larger, and identify $p_1 = 2$ as our first 'sieving number.' We sieve (remove) from our list all higher multiples of p_1, and identify p_2 as the smallest surviving number larger than p_1 (of course $p_2 = 3$). In the second step, we sieve from our remaining list all higher multiples of the sieving number p_2, and identify p_3 as the smallest surviving number which is larger than p_2. Continuing the process, we produce a list $P = \{p_1, p_2, \ldots\}$ of sieving numbers which are precisely the primes.

To produce a list of Hawkins primes, we *randomize* the sieve of Eratosthenes. As before, we start with all natural numbers two and larger, and identify $h_1 = 2$ as our first 'sieving number.' In the first step we independently sieve numbers from our list with probability $1/h_1$, and identify h_2 as the smallest surviving number greater than h_1. In the second step, we sieve numbers from our remaining list with probability $1/h_2$, and identify h_3 as the smallest surviving number which is larger than h_2. If we carry on with the process, we produce a list $H = \{h_1, h_2, \ldots\}$ of sieving numbers which we call a set of *Hawkins primes*. (A computer program producing lists of Hawkins primes may be desirable for simulation purposes, or just for fun. See www.cs.bsu.edu/homepages/jdlorch/preps.htm for an example in *Mathematica*.)

One of the first things we observe about the Hawkins random sieve (as opposed to the sieve of Eratosthenes) is that it is *non-deterministic*. That is, each time we run the random sieve, we almost surely obtain a different list H of Hawkins primes. In fact, every strictly increasing sequence of natural numbers beginning with 2 is a possible (though not necessarily likely) list of Hawkins primes. At first this non-determinism seems exceedingly complex, standing in direct opposition to the principle that a model should be simpler than the object it represents. After all, we have traded in our problematic 'tree' of real primes for a whole 'forest' of Hawkins primes. However, upon closer inspection we see that the Hawkins model has significant virtues. Importantly, the whole notion of divisibility, the testing of which is complex and time-consuming, is swept aside in favor of random elimination. Further, the large number of possible lists of Hawkins primes is advantageous from a probabilistic standpoint. Here's the idea: Imagine the set of 'real' primes to be an individual fish among a school of fish. The primes are too smart to be caught with hook and line, but if we net nearly all of the fish, then there is a good chance we have netted the primes as well. The moral is that any conclusion one can draw about nearly all sets of Hawkins primes will likely be true for the 'real' primes. We will see later that this approach has been fruitful.

Is the model reasonable?

For the Hawkins primes to be a believable model, it seems reasonable to expect them to resemble the 'real' primes fundamentally. Here we focus on two important ways in which the Hawkins primes are similar to the real primes.

Density. The *Prime Number Theorem* (1896, Hadamard and Valée Poussin) is perhaps the best-known established result concerning the distribution of primes. It says that the number of primes less than or equal to n (denoted $\pi(n)$) is approximately $n/\log n$, in the sense that $\pi(n)[n^{-1}\log n] \to 1$ as $n \to \infty$. The weakness of this approximation theorem says a lot about the difficulty of prime distribution problems.

Locally, the Prime Number Theorem implies that for large integers n, the probability that n is prime is approximately $1/\log n$. Hawkins **[11]** showed that an analogous local result holds for Hawkins primes: Letting S_n denote the event that a natural number $n > 2$ is a Hawkins prime, and $P(S_n)$ the probability that S_n occurs, we have

THEOREM 1. *$P(S_n)\log n \to 1$ as $n \to \infty$.*

Therefore, the Hawkins primes have (more or less) the same density in the natural numbers as the 'real' primes.

A proof of Theorem 1, which is very tersely sketched in **[11]**, requires only elementary calculus and probability. We first establish the recursion

$$P(S_{n+1}) = P(S_n) - \frac{P(S_n)^2}{n}, \tag{1}$$

which, with T_n denoting the event complementary to S_n, is equivalent to

$$\begin{aligned} P(T_{n+1}) &= P(T_n) + \frac{P(S_n)^2}{n} \\ &= P(T_n)(P(T_n) + P(S_n)) + \frac{P(S_n)^2}{n}. \end{aligned} \tag{2}$$

To obtain (2) we observe

$$P(T_{n+1}) = P(T_{n+1}|T_n)P(T_n) + P(T_{n+1}|S_n)P(S_n), \tag{3}$$

and use elementary facts about conditional probability together with the definition of the sieve to show

$$P(T_{n+1}|T_n) = P(T_n) \qquad \text{and} \qquad P(T_{n+1}|S_n) = P(T_n) + \frac{P(S_n)}{n}. \tag{4}$$

The results of (3) and (4) together imply (2).

Next, if we put $g_n = 1/P(S_n)$, then using (1) we obtain

$$\begin{aligned} g_{n+1} &= \frac{g_n}{1-\frac{1}{ng_n}} = \frac{n^2g_n^2}{n^2g_n - n} = \frac{n^2g_n^2 - ng_n + ng_n - 1 + 1}{n^2g_n - n} \\ &= g_n + \frac{1}{n} + \frac{1}{n^2g_n - n}. \end{aligned}$$

Therefore,

$$g_n = g_2 + \sum_{k=2}^{n-1}(g_{k+1} - g_k) = g_2 + \sum_{k=2}^{n-1}\frac{1}{k} + \sum_{k=2}^{n-1}\frac{1}{k^2g_k - k}. \tag{5}$$

Since the righthand-most sum in (5) converges as $n \to \infty$, from (5) we have

$$\lim_{n\to\infty} P(S_n)\log n = \lim_{n\to\infty} g_n^{-1}\log n = \lim_{n\to\infty} g_n^{-1}\left[\sum_{k\le n}\frac{1}{k}\right] = 1,$$

concluding the proof.

Interdependence. In 1937, Cramér [**4**] introduced a simple (and today widely known) model for the primes which assumes the local version of the Prime Number Theorem mentioned above. Consider an infinite collection of urns $U_2, U_3, \ldots$ containing black and white balls, such that a random draw from U_n will result in a white ball with probability $1/\log n$. Upon drawing balls, the numbers of those urns from which a white ball was selected form a sequence of *Cramér primes*.

Observe that in Cramér's model, the probability of selecting a white or black ball from U_n does not depend on the selections made earlier. However, this assumption of independence does not appear to be consistent with the behavior of the primes. For example, for the event "n is a prime number" the number of prime numbers less than $\sqrt{n}$ may affect the probability of this event. (We can argue that if there are more than the expected composite numbers between 2 and $\sqrt{n}$, then the probability that "n is a prime number" should be larger, since there are fewer potential prime divisors.) Hawkins' model takes a first step in incorporating the *interdependency* of prime numbers. Since each sieving number n eliminates subsequent numbers with probability $1/n$, clearly the probability that an integer m is a sieving number (that is, a Hawkins prime) depends on the sieving numbers less than m.

Fishing revisited: A sampler of applications

Here we present a short chronological sample of some ways in which the Hawkins model has been used to investigate the primes and other interesting sequences. While no details are provided, perhaps it suffices to say that in almost every instance the game plan is to use limit theorems from probability (along the lines of the Strong Law of Large Numbers) to make statements that will be true in the strongest probabilistic sense for sets of Hawkins primes. (Our collection of applications is not comprehensive. For example, we omit the connections with diffusion and Brownian motion addressed in [**20**] and [**6**], respectively.)

For later use we declare p_n to be the nth 'real' prime and h_n to be the random variable giving the nth Hawkins prime. For real-valued functions f, g on $\mathbb{R}$, we write $f \sim g$ whenever $f(x)/g(x) \to 1$ as $x \to \infty$. (We say f is *asymptotic* to g.) Finally, an event occurring with probability 1 is said to occur *almost surely*. (We let *a.s.* stand for almost surely. Recall that an event occurring with probability 1 is *not* necessarily certain to occur!)

Lucky numbers. Perhaps surprisingly, one of the earliest applications of the Hawkins model was not to the primes, but rather to the *lucky numbers*. First introduced in 1956 by Ulam et al [**8**], the lucky numbers are given by a sieving process similar to the sieve of Eratosthenes. One begins with the sequence $L_2 = 2, 3, 5, 7, 9, \ldots,$ and, given that L_n has been defined and that $t_{m,n}$ is the nth element of L_m, we obtain L_{n+1} by crossing out every $t_{n,n}$th element from L_n. The lucky numbers $L = 2, 3, 7, 9, 13, 15, 21, \ldots$ are obtained from L_n by letting n tend to infinity.

Based on Theorem 1, Hawkins conjectured that

$$l_n \sim n \log n \qquad \text{and} \qquad \sigma_n \sim \frac{1}{\log n}$$

where l_n denotes the n-th lucky number and $\sigma_n = (1 - 1/2)(1 - 1/3) \ldots (1 - 1/l_n)$. These lucky number analogues of the Prime Number Theorem and Mertens' Theorem, respectively, were later proved by Hawkins and Briggs [**3**] (see below for more on Mertens' Theorem).

Strong probabilistic versions of PNT and Mertens' Theorem. If $\pi_H(n)$ denotes the random variable counting the number of Hawkins primes less than n in a set of Hawkins primes, then it follows immediately from Theorem 1 that the expected value of $\pi_H(n)$ is asymptotic to $n/\log n$. However, in the 1970's Nuedecker and Williams [**18**], as well as Wunderlich ([**21**], [**22**]), were able to obtain much stronger probabilistic results:

THEOREM 2. $\pi_H(n) \sim n/\log n$ a.s. *and* $\prod_{k \le n}(1 - 1/h_k)^{-1} \sim \log n$ a.s.

The first result of Theorem 2 is a strong probabilistic version of the (global) Prime Number Theorem, while the second result is an imperfect analogue of *Mertens' Theorem.* Understanding Mertens' Theorem begins with the fact (due to the Fundamental Theorem of Arithmetic) that

$$\sum_{m=1}^{\infty} \frac{1}{m^s} = \prod_{k=1}^{\infty}(1 - 1/p_k^s)^{-1} \qquad \text{for } s > 1. \tag{6}$$

The key question is: What analog of (6) holds at $s = 1$? In light of (6), we are tempted to compare the growth of the partial products in the right side of (6) to the partial sums of the harmonic series. Since $\log n \sim \sum_{m \le n} \frac{1}{m}$, a first guess might be

$$\prod_{k \le n}(1 - 1/p_k)^{-1} \sim \log n,$$

and this is precisely what is predicted by the Hawkins model in Theorem 2. Unfortunately, this prediction is off by a factor of e^γ, where γ is Euler's constant. The reason for the discrepancy is that the Hawkins model is blind to divisibility, and it is this special nature of the primes which gives rise to the tantalizing constant e^γ in the *correct* version of Mertens' Theorem:

$$\prod_{k\leq n}(1-1/p_k)^{-1} \sim e^\gamma \log n.$$

(See Section 1.8 of **[7]** for a proof of Mertens' Theorem.)

The Riemann Hypothesis is true! Well, almost surely ... It is natural to wonder what the Hawkins model has to say about open problems on the primes. Among these open problems, there is none more famous than the *Riemann Hypothesis.* Using tables, Gauss observed that the logarithmic integral function $Li(n) = \int_2^n \frac{1}{\log t}\,dt$ does a fairly good job of counting primes. (For us this shouldn't be too surprising; we've seen that the density function for the primes is approximately $1/\log n$. Integration by parts indicates $Li(n)$ has $n/\log n$ as a leading term, so the Prime Number Theorem can be rephrased as $\pi(n) \sim Li(n)$.) If the logarithmic integral $Li(x)$ did a perfect job of counting primes, then $Li(p_n)$ would be n on the nose. However, $Li(x)$ is merely an approximation, and the Riemann Hypothesis (first formulated by Riemann in 1859) is principally a conjecture about the growth of the associated error term $Li(p_n) - n$. It states

$$Li(p_n) = n + O(n^{1/2+\epsilon}) \qquad \text{for every } \epsilon > 0.$$

In the 1970's and 80's, several strong probabilistic results cropped up in favor of the Riemann Hypothesis. Williams **[18]** showed that the Riemann Hypothesis is true almost surely for the Hawkins primes, and further improvements on the error term have been given by Heyde (**[13]**, **[14]**) and Deheuvels **[5]**. Van der Poorten **[19]** finds this to be the most compelling evidence in favor of the Riemann Hypothesis.

Our investigation of open problems need not end with the Riemann Hypothesis. For example, Gauss and Riemann believed that $\pi(n) < Li(n)$ for all n, so it was a surprise when Littlewood showed that $\pi(n) - Li(n)$ must change sign infinitely often. It is unknown exactly where $\pi(n) - Li(n)$ first changes sign, although Bays and Hudson **[2]** give an upper bound of $n \approx 1.39 \times 10^{316}$. Nor is the original Hawkins model of much help: Ingredients in the proof of Theorem 1 imply $P(S_n) < 1/\log n$ for all n, so the model incorrectly predicts that $\pi(n) \leq Li(n)$. However, conjectures about the sign of $\pi(n) - Li(n)$ can be made using the *square-root model.* First introduced by Hawkins in **[12]**, the square root model consists of the original Hawkins model together with the added condition that a sieving number n only strike out numbers larger than n^2. Overall, this has the effect of raising the probability of primality. In 1998, Mahajan **[16]** used the theory of delay differential equations to show that the probability that n is 'prime' in the square-root model oscillates around $1/\log n$, and, using leading terms of her solutions, conjectured that $\pi(n) - Li(n)$ first changes sign around $n = 10^{27}$.

A generalized Hawkins sieve and prime k-tuples

Mahajan's work on the sign of $\pi(n) - Li(n)$ shows that making alterations to Hawkins' original model can be a smart thing to do. In this section, we show that certain adjustments to the model yield results bearing resemblance to those for the original model, and provide the opportunity to investigate various interesting sequences.

Expanding your probabilities: A generalized sieve. In Hawkins' original model, a sieving number n (i.e., a Hawkins prime) sieves subsequent numbers with probability $\mathfrak{p}(n) = 1/n$. A simple way to generalize the model is to allow $\mathfrak{p}(n)$ to vary. Sequences generated by $\mathfrak{p}(n)$ will be called *Hawkins* $\mathfrak{p}$*-primes*. Given certain conditions on the decay of $\mathfrak{p}(n)$ (see [**15**]), several asymptotic results hold for the general model which are reminiscent of those for the original. For instance, letting S_n denote the event that a natural number $n > 2$ is a Hawkins $\mathfrak{p}$-prime, and $P(S_n)$ the probability that S_n occurs, we have the following density result parallel to Theorem 1:

THEOREM 3. *If* $\sum \mathfrak{p}^2(k)$ *converges while* $\sum \mathfrak{p}(k)$ *diverges, then*

$$P(S_n) \sim \left(\sum_{k\le n} \mathfrak{p}(k)\right)^{-1}.$$

With further conditions on the decay of $\mathfrak{p}(n)$ and quite a bit more effort, [**15**] provides strong probabilistic analogs of the Prime Number Theorem and Mertens' Theorem paralleling Theorem 2. Specifically, if we let H_n denote the nth Hawkins $\mathfrak{p}$-prime, $Y_n = \prod_{k\le n}(1 - \mathfrak{p}(H_k))^{-1}$, and $I(t) = \int_2^t \mathfrak{p}(t)\,dt$, then we have

THEOREM 4. $\frac{1}{n} H_n \sim Y_n \sim I(H_n)$ a.s. *provided that* $\mathfrak{p}(t)$ *decays 'appropriately.'*

Here, the asymptotic equivalence $n^{-1}H_n \sim I(H_n)$ is an analogue of the Prime Number Theorem, while $Y_n \sim I(H_n)$ is an analogue of Mertens' Theorem. An application of Theorems 3 and 4 is given below.

Application to prime k-tuples. A pair p_k, p_{k+1} of consecutive primes with $p_{k+1} - p_k = 2$ forms a set of *twin primes*. The *Twin Primes Conjecture* asserts the infinitude of the twins along with the asymptotic formula

$$t(n) \sim C\frac{n}{\log^2 n}, \tag{7}$$

where $t(n)$ is the number of sets of twins less than n, and C is a constant.

The asymptotic formula (7) is commonly motivated using Cramér's model: Natural numbers n and $n+2$ are both prime with probability

$$\frac{1}{\log n \cdot \log(n+2)} \approx (\log n)^{-2}.$$

Compiling these probabilities we guess that $t(n) \approx n \log^{-2} n$. Hardy and Littlewood [**10**] have conjectured that the constant

$$C = 2\prod_{p>2} \frac{p(p-2)}{(p-1)^2},$$

and Neudecker [**17**] showed that the Twin Primes Conjecture (with $C = 1$) is true almost surely for the Hawkins primes.

There is a generalization of the Twin Primes Conjecture called the *k-Tuple Conjecture.* Given natural numbers $0 < a_1 < a_2 \cdots < a_{k-1}$, the k-Tuple Conjecture asserts that the number of primes $p \leq n$ such that each of $p, p+2a_1, \ldots, p+2a_{k-1}$ is prime approaches

$$\frac{C_{a_1,\ldots,a_{k-1}} n}{\log^k n}$$

asymptotically, where $C_{a_1,\ldots,a_{k-1}}$ is a constant depending on $a_1, \ldots, a_{k-1}$, and where we assume that there are no divisibility conditions preventing all of $p, p+2a_1, \ldots, p+2a_{k-1}$ from being prime infinitely often. (For details, including specific values for $C_{a_1,\ldots,a_{k-1}}$, see **[9]**.)

We can use the generalized Hawkins sieve to model k-tuples. If in Theorem 3 we set $\mathfrak{p}(n) = n^{-1}\log^k n$ for some fixed positive integer k, then

$$P(S_n) = \frac{(k+1)}{\log^{k+1} n}.$$

This is (more or less) the conjectured asymptotic density of sets of the prime $(k+1)$-tuples. So, the probabilities $\mathfrak{p}(n) = n^{-1}\log^k n$ provide us with a reasonable new sieving model for prime $(k+1)$-tuples. Further, by applying Theorem 4 with these sieving probabilities, we obtain

$$H_n \sim \frac{n}{k+1}\log^{k+1} n \text{ a.s.} \qquad \text{and} \qquad Y_n \sim \frac{1}{k+1}\log^{k+1} n \text{ a.s.,} \tag{8}$$

which are strong probabilistic analogs of the Prime Number Theorem and Mertens' Theorem, respectively. The asymptotic formula for H_n in (8) translates to a strong probabilistic version of the k-Tuple Conjecture.

REFERENCES

1. T. Apostol, *An Introduction to Analytic Number Theory*, Springer-Verlag, 1976.
2. C. Bays and R. Hudson, A new bound for the smallest x with $\pi(x) > \mathrm{li}(x)$, *Math. Comp.* **69** (2000) 1285–1296.
3. W. Briggs and D. Hawkins, *The lucky number theorem*, this MAGAZINE **31** (1957) 277–280.
4. H. Cramér, On the order of magnitude of the difference between consecutive prime numbers, *Acta Arith.* **2** (1937) 23–46.
5. P. Deheuvels, *Asymptotic results for the psuedo-prime sequence generated by Hawkins' random sieve: Twin primes and Riemann's hypothesis*, Proceedings of the seventh conference on probability theory (Brasov, 1982), 109–115, VNU Sci. Press, Utrecht, 1985.
6. D. Foster and D. Williams, The Hawkins sieve and Brownian motion, *Compositio Math.* **37** (1978) 277–289.
7. M. France and G. Tenenbaum, *The Prime Numbers and Their Distribution*, American Mathematical Society, 2001.
8. V. Gardiner, R. Lazarus, M. Metropolis, and S. Ulam, *On certain sequences of integers defined by sieves*, this MAGAZINE **29** (1956) 117–122.
9. H. Halberstam and H. Richert, *Sieve Methods*, Academic Press, 1974.
10. G. Hardy and J. E. Littlewood, Partitio numerorum III: On the expression of a number as a sum of primes, *Acta Math.* **44** (1923) 1–70.
11. D. Hawkins, *The random sieve*, this MAGAZINE **31** (1957) 1–3.
12. D. Hawkins, Random sieves, II, *J. Number Theory* **6** (1974) 192–200.
13. C. Heyde, On asymptotic behavior for the Hawkins random sieve, *Proc. Amer. Math. Soc.* **56** (1976) 277–280.
14. C. Heyde, A log log improvement to the Riemann hypothesis for the Hawkins random sieve, *Ann. Probab.* **6** (1978) 870–875.
15. J. Lorch, A generalized Hawkins sieve and prime k-tuplets, *Rocky Mountain J. Math.*, to appear. (See `www.cs.bsu.edu/homepages/jdlorch/lorchrmj.pdf`)
16. S. Mahajan, *Order of Magnitude Physics: A Textbook with Applications to the Retinal Rod and to the Density of the Prime Numbers*, Ph.D. thesis, California Institute of Technology, 1998.

17. W. Neudecker, On twin 'primes' and gaps between successive 'primes' for the Hawkins random sieve, *Math. Proc. Camb. Phil. Soc.* **77** (1975) 365–367.
18. W. Neudecker and D. Williams, The 'Riemann hypothesis' for the Hawkins random sieve, *Composito Math.* **29** (1974) 197–200.
19. A. van der Poorten, *Notes on Fermat's Last Theorem*, Canadian Mathematical Society, 1996.
20. D. Williams, A study of a diffusion process motivated by the sieve of Eratosthenes, *Bull. London Math. Soc.* **6** (1974) 192–200.
21. M. Wunderlich, A probabilistic setting for prime number theory, *Acta Arith.* **26** (1974) 59–81.
22. M. Wunderlich, The prime number theorem for random sequences, *J. Number Theory* **8** (1976) 369–371.

NOTES

For All Possible Distances Look to the Permutohedron

JOE SANTMYER
Bethany College
Bethany, WV 26032
jsantmyer@bethanywv.edu

A *permutohedron* is a geometric object whose vertices are the $n!$ points in R^n consisting of $(1, 2, 3, \dots, n)$ and all points obtained by permuting the coordinates of this point. Consider all possible Euclidean distances between the vertices of a permutohedron. Now square each distance. It is rather easy to see that the square of any distance is *even*. What is surprising is that for $n \geq 4$, the squared distances constitute *every* even integer up through the maximum possible value.

Polytope infatuation

A permutahedron is an example of a convex polytope. Convex polytopes have been studied since ancient times [**6**]. Plato considered the five regular polytopes, tetrahedron, cube, octahedron, icosahedron and dodecahedron as ingredients from which the world was constructed [**5**]. Later in this century, physicists made a similar conjecture when considering the symmetry in equations of quantum mechanics [**8**]. Convex polytopes have provided an area where combinatorics, geometry, topology and optimization interact [**4**]. It is not surprising that permutahedra were among the topics of one of the talks given in honor of Richard Stanley's 60th birthday [**7**]. Convex polytopes continue to attract our curiosity [**3**].

Geometry of a permutohedron

Consider the permutohedron whose vertex set is $P_3 = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$, viewed as points in R^3. Is there anything geometrically interesting about them? After a little thought, it is easy to see that they all lie on the plane $x + y + z = 6$. In fact, they lie on the intersection of this plane and the sphere $x^2 + y^2 + z^2 = 14$, that is, they lie on a circle in 3-space. Moreover, the points in P_3 are the vertices of a regular hexagon inscribed in the circle.

As just observed, the six points in P_3 resulting from the six permutations of 1, 2 and 3 provide some interesting geometric properties. The points in P_3 are the vertices of a regular hexagon inscribed in a circle, a 2D object, residing in 3-space. Consider the analogous situation for P_4, the set of 4-tuples corresponding to the 24 permutations of 1, 2, 3 and 4.

FIGURE 1 shows the resulting geometric object, referred to as a *permutohedron* in Berge [**2**, p. 136]. A permutohedron is inscribed in a sphere, whose vertices are

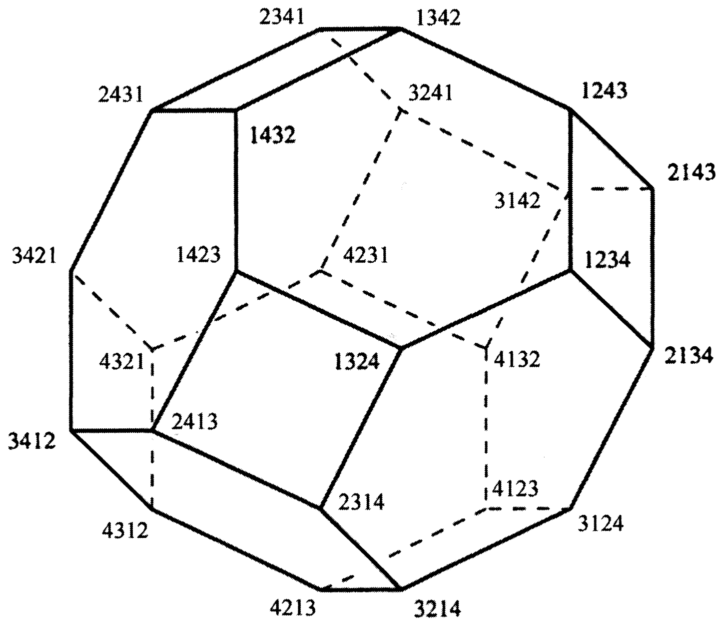


Figure 1 Permutohedron with $4! = 24$ vertices

4-tuples. A permutohedron can therefore be considered as a 3D object residing in 4-space. The vertices of the permutohedron belong to the intersection of the hyperplane $x + y + z + w = 10$ and the 4-sphere $x^2 + y^2 + z^2 + w^2 = 30$, that is, they lie on a sphere residing in 4-space.

In general, if P_n is the set of points in n-space corresponding to the $n!$ permutations of $1, 2, 3, \ldots, n$, they form the vertices of a permutohedron and they lie on the intersection of the hyperplane

$$x_1 + x_2 + x_3 + \cdots + x_n = \frac{n(n+1)}{2}$$

and the n-sphere

$$x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2 = \frac{n(n+1)(2n+1)}{6}.$$

That is, they lie on a $n - 1$ sphere residing in n-space.

Neighborhoods

Let $u = (u_1, u_2, u_3, \ldots, u_n)$ and $v = (v_1, v_2, v_3, \ldots, v_n)$ be two vertices in P_n. The Euclidean distance $d(u, v)$ between u and v is defined in the usual way as $d(u, v) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2}$. If $\epsilon \geq 0$ a ϵ-neighborhood $N(u, \epsilon)$ of vertex u is the set of vertices $N(u, \epsilon) = \{v \in P_n \mid d(u, v) \leq \epsilon\}$. The vertices of P_4 in FIGURE 1 were labeled so that the more distant one is from a given vertex, the larger the ϵ-neighborhood. Examples of ϵ-neighborhoods include $N(1324, \epsilon) = \{1324\}$ for $0 \leq \epsilon < \sqrt{2}$, $N(1324, \epsilon) = \{1324, 1234, 2314, 1423\}$ for $\sqrt{2} \leq \epsilon < \sqrt{4}$, $N(1324, \epsilon) =$

$\{1324, 1234, 2314, 1423, 2413\}$ for $\sqrt{4} \le \epsilon < \sqrt{6}$,

$$N(1324, \epsilon) = \{1324, 1234, 2314, 1423, 2413, 1243, 2134, 3214, 1432\}$$

for $\sqrt{6} \le \epsilon < \sqrt{8}$, etc. The expressions $\sqrt{4}$ and $\sqrt{8}$ were purposely not simplified to 2 and $2\sqrt{2}$ to reveal a pattern.

Now imagine standing on any vertex u of a permutohedron, like the one in FIGURE 1. Looking around at neighboring vertices within an ϵ-neighborhood the view would be the same. There would be $n - 1$ "adjacent" vertices (here vertex v is adjacent to u if $d(u, v) = \sqrt{2}$) evenly distributed around the observation vertex. Moving to another observation vertex yields the same view. The number of vertices and their relative positions within an ϵ-neighborhood looks exactly the same. The fact that the ϵ-neighborhood is *independent* of the observation vertex is part of exercise 14 in [**1**, p. 25]. This exercise provided a starting point for examining and counting distances between vertices in a permutohedron. However, rather than using the metric in exercise 14 the more familiar Euclidean metric was used.

Squared distances must be *even*

Before showing that the squared distances between the vertices of a permutohedron constitute *all* even integers up to a maximum, you are invited to convince yourself why the values are, in fact, even. The following two facts, whose verifications are left for your enjoyment, show that the squared distances must be even.

FACT 1. If $\{d_1, d_2, d_3, \dots, d_n\}$ is a set of integers such that $\sum_{i=1}^{n} d_i = 0$ then $\sum_{i=1}^{n} d_i^2$ is an even integer.

FACT 2. If $X = \{x_1, x_2, x_3, \dots, x_n\}$ is a set of real numbers,

$$P = \{x_{i_1}, x_{i_2}, x_{i_3}, \dots, x_{i_n}\}$$

is a permutation of the elements of X and $d_j = x_j - x_{i_j}$ for $1 \le j \le n$ then

$$\sum_{j=1}^{n} d_j = 0.$$

That the squared distance between two vertices of a permutohedron is an even integer is an immediate consequence of the above two facts.

Inversions of a permutation

Let S_n be the set of permutations of $N_n = \{1, 2, 3, \dots, n\}$. The approach to deriving all distinct distances between vertices was to define the notion of an *inversion difference* for the set of inversions of a particular permutation f of S_n. First, recall the definition of an inversion of a permutation.

DEFINITION 1. Let $f = (a_1, a_2, a_3, \dots, a_{n-1}, a_n)$ be a permutation of S_n. A pair (a_i, a_j) with $a_i > a_j$ and $i < j$ is called an *inversion* of f. The set of inversions of f and the size of this set are denoted as follows

$$I(f) = \{(a_i, a_j) \mid a_i > a_j \text{ and } i < j\}, \qquad i(f) = |I(f)|.$$

EXAMPLE 1. Let $f = (5, 6, 2, 1, 7, 4, 3)$ be a permutation of S_7. Then

$$\begin{aligned} I(f) = \{&(2,1), (4,3), (5,1), (5,2), (5,3), (5,4), \\ &(6,1), (6,2), (6,3), (6,4), (7,3), (7,4)\}, \\ i(f) = |&I(f)| = 12. \end{aligned}$$

DEFINITION 2. Let $f = (a_1, a_2, a_3, \ldots, a_{n-1}, a_n)$ be a permutation of S_n. The *inversion difference* of the set $I(f)$ is the sum $d_{I(f)} = \sum_{(x,y)\in I(f)} (x - y)$.

EXAMPLE 2. Let $f = (5, 6, 2, 1, 7, 4, 3)$ be a permutation of S_7. Then

$$\begin{aligned} d_{I(f)} &= (2-1) + (4-3) + (5-1) + (5-2) + (5-3) + (5-4) \\ &\quad + (6-1) + (6-2) + (6-3) + (6-4) + (7-3) + (7-4) \\ &= 33. \end{aligned}$$

Distances and inversions are related

At first sight it may seem that distances between permutations and their inversions have little in common. However, consider the following result.

FACT 3. Let $f \in S_n$ and $d^2(f)$ be the squared distance between f and the identity permutation $\iota = (1, 2, 3, \ldots, n-1, n)$. Then $d^2(f) = 2d_{I(f)}$.

Notation

In Fact 3, permutation $f \in S_n$ and the vertex it represents in P_n are considered the same. Moreover, $d(f, \iota)$ is abbreviated as $d(f)$ since all distances are measured from ι.

Before proving Fact 3 the following lemma is established. Fact 3 will easily follow from the lemma.

LEMMA. *If $g \in S_n$ satisfying $d^2(g) = 2d_{I(g)}$ and $f \in S_n$ is obtained from g by interchanging two consecutive entries in g then $d^2(f) = 2d_{I(f)}$.*

Proof. Suppose f is formed from g by interchanging entries i and j in positions k and $k+1$. If $i < j$, the inversion difference is increased by $j - i$, and if $i > j$, the inversion difference is decreased by $i - j$. In either case, we have $d_{I(f)} = d_{I(g)} + j - i$.

Now consider how the square of the Euclidean distance changes. From the way f is obtained from g it follows that

$$\begin{aligned} d^2(f) &= d^2(g) + (k-j)^2 - (k-i)^2 + (k+1-i)^2 - (k+1-j)^2 \\ &= d^2(g) + 2(j-i). \end{aligned}$$

Since g satisfies $d^2(g) = 2d_{I(g)}$ we have

$$\begin{aligned} d^2(f) &= 2d_{I(g)} + 2(j-i) \\ &= 2d_{I(f)}. \end{aligned}$$

■

Proof of Fact 3. Starting with the identity permutation ι we have $d^2(\iota) = 2d_{I(\iota)} = 0$. From ι any other permutation f can be reached by a sequence of consecutive interchanges. Repeated applications of the lemma to the intermediate permutations g in the sequence eventually yields the desired result $d^2(f) = 2d_{I(f)}$, thereby proving Fact 3. ∎

Finding all distances

According to Fact 3 the determination of all distinct distances reduces to finding all distinct inversion difference values. *Mathematica* was used to obtain the number of distinct (nonzero) distances in the cases $n = 4, 5, 6, 7, 8$, and 9. From these values it soon became clear that the general formula was given by the binomial coefficient $\binom{n+1}{3}$ for $n \geq 4$. Its easy to verify the formula for $n = 1, 2$ assuming that $\binom{2}{3} = 0$. However, note that the formula does *not* hold for $n = 3$ since there are 3 nonzero distances, namely, $\sqrt{2}$, $\sqrt{6}$ and $\sqrt{8}$ but $\binom{3+1}{3} = 4$.

Using *Mathematica* to generate sample data for the case $n = 5$ and organizing it led to the discovery of Fact 3 and a way to prove the next result. The distinct squared distances extracted from all $5! = 120$ values produced by *Mathematica* were $0, 2, 4, \ldots, 38, 40$. This easily led to the observation that *all* even values up through a maximum value were generated. Examining squared distances and their corresponding permutations motivated the concept of an inversion difference and revealed the equation in Fact 3.

The data showed that permutations of the form $(a_1, a_2, a_3, a_4, 5)$, i.e., 5 is a fixed point, produced the distinct squared distances $0, 2, 4, \ldots, 20$. Reorganizing the data and separating out permutations of the form $(5, a_2, a_3, a_4, a_5)$ gave the remaining squared distances $22, 24, 26, \ldots, 40$. The next result *and* its proof by mathematical induction are generalizations of the case $n = 5$.

FACT 4. For $n \geq 4$, the distinct inversion difference values for the permutations of S_n are $0, 1, 2, \ldots, \binom{n+1}{3}$.

Proof. *Mathematica* can be used to establish the case $n = 4$. Assume $n \geq 5$ and consider all permutations $f \in S_n$ of the form $f = (a_1, a_2, a_3, \ldots, a_{n-1}, n)$, that is, n is a fixed point. All such permutations correspond to permutations $g = (a_1, a_2, a_3, \ldots, a_{n-1}) \in S_{n-1}$. Moreover, it is obvious that the set of distinct inversion difference values $d_{I(f)}$ is the same as the set of distinct inversion difference values $d_{I(g)}$.

Next, consider all permutations $f \in S_n$ of the form $f = (n, a_2, a_3, \ldots, a_{n-1}, a_n)$. The inversion difference value of f can be written as

$$d_{I(f)} = \sum_{i=2}^{n}(n - a_i) + \sum_{2 \leq i < j}(a_i - a_j)$$

where it is understood that the only differences $a_i - a_j$ that appear in the second sum are when $a_i > a_j$. The values a_i and a_j used to calculate differences in the two sums come from the set N_{n-1}. Hence, the first sum calculates all the (positive) differences $n - i$ for $i \in N_n$, but, perhaps, in a different order. Therefore, the value of the first sum is $\binom{n}{2}$.

The differences $a_i - a_j$ that appear in the second sum use values a_i and a_j from the set N_{n-1} where $a_i > a_j$. Hence, the second sum corresponds to an inversion difference for some permutation $g \in S_{n-1}$. Consequently, for each $f \in S_n$ of the form

$f = (n, a_2, a_3, \dots, a_{n-1}, a_n)$ we have

$$d_{I(f)} = \binom{n}{2} + d_{I(g)}.$$

By the induction hypothesis, the distinct inversion difference values $d_{I(g)}$ for all $g \in S_{n-1}$ are 0, 1, 2, ..., $\binom{n-1+1}{3}$. Thus there are permutations g_i in S_{n-1}, for $1 \le i \le \binom{n}{2}$, such that $d_{I(g_i)} = \binom{n}{3} - \binom{n}{2} + i$. Corresponding to these g_i we have permutations f_i in S_n, for $1 \le i \le \binom{n}{2}$, such that

$$d_{I(f_i)} = \binom{n}{3} + i.$$

This yields the remaining inversion difference values and proves Fact 4. ■

Related avenues of investigation

Studying a permutohedron using other metrics to measure distance or focusing on permutations in a *subgroup* of S_n are topics that may uncover some interesting results. Cayley's theorem tells us that every finite group G is isomorphic to a subgroup of S_n. Let $\varphi : G \to S$ be such an isomorphism where S is a subgroup of S_n. If G is a finite group and $x, y \in G$, define a metric $d(x, y) = \|\varphi(x) - \varphi(y)\|$ where $\|\cdot\|$ is the Euclidean distance between vertices $\varphi(x)$ and $\varphi(y)$ of the permutohedron. It would be interesting to count distinct distances, measure ϵ-neighborhoods, etc., for those vertices of the permutohedron corresponding to the group elements under φ.

A departing challenge

Consider the $n! \times n!$ distance matrix D_n where $d_{uv} = d(u, v)$ the Euclidean distance between vertices u and v in P_n. Its easy to see that D_n is symmetric. Find a formula for $det(D_n)$, the determinant of D_n. Enjoy!

REFERENCES

1. M. Aigner, *Combinatorial Theory*, Springer-Verlag, New York, 1979.
2. C. Berge, *Principles of Combinatorics*, Academic Press, New York, 1971.
3. H. Coxeter *Regular Polytopes*, Dover, New York, 1973.
4. V. Emelichev, M. Kovalev, and M. Kravtsov, *Polytopes, Graphs and Optimization*, Cambridge University Press, New York, 1984.
5. M. Gardner, "The Five Platonic Solids," *The Second Scientific American Book of Mathematical Puzzles and Diversions: A New Selection*, Simon and Schuster, New York, 1961.
6. T. Heath, *A History of Greek Mathematics, Vol. 1: From Thales to Euclid*, Dover, New York, 1981.
7. A. Postnikov, *Permutohedra, Associahedra and Beyond*, MIT website `http://www-math.mit.edu/~apost/talks/perm-slides.pdf`, 2004.
8. G. Ziegler, *Lectures on Polytopes*, Springer-Verlag, Graduate Texts in Mathematics, Berlin, 1994.

Finite Mimicry of Gödel's Incompleteness Theorem

PAUL FJELSTAD
32991 Dresden Ave.
Northfield, MN 55057
fjelstap@stolaf.edu

This paper attempts to come up with the minimal structure needed to mimic Gödel's formal system with regard to self-referring sentences, incompleteness, the futility of striving for completeness, and the inability to define or name truth. It takes the approach used by Raymond Smullyan [**2**, pp. 169–192], who chooses for consideration a special set of sentences involving the natural numbers, which numbers are used in turn as code names for these sentences. That system is further simplified here, most essentially by replacing the set of natural numbers with an arbitrary set that can have as few as two members. It will be called a *G-system*, but first let us review Gödel's result.

Gödel systems Suppose we are given a formal axiom system that includes axioms for the basic properties of the natural numbers and their arithmetic. A *Gödel system* for these formal axioms is a code that uses natural numbers to represent not only the axioms, but the formal language (symbols) of the system. This enables the construction of sentences in the arithmetic, which, upon decoding, refer to themselves. In Gödel's proof of his theorem, one such sentence asserts its own unprovability within the system. By reasoning outside the system and assuming this sentence is false, it follows the sentence is provable, which means either the system is not consistent, or, if consistent, then the assumption is wrong and the sentence is true. Since a consistent system in which all true sentences are provable is called *complete*, Gödel's theorem can be expressed simply as

A Gödel system, if consistent, is incomplete.

G-systems In order to construct sentences that refer to themselves, consider a set I and let X be a subset of I. Then for $i \in I$, the sentence "$i \in X$" has i for the subject and X for the predicate. Next, one assigns the elements $i \in I$ as code names for some of the subsets by means of a *naming function* N, where

$$N : I \to \{ X \mid X \subseteq I \}, \quad i \mapsto N_i .$$

The sentences "$i \in N_i$" are then selected for consideration, and, by letting them inherit the same code names as the N_i, they can be decoded to assert something about themselves. To make this explicit, define for each subset X the set of sentences coded by the $i \in X$,

$$S(X) = \{\text{"}i \in N_i\text{"} \mid i \in X \} .$$

For the purposes of this paper, sentences not in $S(I)$ will be excluded from consideration. For sentences in $S(I)$, note that the following assertion, call it G_i, is a direct consequence of the above definition for the case $X = N_i$.

$$G_i : \quad i \in N_i \quad \text{if and only if} \quad \text{"}i \in N_i\text{"} \in S(N_i) .$$

Thus, the sentence "$i \in N_i$" can be interpreted, or decoded, as asserting something about itself, namely that it is a member of $S(N_i)$.

An interesting subset to consider is the *truth subset*,

$$T = \{i \mid i \in N_i\},$$

so $S(T)$ is the set of true sentences. If T is nameable with a code name, say t, so $T = N_t$, then "$t \in N_t$" can be interpreted as saying of itself that it is true.

Another subset is the *proof set* P, consisting of the code names of the sentences that are provable from the axioms, so $S(P)$ is the set of provable sentences. Then if P is nameable, say $P = N_p$, the sentence "$p \in N_p$" says of itself that it is provable. Similarly, if P', the complement of P, is nameable, say by $P' = N_{p'}$, the sentence "$p' \in N_{p'}$" becomes a form of the classic Gödel sentence that says of itself that it is unprovable. For a G-system, which we are now ready to define, we want an axiom asserting that if a subset is nameable, so is the complement. To this end, we will use a *complement function*,

$$c : I \to I, \quad i \mapsto c(i),$$

and then demand that i and $c(i)$ name complements of each other, as in Axiom c, which follows.

DEFINITION. A *G-system* $I(N, c, P)$ consists of a set I, a naming function N, a complement function c, and a specially designated subset P. It satisfies the following two axioms.

AXIOM c. $N_{c(i)} = N_i{}'$,

AXIOM P. *If* $i \in P$ *then* $i \in N_i$.

After showing that P is a proof set (in the next paragraph), axiom P can be decoded as saying that if a sentence is provable then it is true. This is simply a property of a consistent Gödel system, where Gödel's great accomplishment was constructing a predicate for natural numbers in the system's formal language such that, upon decoding, it corresponded to the predicate "provable" for the sentences coded by natural numbers.

The sentences of $S(P)$ are just those of $S(I)$ that are provable. By axiom P it follows directly that the sentences in $S(P)$ are provable. On the other hand the sentences not in $S(P)$ are not provable, since if $i \notin P$, there is no way to decide from the axioms whether "$i \in N_i$" is true or false. They are in fact undecidable, since both cases are possible. Consider the example $I = \{1, 2\}$, $P = \{1\}$, $c(1) = 2$, $c(2) = 1$ with two different naming functions N and M.

$$N_1 = I, \; N_2 = \emptyset, \qquad M_1 = P, \; M_2 = P'.$$

The axioms hold in both cases, so both are G-systems, and "$2 \in N_2$" is false while "$2 \in M_2$" is true. The reader is encouraged to consider other examples of G-systems, which are many and varied. In particular, P can be any subset of I, and N and c needn't be 1-1 mappings. A table of examples with various properties appears at the MAGAZINE website.

The fact that there is a model for a G-system demonstrates that its axioms are consistent. Consistency in turn means that any provable sentence is true, that is, $S(P) \subseteq S(T)$. Note that in the N case above $S(P) = S(T)$, while $S(P) \neq S(T)$ for M, so one is complete and the other incomplete. Note also that P is named by M and not named by N. This leads to the main result.

THEOREM P. *A G-system, if P-nameable, is incomplete.*

Proof. Assume $N_p = P$ for some p. Then, by axioms P and c,

$$\text{If } c(p) \in P, \text{ then } c(p) \in N_{c(p)}, \text{ but } N_{c(p)} = N_p' = P',$$

which is a contradiction. Thus $c(p) \notin P$ and "$c(p) \in N_{c(p)}$" $\notin S(P)$. However, the following are equivalent,

$$c(p) \notin P, \quad c(p) \in P', \quad c(p) \in N_{c(p)},$$

so $c(p) \in N_{c(p)}$ and "$c(p) \in N_{c(p)}$" $\in S(T)$. ∎

In an attempt to gain completeness for the system, one can add the troublesome sentence as an axiom, or, what is equivalent, one can consider the G-system $I(N, c, Q)$, where $Q = P \cup \{c(p)\}$, since then the new set of provable sentences is

$$S(Q) = S(P) \cup \{\text{``}c(p) \in N_{c(p)}\text{''}\}.$$

If Q is nameable, say by $Q = N_q$, then the new system satisfies Theorem P and the sentence "$c(q) \in N_{c(q)}$" is true but not provable. If Q is not nameable, one can augment the set I by adding new elements in order to have names for Q and its complement, which means there will once again be a true but unprovable sentence. One can accordingly continue indefinitely by alternately augmenting the set of provable sentences and the set of names.

A result concerning the concept of truth, which corresponds to Tarski's Theorem on the formal undefinability of truth for a Gödel system is the following.

THEOREM T. *For a G-system, T is unnameable.*

Proof. Assume $N_t = T$ for some $t \in I$. Then $N_{c(t)} = N_t' = T'$ and the following are all equivalent (remember the definition of T).

$$c(t) \notin N_{c(t)}, \quad c(t) \notin T', \quad c(t) \in T, \quad c(t) \in N_{c(t)},$$

which produces a contradiction. ∎

Another way to arrive at this contradiction is to consider $G_{c(t)}$, which asserts the truth of the last sentence above if and only if this sentence is a member of the set of untrue sentences, that is, this sentence is a form of the classic sentence "This sentence is false."

Comparison

In comparing a Gödel system with a G-system, there is a tradeoff between consistency and nameability. In a Gödel system, all sentences are nameable, but the system may not be consistent. If it is consistent, then it is incomplete. A G-system, on the other hand, is consistent, but P may not be nameable. If it is nameable, then it is incomplete.

There is a kind of uncertainty principle here. To guarantee consistency, you can go finite, but then you don't have enough names to code all the sentences in your language. To have enough names, you can go infinite, but you no longer can prove consistency.

Acknowledgment. Thanks to Loren Larson for many discussions and helpful suggestions.

REFERENCES

1. Kurt Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik* **38** (1931) 173–198.
2. Raymond Smullyan, *The lady or the tiger? and other logical puzzles, including a mathematical novel that features Gödel's great discovery*, Knopf, New York, 1982.

Digit Reversal Without Apology

LARA PUDWELL
Rutgers University
Piscataway, NJ 08854
lpudwell@math.rutgers.edu

In *A Mathematician's Apology* [**1**] G. H. Hardy states, "8712 and 9801 are the only four-figure numbers which are integral multiples of their reversals"; and, he further comments that "this is not a serious theorem, as it is not capable of any significant generalization."

However, Hardy's comment may have been short-sighted. In 1966, A. Sutcliffe [**2**] expanded this obscure fact about reversals. Instead of restricting his study to base 10 integers and their reversals, Sutcliffe generalized the problem to study all integer solutions of

$$k(a_h n^h + a_{h-1}n^{h-1} + \cdots + a_1 n + a_0) = a_0 n^h + a_1 n^{h-1} + \cdots + a_{h-1}n + a_h$$

with $n \geq 2$, $1 < k < n$, $0 \leq a_i \leq n-1$ for all i, $a_0 \neq 0$, $a_h \neq 0$. We shall refer to such an integer $a_0 \ldots a_h$ as an $(h+1)$*-digit solution for* n and write

$$k(a_h, a_{h-1}, \ldots, a_1, a_0)_n = (a_0, a_1, \ldots, a_{h-1}, a_h)_n.$$

For example, 8712 and 9801 are 4-digit solutions in base $n = 10$ for $k = 4$ and $k = 9$ respectively. After characterizing all 2-digit solutions for fixed n and generating parametric solutions for higher digit solutions, Sutcliffe left the following open question: Is there any base n for which there is a 3-digit solution but no 2-digit solution?

Two years later T. J. Kaczynski* [**3**] answered Sutcliffe's question in the negative. His elegant proof showed that if there exists a 3-digit solution for n, then deleting the middle digit gives a 2-digit solution for n. Together with Sutcliffe's work, this proved that there exists a 2-digit solution for n if and only if there exists a 3-digit solution for n.

Given the nice correspondence between 2- and 3-digit solutions described by Sutcliffe and Kaczynski, it is natural to ask if there exists such a correspondence for higher digit solutions. In this paper, we will explore the relationship between 4- and 5-digit solutions. Unfortunately, there is not a bijection between these solutions, but there is a nice family of 4- and 5-digit solutions which have a natural one-to-one correspondence.

*Better known for other work.

A second extension of Sutcliffe and Kaczynski's results is to ask, "Is there any value of n for which there is a 5-digit solution but no 4-digit solution?" We will answer this question in the negative; and, furthermore, we will show that there exist 4- and 5-digit solutions for every $n \geq 3$.

An attempt at generalization. In the case of 3-digit solutions, Kaczynski proved that if $n+1$ is prime and $k(a, b, c)_n = (c, b, a)_n$ is a 3-digit solution for n, then $k(a, c)_n = (c, a)_n$ is a 2-digit solution. Thus, we consider the following:

QUESTION 1. *Let $k(a, b, c, d, e)_n = (e, d, c, b, a)_n$ be a 5-digit solution for n. If $n+1$ is prime, then is $k(a, b, d, e)_n = (e, d, b, a)_n$ a 4-digit solution for n?*

First, following Kaczynski, let $p = n + 1$. We have

$$k(an^4 + bn^3 + cn^2 + dn + e) = en^4 + dn^3 + cn^2 + bn + a. \tag{1}$$

Reducing this equation modulo p, we obtain

$$k(a - b + c - d + e) \equiv e - d + c - b + a = a - b + c - d + e \bmod p.$$

Thus, $(k-1)(a - b + c - d + e) \equiv 0 \bmod p$, and

$$p \mid (k-1)(a - b + c - d + e). \tag{2}$$

If $p \mid (k-1)$, then $k - 1 \geq p$, which is impossible because $k < n$. Therefore,

$$p \mid (a - b + c - d + e).$$

But $-2p < -2n < a - b + c - d + e < 3n < 3p$, so there are four possibilities:

(i) $a - b + c - d + e = -p$,

(ii) $a - b + c - d + e = 0$,

(iii) $a - b + c - d + e = p$,

(iv) $a - b + c - d + e = 2p$.

Write $a - b + c - d + e = fp$, where $f \in \{-1, 0, 1, 2\}$. Substituting $c = -a + b + d - e + fp$ into equation 1 gives:

$$\begin{aligned} &k[n^2(n^2-1)a + n^2(n+1)b + fpn^2 + n(n+1)d - (n^2-1)e] \\ &\quad = n^2(n^2-1)e + n^2(n+1)d + fpn^2 + n(n+1)b - (n^2-1)a. \end{aligned}$$

After substituting for p, dividing by $n+1$, and rearranging, one sees that $k[an^3 + (b - a + f)n^2 + (d - e)n + e] = en^3 + (d - e + f)n^2 + (b - a)n + a$. Indeed, this is a 4-digit solution for n if $f = 0$, $b - a \geq 0$, and $d - e \geq 0$, but not necessarily a 4-digit solution of the form conjectured in Question 1.

As in Kaczynski's proof for 2- and 3-digit solutions, it would be ideal if three of the four possible values for f lead to contradictions and the fourth leads to a "nice" pairing of 4- and 5-digit solutions. Unlike Kaczynski, we now have the added advantage of exploring these cases with computer programs such as Maple. Experimental evidence suggests that the cases $f = -1$ and $f = 2$ are impossible. The cases $f = 0$ and $f = 1$ are discussed below.

A counterexample. Unfortunately, Kaczynski's proof does not completely generalize to higher digit solutions. Most 5-digit solutions do, in fact, yield 4-digit solutions in the manner described in Question 1, but for sufficiently large n there are examples where $(a, b, c, d, e)_n$ is a 5-digit solution but $(a, b, d, e)_n$ is not a 4-digit solution.

A computer search shows that the smallest such counterexamples appear when $n = 22$:

$$7(2, 8, 3, 13, 16)_{22} = (16, 13, 3, 8, 2)_{22}, 3(2, 16, 11, 5, 8)_{22} = (8, 5, 11, 16, 2)_{22}.$$

However, there is no integer k for which $k(2, 8, 13, 16)_{22} = (16, 13, 8, 2)_{22}$ or $k(2, 16, 5, 8)_{22} = (8, 5, 16, 2)_{22}$. Note that $-2 + 8 + 13 - 16 = 3$ and $-2 + 16 + 5 - 8 = 11$; that is, both of these counterexamples to Question 1 occur when $f = 0$. The next smallest counterexamples are

$$3(3, 22, 15, 7, 11)_{30} = (11, 7, 15, 22, 3)_{30}, 8(2, 13, 8, 16, 19)_{30} = (19, 16, 8, 13, 2)_{30},$$

which occur when $f = 0$ and $n = 30$.

A family of 4- and 5-digit solutions. Although Kaczynski's proof does not generalize entirely, there exists a family of 5-digit solutions when $f = 1$ that has a nice structure.

THEOREM 1. *Fix $n \geq 2$ and $a > 0$. Then*

$$k(a, a - 1, n - 1, n - a - 1, n - a)_n = (n - a, n - a - 1, n - 1, a - 1, a)_n$$

is a 5-digit solution for n if and only if $a \mid (n - a)$.

Proof. We have

$$\frac{(n - a)n^4 + (n - a - 1)n^3 + (n - 1)n^2 + (a - 1)n + a}{an^4 + (a - 1)n^3 + (n - 1)n^2 + (n - a - 1)n + (n - a)}$$
$$= \frac{(n - a)(n^4 + n^3 - n - 1)}{a(n^4 + n^3 - n - 1)} = \frac{n - a}{a},$$

and the result is clear. ∎

Notice that

$$(-a + (a - 1)) + ((n - a - 1) - (n - a)) + p = -1 + -1 + (n + 1) = n - 1.$$

That is, this family of solutions occurs when $f = 1$. Moreover, this family follows the pattern described in Question 1; that is, for each 5-digit solution described in Theorem 1, deleting its middle digit gives a 4-digit solution.

THEOREM 2. *If*

$$k(a, a - 1, n - 1, n - a - 1, n - a)_n = (n - a, n - a - 1, n - 1, a - 1, a)_n$$

is a 5-digit solution for n, then

$$k(a, a - 1, n - a - 1, n - a)_n = (n - a, n - a - 1, a - 1, a)_n$$

is a 4-digit solution for n.

Proof. By Theorem 1, $\frac{n-a}{a} \in \mathbb{N}$. Now

$$\frac{(n-a)n^3 + (n-a-1)n^2 + (a-1)n + a}{an^3 + (a-1)n^2 + (n-a-1)n + (n-a)} = \frac{(n-a)(n^3+n^2-n-1)}{a(n^3+n^2-n-1)} = \frac{n-a}{a}. \quad \blacksquare$$

These 4-digit solutions were first described by Klosinski and Smolarski [**4**] in 1969, but their relationship to 5-digit solutions was not made explicit before now.

It is also interesting to note that 9801 and 8712, the two integers in Hardy's discussion of reversals, are included in this family of solutions.

We conclude with the following corollary.

COROLLARY 1. *There is a 4-digit solution and a 5-digit solution for every* $n \geq 3$.

Proof. Let $a = 1$ in the statements of Theorem 1 and Theorem 2 above. ■

Some open questions. We have shown that there is no n for which there is a 5-digit solution but no 4-digit solution. More specifically, we know that there are 4- and 5-digit solutions for every $n \geq 3$.

Although Kaczynski's proof does not generalize directly to 4- and 5-digit solutions, it does bring to light several questions about the structure of solutions to the digit reversal problem.

First, it would be interesting to completely characterize 4- and 5-digit solutions for n. Namely,

1. All known counterexamples to Question 1 occur when $f = 0$. Are there counterexamples for which $f \neq 0$? Is there a parameterization for all such counterexamples?
2. Theorems 1 and 2 exhibit a family of 4- and 5-digit solutions for $f = 1$ with a particularly nice structure. To date, no other 4- or 5-digit solutions are known for $f = 1$. Do such solutions exist?

More generally,

3. Solutions to the digit reversal problem have not been explicitly characterized for more than 5 digits. Do there exist analogous results to Theorems 1 and 2 for higher digit solutions?

A Maple package for exploring these questions is available from the author's web page at `http://www.math.rutgers.edu/~lpudwell/maple.html`.

Acknowledgment. Thank you to Doron Zeilberger for suggesting this project.

REFERENCES

1. G. H. Hardy, *A Mathematician's Apology*, Cambridge University Press, New York, NY, 1993.
2. Alan Sutcliffe, "Integers That Are Multipled When Their Digits Are Reversed," this MAGAZINE **39** (1966) 282–287.
3. T. J. Kaczynski, "Note on a Problem of Alan Sutcliffe," this MAGAZINE **41** (1968) 84–86.
4. Leonard F. Klosinski and Dennis C. Smolarski, "On the Reversing of Digits," this MAGAZINE **42** (1969) 208–210.
5. N. J. A. Sloane, Sequence A031877 in "The On-Line Encyclopedia of Integer Sequences," `http://www.research.att.com/projects/OEIS?Anum=A031877`.
6. Eric W. Weisstein, "Reversal," From MathWorld—A Wolfram Web Resource, `http://mathworld.wolfram.com/Reversal.html`.

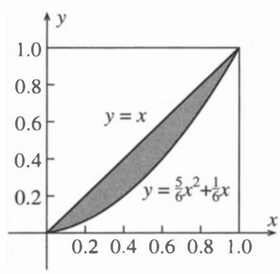
A Quintile Rule for the Gini Coefficient

LEON GERBER
St. John's University
Jamaica NY 11439
gerberle@gmail.com

Undergraduate calculus courses for business and economics majors frequently include the Gini coefficient of income inequality, or simply the "Gini coefficient," as an application of integration. See for example [**1**] and [**2**]. Because the U.S. Census Bureau provides the relevant information in quintiles, Simpson's rule, which requires an even number of intervals, cannot be used to approximate the integral. Usually the trapezoidal rule is used which, because the Lorenz curve is concave upward, underestimates the Gini coefficient. Here we derive a rule which is exact for fifth-degree polynomials and is just as simple as Simpson's rule, which is exact only for third-degree polynomials. We first provide the necessary definitions and a typical textbook problem.

A Lorenz Curve $L(x)$ is the fraction of total income earned by the poorest fraction x, $0 \le x \le 1$, of the population. The number $100x$ is a percentile. $L(0) = 0$, $L(1) = 1$, $L'(x) \ge 0$, $L''(x) \ge 0$. The Lorenz curve was invented by Max Lorenz in an undergraduate essay and quickly became popular.

The Gini Coefficient $G = 2\int_0^1 [x - L(x)]dx$ is the ratio of the area between the line $y = x$ (all incomes equal) and $y = L(x)$ (shaded) to the maximum possible area of $\frac{1}{2}$ (all income goes to one person).



TYPICAL EXAMPLE. Graph the Lorenz curve $L(x) = \frac{5}{6}x^2 + \frac{1}{6}x$. (a) What part of the total income is earned by the poorest fifth? (b) Find the Gini coefficient G.

Solution. See the graph. Note that $x - L(x)$ is 0 at both ends.

(a) $L(.2) = \frac{5}{6}(.2)^2 + \frac{1}{6}(.2) = .067 = 6.7\%$.

(b) $$G = 2\int_0^1 \left[x - \left(\frac{5}{6}x^2 + \frac{1}{6}x\right)\right] dx = 2\int_0^1 \left(\frac{5}{6}x - \frac{5}{6}x^2\right) dx = \frac{5}{3}\int_0^1 (x - x^2)dx$$

$$= \frac{5}{3}\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1 = \frac{5}{3} \cdot \frac{1}{6} = \frac{5}{18} \approx .278$$

We now state and prove the quintile rule and apply it to real Census data.

QUINTILE RULE. Let f be a polynomial of degree ≤ 5 such that $f(a) = f(b) = 0$ and $h = \frac{1}{5}(b - a)$. Then

$$\int_a^b f(x)dx = \frac{125}{288}h[3f(a+h)+2f(a+2h)+2f(a+3h)+3f(a+4h)]$$

Proof. Substitute to change the interval to $[-5, 5]$. Then $h = \frac{10}{5} = 2$ and we seek coefficients A, B, C, D so

$$\int_{-5}^{5} (5+x)(5-x)x^n\, dx = 2[Af(-3) + Bf(-1) + Cf(1) + Df(3)]$$

is satisfied for $n = 0, 1, 2, 3,$ and hence for any linear combination.

$$n = 0: \int_{-5}^{5} (25 - x^2)\, dx = \frac{500}{3} = 2[16A + 24B + 24C + 16D]$$

$$n = 1: \int_{-5}^{5} (25 - x^2)x\, dx = 0 = 2[-48A - 24B + 24C + 48D]$$

$$n = 2: \int_{-5}^{5} (25 - x^2)x^2\, dx = \frac{2500}{3} = 2[144A + 24B + 24C + 144D]$$

$$n = 3: \int_{-5}^{5} (25 - x^2)x^3\, dx = 0 = 2[-432A - 24B + 24C + 432D]$$

From $n = 1$ and $n = 3$ (the integrals vanish because the integrands are odd), we have $D = A$ and $C = B$, so $n = 0$ and $n = 2$ give

$$64A + 96B = \tfrac{500}{3}$$

$$576A + 96B = \tfrac{2500}{3}$$

which yield $A = \frac{125}{96} = 3 \cdot \frac{125}{288}$ and $B = \frac{125}{144} = 2 \cdot \frac{125}{288}$. ■

EXAMPLE. Use Census data and the Quintile Rule to estimate the Gini coefficient of income inequality in 2000: the lowest fifth of U.S. families earned 4.3% of the total, the second fifth 9.8%, the third fifth 15.4%, the fourth fifth 22.7% and the highest fifth earned 47.7%, based on cash income before taxes [**3**].

Solution. $L(x)$ is the cumulative fraction. $f(x) = x - L(x)$ is 0 at 0 and 1. $h = \frac{1}{5}$.

x	fifth	pct	$L(x)$	$x - L(x)$
.200	lowest	4.3	.043	.157
.400	second	9.8	.043 + .098 = .141	.259
.600	third	15.4	.141 + .154 = .295	.305
.800	fourth	22.7	.295 + .227 = .522	.278

$$G = 2\int_0^1 [x - L(x)]\, dx$$

$$\approx 2 \cdot \frac{125}{288} \cdot \frac{1}{5}(3 \times .157 + 2 \times .259 + 2 \times .305 + 3 \times .278) \approx .422$$

The Census Bureau gets .433 using all the data, not just quintiles.

REMARK. The equations $n = 0$ and $n = 1$ are satisfied by $A = B = C = D = \frac{25}{24}$ giving the equal-weight formula for 0-ended cubics

$$\int_a^b f(x)\,dx = \frac{25}{24}h\,[f(a+h) + f(a+2h) + f(a+3h) + f(a+4h)]$$

where $h = \frac{1}{5}(b-a)$.

Acknowledgment. I would like to thank the referee for improving the rule from cubic to quintic polynomials and bringing the Example closer to the Census Bureau value.

REFERENCES

1. L. D. Hoffmann and G. L. Bradley, *Calculus for Business, Economics, and the Social and Life Sciences*, 9th ed., McGraw Hill, New York, 2007.
2. R. A. Barnett et al., *College Mathematics for Business, Economics, Life Sciences, and Social Sciences*, 10th ed., Prentice-Hall, Upper Saddle River, 2005.
3. U.S. Census Bureau, "Historical Income Tables—Families," `http://www.census.gov/hhes/income/histinc/f02.html`, and Note 30. Revised July 8, 2004.

A Short Proof of the Two-sidedness of Matrix Inverses

RAYMOND A. BEAUREGARD
University of Rhode Island
Kingston, RI 02881
beau@math.uri.edu

In this note we offer a short proof that for any $n \times n$ matrices A and C over a field of scalars,

$$AC = I \text{ if and only if } CA = I.$$

An interesting aspect of this biconditional is the fact that it is equivalent to the conditional if $AC = I$ then $CA = I$; one simply interchanges the roles of A and C.

We assume the reader is familiar with the reduction of a matrix to row-echelon form. Recall that if A is $n \times n$ matrix then its reduced row-echelon form is a matrix of the same size with zeros in the pivot columns except for the pivots which are equal to 1. It is achieved by applying elementary row operations (row swapping, row addition, row scaling) to A. An elementary matrix is one obtained by applying a single elementary row operation to the $n \times n$ identity matrix I. Elementary matrices have inverses that are also elementary matrices. Left multiplication of A by an elementary matrix E effects the same row operation on A that was used to create E.

Let H be the reduced row echelon form of A, and let P be the product of those elementary matrices (in the appropriate order) that reduce A to H. P is an invertible matrix such that $PA = H$. Notice that H is the identity matrix if and only if it has n pivots.

The proof. Beginning with $AC = I$, we left multiply this equation by P obtaining $PAC = P$ or $HC = P$. If H is not the identity matrix it must have a bottom row of zeros forcing P to have likewise a bottom row of zeros, and this contradicts the invertibility of P. Thus $H = I, C = P$, and the equation $PA = H$ is actually $CA = I$.

This argument shows at once that (i) a matrix is invertible if and only if its reduced row echelon form is the identity matrix, and (ii) the set of invertible matrices is precisely the set of products of elementary matrices.

REFERENCES

1. J. B. Fraleigh and R. A. Beauregard, *Linear Algebra*, Addison-Wesley, 3rd ed., Reading, 1995.

Root Preserving Transformations of Polynomials

BRANKO ĆURGUS
Western Washington University
Bellingham, WA 98225-9063
curgus@cc.wwu.edu

VANIA MASCIONI
Ball State University
Muncie, IN 47306-0490
vdm@cs.bsu.edu

Consider the real vector space $\mathcal{P}_2$ of all polynomials of degree at most 2. High-school students study the roots of the polynomials in $\mathcal{P}_2$, while linear algebra students study linear transformations on $\mathcal{P}_2$. Is it possible to bring these two groups together to do some joint research?

For example, a linear algebra student chooses a specific linear transformation $T : \mathcal{P}_2 \to \mathcal{P}_2$ and asks others to study the roots of a polynomial

$$p(x) = ax^2 + bx + c, \quad x \in \mathbb{R},$$

and the roots of its image

$$(Tp)(x) = bx^2 + cx + a, \quad x \in \mathbb{R}. \tag{1}$$

Here a, b, and c are arbitrary real numbers. The students may immediately notice that the polynomial $x^2 + x + 1$ is unchanged by this transformation. Hence this particular polynomial and its image have the same (complex) roots. After some "trial and error," a high-school student points out that the polynomial $x^2 + 3x + 2$ has the roots -1 and -2, while its image $3x^2 + 2x + 1$ does not have real roots. Their next interesting discovery is that, with $v \neq 1$, the polynomial $x^2 + (v-1)x - v$ has roots 1 and $-v$, while its image $(v-1)x^2 - vx + 1$ has roots 1 and $1/(v-1)$. This is curious since in this case a polynomial and its image have one common root, namely 1.

After further study the students conclude that there doesn't seem to be any general simple relationship between the roots of a polynomial p and the roots of its image Tp under the linear transformation given by (1). But the obvious fact is that there are plenty of other linear transformations on $\mathcal{P}_2$; will it always be the case that there is no simple relationship between the roots? Clearly, a non-zero multiple of the identity on

$\mathcal{P}_2$ does not change the roots of a polynomial, at all, and so such linear transformations are of no big interest in this study.

In the rest of the note, instead of $\mathcal{P}_2$, we consider the (complex or real) vector space $\mathcal{P}_n$ of all polynomials of degree at most n. To cover both cases, $\mathbb{F}$ stands for $\mathbb{R}$ if we consider $\mathcal{P}_n$ as a real vector space and $\mathbb{F}$ stands for $\mathbb{C}$ if we consider $\mathcal{P}_n$ as a complex vector space. For p in $\mathcal{P}_n$ we denote by $Z(p)$ the set of all roots of p in $\mathbb{F}$.

Inspired by the students' investigations we ask the following question:

Is there a (non-trivial) linear transformation T from $\mathcal{P}_n$ to $\mathcal{P}_n$ such that for each $p \in \mathcal{P}_n$ with a root in $\mathbb{F}$, the polynomials p and Tp have a common root?

Surprisingly, it seems that this question has not been addressed in the literature. The first author of this note has been assigning it at various levels of linear algebra courses. His experience is that students find it quite challenging even in the case $n = 2$. Students often offer "brute force" proofs that are based on calculating the matrix for the transformation T entry by entry.

In the next theorem we give a general answer to the above question. In the proof we use only elementary linear algebra and Taylor polynomials.

THEOREM. *Let $T \neq 0$ be a linear transformation from $\mathcal{P}_n$ to $\mathcal{P}_n$. Then*

$$Z(p) \cap Z(Tp) \neq \emptyset \quad \text{for all} \quad p \in \mathcal{P}_n \quad \text{such that} \quad Z(p) \neq \emptyset \tag{2}$$

if and only if T is a non-zero multiple of the identity on $\mathcal{P}_n$.

Proof. The "if" part of the theorem is obvious. To prove the "only if" part we assume (2).

Let $p \in \mathcal{P}_n$ be arbitrary. To prove that Tp is a constant multiple of p we choose an arbitrary $w \in \mathbb{F}$ and evaluate $(Tp)(w)$. To this end we consider the following $n + 1$ polynomials in $\mathcal{P}_n$

$$e_0(x) := 1, \quad e_{k,w}(x) := (x - w)^k, \quad x \in \mathbb{F}, \quad k = 1, \dots, n. \tag{3}$$

With notation (3), the nth degree Taylor polynomial of p at w is

$$p(x) = p(w)e_0(x) + \sum_{k=1}^{n} \frac{p^{(k)}(w)}{k!} e_{k,w}(x), \quad x \in \mathbb{F}.$$

This equality provides a representation of p as a linear combination of the polynomials in (3). Applying T to both sides of the last equality and using the linearity of T we obtain

$$(Tp)(x) = p(w)(Te_0)(x) + \sum_{k=1}^{n} \frac{p^{(k)}(w)}{k!} (Te_{k,w})(x), \quad x \in \mathbb{F}. \tag{4}$$

Clearly, $Z(e_{k,w}) = \{w\} \neq \emptyset$ for all $k = 1, \dots, n$. Therefore, by assumption (2),

$$\emptyset \neq Z(e_{k,w}) \cap Z(Te_{k,w}) = \{w\} \cap Z(Te_{k,w}).$$

Consequently, $w \in Z(Te_{k,w})$ and thus

$$(Te_{k,w})(w) = 0 \quad \text{for all} \quad k = 1, \dots, n.$$

Now we set $x = w$ in (4) and use the preceding n equalities to get

$$(Tp)(w) = p(w)(Te_0)(w). \tag{5}$$

Notice that $w \in \mathbb{F}$ and $p \in \mathcal{P}_n$ in (5) are arbitrary. Since the degree of the polynomial $Tp \in \mathcal{P}_n$ is less than or equal to n, if we choose $p \in \mathcal{P}_n$ to be of degree n, then (5) implies that the degree of Te_0 must be zero. That is, Te_0 is a constant polynomial: $(Te_0)(w) = c$ for all $w \in \mathbb{F}$, and so (5) implies that T is a multiple of the identity. ■

Now, the next natural (but quite a bit harder) question would be the following:

Characterize those linear transformations T from $\mathcal{P}_n$ to $\mathcal{P}_n$ such that, for some constant $C > 0$ and for all $p \in \mathcal{P}_n$ with $Z(p) \neq \emptyset$, some zeros of polynomials p and Tp are at most "distance C apart."

The notion itself of distance between the zero sets $Z(p)$ and $Z(Tp)$ needs to be clarified, of course, but this question has also been completely answered by the authors and the results will appear in a forthcoming article [**1**]. A similar question was also considered in [**2**].

REFERENCES

1. B. Ćurgus and V. Mascioni, Perturbations of roots under linear transformations of polynomials, to appear in *Constructive Approximation.* Available as arXiv:math.CV/0502027.
2. V. Tulovsky, On perturbations of roots of polynomials, *J. Analyse Math.* **54** (1990) 77–89.

Do Cyclic Polygons Make the Cut?

SAMUEL OTTEN
Grand Valley State University
Allendale, MI 49401
ottensj@student.gvsu.edu

For thousands of years mathematicians have studied the properties of *cyclic polygons*—polygons that can be circumscribed by a circle. There also exists a large number of findings concerning *cyclic product relations* for polygons—products of ratios of segment lengths, as in [**2**] and the theorem of Menelaus (see FIGURE 1). This paper intends to mix the two topics together, offering results reminiscent of but distinct from those found in [**4**] and [**6**]. A product of length-ratios is the primary focus, but rather than dealing with a single polygon we look at the interaction between a pair of cyclic polygons.

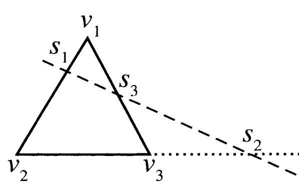


Figure 1 Menelaus' theorem states $\frac{|v_1 s_1|}{|s_1 v_2|} \cdot \frac{|v_2 s_2|}{|s_2 v_3|} \cdot \frac{|v_3 s_3|}{|s_3 v_1|} = 1$

On a circle we find the vertices of a polygon $V = [v_1, v_2, \ldots, v_n]$, where the order of this set indicates the connection of the vertices. Instead of cutting this

n-gon with a transversal, we are going to "cut" it with a second inscribed n-gon $W = [w_1, w_2, \ldots, w_n]$ whose vertices are distinct from V. For now, we assume both polygons are simple (i.e. neither intersects itself) and W has one vertex between every two consecutive vertices of V. Within these polygons, every side necessarily contains two intersection points. Let $S = \{s_1, s_2, \ldots, s_n\}$ be a subset from the set of all intersection points such that every side of each polygon contains exactly one element of S. There are necessarily two such sets. We label the vertices of V and W so that v_i is adjacent to w_i on the circle and s_i so that it marks the intersection between the chords $v_i v_{i+1}$ and $w_i w_{i+1}$. An example with triangles is depicted in FIGURE 2. Using the "cut" and the labeling system just described, we have the following result.

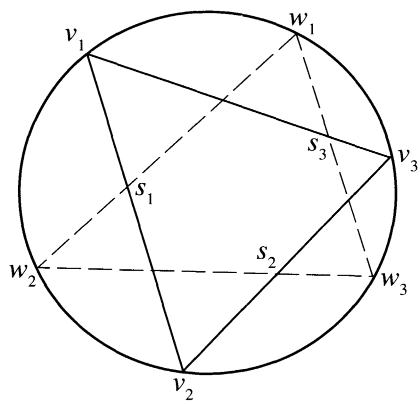


Figure 2 Two cyclic triangles

LEMMA. *If V and W are simple n-gons inscribed in the same circle such that each side of the polygons contains exactly one point from the intersection subset S, then*

$$\prod_{i=1}^{n} \left(\frac{|v_i s_i|}{|s_i v_{i+1}|} \cdot \frac{|w_i s_i|}{|s_i w_{i+1}|} \right) = 1.$$

The proof of this follows from two well-known geometric results—the law of sines and the two-chord theorem, both of which can be found in [**5**]. We construct n triangles using v_i, w_{i+1}, and s_i for $1 \le i \le n$, modulo n (see FIGURE 3). Note that these points are never collinear because the vertices of V are distinct from the vertices of W. We define θ_i to be the inscribed angle with endpoints v_i and w_i, and by the law of sines we have

$$\prod_{i=1}^{n} \frac{|v_i s_i|}{|s_i w_{i+1}|} = \prod_{i=1}^{n} \frac{\sin \theta_i}{\sin \theta_{i+1}}.$$

Note that θ_{n+1} is the same as θ_1, so $\sin \theta_i$ appears in both the numerator and denominator of the right-hand side for $1 \le i \le n$. Thus, we cancel to obtain

$$\prod_{i=1}^{n} \frac{|v_i s_i|}{|s_i w_{i+1}|} = 1.$$

The two-chord theorem states that when chords of a circle intersect, the product of the lengths of the segments formed on one chord equals that on the other chord. This

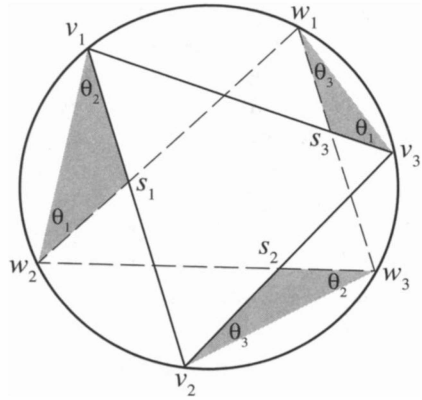


Figure 3 A possible construction for the triangles

gives us n equations using the elements of S as the intersections, and multiplying these together yields

$$\prod_{i=1}^{n}(|v_is_i|\cdot|s_iv_{i+1}|)=\prod_{i=1}^{n}(|w_is_i|\cdot|s_iw_{i+1}|)$$

or

$$\prod_{i=1}^{n}\frac{|v_is_i|}{|s_iw_{i+1}|}=\prod_{i=1}^{n}\frac{|w_is_i|}{|s_iv_{i+1}|}.$$

But we have determined that the left-hand side of this equation is one, so

$$\prod_{i=1}^{n}\frac{|v_is_i|}{|s_iw_{i+1}|}=\prod_{i=1}^{n}\frac{|w_is_i|}{|s_iv_{i+1}|}=1,$$

and the result follows from the multiplication of these expressions.

The requirement that the polygons be simple turns out to be unnecessary. Inscribed polygons that are self-intersecting will also cut each other in such a way that the product from the lemma remains valid.

THEOREM. *Let $V=[v_1,v_2,\ldots,v_n]$ and $W=[w_1,w_2,\ldots,w_n]$ be distinct n-gons inscribed in the same circle such that, in a given direction, w_i is the next vertex on the circle after v_{i+r} for some constant r (mod n). If $S=\{s_1,s_2,\ldots,s_n\}$ such that s_i is the intersection of v_iv_{i+1} and w_iw_{i+1}, then*

$$\prod_{i=1}^{n}\left(\frac{|v_is_i|}{|s_iv_{i+1}|}\cdot\frac{|w_is_i|}{|s_iw_{i+1}|}\right)=1.$$

An interested reader should be able to prove this generalization by closely following the reasoning of the proof of the lemma. As an assistive measure, and inspired by the subject of [**4**], we may use pentagrams as a concrete example (see FIGURE 4). There are many points of intersection between V and W and thus several sets S (each corresponding to a different labeling of W). FIGURE 5 presents the cases where w_i is after v_{i+1} in the clockwise direction and after v_{i+3} in the counterclockwise direction, respectively. The triangles with which to apply the law of sines are highlighted.

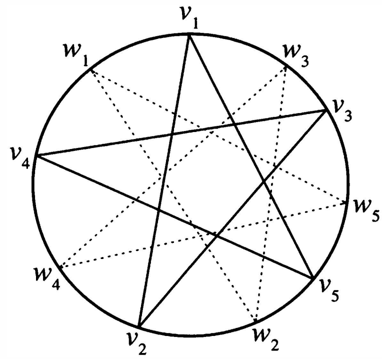


Figure 4 Two cyclic pentagrams

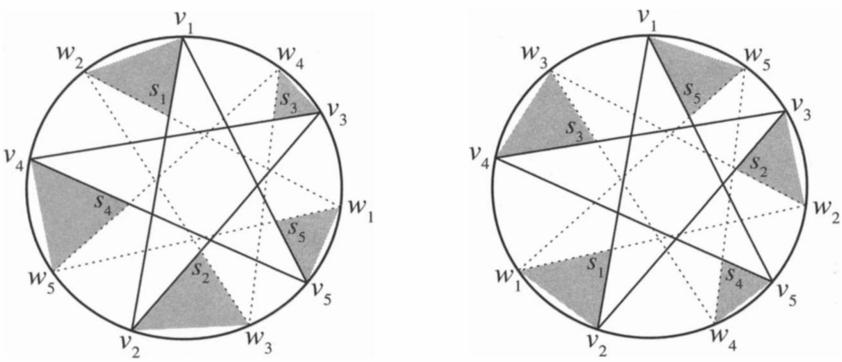


Figure 5 Two possible constructions for the pentagrams

There are further exercises that may be undertaken by the truly motivated reader. For instance, he or she may wish to prove the theorem of [**4**] using these results, or possibly to explore the author's conjecture that the theorem holds for polygons inscribed within an ellipse. For the non-geometer, situations can arise in dynamical systems where two polygons are inscribed within the same circle, specifically when dealing with circle maps [**3**, pp. 125–130]. One such case is outer billiards in the hyperbolic plane [**1**]. (For additional information on outer billiards look to [**7**].) If nothing else, it is always rewarding to uncover mathematical connections and we have found one between cyclic polygons, cyclic product relations, and dynamical systems.

REFERENCES

1. Filiz Doğru and Samuel Otten, On differentiation and classification within outer billiard maps, submitted for publication.
2. Branko Grünbaum and G.C. Shephard, Ceva, Menelaus, and selftransversality, *Geom. Dedicata* **65** (1997) 179–192.
3. Boris Hasselblatt and Anatole Katok, *A First Course in Dynamics*, Cambridge Univ. Pr., Cambridge, UK, 2003.
4. Larry Hoehn, A Menelaus-type theorem for the pentagram, this MAGAZINE **66** (1993) 121–123.
5. David Kay, *College Geometry*, Addison Wesley, Boston, MA, 2000.
6. G.C. Shephard, Pratt sequences and n-gons, *Discrete Math.* **221** (2000) 125–154.
7. Serge Tabachnikov, *Billiards*, Soc. Math. de France, Paris, 1995.

Nonexistence of a Composition Law

ŞABAN ALACA
Carleton University
Ottawa, Ontario, Canada K1S 5B6
salaca@math.carleton.ca

KENNETH S. WILLIAMS*
Carleton University
Ottawa, Ontario, Canada K1S 5B6
kwilliam@connect.carleton.ca

It was known to the ancient Greeks that sums of two squares satisfy the composition law

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = z_1^2 + z_2^2$$

with

$$z_1 = x_1y_1 + x_2y_2, \ z_2 = x_1y_2 - x_2y_1,$$

and to Euler in 1770 that sums of four squares satisfy the composition law

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2$$

with

$$z_1 = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4, \ z_2 = x_1y_2 - x_2y_1 + x_3y_4 - x_4y_3,$$
$$z_3 = x_1y_3 - x_2y_4 - x_3y_1 + x_4y_2, \ z_4 = x_1y_4 + x_2y_3 - x_3y_2 - x_4y_1.$$

Degen in 1822 and Cayley in 1845 gave the corresponding identity for eight squares, see for example [**6**, p. 2]. Sums of three squares however cannot possess an analogous composition law as $3 = 1^2 + 1^2 + 1^2$, $5 = 0^2 + 1^2 + 2^2$ but $15 = 3 \cdot 5 \neq x^2 + y^2 + z^2$ for integers x, y, z. Hurwitz proved in 1898 that there is an identity of the type

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = z_1^2 + \cdots + z_n^2,$$

where the z_k are bilinear functions of the x_i and y_i, if and only if $n = 1, 2, 4, 8$. Dickson [**2**] gave a detailed, amplified form of Hurwitz's proof in four pages. Rajwade [**6**] gave an amplified version of Dickson's proof in six pages. A proof using normed algebras is given in [**1**]. For more on such laws see for example [**6**].

As $2 = 1^2 + 1^2 + 2 \cdot 0^2$, $7 = 1^2 + 2^2 + 2 \cdot 1^2$, and $14 = 2 \cdot 7 \neq x^2 + y^2 + 2z^2$ for integers x, y, z there cannot exist a composition law of the type

$$(x_1^2 + x_2^2 + 2x_3^2)(y_1^2 + y_2^2 + 2y_3^2) = z_1^2 + z_2^2 + 2z_3^2$$

with z_1, z_2, z_3 bilinear functions of x_1, x_2, x_3 and y_1, y_2, y_3 with integer coefficients. However every odd positive integer can always be expressed in the form $x^2 + y^2 + 2z^2$ for some integers x, y, z, see for example [**3**, Theorem 86, p. 96], [**4**], [**5**, Theorem 1].

*Research of the second author was supported by National Sciences and Engineering Research Council of Canada grant A-7233.

Moreover one of x and y is odd and one is even. Thus every positive odd integer is of the form

$$(2x_1 + 1)^2 + 2x_2^2 + 4x_3^2$$

for some integers x_1, x_2, x_3. Let m and n be odd positive integers. Then mn is also an odd positive integer and there exist integers $x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2$ and z_3 such that

$$\begin{aligned} m &= (2x_1 + 1)^2 + 2x_2^2 + 4x_3^2, \\ n &= (2y_1 + 1)^2 + 2y_2^2 + 4y_3^2, \\ mn &= (2z_1 + 1)^2 + 2z_2^2 + 4z_3^2. \end{aligned}$$

Hence

$$\begin{aligned} &((2x_1 + 1)^2 + 2x_2^2 + 4x_3^2)((2y_1 + 1)^2 + 2y_2^2 + 4y_3^2) \\ &\quad = (2z_1 + 1)^2 + 2z_2^2 + 4z_3^2. \end{aligned}$$

The question naturally arises: Is this equality a consequence of some underlying composition law for the polynomial $(2x_1 + 1)^2 + 2x_2^2 + 4x_3^2$? In fact it is not, as can be deduced from Hurwitz's theorem. We show this directly from first principles without recourse to Hurwitz's theorem.

Suppose that there exist integers

$$a_1, a_2, \dots, a_{16}, b_1, b_2, \dots, b_{16}, c_1, c_2, \dots, c_{16}$$

such that

$$\begin{aligned} &((2x_1 + 1)^2 + 2x_2^2 + 4x_3^2)((2y_1 + 1)^2 + 2y_2^2 + 4y_3^2) \\ &\quad = (2z_1 + 1)^2 + 2z_2^2 + 4z_3^2 \end{aligned} \tag{1}$$

is an identity in $\mathbb{Z}[x_1, x_2, x_3, y_1, y_2, y_3]$ with

$$\begin{aligned} z_1 &= a_1x_1y_1 + a_2x_1y_2 + a_3x_1y_3 + a_4x_2y_1 + a_5x_2y_2 + a_6x_2y_3 \\ &\quad + a_7x_3y_1 + a_8x_3y_2 + a_9x_3y_3 + a_{10}x_1 + a_{11}x_2 + a_{12}x_3 \\ &\quad + a_{13}y_1 + a_{14}y_2 + a_{15}y_3 + a_{16}, \end{aligned} \tag{2}$$

$$\begin{aligned} z_2 &= b_1x_1y_1 + b_2x_1y_2 + b_3x_1y_3 + b_4x_2y_1 + b_5x_2y_2 + b_6x_2y_3 \\ &\quad + b_7x_3y_1 + b_8x_3y_2 + b_9x_3y_3 + b_{10}x_1 + b_{11}x_2 + b_{12}x_3 \\ &\quad + b_{13}y_1 + b_{14}y_2 + b_{15}y_3 + b_{16}, \end{aligned} \tag{3}$$

$$\begin{aligned} z_3 &= c_1x_1y_1 + c_2x_1y_2 + c_3x_1y_3 + c_4x_2y_1 + c_5x_2y_2 + c_6x_2y_3 \\ &\quad + c_7x_3y_1 + c_8x_3y_2 + c_9x_3y_3 + c_{10}x_1 + c_{11}x_2 + c_{12}x_3 \\ &\quad + c_{13}y_1 + c_{14}y_2 + c_{15}y_3 + c_{16}. \end{aligned} \tag{4}$$

We equate the coefficients of y_3^2, y_3, $x_2y_3^2$, x_2^2, $x_2^2y_3$, and $x_2^2y_3^2$ in (1) (with z_1, z_2, z_3 given by (2), (3), (4) respectively) to obtain the required contradiction. We have

$[y_3^2]$ $4a_{15}^2 + 2b_{15}^2 + 4c_{15}^2 = 4$

so

$$b_{15} = 0, \ (a_{15}, c_{15}) = (\pm 1, 0) \text{ or } (0, \pm 1); \tag{5}$$

$[y_3]$ $4a_{15}(2a_{16}+1)+4b_{15}b_{16}+8c_{15}c_{16}=0$

so by (5) and division by 4 we have

$$a_{15}(2a_{16}+1)+2c_{15}c_{16}=0,$$

which forces a_{15} to be even and thus, by (5) again

$$a_{15}=0,\ c_{15}=\pm 1; \tag{6}$$

$[x_2 y_3^2]$ $8a_6 a_{15}+4b_6 b_{15}+8c_6 c_{15}=0$

so by (5) and (6)

$$c_6=0; \tag{7}$$

$[x_2^2]$ $4a_{11}^2+2b_{11}^2+4c_{11}^2=2$

so

$$a_{11}=c_{11}=0,\ b_{11}=\pm 1; \tag{8}$$

$[x_2^2 y_3]$ $8a_6 a_{11}+4b_6 b_{11}+8c_6 c_{11}=0$

so by (8)

$$b_6=0. \tag{9}$$

Finally we consider the coefficient of $x_2^2 y_3^2$ in (1). We have

$$4a_6^2+2b_6^2+4c_6^2=8.$$

Appealing to (7) and (9) we obtain the required contradiction $a_6^2=2$.

Panaitopol [**5**] has shown that the only diagonal ternary quadratic forms $ax^2+by^2+cz^2$ $(1\le a\le b\le c)$, which represent every odd positive integer are the forms $x^2+y^2+2z^2$, $x^2+2y^2+3z^2$, and $x^2+2y^2+4z^2$. Our proof shows that the representability of odd integers by $x^2+y^2+2z^2$ and $x^2+2y^2+4z^2$ does not arise from an underlying composition law. We leave it to the reader to show also that $x^2+2y^2+3z^2$ does not possess such a composition law.

REFERENCES

1. A. A. Albert (editor), *Studies in Modern Algebra*, Vol. 2, MAA Studies in Mathematics, 1963.
2. L. E. Dickson, On quaternions and their generalizations and the history of the 8-square theorem, *Annals of Math.* **20** (1919) 155–171.
3. L. E. Dickson, *Modern Elementary Theory of Numbers,* The University of Chicago Press, Chicago, Illinois, 1947.
4. I. Kaplansky, Ternary positive quadratic forms that represent all odd positive integers, *Acta Arith.* **70** (1995) 209–214.
5. L. Panaitopol, On the representation of natural numbers as sums of squares, *Amer. Math. Monthly* **112** (2005) 168–171.
6. A. R. Rajwade, *Squares*, London Mathematical Lecture Note Series 171, Cambridge University Press, 1993.

PROBLEMS

Proposals

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

1766. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let f be differentiable on $(0, \infty)$ and let ω be a positive real number. Prove that if $\lim_{x\to\infty}(f'(x) + \omega f(x)) = A$, then $\lim_{x\to\infty} f(x) = A/\omega$.

1767. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let G be the centroid of $\triangle ABC$. Prove that if $\angle BAC = 60°$ and $\angle BGC = 120°$, then the triangle is equilateral.

1768. *Proposed by G.R.A.20 Problem Solving Group, Rome, Italy.*

For which positive integers n can the set $\{1, 2, \dots, 2n\}$ be partitioned into n two element subsets so that the sum of the two numbers in each subset is a perfect square?

1769. *Proposed by Michel Bataille, Rouen, France.*

For positive integer n, let

$$P_n(x, y) = \sum_{k=0}^{n} \binom{2n+1}{2k+1} x^{n-k}(x+y)^k.$$

Find a closed form expression for the coefficient of $x^i y^j$ when P_n is expanded.

1770. *Proposed by Scott N. Armstrong, University of California, Berkeley, CA, and Christopher J. Hillar, Texas A&M University, College Station, TX.*

Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be nonnegative real numbers summing to 1, and let $a_1, a_2 \dots, a_k$ be complex numbers. For $n > k$, define

$$a_n = \lambda_1 a_{n-1} + \lambda_2 a_{n-2} + \cdots + \lambda_k a_{n-k}.$$

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

Solutions should be written in a style appropriate for this MAGAZINE.

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

Prove that if there is a j, $1 \le j \le n-1$, such that λ_j and λ_{j+1} are both nonzero, then $\lim_{n\to\infty} a_n$ exists. In addition, determine the value of this limit.

Quickies

Answers to the Quickies are on page 151.

Q969. *Proposed by Ovidiu Furdui, Western Michigan University, Kalamazoo, MI.*

Let n be a positive integer and k a natural number. Show that

$$\int_0^1 (x-x^2)^k \{nx\}\,dx = \frac{(k!)^2}{2(2k+1)!},$$

where $\{a\} = a - \lfloor a \rfloor$ denotes the fractional part of a.

Q970. *Proposed by Li Zhou, Polk Community College, Winter Haven, FL.*

Let x be a positive real number and let m and n be integers with $m \le n$. Evaluate

$$\prod_{k=m}^{n} \left(x^{2^{k+1}} - x^{2^k} + 1 \right).$$

Solutions

Switch + and = **April 2006**

1741. *Proposed by Shahin Amrahov, ARI College, Ankara, Turkey.*

Find all positive integer triples (k, m, n) that solve

a. $2^k + 9^m = 7^n$.
b. $2^k = 9^m + 7^n$.

Solution by Lloyd Husbands, Josh Nichols-Barrer, Yanir A. Rubinstein, and Olof Sisask, Massachusetts Institute of Technology, Cambridge, MA.

a. Reducing the equation modulo 7 we find

$$2^k + 2^m \equiv 0 \pmod 7.$$

This is impossible because for positive integer r, $2^r \equiv 1, 2,$ or $4 \pmod 7$.

b. The only solution is $(k, m, n) = (4, 1, 1)$. Considering the equation modulo 3, we find

$$(-1)^k \equiv 2^k \equiv 7^n \equiv 1 \pmod 3.$$

It follows that $k = 2k'$ is even. We can then rewrite the equation as

$$7^n = 2^k - 9^m = (2^{k'} + 3^m)(2^{k'} - 3^m).$$

Because the difference of the two factors, $(2^{k'} + 3^m) - (2^{k'} - 3^m) = 2 \cdot 3^m$ is not a multiple of 7, it is not possible that both factors are multiples of 7. Thus one of the factors is 1 and it must be the smaller factor, that is,

$$2^{k'} - 3^m = 1.$$

If $k' \geq 3$ then modulo 8 this equation becomes $-3^m \equiv 1 \pmod 8$, which is impossible. It is then easy to check that the only solution is $k' = 2$, $m = 1$. This leads to the solution $(k, m, n) = (4, 1, 1)$.

Also solved by Armstrong Problem Solvers, Brian D. Beasley, Robert Calcaterra, Minh Can, John Christopher, Chip Curtis, Robert L. Doucette, Elias Lampakis (Greece), Lenny Jones and Mike Long, Peter W. Lindstrom, David Lovit, David E. Manes, José H. Nieto (Venezuela), Northwestern University Math Problem Solving Group, Gabriel T. Prăjitură, Jeffrey Silver, Albert Stadler (Switzerland), Marian Tetiva (Romania), Gary L. Walls, Li Zhou, and the proposer. Michel Bataille (France), and Paul Weisenhorn (Germany), solved part a. There was one incorrect submission.

Similar triangles **April 2006**

1742. *Proposed by Luz M. DeAlba and Jeffrey Langford (student), Drake University, Des Moines, IA.*

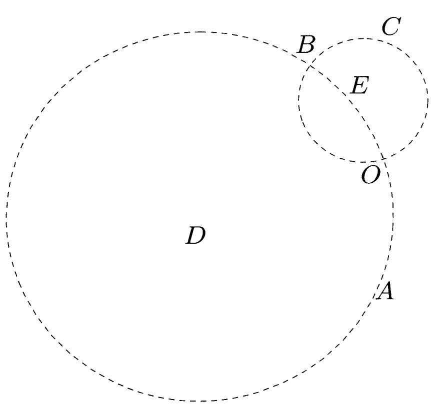
Let $\mathcal{C}$ be a circle with center O and diameter $\overline{AC}$ and let B be any point on $\mathcal{C}$ different from A and C. Let D be the point of intersection of the perpendicular bisectors of $\overline{OA}$ and $\overline{OB}$, and let E be the point of intersection of the perpendicular bisectors of $\overline{OC}$ and $\overline{OB}$. Prove that $\triangle DBE$ is similar to $\triangle ABC$.

Solution by Victor Y. Kutsenok, University of St. Francis, Fort Wayne, IN.

Note that E is the circumcenter of triangle OCB and D is the circumcenter of triangle OBA. Thus

$$\angle DEB = \frac{1}{2}\angle OEB = \angle ACB \qquad \text{and} \qquad \angle EDB = \frac{1}{2}\angle ODB = \angle CAB.$$

It follows that triangles DBE and ABC are similar.



Also solved by ABC Student Problem Solving Group, Herb Bailey, Michel Bataille (France), Jany C. Binz (Switzerland), Robert Calcaterra, Minh Can, Miguel Amengual Covas, Gordon Crandall, Chip Curtis, Robert L. Doucette, Rodrigo Flores (Chile), Michael Goldenberg and Mark Kaplan, John G. Heuver (Canada), Elias Lampakis (Greece), Lau Sai Luk (Hong Kong), Junaid N. Mansuri, Charles McEachern, José H. Nieto (Venezuela), Kees Onneweer, Gabriel T. Prăjitură, Jawad Sadek, Volkhard Schindler (Germany), Jeffrey Silver and Christopher Mackeprang, Raúl A. Simon (Chile), Seshadri Sivakumar (Canada), Albert Stadler (Switzerland), H. T. Tang, Marian Tetiva (Romania), R. S. Tiberio, Michael Vowe (Switzerland), Paul Weisenhorn (Germany), James G. Wendelberger, Doug Wilcock, Paul Wilfong and Harry Wilfong, Stuart V. Witt, Tom Zerger, Li Zhou, and the proposers. There was one solution with no name.

Bounding a sum **April 2006**

1743. *Proposed by David P. Lang, Wentworth Institute of Technology, Boston, MA.*

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers. For positive integer n, define $s_n = \sum_{j=1}^{n} a_j$, and define S_n by

$$S_n = \sum_{k=1}^{n} \frac{a_k}{s_n - a_k}.$$

Prove that if $L = \lim_{n\to\infty} S_n$ exists, then $L \geq 1$.

Many readers gave a solution along the following lines.

For $n \geq 2$ and $k \leq n$, we have $\frac{a_k}{s_n - a_k} \geq \frac{a_k}{s_n}$. Therefore

$$S_n = \sum_{k=1}^{n} \frac{a_k}{s_n - a_k} \geq \sum_{k=1}^{n} \frac{a_k}{s_n} = 1.$$

Thus if $L = \lim_{n\to\infty} S_n$ exists, then $L \geq 1$.

Solved by Michael Andreoli, Michel Bataille (France), Paul Bracken, J. L. Díaz-Barrero and M. Grau-Sánchez (Spain), Robert Calcaterra, Minh Can, Gordon Crandall, Chip Curtis, Prithwijit De (Ireland), Robert L. Doucette, Michael Goldenberg and Mark Kaplan, Peter Gressis, Eugene A. Herman, Elias Lampakis (Greece), Peter W. Lindstrom, S. C. Locke, Behailu Mammo, José H. Nieto (Venezuela), Northwestern University Math Problem Solving Group, Kees Onneweer, Paolo Perfetti (Italy), Gabriel T. Prăjitură, Henry Ricardo, Edward Schmeichel, Albert Stadler (Switzerland), Marian Tetiva (Romania), Michael Vowe (Switzerland), Paul Weisenhorn (Germany), Stuart V. Witt, Li Zhou, and the proposer. There was one solution with no name. There was one incorrect submission.

A sum of products **April 2006**

1744. *Proposed by Michael Goldenberg and Mark Kaplan, The Ingenuity Project, Baltimore Polytechnic Institute, MD.*

For real number $x \geq 1$, define $a_1 = 2x$ and $a_{n+1} = a_n^2 - 2$, $n = 1, 2, 3, \ldots$. Find a closed form expression for

$$S(x) = \sum_{n=1}^{\infty} \prod_{k=1}^{n} a_k^{-1}.$$

Solution by Michel Bataille, Rouen, France.

We show that $S(x) = x - \sqrt{x^2 - 1}$.

The result holds for $x = 1$ because then $a_n = 2$, and $\prod_{k=1}^{n} a_k^{-1} = \frac{1}{2^n}$ for $n = 1, 2, 3, \ldots$, and it follows that $S(1) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$.

Now let $x > 1$ and define $u > 0$ by $\cosh u = x$. Using the formula $\cosh 2\alpha = 2\cosh^2\alpha - 1$, an easy induction argument can be used to prove that $a_n = 2\cosh(2^{n-1}u)$ for all positive integers n. Thus,

$$(\sinh u)a_1 a_2 \cdots a_n = 2^n(\sinh u)(\cosh u)(\cosh(2u)) \cdots (\cosh(2^{n-1}u)).$$

Repeated use of the formula $\sinh 2\alpha = 2\sinh\alpha\cosh\alpha$ then yields

$$a_1 a_2 \cdots a_n = \frac{\sinh(2^n u)}{\sinh u}.$$

Because

$$\frac{1}{\tanh(2^{n-1}u)} - \frac{1}{\tanh(2^n u)} = \frac{1}{\sinh(2^n u)},$$

it follows that

$$S(x) = \sum_{n=1}^{\infty} \frac{1}{a_1 a_2 \cdots a_n} = (\sinh u) \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{\tanh(2^{n-1}u)} - \frac{1}{\tanh(2^n u)} \right)$$
$$= (\sinh u) \left(\frac{1}{\tanh u} - 1 \right) = e^{-u}.$$

Finally, solving $e^u + e^{-u} = 2x$ for e^{-u} gives $S(x) = e^{-u} = x - \sqrt{x^2 - 1}$.

Also solved by Dione Bailey and Elsie Campbell and Charles Diminnie, Paul Bracken and N. Nadeau, Robert Calcaterra, Minh Can, Knut Dale (Norway), Prithwijit De (Ireland), Jim Delany, Robert L. Doucette, G.R.A.20 Problem Solving Group (Italy), Peter W. Lindstrom, José H. Nieto (Venezuela), Northwestern University Math Problem Solving Group, Volkhard Schindler (Germany), Nicholas C. Singer, Albert Stadler (Switzerland), Marian Tetiva (Romania), Michael Vowe (Switzerland), Li Zhou, and the proposer. There were three incorrect submissions.

A continued fraction and e — April 2006

1745. *Proposed by Gerald A. Edgar, The Ohio State University, Columbus, OH.*

For positive real number r, define

$$A_0(r) = r, \qquad A_{n+1}(r) = r + \frac{A_n(r)}{A_n(r+1)}, \qquad n = 0, 1, 2, \ldots.$$

Show that $\lim_{n\to\infty} A_n(1)$ exists and find the value of the limit.

Note: Expanding the recursive definition when $r = 1$ one leads to the elaborate continued fraction-like configuration

$$1 + \cfrac{1 + \cfrac{1 + \frac{1+\cdots}{2+\cdots}}{2 + \frac{2+\cdots}{3+\cdots}}}{2 + \cfrac{2 + \frac{2+\cdots}{3+\cdots}}{3 + \frac{3+\cdots}{4+\cdots}}} \Bigg/ \cdots$$

$$1 + \cfrac{1 + \cfrac{1 + \cfrac{1 + \frac{1+\cdots}{2+\cdots}}{2 + \frac{2+\cdots}{3+\cdots}}}{2 + \cfrac{2 + \frac{2+\cdots}{3+\cdots}}{3 + \frac{3+\cdots}{4+\cdots}}}}{2 + \cfrac{2 + \cfrac{2 + \frac{2+\cdots}{3+\cdots}}{3 + \frac{3+\cdots}{4+\cdots}}}{3 + \cfrac{3 + \frac{3+\cdots}{4+\cdots}}{4 + \frac{4+\cdots}{5+\cdots}}}}.$$

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

We prove that $\lim_{n\to\infty} A_n(1) = \frac{e}{e-1}$. The proof follows from three claims, each of which is proved by induction.

CLAIM 1. For every $r > 0$ and integer $n \geq 0$, $r \leq A_n(r) < r + 1$.

The claim is true for $n = 0$. If the claim is true for $n = k$, then

$$0 < \frac{A_k(r)}{A_k(r+1)} < 1 \qquad \text{and hence,} \qquad r < A_{k+1}(r) < r + 1.$$

CLAIM 2. For every $r \geq \frac{3}{2}$ and integer $n > 0$, $|A_n(r) - A_{n-1}(r)| \leq (\frac{4}{5})^{n-1}$.

It follows from Claim 1 that $|A_1(r) - A_0(r)| < 1$. If $n > 1$, then

$$
\begin{aligned}
|A_{n+1}(r) - A_n(r)| &= \left|\frac{A_n(r)}{A_n(r+1)} - \frac{A_{n-1}(r)}{A_{n-1}(r+1)}\right| \\
&= \left|\frac{A_n(r) - A_{n-1}(r)}{A_n(r+1)} - \frac{A_{n-1}(r)}{A_{n-1}(r+1)} \cdot \frac{A_n(r+1) - A_{n-1}(r+1)}{A_n(r+1)}\right| \\
&\le \frac{|A_n(r) - A_{n-1}(r)|}{A_n(r+1)} + \frac{|A_n(r+1) - A_{n-1}(r+1)|}{A_n(r+1)}.
\end{aligned}
$$

Because $A_n(r+1) \ge r+1 \ge \frac{5}{2}$, Claim 2 follows by induction on n.

Note that by applying Claim 2 to a telescoping sum, we can conclude that

$$|A_j(r) - A_k(r)| \le 5\left(\frac{4}{5}\right)^n$$

for $j \ge k \ge n$. It then follows from Cauchy's Criterion that $\lim_{n\to\infty} A_n(r)$ exists for all $r \ge \frac{3}{2}$.

CLAIM 3. Let $L(r) = \lim_{n\to\infty} A_n(r)$, $c = L(2)$, and $b_n = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$. Then for integer $r \ge 2$,

$$L(r) = (-1)^r \frac{c}{(r-1)!(1 - cb_{r-1})}.$$

Claim 3 is true for $r = 2$. Because

$$L(r) = r + \frac{L(r)}{L(r+1)}, \qquad \text{it follows that} \qquad L(r+1) = \frac{L(r)}{L(r) - r}.$$

Claim 3 can now be proved by a routine induction argument.

Now consider

$$\frac{L(r)}{r} = (-1)^r \frac{c}{r!(1 - cb_{r-1})}.$$

By Claim 1,

$$\lim_{r\to\infty} \frac{L(r)}{r} = 1.$$

Combining this with

$$\lim_{r\to\infty} \frac{c}{r!} = 0 \qquad \text{and} \qquad \lim_{r\to\infty} (1 - cb_{r-1}) = 1 - \frac{c}{e},$$

we conclude that $L(2) = c = e$. Hence, for n sufficiently large, $A_n(2)$ is larger than $\frac{5}{2}$, so an argument similar to that used to prove Claim 2 can be used to prove that $L(1)$ exists. Finally, from $L(1) = 1 + \frac{L(1)}{L(2)}$ we find $L(1) = \frac{e}{e-1}$.

Also solved by Peter W. Lindstrom, S. C. Locke and A. D. Meyerowitz, Dave Trautman, Paul Weisenhorn (Germany), and the proposer. There were two incorrect submissions.

Answers

Solutions to the Quickies from page 146.

A969. First observe that if n is a natural number, then $\{n(1-y)\} = 1 - \{ny\}$ for all $y \in [0, 1]$, except for $y = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$. Let I be the value of the integral and make the substitution $x = 1 - y$. Because the value of a Riemann integral is unchanged if the value of the integrand is changed on a discrete set, we obtain

$$I = \int_0^1 (x - x^2)^k \{nx\}\, dx = \int_0^1 (y - y^2)^k \{n(1-y)\}\, dy = \int_0^1 (y - y^2)^k \,(1 - \{ny\})\, dy$$
$$= \int_0^1 (y - y^2)^k\, dy - I = \beta(k+1, k+1) - I = \frac{(k!)^2}{(2k+1)!} - I,$$

where β denotes the beta function. The result follows.

A970. Observe that

$$x^{2^{k+1}} - x^{2^k} + 1 = \frac{x^{2^{k+2}} + x^{2^{k+1}} + 1}{x^{2^{k+1}} + x^{2^k} + 1}$$

for any integer k. With this substitution, the product becomes a telescoping product and simplifies to

$$\frac{x^{2^{n+2}} + x^{2^{n+1}} + 1}{x^{2^{m+1}} + x^{2^m} + 1}.$$

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.

Fraga, Robert (ed.), *War Stories from Applied Math: Undergraduate Consultancy Projects*, MAA, 2007; vii + 147 pp, $48.95 (P) (member: $39.50). ISBN 978-0-88385-181-4.

At most institutions, the pure mathematicians long ago won the skirmish with the applied mathematicians for control of the department's mission. However, the department as a whole has been losing the struggle for students and for recognition in the institution. One result is that more applied mathematics is taught outside of mathematics departments than inside. Departments seeking to change the balance find that applied mathematicians are rare and often more expensive than the dean will afford. The "war stories" in this collection are from a different front, the one where mathematics meets the "real world." They recount very successful campaigns to teach undergraduates consulting and research in "industrial mathematics," at a variety of institutions—a liberal arts university (Baker University), a management training program for liberal arts students (Indiana University), a private urban university (Marquette), a public university (Towson), a science college (Harvey Mudd), and an engineering school (Milwaukee School of Engineering). The contributions address finding industrial contacts, integration of project work into course work, project deliverables, and management of such a program (including funding). A discouraging confirmatory note: One author moved from his department of mathematical sciences to the department of electrical and computer engineering, "hoping to find the climate in engineering a bit more encouraging for applications and for students' careers." (Disclosure: I have a contribution in this book, but I like the others better.)

Hayes, Brian, Foolproof, *American Scientist* 95 (1) (January–February 2007) 10–15; `http://www.americanscientist.org/template/AssetDetail/assetid/54428`.

Not many articles begin, "I was a teenage trisector" (though not the kind you may think). Hayes tried to convince a fellow worker drawing scales on pointer meters that exact trisection of most angles is impossible. That Hayes could not convince his colleague left the question hanging, "if proof is a magic wand that works only in the hands of wizards, what is its utility to the rest of us?" Hayes recounts Thomas Hobbes's remarkable "epiphany" with Euclid's *Elements* (whose logic Hobbes unfortunately failed to master, becoming a mathematical crank). Hayes relates the story of the proof of the four-color theorem and cites Perelman's proof of the Poincaré conjecture, caricatures (at his own expense) Plato's dialogue *Meno* with the slave boy, and concludes with an outline of Wantzel's proof of nontrisectability. Along the way, Hayes (who characterizes himself as "not a mathematician" but "an embedded journalist in the math corps") claims that there is a kind of crisis going on about proof in mathematics, "but only because the entire history of mathematics is just one crisis after another. The foundations are *always* crumbling, and the barbarians are *always* at the gate." Gee, that, added to students' fear of mathematics and hatred of science (or is it the other way round?), should let you sleep easy.

Langville, Amy N., and Carl D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, 2006; x + 224 pp, $35. ISBN 0-691-12202-4, 978-0-691-12202-1. Chartier, Timothy, Googling Markov, *The UMAP Journal* 27 (1) (2006) 17–30. Wills, Rebecca, Google's PageRank: The math behind the search engine, *Mathematical Intelligencer* 28 (4) (Fall 2006) 6–11.

Would Google have achieved search-engine supremacy without its PageRank algorithm? No way. Students and the general population should know that this algorithm, which presents search results in order of "importance" (number of links to the result), is an application of pure mathematics from long ago that touches daily lives today. Mathematics majors ought to know the linear algebra behind the algorithm (Perron-Frobenius theorem applied to a constantly-growing matrix of perhaps 25 billion rows)—and they can learn it in varying depth from these sources.

Alsina, Claudi, and Roger B. Nelsen, *Math Made Visual: Creating Images for Understanding Mathematics*, MAA, 2006; xv + 173 pp, $49.95 (member: $39.95). ISBN 0-88385-746-4. Casselman, Bill, *Mathematical Illustrations: A Manual of Geometry and PostScript*, Cambridge University Press, 2005; ix + 318 pp, $90, $39 (P). ISBN 0-521-83921-1, 0-521-54788-1.

"Manuscripts should be decorated so that their appearance alone will induce perusal" (Johannes Trithemius, c. 1492). Once upon a time, mathematicians sent typescript to a journal, which typeset it and had artists render the author's sketched figures into publication-quality illustrations. The forced compound interest of ever-greater productivity has abolished those practices. Fortunately, Donald Knuth gave us $\TeX$ for producing beautiful mathematical text on our own. But from handouts to lecture notes to published papers, mathematicians now must also generate their own figures (pity the geometers, whose need is the greatest!). The first of the two books listed focuses on a wealth of wonderful suggestions on *how to visualize* a mathematical idea, independent of technology (in fact, the authors give no hint about how their illustrations were prepared). Casselman's book gives details of *how to produce mathematical graphics*, specifically via PostScript, which gives complete control over the result; so does MetaPost, but Casselman finds it less intuitive. He also admits that integrating text (labels) into PostScript figures has not yet been made as easy as it should be.

Newman, Mark, Albert-László Barabási, and Duncan J. Watts, *The Structure and Dynamics of Networks*, Princeton University Press, 2006; x + 582 pp, $89.50, $49.50 (P). ISBN 10: 0-691-11356-4, ISBN 13: 978-0-691-11356-2; ISBN 10: 0-691-11357-2, ISBN 13: 978-0-691-11356-9.

Many systems, both natural and artificial, from metabolism to food webs, from friendship to terrorism, from can be represented as networks. This book reprints fundamental papers on aspects of networks: historical development, empirical studies, models (including random, small-world, and scale-free networks), and applications (epidemics and rumors, robustness, and searching). The editors emphasize that the "new" science of networks goes beyond graph theory in being concerned with empirical as well as theoretical questions, in approaching networks as dynamic objects, and in aiming to understand networks "not just as topological objects, but also as the framework upon which distributed dynamical systems are built."

Aczel, Amir D., *Descartes' Secret Notebook: A True Tale of Mathematics, Mysticism, and the Quest to Understand the Universe*, Broadway Books, 2005; xiv + 274 pp, $24.95. ISBN 0-7679-2033-3.

This is a cleverly-told tale of a notebook left in a strongbox by Descartes, which has been lost except for 10% that was copied by Leibniz. I am not a fan of mixing history with historical fiction (made-up dialogue, insinuated emotions); but I enjoyed what amounts to a popular biography of Descartes, including personal delight in learning where he had the famous dreams that set him on the path to mathematics (though neither "Descartes, dreams" nor "Neuburg an der Donau" occurs in the index). I won't spoil the ending but must say that, despite Rosicrucian echoes, many readers will find the final "cosmic formula" not nearly as sensational as the suppositions of Dan Brown's *The Da Vinci Code*. (In the text and in the index, "Lull" should be "Llull.")

NEWS AND LETTERS

67th Annual William Lowell Putnam Mathematical Competition

Editors Note: Additional solutions will be printed in the *Monthly* later in the year.

PROBLEMS

A1. Find the volume of the region of points (x, y, z) such that

$$(x^2 + y^2 + z^2 + 8)^2 \leq 36(x^2 + y^2).$$

A2. Alice and Bob play a game in which they take turns removing stones from a heap that initially has n stones. The number of stones removed at each turn must be one less than a prime number. The winner is the player who takes the last stone. Alice plays first. Prove that there are infinitely many n such that Bob has a winning strategy. (For example, if $n = 17$, then Alice might take 6 leaving 11; then Bob might take 1 leaving 10; then Alice can take the remaining stones to win.)

A3. Let 1, 2, 3, ..., 2005, 2006, 2007, 2009, 2012, 2016, ... be a sequence defined by $x_k = k$ for $k = 1, 2, \ldots, 2006$ and $x_{k+1} = x_k + x_{k-2005}$ for $k \geq 2006$. Show that the sequence has 2005 consecutive terms each divisible by 2006.

A4. Let $S = \{1, 2, \ldots, n\}$ for some integer $n > 1$. Say a permutation π of S has a *local maximum* at $k \in S$ if

(i) $\pi(k) > \pi(k+1)$	for $k = 1$
(ii) $\pi(k-1) < \pi(k)$ and $\pi(k) > \pi(k+1)$	for $1 < k < n$
(iii) $\pi(k-1) < \pi(k)$	for $k = n$

(For example, if $n = 5$ and π takes values at 1, 2, 3, 4, 5 of 2, 1, 4, 5, 3, then π has a local maximum of 2 at $k = 1$, and a local maximum of 5 at $k = 4$.) What is the average number of local maxima of a permutation of S, averaging over all permutations of S?

A5. Let n be a positive odd integer and let θ be a real number such that θ/π is irrational. Set $a_k = \tan(\theta + k\pi/n)$, $k = 1, 2, \ldots, n$. Prove that

$$\frac{a_1 + a_2 + \cdots + a_n}{a_1 a_2 \cdots a_n}$$

is an integer, and determine its value.

A6. Four points are chosen uniformly and independently at random in the interior of a given circle. Find the probability that they are the vertices of a convex quadrilateral.

B1. Show that the curve $x^3 + 3xy + y^3 = 1$ contains only one set of three distinct points A, B, and C, which are the vertices of an equilateral triangle, and find its area.

B2. Prove that, for every set $X = \{x_1, x_2, \ldots, x_n\}$ of n real numbers, there exist a non-empty subset S of X and an integer m such that

$$\left| m + \sum_{s \in S} s \right| \leq \frac{1}{n+1}.$$

B3. Let S be a finite set of points in the plane. A *linear partition* of S is an unordered pair $\{A, B\}$ of subsets of S such that $A \cup B = S$, $A \cap B = \emptyset$, and A and B lie on opposite sides of some straight line disjoint from S (A or B may be empty). Let L_S be the number of linear partitions of S. For each positive integer n, find the maximum of L_S over all sets S of n points.

B4. Let Z denote the set of points in $\mathbb{R}^n$ whose coordinates are 0 or 1. (Thus Z has 2^n elements, which are the vertices of a unit hypercube in $\mathbb{R}^n$.) Given a vector subspace V of $\mathbb{R}^n$, let $Z(V)$ denote the number of members of Z that lie in V. Let k be given, $0 \le k \le n$. Find the maximum, over all vector subspaces $V \subseteq \mathbb{R}^n$ of dimension k, of the number of points in $V \cap Z$.

B5. For each continuous function $f : [0, 1] \to \mathbb{R}$, let $I(f) = \int_0^1 x^2 f(x)\,dx$ and $J(f) = \int_0^1 x(f(x))^2\,dx$. Find the maximum value of $I(f) - J(f)$ over all such functions f.

B6. Let k be an integer greater than 1. Suppose $a_0 > 0$, and define $a_{n+1} = a_n + (1/\sqrt[k]{a_n})$ for $n \ge 0$. Evaluate $\lim_{n\to\infty} a_n^{k+1}/n^k$.

SOLUTIONS

Solution to A1. The answer is $6\pi^2$.

Convert to cylindrical coordinates by substituting $r^2 = x^2 + y^2$ so the inequality becomes $(r^2 + z^2 + 8)^2 \le 36r^2$, which is equivalent to $(r-3)^2 + z^2 \le 1$. The region is obtained by rotating the area in the x-z plane defined by $(x-3)^2 + z^2 \le 1$ about the z-axis. A standard method of shells integration gives the answer,

$$V = \int_2^4 2\sqrt{1-(x-3)^2} \cdot 2\pi x\,dx = 4\pi \int_{-1}^1 \sqrt{1-u^2} \cdot (u+3)\,du$$

$$= 4\pi \left(\int_{-1}^1 u\sqrt{1-u^2}\,du + 3\int_{-1}^1 \sqrt{1-u^2}\,du \right) = 4\pi \left(0 + 3\frac{\pi}{2}\right) = 6\pi^2$$

substituting $u = x - 3$ and because the the first integral is of an odd function while the second is the area of a unit semicircle.

Solution to A2. Suppose there are only finitely many n such that Bob will win if Alice starts with n stones. The smallest such n is 3, and let us say the largest is N where $N \ge 8$. Consider the game starting with $N! - 1$ stones. Alice must remove $p - 1$ stones, for p a prime number. Since $N!$ isn't prime, she can't take all the stones, and she must leave $m = N! - 1 - (p-1) = N! - p$ stones where $3 \le m \le N$. But this implies that m divides $p = N! - m$, contradicting the choice of p to be prime. Thus no such largest N can exist. That is, there must be infinitely many n for which Bob has a winning strategy.

Solution to A3. We need only consider the sequence mod 2006. The continuation of the sequence is determined by any consecutive 2006 elements $x_{k-2005}, x_{k-2004}, \ldots, x_{k-1}, x_k$ by the recursion relation. But there are only a finite number of such subsequences mod 2006, so the same subsequence must appear again eventually and so the sequence is eventually periodic say with a period L. On the other hand, the recursion relation can be rewritten to determine the elements of the sequence before a given block of 2006 elements $x_{k-2004}, \ldots, x_k, x_{k+1}$, as $x_{k-2005} = x_{k+1} - x_k$. Thus if $x_{k-2005} = x_{k-2005+L}, \ldots, x_{k-1} = x_{k-1+L}, x_k = x_{k+L}$, then working backward to the starting values, $1 = x_1 = x_{L+1}, 2 = x_2 = x_{L+2}, \ldots, 2006 = x_{2006} = x_{L+2006}$ and the sequence is periodic from the beginning. Extend the sequence further backward to define $x_0 = x_{2006} - x_{2005} = 1, x_{-1} = x_{2005} - x_{2004} = 1, \ldots, x_{-2004} = x_2 - x_1 = 1$ and then $x_{-2005} = x_1 - x_0 = 0, x_{-2006} = x_0 - x_{-1} = 0, \ldots, x_{-4009} = x_{-2003} - x_{-2004} = 0$. But then $L > 4009$ and $x_{L-4009} \equiv \cdots \equiv x_{L-2005} \equiv 0 \bmod 2006$, is a subsequence of 2005 elements of the original sequence each divisible by 2006.

Solution to A4. The answer is $(n+1)/3$. We need to determine the total number of local maxima of permutations of S and divide by the total number $n!$ of such permutations. We total the number of permutations that have a local maximum at k for $1 \le k \le n$. For $k = 1$, there are

$\binom{n}{2}(n-2)! = n!/2$ such obtained by choosing which two elements are the first two values of the permutation, knowing that the larger will be first if and only if the permutation has a local maximum at 1, followed by any order of the remaining $n-2$ values. The same logic applies for the case $k=n$. If $1 < k < n$, first choose the three values for $\{\pi(k-1), \pi(k), \pi(k+1)\}$. This π will have a relative maximum at k iff these three values occur in one of two allowable orders out of the six possible orders, the highest value must be $\pi(k)$ and the second highest can be either $\pi(k-1)$ or $\pi(k+1)$. In any case, there are $(n-3)!$ ways of ordering the remaining the $n-3$ values. Hence there $2\binom{n}{3}(n-3)! = n!/3$ permutations having a local maximum at an k with $1 < k < n$. The total number of local maxima at all positions is thus $2n!/2 + (n-2)n!/3 = (n+1)n!/3$. Thus the average number of local maxima for permutations of S is $(n+1)/3$.

Solution to A5. The answer is $n(-1)^{(n-1)/2}$. Consider the complex number $\omega = \cos(\theta) + i\sin(\theta)$. The roots of the equation

$$\left(\frac{1+ix}{1-ix}\right)^n = \omega^{2n}$$

are precisely the $a_k = \tan(\theta + \frac{k\pi}{n})$, for $k = 1, \ldots, n$, since we check these are roots

$$\left(\frac{1+ia_k}{1-ia_k}\right)^n = \left(\frac{\cos\left(\theta+\frac{k\pi}{n}\right) + i\sin\left(\theta+\frac{k\pi}{n}\right)}{\cos\left(\theta+\frac{k\pi}{n}\right) - i\sin\left(\theta+\frac{k\pi}{n}\right)}\right)^n = \left(\frac{\omega}{\bar{\omega}}\right)^n = \frac{\omega^{2n}}{|\omega|^{2n}} = \omega^{2n},$$

there are n of them, and the equation expands to a degree n polynomial equation

$$\begin{aligned} 0 &= (1+ix)^n - \omega^{2n}(1-ix)^n \\ &= (1-\omega^{2n}) + ni(1+\omega^{2n})x + \cdots + ni^{n-1}(1-\omega^{2n})x^{n-1} + i^n(1+\omega^{2n})x^n. \end{aligned}$$

The sum of the zeros of this polynomial is the negative of the coefficient on x^{n-1} divided by the coefficient on x^n

$$\frac{-ni^{n-1}(1-\omega^{2n})}{i^n(1+\omega^{2n})}$$

and the product of the zeros is the negative (since n is odd) of the constant coefficient divided by the coefficient of x^n

$$\frac{-(1-\omega^{2n})}{i^n(1+\omega^{2n})}.$$

Thus

$$\frac{a_1 + a_2 + \cdots + a_n}{a_1 a_2 \cdots a_n} = \frac{-ni^{n-1}(1-\omega^{2n})}{-(1-\omega^{2n})} = n(-1)^{\frac{n-1}{2}}.$$

Solution to A6 (from a student paper). The answer is $1 - \frac{35}{12\pi^2}$.

We may assume that our circle has unit radius and is centered at the origin. Suppose that three points P_1, P_2, P_3 are chosen with coordinates $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, going in counterclockwise order about the origin. Let T be the triangle formed by $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$. Then

$$\text{Area}(T) = \left|\tfrac{1}{2}\left((\mathbf{X}_1 \times \mathbf{X}_2)\cdot\mathbf{k} + (\mathbf{X}_2 \times \mathbf{X}_3)\cdot\mathbf{k} + (\mathbf{X}_3 \times \mathbf{X}_1)\cdot\mathbf{k}\right)\right|.$$

The enclosed quantity is positive unless one of the points is inside the triangle formed by the other two points and the origin.

Let $\theta_{1,2}$ denote the counterclockwise angle from $\mathbf{X}_1$ to $\mathbf{X}_2$. From probability theory, the distribution of $\theta_{1,2}$, the smallest order statistic, is $\dfrac{2}{(2\pi)^2}(2\pi - \theta)$.

LEMMA 1. $E[\sin(\theta_{1,2})] = \frac{1}{\pi}$.

Proof. We have

$$E[\sin(\theta_{1,2})] = \int_0^{2\pi} \frac{2}{(2\pi)^2}(2\pi - \theta)\sin\theta\, d\theta$$
$$= \frac{2}{(2\pi)^2}\left[-2\pi\cos\theta + \theta\cos\theta - \sin\theta\right]_0^{2\pi} = \frac{2}{(2\pi)^2}(2\pi) = \frac{1}{\pi}.$$

Let $r_i = |\mathbf{X}_i|$, the distance between P_i and the origin.

LEMMA 2. $E[r_i] = \frac{2}{3}$.

Proof. The probability distribution function for r_i is $2r$, so

$$E[r_i] = \int_0^1 r \cdot 2r\, dr = \int_0^1 2r^2\, dr = \left.\frac{2r^3}{3}\right|_0^1 = \frac{2}{3}.$$

If none of P_1, P_2, P_3 is inside the triangle defined by the origin and the other two points, and T' is the triangle formed by P_1, P_2, P_3,

$$E[\text{Area}(T')] = \tfrac{1}{2}E\big[(\mathbf{X}_1 \times \mathbf{X}_2)\cdot\mathbf{k} + (\mathbf{X}_2 \times \mathbf{X}_3)\cdot\mathbf{k} + (\mathbf{X}_3 \times \mathbf{X}_1)\cdot\mathbf{k}\big].$$

The expectation of the sum is the sum of the expectations, and these are equal by symmetry. Furthermore, $\theta_{1,2}$, r_1, r_2 are independent of each other, so we have

$$E[\text{Area}(T')] = \frac{3}{2}E[(\mathbf{X}_1 \times \mathbf{X}_2)\cdot\mathbf{k}] = \frac{3}{2}E[\sin\theta_{1,2}\, r_1 r_2] = \frac{3}{2}E[\sin\theta_{1,2}]\, E[r_1]E[r_2]$$
$$= \frac{3}{2}\cdot\frac{1}{\pi}\cdot\left(\frac{2}{3}\right)^2 = \frac{2}{3\pi}.$$

Now consider the case in which one of the points is in the convex hull of the other two and the origin. The expected value for the area of such a triangle is three times the expected value when we consider only P_3 inside the triangle formed by the other two and the origin. If S is the triangle formed by P_1, P_2 and the origin, the probablility that P_3 is inside S is $\frac{\text{Area}(S)}{\pi}$. Furthermore, the expected area of $\triangle P_1P_2P_3$, for P_3 chosen within S is $\frac{1}{3}\text{Area}(S)$, since the expected ratio

$$E\left[\frac{\text{Area}(T)}{\text{Area}(S)}\right] = E\left[\frac{d(P_3, P_1P_2)}{d(0, P_1P_2)}\right] = \frac{1}{3}.$$

(because this is linear in P_3 and is $1/3$ for the center of mass). Thus, the expected area for a triangle in this case is

$$3E\left[\frac{(\text{Area}(S))^2}{3\pi}\right] = \frac{1}{\pi}E\big[(\text{Area}(S))^2\big].$$

If θ denotes the angle between P_1 and P_2, then θ is uniform on $[0, \pi]$ and independent of r_1 and r_2, so we have

$$\frac{1}{\pi}E\big[(\text{Area}(S))^2\big] = \frac{1}{\pi}E\left[\frac{1}{4}\sin^2\theta\, r_1^2 r_2^2\right] = \frac{1}{4\pi}\, E[\sin^2\theta]\, E[r_1^2]\, E[r_2^2]$$
$$= \frac{1}{4\pi}\left(\frac{1}{\pi}\int_0^{\pi}\sin^2\theta\, d\theta\right)\left(\int_0^1 2r^3\, dr\right)^2 = \frac{1}{4\pi}\cdot\frac{1}{2}\cdot\left(\frac{1}{2}\right)^2 = \frac{1}{32\pi}.$$

Putting this together with the preceding case,

$$E[\text{Area}(T)] = \frac{2}{3\pi} + 2\cdot\frac{1}{32\pi}.$$

It now follows that the probability that P_4 is in triangle T once P_1, P_2, P_3 are chosen, is

$$E\left[\frac{\text{Area}(T)}{\pi}\right] = \frac{E[\text{Area}(T)]}{\pi} = \frac{2}{3\pi^2} + \frac{1}{16\pi^2} = \frac{35}{48\pi^2}.$$

Finally, the probability that P_1, P_2, P_3, P_4 forms a convex quadrilateral is 1 minus the probability that one of the points is in the convex hull of the triangle formed by the other three, and this is

$$1-4\left(\frac{35}{48\pi^2}\right)$$

as claimed.

Solution to B1. We check

$$\begin{aligned}0&=x^3+3xy+y^3-1\\&=\frac{x+y-1}{2}\Big((x+1)^2+(x-y)^2+(y+1)^2\Big),\end{aligned}$$

so there are two parts of the curve, the straight line with equation $x+y=1$ and the single point $(-1,-1)$. The three points can only be $A=(-1,-1)$ together with two symmetrically placed points $B=(a,1-a)$ and $C=(1-a,a)$ so $\triangle ABC$ will be equilateral. The altitude of this triangle is the distance from A to $D=(1/2,1/2)$, $AD=3\sqrt{2}/2$, so the side of the triangle is $\sqrt{6}$, and the area is $3\sqrt{3}/2$.

Solution to B2. Let $y_k=x_1+x_2+\cdots+x_k$ for $k=1,2,\ldots,n$. Let $m_k=-\lfloor y_k\rfloor$ be the negative of the greatest integer less than or equal to y_k, so $0\le m_k+y_k<1$. If $m_k+y_k\le\frac{1}{n+1}$ then $S=\{x_1,x_2,\ldots,x_k\}$ and $m=m_k$ satisfy the conditions. If $\frac{n}{n+1}\le m_k+y_k<1$ then $S=\{x_1,x_2,\ldots,x_k\}$ and $m=m_k-1$ satisfy the conditions. Otherwise $\frac{1}{n+1}<m_k+y_k<\frac{n}{n+1}$ for each k. Since the $n-1$ intervals $[\frac{j}{n+1},\frac{j+1}{n+1}]$ for $j=1,2,\ldots,n-1$ together contain the n numbers m_k+y_k, some interval contains two of these, say y_k and y_ℓ with $\ell>k$. In this case, $S=\{x_{k+1},x_{k+2},\ldots,x_\ell\}$ and $m=m_\ell-m_k$ satisfy the conditions since

$$\left|m+\sum_{s\in S}s\right|=|m_\ell-m_k+x_{k+1}+\cdots+x_\ell|=|(m_\ell+y_\ell)-(m_k+y_k)|\le\frac{1}{n+1}.$$

Solution to B3. We will show that the maximum of L_S is $\binom{n}{2}+1$, and that this is attained if and only if the set S is in general position. That is, it does not have any three points lying on a single line.

First, consider a set of points in general position. Let H denote the convex hull of this set. The partition having one empty set results whenever we select a line that does not intersect H. With this partition as an exception to the general rule, we will show that the other partitions are in one-to-one correspondence with the lines through pairs of points in S, thus giving the total of $\binom{n}{2}+1$ partitions.

For any line L separating S, we shall rotate it to obtain a line contacting one point from set A and one point from set B. First rotate the plane to view L as a vertical line with A on its left and B on its right. Rotate L in a positive direction (counterclockwise) about its lowest point of intersection with the hull until we contact the first point (or possibly two points) of A. Continue to rotate about the higher contact point until we contact a new point. If the new contact is in B, we stop, and if it is again in A, we continue to rotate about the new contact point. This procedure must conclude with a new line L' that has an A point contacted on the left above a B point contacted on the right.

The original line L producing the partition is not unique but the contact line L' is unique to the partition. To verify this, suppose that a single partition has two partitioning lines L_1 and L_2 that give rise to two different contact lines L_1' and L_2'. What can we say about these two lines? If they do not intersect inside the hull, the strip between them must not contain any points of S, making it impossible to locate two of the four contact points. And if they do intersect, then two opposite sectors must be devoid of points from S, and again we cannot locate the four contact points (or perhaps three if two happen to coincide with the intersection point) while avoiding the forbidden sectors. The only conclusion we can reach is that both lines must transform to the same contact line. Thus, each partition has a unique contact line.

But could, perhaps, two different partitions transform to the same contact line? No, because as we rotate the separating line, points never change sides, so when two lines transform to the same contact line they necessarily have the same partition. So, the number of partitions cannot exceed $\binom{n}{2} + 1$.

It remains to show that every line L' formed by a pair of points x and y in S is a contact line for some partition. To see this, rotate L' about the midpoint of the segment xy in the negative (clockwise) direction by a very tiny angle so that no other point in S is encountered during the rotation. The new line L is thus a partitioning line that will revert to L' as its contact line. Consequently, we have shown that any set of points in general position has exactly $\binom{n}{2} + 1$ partitions.

But what happens for sets that have three or more points on a single line? Say we have k points $x_1, x_2, \ldots, x_k$ of S on a single line L. Now the $\binom{k}{2}$ pairs all generate the same line. How many partitions can have this as the contact line? The separating line must intersect L at a point between a pair of the k points. Any two separating lines L_1 and L_2 that both intersect between x_j and x_{j+1} and then rotate to L as the contact line have to give the same partition. Thus, the line L can serve as the contact line for at most $k - 1$ partitions. We have lost $\binom{k-1}{2}$ of the $\binom{k}{2}$ potential pairs, so we cannot attain the prior maximum of $\binom{n}{2} + 1$. Thus, any set with three or more points on a single line cannot attain the maximum.

Solution to B4. The answer is 2^k.

Since 2^k is realized when V is the k-dimensional space $\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid x_i = 0$ for $i > k\}$, the maximum value of $Z(V)$ is $\geq 2^k$. We must show that $Z(V) \leq 2^k$ for all V of dimension k.

Proof 1. Form a $k \times n$ matrix M whose rows comprise a basis of V. We may assume that M is fully row-reduced and, in fact, by permuting coordinates of $\mathbb{R}^n$, that M is of the form $\left(I_k \mid M' \right)$ where I_k is the $k \times k$ identity matrix. Then if $\mathbf{x} = (x_1, \ldots, x_n) \in Z \cap V$, then $\mathbf{x} = \sum_{i=1}^{k} x_i \mathbf{r_i}$, where $\mathbf{r_i}$ is the ith row of M. Since, for each i, $x_i = 0$ or 1, there are at most 2^k such $\mathbf{x}$.

Proof 2. By induction on n. If $n = 0$ then $k = 0$ and $V \cap Z = V = \mathbf{0}$, so that $Z(V) = 1 = 2^k$. Let $n > 0$ and assume the result holds for subspaces of $\mathbb{R}^{n-1}$ and all k, $0 \leq k \leq n - 1$. Suppose V is a k-dimensional subspace of $\mathbb{R}^n$. Let $p : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be the projection mapping $(x_1, x_2, \ldots, x_{n-1}, x_n)$ to $(x_1, x_2, \ldots, x_{n-1})$.

Case 1. $\ker(p|_V) = \mathbf{0}$. Then $p|_V$ is injective and so $Z(V) = Z(p(V)) \leq 2^k$ by the induction hypothesis.

Case 2. $\ker(p|_V)$ is spanned by $e_n = (0, 0, \ldots, 0, 1)$ and so $\dim(p(V) = k - 1$. By the induction hypothesis $Z(p(V)) \leq 2^{k-1}$. But for any $w = (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$, $p^{-1}(w) \cap Z = \{(x_1, \ldots, x_{n-1}, 0), (x_1, \ldots, x_{n-1}, 1)\}$. Hence, $Z(V) \leq Z(p^{-1}(p(V)) \leq 2Z(p(V)) \leq 2^k$.

Solution to B5. The maximum is $1/16$. We have

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

Integrating, we have

$$J(f) - I(f) + \int_0^1 \frac{x^3}{4}\, dx = J(f) - I(f) + \frac{1}{16} \geq 0$$

so $I(f) - J(f) \leq 1/16$ and this value is achieved for $f(x) = x/2$.

Solution to B6. The limit is

$$\left(1 + \frac{1}{k}\right)^k.$$

First we observe that a_n is an increasing sequence of positive terms. If there was a finite limit L to this sequence, then we would have $L = L + 1/L^{1/k}$, but since this can't hold, there can be no finite limit and the sequence increases without bound.

LEMMA. *If* $\lim_{n\to\infty} x_{n+1} - x_n = L$ *then* $\lim_{n\to\infty} x_n/n = L$.

Proof. Let $\epsilon > 0$ be given and take N so $L - \epsilon/2 < x_{n+1} - x_n < L + \epsilon/2$ for $n \geq N$. Then for $n > N$,

$$x_n - x_{n-1} + x_{n-1} - x_{n-2} + \cdots + x_{N+1} - x_N = x_n - x_N,$$

so

$$(L - \epsilon/2)(n - N) < x_n - x_N < (L + \epsilon/2)(n - N).$$

Thus

$$\begin{aligned} L - \frac{\epsilon}{2} + \frac{x_N - LN + N\epsilon/2}{n} &= \frac{(L - \epsilon/2)(n - N) + x_N}{n} \\ &< \frac{x_n}{n} < \frac{(L + \epsilon/2)(n - N) + x_N}{n} \\ &= L + \frac{\epsilon}{2} + \frac{x_N - LN - N\epsilon/2}{n}. \end{aligned}$$

Taking $N_2 > \max(2|x_N - LN + N\epsilon/2|/\epsilon, 2|x_N - LN - N\epsilon/2|/\epsilon, N)$, for all $n \geq N_2$ we have

$$L - \epsilon < \frac{x_n}{n} < L + \epsilon.$$

Hence $\lim_{n\to\infty} x_n/n = L$.

Apply this lemma to the limit of $a_n^{1+1/k}/n$. (The difference between successive numerators is almost given by the recurrence relation, thus the observation that such a lemma should be useful.) Let $r_n = (a_{n+1}/a_n)^{1/k}$. Then $r_n^k - 1 = 1/a_n^{1+1/k} \to 0$, $\lim_{n\to\infty} r_n^k = 1$ and $\lim_{n\to\infty} r_n = 1$. Also,

$$\begin{aligned} a_{n+1}^{1+1/k} - a_n^{1+1/k} &= a_{n+1}^{1+1/k} - a_{n+1}a_n^{1/k} + a_{n+1}a_n^{1/k} - a_n^{1+1/k} \\ &= a_{n+1}(r_n - 1)a_n^{1/k} + (a_{n+1} - a_n)a_n^{1/k} \\ &= \frac{a_{n+1}(r_n^k - 1)a_n^{1/k}}{r_n^{k-1} + r_n^{k-2} + \cdots + 1} + 1 \\ &= \frac{r_n}{r_n^{k-1} + r_n^{k-2} + \cdots + 1} + 1. \end{aligned}$$

Hence

$$\lim_{n\to\infty} a_{n+1}^{1+1/k} - a_n^{1+1/k} = 1 + \frac{1}{k},$$

so by the lemma

$$\lim_{n\to\infty} \frac{a_n^{1+1/k}}{n} = 1 + \frac{1}{k},$$

and hence

$$\lim_{n\to\infty} \frac{a_n^{k+1}}{n^k} = \left(1 + \frac{1}{k}\right)^k.$$

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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