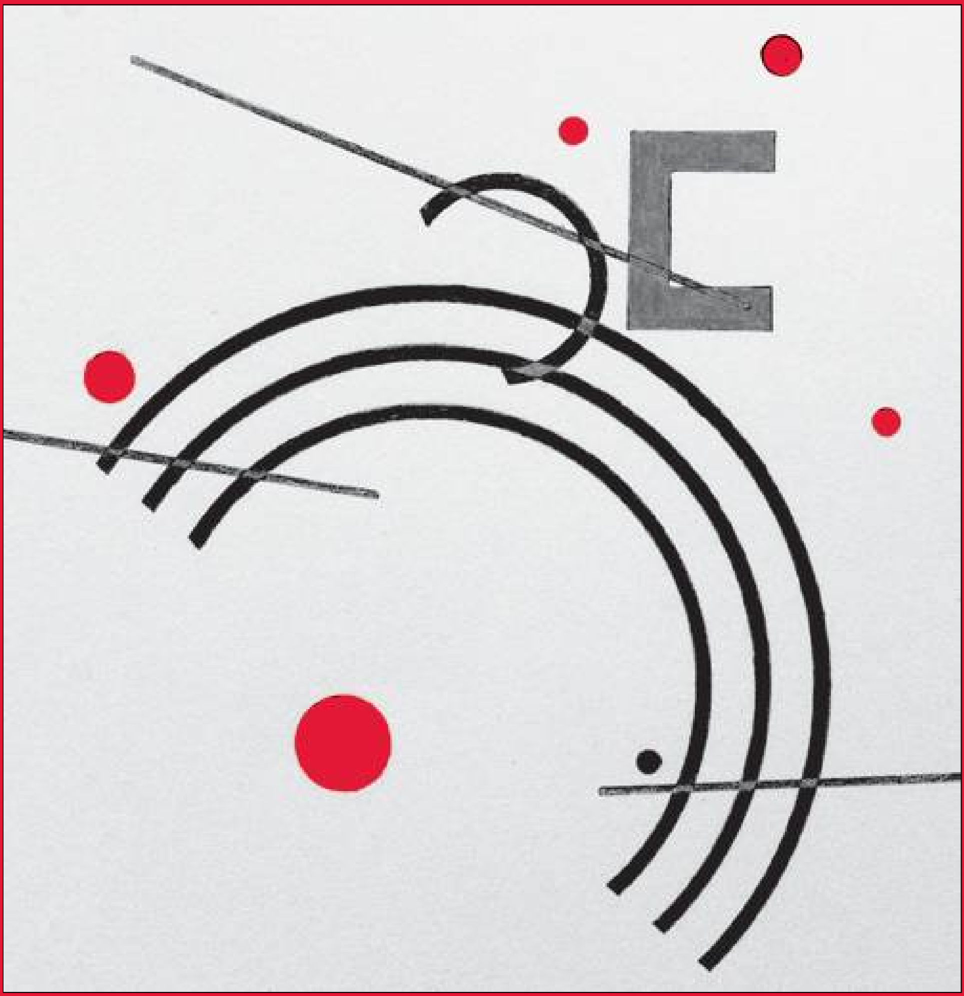
Vol. 87, No. 5, December 2014

MAA

MATHEMATICS MAGAZINE



- Conway's subprime Fibonacci sequences
- $\sin(x)$, $\sin(\sin x)$, $\sin(\sin(\sin x))$, . . .
- Straightedge-and-compass constructions on the sphere
- Patterns in Pascal's (four-dimensional) triangle

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. 83, at pages 73–74, and is available at the *Magazine*'s website 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.

Please submit new manuscripts by email directly to the editor at mathmag@maa.org. A brief message containing contact information and with an attached PDF file is preferred. Word-processor and DVI files can also be considered. Alternatively, manuscripts may be mailed to Mathematics Magazine, 132 Bodine Rd., Berwyn, PA 19312-1027. If possible, please include an email address for further correspondence.

Cover image by Annie Stromquist (`http://www.anniestromquist.com`). The image is adapted from a pencil drawing called "Compositions: One After Another (3)."

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Lancaster, PA, 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
1529 Eighteenth St. NW
Washington DC 20036

Phone: (877) 622-2373
E-mail: tmarmor@maa.org

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

Change of address, missing issue inquiries, and other subscription correspondence:

MAA Service Center, *maahq@maa.org*

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

Copyright © by the Mathematical Association of America (Incorporated), 2014, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. Permission to make copies of individual articles, in paper or electronic form, including posting on personal and class web pages, for educational and scientific use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the following copyright notice:

Copyright the Mathematical Association of America 2014. All rights reserved.

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

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

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

Printed in the United States of America

Vol. 87, No. 5, December 2014

MAA

MATHEMATICS MAGAZINE

EDITOR
Walter Stromquist

ASSOCIATE EDITORS

Bernardo M. Ábrego
California State University, Northridge

Paul J. Campbell
Beloit College

Annalisa Crannell
Franklin & Marshall College

Deanna B. Haunsperger
Carleton College

Warren P. Johnson
Connecticut College

Victor J. Katz
University of District of Columbia, retired

Keith M. Kendig
Cleveland State University

Roger B. Nelsen
Lewis & Clark College

Kenneth A. Ross
University of Oregon, retired

David R. Scott
University of Puget Sound

Paul K. Stockmeyer
College of William & Mary, retired

MANAGING EDITOR

Beverly Ruedi
MAA, Washington, DC

LETTER FROM THE EDITOR

Are you an algebraist, geometer, analyst, combinatorist, or a problem solver? Then there is something here for you.

Perhaps I should warn you about the problems posed by Guy, Khovanova, and Salazar. They are about "subprime sequences," and they share the same attractions as the $3x + 1$ problem—including the risk of being intractable. But they're fun! And if you can shed light on what makes some sequences predictable, while others are so spectacularly not, then all your efforts will be worthwhile.

What lengths are constructible, using straightedge and compass? That's a rich subject in Euclidean geometry—as we saw in our June issue, for example. In this issue, Daniel Heath addresses the same question in spherical geometry. There are some interesting differences!

Which roots can be computed by radicals? Starting from the the coefficients of a cubic equation, Cardano's formula lets you find the roots using just $+$, $-$, $\times$, $\div$, and radicals. But even when all of the coefficients and roots are real, you may need to use complex numbers along the way. Matt Lunsford describes this result, called the *Casus Irreducibilis*, and then shows us an analogy in finite fields. If your cubic has coefficients in the 5-element field and all of its roots are in the 125-element field, then in order to compute the roots by radicals, you may work in the 15625-element field.

How are you at visualizing four dimensions? It would be hard to find better practice then the article by Garcia and Pedersen. It is a sequel to a 2012 article in this MAGAZINE. That article started with some nice Pascal's-triangle identities, and extended them to three dimensions. This article takes the next step.

Chris Towse's article starts with the functions $\sin x$, $\sin(\sin x)$, $\sin(\sin(\sin x))$, etc. Do they converge to a limit function? Yes. But they converge to a more interesting limit if they are properly scaled. In that case, the outcome doesn't depend on the function you start with, as long as your function have the same kind of inflection point at zero as $\sin x$ does.

We have some echoes from past issues. Wijarn Sodsiri's note corrects an error about subgroups of $\mathbb{C}$ in this MAGAZINE from 1993; it turns out that a more careful argument gives a richer result. The article by Bayer and Brandt starts from a 1991 *Monthly* problem and takes it in a new direction, using Catalan numbers.

Transition. Each December we acknowledge our referees, whose efforts are so important to all of the MAA publications. We honor those listed on pages 406–407.

Search engines being as they are, we are no longer printing an annual index of articles and notes. You can still find the index for Volume 87 at the MAGAZINE's website and (temporarily) at http://www.mathematicsmagazine.org. The crossword in this issue is becoming a new tradition, and as usual you can find a fresh pdf version at those same sites.

The next issue will mark the start of Mike Jones's term as Editor. As my term ends, I am grateful for all of the help I have had from too many people to name. I thank particularly the support of the Associate Editors, the referees, and all of the people involved in the production process. I continue to admire the authors who submit more excellent papers than we have room to print, and I thank them all for their efforts and good will. This is a time of change for all publications. I am eager to see the MAGAZINE change and grow under Mike's leadership.

Walter Stromquist, Editor

ARTICLES

Conway's Subprime Fibonacci Sequences

RICHARD K. GUY
University of Calgary
Calgary, Alberta T2N 1N4
rkg@cpsc.ucalgary.ca

TANYA KHOVANOVA
Massachusetts Institute of Technology
Cambridge, MA 02139
tanyakh@yahoo.com

JULIAN SALAZAR
Harvard University
Cambridge, MA 02138
jsalazar01@college.harvard.edu

In memory of Martin Gardner

When John Conway last visited the first author, he passed the time on the plane by calculating what we now call *subprime Fibonacci sequences*. They are just the sort of thing Martin Gardner would have featured in his column. There is some risk of their becoming as notorious as the $3x + 1$ (Collatz) problem [**10**], with which they seem to have something in common, and of which Erdős has said, "Mathematics is not yet ripe for such problems."

The $3x + 1$ sequences take a positive integer and iteratively apply the following rule: If a number is odd, triple it and add one; if even, halve it. The sequences produced by this rule always appear to reach an infinite string of 4, 2, 1, 4, 2, 1, etc., and the problem is whether all sequences reach this *cycle*; that is, whether for all t_0, there is some n where $t_n = 1$. Here are some examples:

6, 3, 10, 5, 16, 8, **4, 2, 1**, 4, . . . ,

17, 52, 26, 13, 40, 20, 10, 5, 16, 8, **4, 2, 1**, 4, . . . ,

30, 15, 46, 23, 70, 35, 106, 53, 160, 80, 40, 20, 10, 5, 16, 8, **4, 2, 1**, 4, ...

Despite the simple rule, the paths of the sequences are rather unpredictable. Starting with 33 takes 26 steps and climbs to 100 before reaching 1, while 27 takes 111 steps and climbs to over 9000 before reaching 1. Such behavior has made this and other similar problems seem intractable [**6**]; we cannot even show that such sequences could not go to infinity! As Lagarias introduces the problem in his $3x + 1$ compendium [**10**], he states that it touches number theory, ergodic theory, stochastic processes, and more, while not lying squarely in any of their domains.

A more recreational example is given by Conway's RATS sequences, one of many base-dependent "reversal" sequences [**7**]. RATS stands for Reverse, Add, Then Sort:

Math. Mag. **87** (2014) 323–337. doi:10.4169/math.mag.87.5.323. © Mathematical Association of America

take a number with digits in increasing order, reverse it, add to the original number, and then sort the result's digits in increasing order. Here are some base-10 examples:

12334444, 55667777, 123334444, 556667777, 1233334444, 5566667777, ...,

123, 444, 888, 1677, 3489, **12333, 44556, 111, 222, 444, 888, 1677, 3489**, 12333, ...

The first sequence is known as the *creeper* (A164338 in OEIS [**1**]). It provably diverges in this regular pattern, and is reached by various starting terms such as 1. Conway's conjecture is that all base-10 RATS sequences enter cycles (as in the second sequence) or enter the creeper and diverge.

One natural approach in tackling problems like these involves restricting possible end behaviors of such sequences; their destinies, so to speak [**8**]. These two classes of sequence have rather different fates. For example, Oliveira e Silva has verified that $3x + 1$ sequences reach 1 for starting numbers less than 5.76×10^{18} [**10**], and Simons and de Weger proved that if there were another $3x + 1$ cycle it would have at least 69 terms [**11**]. By contrast, Cooper and Kennedy have shown the existence of base-10 RATS cycles of every length greater than 1 [**5**]. Regardless, there are limits on potential analysis: Kurtz and Simon, building on Conway, proved that a natural generalization of the $3x + 1$ problem is undecidable [**9**].

To help develop the terminology and flavor of these sequences, we can plot their trajectories on a directed graph. This visual approach is often used in expositions of the $3x + 1$ problem [**10**, p. 62], with sequences as paths in an infinite digraph (FIGURE 1). The problem is whether this digraph is weakly connected (connected when viewed as an undirected graph).

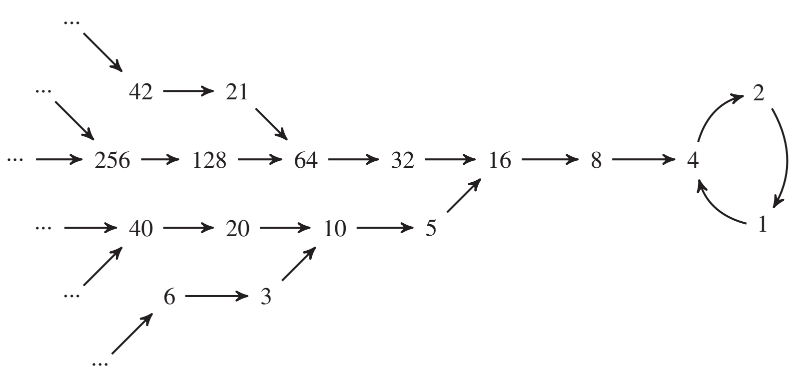


Figure 1 Digraph generated by the $3x + 1$ sequences

It is easy to discount these results as too problem-specific, and to say that such sequences could never lead to "useful" mathematics. Yet the appeal of such problems has always lain in the contrast between how easy they are to play with and how hard it is to answer their questions. We hope the subprime Fibonacci sequences continue this tradition.

Subprime Fibonacci sequences

Start like the Fibonacci sequence 0, 1, 1, 2, 3, 5, ..., but before you write down a composite term, divide it by its least prime factor so that this next term is not 8, but rather $8/2 = 4$. After that the sum gives us $5 + 4 = 9$, but we write $9/3 = 3$; then

$4+3=7$, which is okay since it is prime; then $3+7=10$, but we write $10/2=5$; and so on:

0 1 1 2 3 5 4 3 7 5 6 11 17 14 31 15 23 19
21 20 41 61 51 56 107 163 135 149 142 97 239 168 37 41 39 40
79 17 48 **13 61 37 49 43 46 89 45 67 56 41 97 69 83 76**
53 43 48 13 61 37 ...

and we are in an 18-cycle. If we start with 1, 1 or 1, 2, it follows that we get the same result. But we may start with any pair of numbers, and you may like to try starting with 2, 1, or 1, 3, or 3, 9, or 13, 11, etc.

One might suspect that every such sequence enters this 18-cycle, similar to the $3x+1$ problem's conjecture. After all, since our sequences are bounded or unbounded, they must either enter a cycle or increase indefinitely. We do not believe the latter happens, and we provide a heuristic argument below. But is the 18-cycle the only "non-trivial" cycle? Wait and see.

First, note that a, a, where $a \neq \pm 1$, gives the sequence $a, a, a, a, \ldots$. This is a *trivial cycle*. Sequences that end in trivial cycles are *trivial sequences*, such as 5, 15, 10, 5, 5, 5, ..., or −143, 39, −52, −13, −13, −13, ... If two consecutive terms have the same sign, then so do all subsequent terms. If they have opposite sign or include a zero, they bound further terms until two consecutive terms of the same sign appear, e.g., $-17,\ 7, -5,\ 2, -3, -1, -2, \ldots$, after which the sign remains constant.

Next, two terms of opposite parity are followed by an odd term, and two odd terms are followed by an even or an odd term depending on whether their sum is a multiple of 4. One can have arbitrarily long strings of even terms, but they must terminate since the power of 2 in consecutive terms must eventually decrease, e.g., 128, 160, 144, 152, 148, 150, 149, ...; once we have an odd term (unless this sequence is trivial), subsequent even terms are isolated with each followed by at least two odd terms. Therefore, *we need only consider sequences of positive terms, comprised of "runs" of odd terms separated by even terms*.

Finally, let the *shape* of a sequence be the string of its terms' parities (O for odd, E for even). The Fibonacci sequence has shape EOOEOOEOOEOO.... Our first subprime Fibonacci sequence had shape EOOEOOEOOOEOO.... The "extra" odd term here came from where the sum of the previous two odd terms had only one factor of 2. The example, starting at 13, 61 inclusive, gives the shape OOOOOEOOOEOOOOEOOE that repeats with the 18-cycle.

Nodes and other cycles

Our sequences cannot immediately be represented in the fashion of FIGURE 1 because of the second-order nature of our recurrence. We must carefully define vertices for our sequences, and so we introduce two important terms. The *nodes* of a sequence are ordered pairs of positive, odd coprimes that either begin the sequence or immediately follow the even terms of a sequence. *Runs* are the strings that begin with a node and consist of odd terms together with a single terminating even term.

In the next section we will see that every non-trivial sequence becomes composed of runs after some point. Here is our initial sequence with nodes parenthesized:

0 (1, 1) 2 (3, 5) 4 (3, 7) 5 6 (11, 17) 14
(31, 15) 23 19 21 20 (41, 61) 51 56 (107, 163) 135 149
142 (97, 239) 168 (37, 41) 39 40 (79, 17) 48 **(13, 61) 37**
49 43 46 (89, 45) 67 56 (41, 97) 69 83 76 (53, 43)
48 (13, 61) 37 ...

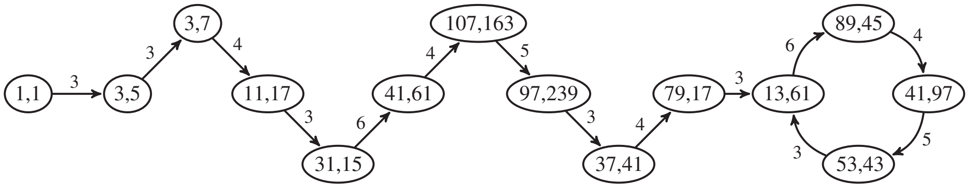


Figure 2 Path generated by the 0, 1 sequence

Taking the sequence's shape, we can treat each substring of the form O...OE as a unit, starting when the first two terms of such a substring are coprime and not preceded by an odd term. The corresponding terms comprise a run, and the first two terms of the run comprise a node. Let us now construct our first sequence path (FIGURE 2). For notational convenience we weight the digraph by assigning to each arc the length of the run generated by the node at the arc's tail.

We could then imagine the infinite digraph generated by all non-trivial subprime Fibonacci sequences, as we have done for the $3x + 1$ sequences in FIGURE 1. If the 18-cycle were the only non-trivial cycle, the subprime Fibonacci digraph would look like FIGURE 3.

One reason this digraph is a nice representation is that it shows how many nodes are *direct predecessors* to a single node. If a node is a predecessor (not necessarily direct) to a node or cycle, we say it is *tributary* to the node or cycle. How could we grow this graph? One way is to go outwards from known nodes. This would require a way of enumerating a node's direct predecessors, which can be done with some work:

1. For example, with the node $(89, 45)$ of the 18-cycle, start with the preceding even term t, which must satisfy $t + 89 = 45q$, where q is 1 or 3 ($q = 2$ makes t odd, and q cannot exceed a prime factor of 45). This gives $t = -44$ or 46, so the node must always be preceded by 46.
2. Let the positive odd term before t, if it exists, be s. Then $s + t = 89p$, where p is 1 or an odd prime ≤ 89. For $t = 46$, possible values of s are 43, 221, 399, etc. The term before s must also be odd. If this term is r, it must satisfy $r + s = 2t$ since $r + s$ is even. For example, $s = 43$ gives $r = 49$, and none of the other possibilities for s would work, since they would make $r \leq 0$.
3. Since we are only looking for possible direct predecessors (positive, odd coprimes), we can assume that each prior step involved division by two. Working backwards gives $\ldots, -83, 109, 13, 61, 37, 49, 43, 46$. Thus, our direct predecessors are exactly $(109, 13)$, $(13, 61)$, $(61, 37)$, $(37, 49)$, and $(49, 43)$, only two of which are depicted in FIGURE 3.

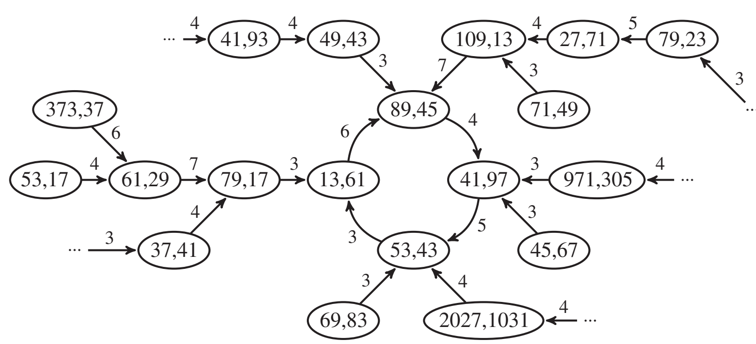


Figure 3 Some paths leading to the 18-cycle

Contrast this with FIGURE 1, where there are at most two direct predecessors as a result of the sequence definition. This procedure for constructively generating nodes is quite finicky, however, and discourages a graph-theoretic approach to analysis. Not to say that it is impossible; there exist reductions and results on the $3x + 1$ graph [**2, 12**], and we encourage the reader to explore the possibility of deriving properties for the subprime digraph from this perspective.

Do sequences all enter the 18-cycle we have already seen? In other words, is the subprime digraph weakly connected? Let us start at the node (151, 227):

(151, 227) 189 208 (397, 121) 259 190 (449, 213) 331 272 (201, 43) 122
(55, 59) 57 58 **(23, 27) 25 26 (17, 43) 30 (73, 103) 88**
(191, 93) 142 (47, 63) 55 59 57 58 (23, 27) ...

and we are in a 19-cycle whose first repeated node is (23, 27). Note that though 55, 59 are the first two repeated terms, they only act as a node the first time through; thus (47, 63) being a node with the terms 55, 59 in its run does not preclude (55, 59) from being a node in another context. Both nodes are tributary to the node (23, 27).

Furthermore, if you start with 5, 13 you will enter a 136-cycle through node (47, 23) (though simpler starting terms like 1, 4 suffice). If you start with 5, 23 you will enter a 56-cycle through node (119, 109) with 5693 as its largest term. Finally, the nodes (37, 199) and (127, 509) generate an 11-cycle and a 10-cycle, respectively.

FIGURE 4 displays the nodes in these non-trivial cycles. We checked sequences that start with two numbers below 10^6 and found no non-trivial cycles other than these six. That bound is easily extendable by our more computationally-minded readers.

In TABLE 1 the headings indicate the range for the first two terms of the sequence, and the entries are the number of occurrences for each cycle length. The proportion

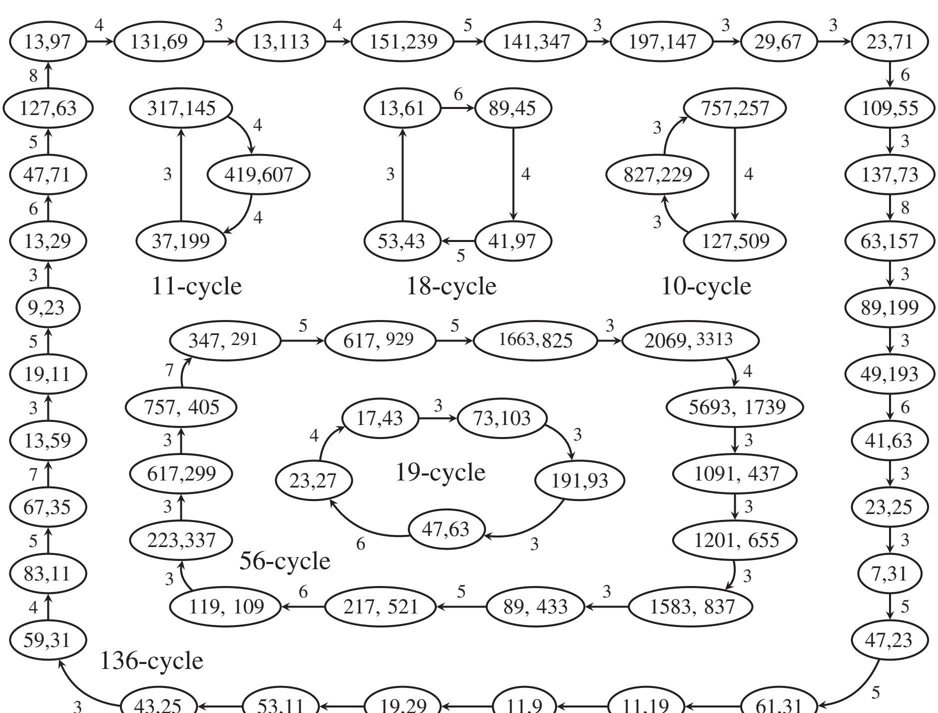


Figure 4 Digraphs of the six known non-trivial cycles

TABLE 1: Distribution of final cycle lengths generated by starting node (a, b)

Cycle length	$a, b \leq 10$	$a, b \leq 10^2$	$a, b \leq 10^3$	$a, b \leq 10^4$	$a, b \leq 10^5$
1	14	348	10022	320531	11588563
10	0	0	33	6310	668764
11	0	0	390	34520	3479974
18	63	4837	467014	46985673	4709133000
19	0	249	30490	3090886	307710709
56	0	188	21990	2238493	224936180
136	23	4378	470061	47323587	4742482810

of pairs that generate each non-trivial cycle stabilizes as the range for starting terms increases. Additionally, non-trivial cycles appear to be distributed among the starting pairs rather arbitrarily. However, trivial cycles decrease in proportion since a cycle $a, a \ldots$ requires all earlier terms to be multiples of a. Applying the "direct predecessor" method shows why this is, and how this makes relatively few starting conditions lead to a given trivial cycle.

Generally, non-trivial sequences seem to exhibit pseudo-random behavior in their terms and their digraphs, regarding the length of their paths, the nodes they pass, and their associated cycles. We believe this is due partly to the construction, which relies on prime factorizations (the relationship between these factorizations and addition is not well understood). A similar difficulty is seen in the earlier RATS sequences, where the relationship between base-dependent reversal/sort and addition is essential to analysis.

However, another source of apparent randomness seems to be the iteration's conditionality itself, as with the $3x + 1$ sequences. For example, if one considers a variant of subprime Fibonacci where only division by 5 occurs (when the sum is divisible by 5), similar observations as the above arise. We begin to feel the apparent intractability mentioned earlier of proving results on "destinies" of sequences like these.

End conditions

A sequence must either end in a trivial cycle, end in a non-trivial cycle, or increase indefinitely. These *end conditions* are of interest; however, it seems more likely here than in the $3x + 1$ problem that sequences do not increase indefinitely. Here is an informal argument that supports such a conjecture. We rely on the following observation:

PROPOSITION 1. *The terms of the run defined by (a, b) are bounded, above and below, in the interval $[a, b]$, and terms after the node are bounded in the interval $[M/4, M]$, where $M = \max(a, b)$. In general, two consecutive run terms bound the rest of the run.*

This is because after the first two odds a, b (the node terms) of a run, every successive term up to the terminating even is due to a division by 2; within the run we are averaging two consecutive terms at a time. Now, given the maximum of the current run M, what can we say about the maximum of the next run? Let $s, t \leq M$ be the last two terms of the current run, and let c, d be the first two terms of the next run. By Proposition 1, $t \geq M/4$. Remembering that t is even and s is odd, we do casework:

1. $c = s + t$ is prime, so $c < 2M$ (strict since $s + t$ is odd).
   (a) $t + c$ is prime, so $d = t + c < 3M$. Then the next run is bounded by $\max(c, d) < 3M$.
   (b) $t + c$ is composite, so $d \leq (t + c)/3 < M$. Then the next run is bounded by $\max(c, d) < 2M$.
2. $c = s + t$ is composite, so $c \leq (s + t)/3 < 2M/3$ (strict since $s + t$ is odd).
   (a) $t + c$ is prime, so $d = t + c < 5M/3$. Then the next run is bounded by $\max(c, d) < 5M/3$.
   (b) $t + c$ is composite, so $d \leq (t + c)/3 < 5M/9$. Then the next run is bounded by $\max(c, d) < 2M/3$.

What is the probability that $s + t$ or $t + c$ is prime? Assuming that our values are random and independent, the probability that each individually is prime is at most $1/(\ln t) \leq 1/(\ln(M/4))$ by the prime number theorem. Then the next run is bounded by $(2/3)M$ with probability at least $(1 - 1/(\ln(M/4)))^2$, which approaches 1 as M increases.

Considering that starting terms 5, 23 produce a term as large as 5693, proving the non-existence of divergent sequences seems comparable in difficulty to the same problem for the $3x + 1$ sequences, and so we do not dwell on it further. However, there are other types of result we can prove about sequences:

PROPOSITION 2. *A non-trivial sequence contains infinitely many primes (not necessarily distinct), each greater than its two preceding terms.*

Proof. If, after some point, the sequence contains no two consecutive terms that sum to a prime, then at each step a division happens and so the maximum value of any two consecutive terms decreases over time. Since this value cannot decrease forever, we get a contradiction. This proposition is stronger than only asserting infinitely many primes, as a prime could be generated after the division by a prime factor. ■

PROPOSITION 3. *After some point consecutive terms of a non-trivial sequence are always coprime.*

Proof. The greatest common divisor of two consecutive terms of the sequence cannot increase, since $\gcd(a, b) = \gcd(b, a + b)$, and so $\gcd(b, (a + b)/p) \leq \gcd(a, b)$. A non-trivial sequence must include a prime larger than its two preceding terms. Thereafter the GCD of any two consecutive terms is 1. ■

COROLLARY. *If the two starting terms of a sequence are coprime, the sequence is non-trivial.*

These results help justify our earlier definitions of node and run, as they ensure that only non-trivial sequences produce digraphs and that digraphs are unique representations (since a run uniquely leads into another run with no intervening terms, as evens cannot appear consecutively once nodes come into play).

Proposition 3 and its corollary also simplify the search for cycles via starting conditions. Since all starting terms a, b with $\gcd(a, b) > 1$ produce trivial sequences or reach $\gcd(a, b) = 1$ for consecutive terms, it suffices to study starting conditions with $\gcd(a, b) = 1$ to enumerate all non-trivial end conditions. Caching certain intermediate nodes and using a lookup table of primes provides a more efficient search method for new cycles than testing all positive integer ordered pairs.

The general system

We devote the rest of our paper to the cycles that non-trivial sequences generate. By the definition of a run, a non-trivial cycle must consist of a concatenation of runs. It follows from Proposition 3 that any two consecutive terms in a non-trivial cycle are coprime.

When we build a subprime Fibonacci sequence we add two numbers first, then divide by a prime number or 1. Let us correspond to each term of a sequence or a cycle the smallest prime divisor (or 1) by which the sum of the two prior terms was divided. These divisors are the sequence's or cycle's *signature*. For example, the 10-cycle 127, 509, 318, 827, 229, 528, 757, 257, 507, 382 has signature 7, 1, 2, 1, 5, 2, 1, 5, 2, 2; the initial 7 is the divisor to get 127, after adding the preceding cycle terms 507, 382.

Signature terms can be relatively large; the 11-cycle has signature 29, 3, 2, 1, 3, 2, 2, 1, 1, 2, 2 (since one of the intermediate sums is $29 \times 37 = 1073$). Runs consist of consecutive averages, so given a run within a cycle, only its node (first two terms) has signature values not equal to 2. See this in action by noting the shape of the 10-cycle (OOEOOEOOOE) and comparing with the signature given earlier. Using Proposition 2, this result on signatures follows:

PROPOSITION 4. *The largest term of a cycle must be prime with a signature value of 1.*

Proof. Let a, b, c be consecutive terms of a cycle. Then $c > \max(a, b)$ if and only if $c = a + b$, i.e., the signature value of c is 1 and c is prime. ■

With the terms and the signature of a cycle, we can establish a homogeneous linear system. Let $t_1, \dots, t_m$ be the terms of the cycle and $s_1, \dots, s_m$ be the corresponding signature. Then

$$\dots, t_{m-1} + t_m = s_1 t_1, \quad t_m + t_1 = s_2 t_2, \quad t_1 + t_2 = s_3 t_3, \quad \dots, \quad t_{i-2} + t_{i-1} = s_i t_i, \quad \dots.$$

In matrix form,

$$\begin{bmatrix} s_1 & 0 & 0 & \cdots & 0 & -1 & -1 \\ -1 & s_2 & 0 & \cdots & 0 & 0 & -1 \\ -1 & -1 & s_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & s_{m-2} & 0 & 0 \\ 0 & 0 & 0 & \cdots & -1 & s_{m-1} & 0 \\ 0 & 0 & 0 & \cdots & -1 & -1 & s_m \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ \vdots \\ t_{m-2} \\ t_{m-1} \\ t_m \end{bmatrix} = \mathbf{0}. \qquad (1)$$

We can now relate signatures to cycles and begin restricting potential cycles.

THEOREM 1. *No two cycles have the same signature.*

Proof. This is equivalent to showing that a potential signature $s_1, \dots, s_m$ defines at most one cycle. Given a potential signature, consider solutions for $t_1, \dots, t_m$ over the reals. We have a system of m linear homogeneous equations in m variables. In this particular set of equations, all of the variables are expressible through exactly two consecutive ones, so the space of real solutions is at most 2-dimensional. Given consecutive terms t_i, t_{i+1} and positive signature values, the equations must reduce to $At_i + Bt_{i+1} = t_i$ and $Ct_i + Dt_{i+1} = t_{i+1}$ for some positive A, B, C, D. Thus t_i is expressible through t_{i+1}, and the solution space is at most 1-dimensional.

If the solution is 1-dimensional, let one of the terms equal 1. The terms are in constant rational proportion to each other, so we can scale all the terms until the smallest set of integer solutions is produced. The largest term may be prime; this solution is potentially a cycle. Further scaling cannot produce another cycle since the largest term would not be prime (Proposition 4). ■

THEOREM 2. *There are no non-trivial cycles of one run (i.e., one even term).*

Proof. Let (t_1, t_2) be the node of the run. Sum all the row equations of Equation (1) to get

$$2(t_1 + \cdots + t_m) = s_1 t_1 + \cdots + s_m t_m.$$

By definition, $s_3, \ldots, s_m = 2$, so the equation becomes $2(t_1 + t_2) = s_1 t_1 + s_2 t_2$. Suppose first that t_1 is the largest prime with $s_1 = 1$. Then $t_1 = (s_2 - 2)t_2$. Since $t_2 > 1$ (dividing a composite number by its smallest prime factor will never produce 1) and t_1 is prime, $s_2 = 3$ and $t_1 = t_2$, which is a contradiction since this is a non-trivial cycle. The argument is the same if t_2 is the largest prime. ■

Since each run has at least 3 terms:

COROLLARY. *There are no non-trivial cycles of length below 6. If a cycle of length 6 exists, its shape must be* OOEOOE.

The trick of Theorem 2 does not generalize to helping find cycles of more than one run. In this regard, we look to Theorem 1 because it shows that results on signatures are necessarily results on cycles, which makes it desirable to relate signature terms within a cycle in a meaningful way. A signature is only useful if it produces a 1-dimensional solution space, requiring a determinant of 0.

One such relation could involve finding the general expression for the determinant of an m-cycle in terms of $s_1, \ldots, s_m$, which we leave as an exercise to the reader. Our issue with this approach is that it ignores the run-based structure of cycles, and so we present a reduction where the only signature terms of interest are those corresponding to the node of each run (divisors not equal to 2).

The run-centric system

Constructing such a relation is powerful as it cements the correspondence between cycles and nodes, providing a more natural categorization of cycles. Later on, it lets us demonstrate an algorithm that disqualifies entire classes of cycle. This, combined with related signature restrictions in the next section, contributes to bounding future cycles by their lengths and shapes, as opposed to bounding the size of their terms. This is analogous to the two types of bound for the $3x + 1$ problem: Simons and de Weger's lower bound on cycle lengths versus Oliveira e Silva's lower bound on cycle term size, which restrict cycle classes and magnitudes, respectively.

Before we begin, let us define the Jacobsthal numbers. They are defined by the recurrence $J_n = J_{n-1} + 2J_{n-2}$, where $J_0 = 0$, $J_1 = 1$ (A001045 on OEIS [**1**]). The next few are $J_2 = 1$, followed by 3, 5, 11, 21, ... Solving the recurrence gives $J_n = \frac{1}{3}(2^n - (-1)^n)$, and so apart from J_0 they are all odd. Using these numbers, we relate a run's terms with its node:

THEOREM 3. *Given a node (a, b) where $a, b > 0$ are odd, let $b = a + 2^{k-2}d$ with odd (but not necessarily positive) $d > -a/2^{k-2}$. The corresponding run is then $\{a + 2^{k-i} J_{i-1} d\}$, where i goes from 1 to k. The run has length k, consisting of $k - 1 \geq 2$ odd terms followed by a single even term $a + J_{k-1} d$.*

Proof. We justify the exponent $k-2$ in $b = a + 2^{k-2}d$ as it counts the number of divisions by 2, which occur for all terms except the two node terms. Thus k denotes run length. We conclude that the first $k-1$ members

$$a + 2^{k-1}J_0 d = a, \quad a + 2^{k-2}J_1 d = a + 2^{k-2}d, \quad \ldots, \quad a + 2^1 J_{k-2}d \tag{2}$$

are all odd since a and d are odd by definition, while the last (kth) term $a + 2^0 J_{k-1}d$ is even. By the recurrence, each term after the first two is the average of the two previous ones, a consequence of the definition of a subprime Fibonacci sequence. The condition $d > -a/2^{k-2}$ ensures that all our terms are positive. ∎

We refer to Theorem 3 for a more run-based system for a cycle. Write out two runs in the style of Equation (2):

$$a_1,\ a_1 + 2^{k_1-2}d_1,\ a_1 + 2^{k_1-3}d_1,\ \ldots,\ a_1 + 2J_{k_1-2}d_1,\ a_1 + J_{k_1-1}d_1$$

$$a_2,\ a_2 + 2^{k_2-2}d_2,\ a_2 + 2^{k_2-3}d_2,\ \ldots,\ a_2 + 2J_{k_2-2}d_2,\ a_2 + J_{k_2-1}d_2$$

Concatenate the two runs. The two terms after the first run will be the first two terms (the node) of the second run. Remembering that $J_n = J_{n-1} + 2J_{n-2}$, we can express these terms as

$$\frac{2a_1 + J_{k_1}d_1}{p_2} \quad \text{and} \quad \frac{(p_2+2)a_1 + (J_{k_1-1}p_2 + J_{k_1})d_1}{p_2 q_2},$$

respectively, where p_2, q_2 are the least prime *divisors* of the node of the second run. We will reserve the use of "divisors" to the signature terms of nodes.

Curve the two runs into a cycle and denote the divisors of the first run as p_1 and q_1. As with Equation (1) we fix the length m of the cycle, which is done by fixing the individual run lengths k_1, k_2. We now have the four equations

$$\begin{aligned} p_2 a_2 &= 2a_1 + J_{k_1}d_1 \\ p_2 q_2 (a_2 + 2^{k_1-2}d_2) &= (p_2+2)a_1 + (J_{k_1-1}p_2 + J_{k_1})d_1 \\ p_1 a_1 &= 2a_2 + J_{k_2}d_2 \\ p_1 q_1 (a_1 + 2^{k_2-2}d_1) &= (p_1+2)a_2 + (J_{k_2-1}p_1 + J_{k_2})d_2 \end{aligned}$$

giving four linear homogeneous equations as viewed in terms of a_1, d_1, a_2, d_2. Subtracting the first and third equations from the second and fourth, then removing a factor p_i from each, gives

$$\begin{vmatrix} 2 & J_{k_1} & -p_2 & 0 \\ 1 & J_{k_1-1} & -q_2+1 & -2^{k_2-2}q_2 \\ -p_1 & 0 & 2 & J_{k_2} \\ -q_1+1 & -2^{k_1-2}q_1 & 1 & J_{k_2-1} \end{vmatrix} = 0. \tag{3}$$

Expanding and applying the identity $2^{k-1} - J_k = J_{k-1}$ reduces Equation (3) to

$$\begin{aligned} 2^{k_1+k_2-4}p_1 q_1 p_2 q_2 = {} & J_{k_1-1}J_{k_2-1}p_1 p_2 + J_{k_1}J_{k_2-2}p_1 q_2 + J_{k_1-2}J_{k_2}q_1 p_2 \\ & + J_{k_1-1}J_{k_2-1}q_1 q_2 + J_{k_1}J_{k_2-1}p_1 + J_{k_1-1}J_{k_2}q_1 \\ & + J_{k_1-1}J_{k_2}p_2 + J_{k_1}J_{k_2-1}q_2 + J_{k_1}J_{k_2} - (-1)^{k_1+k_2-4}. \end{aligned}$$

Note that *any* 2-run cycle's node divisors must satisfy this relation.

One can similarly reconstruct formulas for n-run cycles, where the only inputs are n and the associated *run configuration*, which is the n-tuple of run lengths and thus a concise version of the shape. For example, the run configuration of the 10-cycle of shape OOEOOEOOOE is $(k_1, k_2, k_3) = (3, 3, 4)$, where $\sum k_i = 10$. It is important to note that, as with the shapes of cycles, run configurations $(3, 4, 3)$ and $(4, 3, 3)$ are identical to $(3, 3, 4)$ since cycles have no definitive starting nodes; what matters is that the order of run lengths within the cycle is preserved.

Signature restrictions

Let the terms of an arbitrary cycle be

$$a_1, b_1, \frac{a_1 + b_1}{2}, \frac{a_1 + 3b_1}{4}, \ldots, a_n, b_n, \frac{a_n + b_n}{2}, \frac{a_n + 3b_n}{4}, \ldots$$

where (a_i, b_i) are the cycle's nodes. Let the respective signature be

$$p_1, q_1, 2, 2, \ldots, p_n, q_n, 2, 2, \ldots, \text{ etc.},$$

where p_i corresponds to a_i and q_i corresponds to b_i. Each p_i, q_i is either 1 or an odd prime. We will refer to these, the divisors of the cycle's nodes, collectively as *the cycle's divisors*. As already established, $n \geq 2$.

We were able to rule out 1-run cycles and relate the divisors of n-run cycles to one another given the run configuration. But even with a given 2-run configuration, we are left with a relation in four unknown variables p_1, q_1, p_2, q_2. (Remember that a cycle is uniquely determined by its signature, from which every (a_i, b_i) can be recovered.) Can we further restrict these variables? We already know that at least one of $p_1, q_1, \ldots, p_n, q_n$ is 1, and that $p_1, q_1, \ldots, p_n, q_n$ cannot all equal 1 since cycle terms cannot increase indefinitely.

Can we strengthen these results? To motivate another approach, consider the following diagram of term vs. index for the 18-cycle (FIGURE 5), which is composed of four runs:

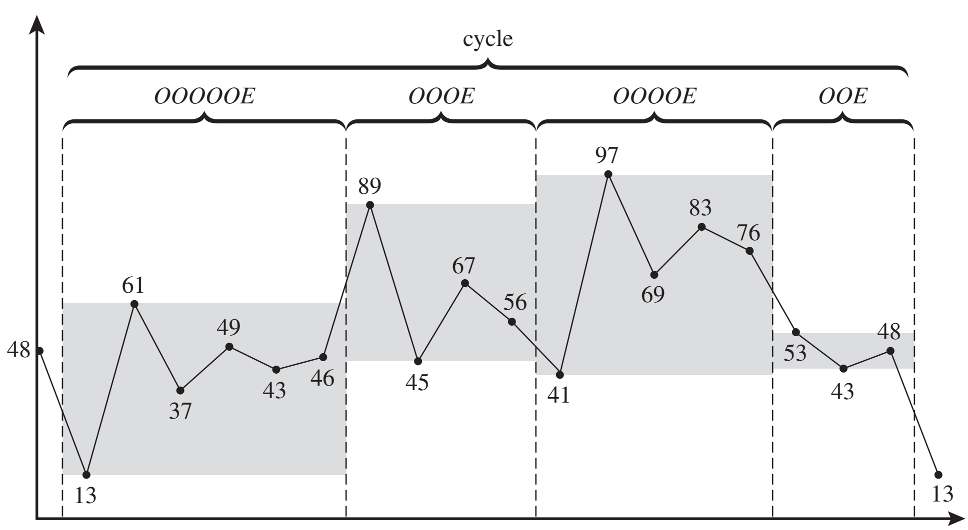


Figure 5 18-cycle with runs labeled and local run bounds shaded

The shaded areas represent bounds on each run and are a consequence of Proposition 1. Remembering Proposition 4 and that if some p_i or $q_i \neq 1$ then the sum of the two terms preceding the corresponding a_i or b_i was divided by at least 3, one can prove two propositions:

PROPOSITION 5. *At least two of $p_1, q_1, \ldots, p_n, q_n$ do not equal 1.*

PROPOSITION 6. *At least two of $p_1, q_1, \ldots, p_n, q_n$ equal 1.*

These results are particularly restrictive on cycles of only two nodes, that is, where the only divisors are p_1, q_1, p_2, q_2. One might conjecture the following:

CONJECTURE. *There are no non-trivial cycles of two runs (i.e., two even terms).*

Because of these results, all that is needed to prove this conjecture is a similar argument against the cases where one of p_1, q_1 and one of p_2, q_2 are 1. However, consider the cycle signature 7, 1, 2, 1, 5, 2, 2. Using these values for $s_1, \ldots, s_n$ in the earlier system and scaling as in Theorem 1 gives the cycle candidate 13, 51, 32, 83, 23, 53, 38, which would work if 51 were prime. Thus, to prove that other "cycles" like this similarly fail, the primality test for an unknown set of numbers may be required.

However, it also seems possible that, with a requirement of exactly two runs, primes in a signature and terms in a cycle are bounded in some way. After all, longer instances of such a cycle only means that runs take longer to terminate; but since runs are recurrences of averages, the cycle's two nodes' positions relative to each other should be fairly restricted.

Cycles of a given length

Regardless of whether the preceding argument can be formalized and generalized to cycles of any number of runs, it is still important that the cases involving cycles of shorter lengths are exhausted. How can we do this? Consider what we know:

- Relationships between signature terms and between divisors (Equations (1) and (3))
- Each signature corresponds to a unique potential cycle (Theorem 1)
- Non-existence of 1-run cycles (Theorem 2)
- Restrictions on possible signatures (Propositions 5 and 6)

These give a way to determine whether 2-run cycles of a given length exist, which for cycle lengths of 6 to 8 exhaust all possible cycles of that length:

THEOREM 4. *There are no 6-cycles.*

Proof. A 6-cycle must have shape OOEOOE and therefore a signature $p_1, q_1, 2, p_2, q_2, 2$. Using Equation (3) with run configuration $(k_1, k_2) = (3, 3)$, we get

$$4p_1q_1p_2q_2 = p_1p_2 + q_1q_2 + 3(p_1q_2 + q_1p_2 + p_1 + q_1 + p_2 + q_2) + 8.$$

By our previous results, exactly two of p_1, q_1, p_2, q_2 must be 1. One works through 4 cases to restrict the signature possibilities, which can be substituted into Equation (1) to solve for cycle terms. Since the solution space is 1-dimensional, we can express all the cycle terms $t_1, \ldots, t_n$ in terms of t_1 and then multiply by the common denominator to get the unique cycle candidate, each of which can be shown to fail (see TABLE 2). ∎

On a minor note, all the exceptional cases are also eliminated by the following proposition, which one proves using casework that disqualifies 1, 3, and 5 by looking at preceding terms:

TABLE 2: Candidates for a 6-cycle

(p_1, q_1, p_2, q_2)	Cycle candidate	It should be...
(1, 1, 3, 7)	3, 5, 4, 3, 1, 2	4, 3, 7, not 4, 3, 1
(1, 1, 7, 3)	13, 19, 16, 5, 7, 6	19, 16, 7, not 19, 16, 5
(1, 3, 1, 11)	5, 3, 4, 7, 1, 4	4, 7, 11, not 4, 7, 1
(1, 5, 1, 5)	3, 1, 2, 3, 1, 2	2, 3, 5, not 2, 3, 1
(1, 5, 37, 1)	41, 11, 26, 1, 27, 14	11, 26, 37, not 11, 26, 1
(1, 37, 5, 1)	27, 1, 14, 3, 17, 10	10, 27, 37, not 10, 27, 1
(3, 1, 11, 1)	9, 19, 14, 3, 17, 10	19, 14, 11, not 19, 14, 3
(5, 1, 5, 1)	1, 3, 2, 1, 3, 2	3, 2, 5, not 3, 2, 1

PROPOSITION 7. *The smallest term in a non-trivial cycle must be a node term (and thus odd), and at least 7 (which appears in the 136-cycle).*

Overall, the problem is that the linear system takes divisibility into account—but not divisibility by the *smallest* prime factor—or no division if a sum is already prime. Note that the symmetries above do not always occur; here they arise from both the runs being of shape OOE. The method of generating candidates in Theorem 4's proof and showing why each fails can be applied in general to disqualify longer 2-run configurations. For 7-cycles, the only possible run configuration is $(3, 4)$; for 8-cycles, the two possible configurations are $(3, 5)$ and $(4, 4)$. By considering all configuration cases and proceeding with the method programmatically, Andrew Bremner has shown [**3**] that:

THEOREM 5. *There are no 2-run cycles of length 30 or less.*

One can also consider cycles of more than two runs, though they require significantly more casework. Consider $(3, 3, 4)$, the only valid 3-run configuration for 10-cycles. Bremner has provided the following argument. In the manner of Equation (3) (but with three runs), we get

$$\begin{aligned}16p_1q_1p_2q_2p_3q_3 = {}& 3p_1p_2p_3 + 3p_1p_2q_3 + 9p_1q_2p_3 + 3p_1q_2q_3 + 5q_1p_2p_3 \\ & + 9q_1p_2q_3 + 5q_1q_2p_3 + 3q_1q_2q_3 + 9p_1p_2 + 9p_1q_2 + 9p_1p_3 \\ & + 9p_1q_3 + 15q_1p_2 + 5q_1q_2 + 5q_1p_3 + 9q_1q_3 + 5p_2p_3 + 9p_2q_3 \\ & + 15q_2p_3 + 9q_2q_3 + 27p_1 + 15q_1 + 15p_2 + 15q_2 + 15p_3 \\ & + 27q_3 + 44.\end{aligned}$$

Observe that each summand can be bounded in terms of $p_1q_1p_2q_2p_3q_3$ as long as an equivalent condition holds, e.g.,

$$3p_1p_2p_3 < (43/70)p_1q_1p_2q_2p_3q_3 \iff q_1q_2q_3 > 210/43.$$

Construct 26 such conditions (one for each summand) so that if they all hold, then

$$16p_1q_1p_2q_2p_3q_3 < 26(43/70)p_1q_1p_2q_2p_3q_3 + 44,$$

which is equivalent to $p_1q_1p_2q_2p_3q_3 < 1540$ (attaining this low but not fully restrictive bound motivated our 43/70 coefficient). Either this holds, or one of the 26 conditions is false; for example, $q_1q_2q_3 > 210/43$ might not hold. Equivalently, this means at least one of 27 upper-bounding conditions must be satisfied.

Casework across a finite number of signature candidates recovers the 10-cycle discovered earlier as the only 10-cycle of configuration $(3, 3, 4)$. Bremner also showed that there are no 9-cycles of configuration $(3, 3, 3)$ in a similar manner [**3**]. Since no cycles of length 8 or less have 3 runs, and because the only 3-run configurations for 9-cycles and 10-cycles are equivalent to $(3, 3, 3)$ and $(3, 3, 4)$, respectively, we can definitively state that:

THEOREM 6. *There are no (non-trivial) cycles of length 9 or less. There is only one cycle of length 10, generated by* $(127, 509)$.

Conclusion

This paper has explored relatively cursory properties of the subprime Fibonacci sequences: Most of our deductions have relied only on elementary number theory, algebra, empirical observations, and diagrams. Of course, this is the way we prefer it; to write the first exposition and let others prove the hard results!

It all returns to the low barrier of playing with these sequences. Surely, similar manipulations will yield new results, but we expect that significantly deeper mathematics will be needed to answer the difficult question of the (non-)existence of divergent sequences and the finitude of cycles, perhaps the kind of mathematics necessary to solve the notorious $3x + 1$ problem.

However, there are plenty of questions that seem both computationally and mathematically tractable. Here are some of the more obvious ones:

- We have shown that there are no 2-run cycles of length 30. How far can this be extended computationally? There are no 3-run cycles of configuration $(3, 3, 3)$ and only one of $(3, 3, 4)$. Can you also extend the 3-run procedure and show that the 10-cycle and the 11-cycle are the only 3-run cycles less than a certain length? Procedures to exhaust 4-run or greater cases would also be welcome.
- Are there *any* other non-trivial cycles? We have found six non-trivial cycles using starting values a, b within the range $1 \leq a, b \leq 10^6$. This could be attacked by increasing the search range or considering more classes (run configurations or otherwise) of cycles. Relevant code can be found at `http://github.com/j-salazar/subprime`.
- We did not explore if/how divisors and terms are bounded based on the number of runs. Maybe one can prove cycle results in this manner. Similarly, don't immediately accept our abstractions of runs, nodes, and signatures if other approaches are fruitful!

We leave the reader with a recent article by Conway on unsettleable arithmetical problems, featuring the $3x + 1$ problem and "Collatzian games" [**4**]. It's a casual warning not to be too occupied with answering the big questions. Regardless, have fun and let us know what you discover.

REFERENCES

1. The On-Line Encyclopedia of Integer Sequences. `http://oeis.org`. Conway's creeper sequence (A164338), Jacobsthal numbers (A001045).
2. Paul J. Andaloro, The $3x + 1$ problem and directed graphs, *Fibonacci Quart.* **40** (2002) 43–54. `http://www.fq.math.ca/Scanned/40-1/andaloro.pdf`

3. Andrew Bremner, personal communication, 2013.
4. John H. Conway, On unsettleable arithmetical problems, *Amer. Math. Monthly* **120** (2013) 192–198. http://dx.doi.org/10.4169/amer.math.monthly.120.03.192
5. Curtis Cooper and Robert E. Kennedy, Base 10 RATS cycles and arbitrarily long base 10 RATS cycles, *Applications of Fibonacci Numbers*, vol. 8, Kluwer, Dordrecht, 1999, 83–93. http://dx.doi.org/10.1007/978-94-011-4271-7_9
6. Richard K. Guy, Unsolved problems: Don't try to solve these problems, *Amer. Math. Monthly* **90** (1983) 35–38 and 39–41. http://dx.doi.org/10.2307/2975688
7. Richard K. Guy, Conway's RATS and other reversals, *Amer. Math. Monthly* **96** (1989) 425–428. http://dx.doi.org/10.2307/2325149
8. Tanya Khovanova, Destinies of numbers. http://blog.tanyakhovanova.com/?p=155, accessed July 2009.
9. Stuart A. Kurtz and Janos Simon, The undecidability of the generalized Collatz problem, *Theory and Applications of Models of Computation*, Springer, Berlin, 2007, 542–553. http://dx.doi.org/10.1007/978-3-540-72504-6_49
10. Jeffrey C. Lagarias (ed.), *The Ultimate Challenge: The* $3x+1$ *Problem*, American Mathematical Society, Providence, RI, 2010.
11. John Simons and Benne de Weger, Theoretical and computational bounds for m-cycles of the $3n+1$-problem, *Acta Arith.* **117** (2005) 51–70. http://dx.doi.org/10.4064/aa117-1-3
12. Tanguy Urvoy, Regularity of congruential graphs, *Mathematical Foundations of Computer Science 2000* (Bratislava), Springer, Berlin, 2000, 680–689. http://dx.doi.org/10.1007/3-540-44612-5_63

Summary It's the age-old recurrence with a twist: sum the last two terms and *if the result is composite, divide by its smallest prime divisor* to get the next term (e.g., 0, 1, 1, 2, 3, 5, 4, 3, 7, . . .). These sequences exhibit pseudo-random behavior and generally terminate in a handful of cycles, properties reminiscent of $3x+1$ and related sequences. We examine the elementary properties of these "subprime" Fibonacci sequences.

RICHARD GUY has enjoyed teaching mathematics at all levels from kindergarten to graduate students in Britain, Singapore, India, and Canada. He believes that math is fun and available to everyone, and attributes his longevity to his love of mathematics.

TANYA KHOVANOVA received her Ph.D. in mathematics from Moscow State University. She currently does research in recreational mathematics. In addition to being successful in Olympiads herself, she now coaches gifted high school children for math Olympiads. In addition, she is a head mentor at the PRIMES and RSI programs, helping high school students conduct research in mathematics. She also has a popular math blog that emphasizes puzzles and interesting math problems.

JULIAN SALAZAR is an undergrad studying mathematics and computer science at Harvard University. His academic interests include recreational math, algebraic number theory, algebraic geometry, and data science. This paper represents independent research he did in his senior year of high school.

Here is another cool identity to celebrate the new year.

$$2014 = 2015 - \cfrac{1 \times 2015}{2017 - \cfrac{2 \times 2016}{2019 - \cfrac{3 \times 2017}{2021 - \cfrac{4 \times 2018}{\ddots}}}}$$

Contributed by Joseph Tonien,
University of Wollongong, Australia

Iteration of Sine and Related Power Series

CHRISTOPHER TOWSE
Scripps College
Claremont, CA 91711
ctowse@scrippscollege.edu

We begin with the familiar power series

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \cdots$$

which we can take to be the definition of the sine function. A careful second-semester calculus student could iterate sine and come up with

$$\sin^{\circ 2} x := \sin(\sin(x)) = x - \frac{1}{3}x^3 + \frac{1}{10}x^5 - \frac{8}{315}x^7 + \cdots ,$$

where here it is no longer clear what pattern the coefficients follow other than the fact that all the even-degree coefficients are zero. (Is even this small fact completely obvious at first glance? Maybe or maybe not.) Perhaps our intrepid calculus student could also compute (term-by-term)

$$\sin^{\circ 3} x = x - \frac{1}{2}x^3 + \frac{11}{40}x^5 - \frac{731}{5040}x^7 + \cdots ,$$

and maybe even

$$\sin^{\circ 4} x = x - \frac{2}{3}x^3 + \frac{8}{15}x^5 - \frac{3}{7}x^7 + \cdots .$$

In general, we claim that the nth iterate of sine is

$$\sin^{\circ n} x = x - \frac{n}{6}x^3 + \left(\frac{1}{24}n^2 - \frac{1}{30}n\right) x^5 - \left(\frac{5}{432}n^3 - \frac{1}{45}n^2 + \frac{41}{3780}n\right) x^7 + \cdots . \tag{1}$$

Of course, the ellipsis is a bit misleading: The pattern is not at all clear. What can we even say about the next term? We might expect an x^9 term with coefficient equal to a degree-4 polynomial in n. A method for calculating each coefficient will follow from the discussion below, but we will not give a formula for the general term. We will, however, establish the degree (in n) of the coefficients and a bit more.

Let us introduce some notation for these coefficients. Writing

$$\sin^{\circ n} x = x + b_3^{\circ n} x^3 + b_5^{\circ n} x^5 + b_7^{\circ n} x^7 + \cdots$$

and considering each $b_k^{\circ n}$ as a function of n, we are claiming that

$$b_3^{\circ n} = -\frac{n}{6},$$

$$b_5^{\circ n} = \frac{1}{24}n^2 - \frac{1}{30}n,$$

$$b_7^{\circ n} = -\frac{5}{432}n^3 + \frac{1}{45}n^2 - \frac{41}{3780}n,$$

and so on.

Math. Mag. **87** (2014) 338–349. doi:10.4169/math.mag.87.5.338. © Mathematical Association of America

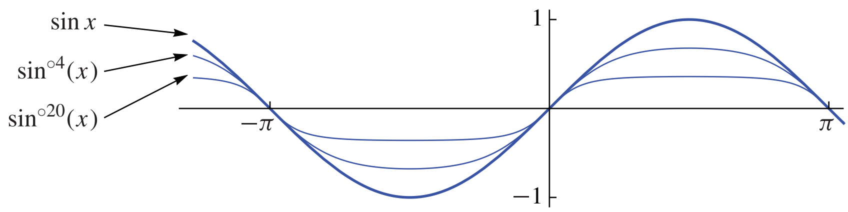


Figure 1 $y = \sin x$, $y = \sin^{\circ 4}(x)$, and $y = \sin^{\circ 20}(x)$

It will be easier to work in a bit more generality. Throughout the rest of this paper, we will consider a function f given by a power series of the form

$$f(x) = x + a_3x^3 + a_4x^4 + \cdots .$$

This is a quite general function overall; it has a fixed point of 0, which allows us to iterate easily, and it is normalized so that $f'(0) = 1$. The only real restriction is that we have set the quadratic term of f to be 0. We will usually assume that $a_3 \neq 0$.

We are interested in $f^{\circ n}(x)$, the nth iterate of f. As before, we write the kth coefficient of $f^{\circ n}(x)$ as $a_k^{\circ n}$ so that

$$f^{\circ n}(x) = x + a_3^{\circ n}x^3 + a_4^{\circ n}x^4 + \cdots .$$

(Of course, $a_k^{\circ 1} = a_k$, for all k.)

We could initially suspect that taking the limit as $n \to \infty$ might give us a hint about the behavior of the nth iterates. Let us go back to sine. FIGURE 1 shows the 1st, 4th, and 20th iterates of sine. Interestingly, although equation (1) shows that each individual coefficient goes to infinity, the iterate functions appear to shrink monotonically toward the zero function. In fact, if we consider our function pointwise, we realize that $\lim_{n\to\infty} \sin^{\circ n}(x) = 0$ for all x. The convergence is quite slow. The recursion spiderweb diagram (FIGURE 2) may look familiar to many readers.

However, once we establish that the coefficients $a_k^{\circ n}$ are indeed polynomials in n, we can attempt to analyze them. We will show that their degrees as polynomials are what we expect, and then a particular conjugate-scaling process presents itself. To see this in action, consider

$$\sqrt{n}\,\sin^{\circ n}(x/\sqrt{n}) = x - \frac{1}{6}x^3 + \left(\frac{1}{24} - \frac{1}{30n}\right)x^5 - \left(\frac{5}{432} - \frac{1}{45n} + \frac{41}{3780n^2}\right)x^7 + \cdots .$$

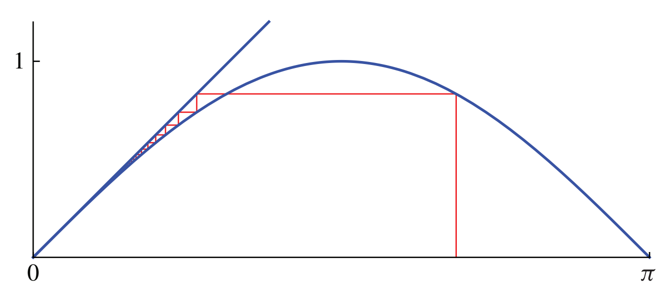


Figure 2 A spiderweb diagram for $x_{n+1} = \sin(x_n)$

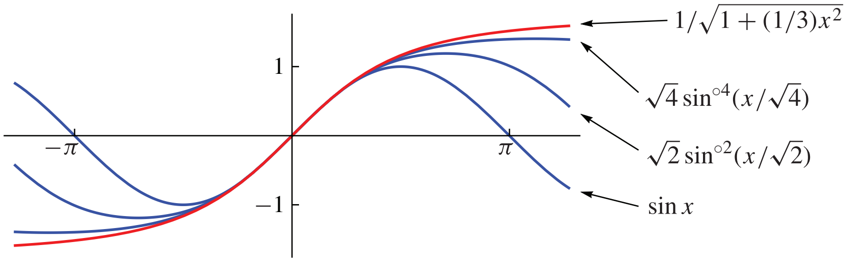


Figure 3 $y = \sin x$, $y = \sqrt{2}\sin^{\circ 2}(x/\sqrt{2})$, $y = \sqrt{4}\sin^{\circ 4}(x/\sqrt{4})$, and the limit function

As we can observe, after this rescaling we are able to take the limit as n goes to infinity without the coefficients going to infinity. (That is precisely why we chose to conjugate by $\sqrt{n}$.) This will allow us to identify, for all k, the leading coefficient of $a_k^{\circ n}$, as we will see. Graphically (FIGURE 3), the scaled iterates appear to converge slowly to the limit function (which we will establish later).

A diversion: formal power series

Before continuing, it may be interesting (but not necessary to what follows) to make some side observations regarding $f^{\circ n}$.

First, note that $f^{\circ m} \circ f^{\circ n} = f^{\circ(m+n)}$, so iteration of the functions corresponds to addition in the natural numbers. But now consider the tth "iterate" of f:

$$f^{\circ t}(x) = x + a_3^{\circ t}x^3 + a_5^{\circ t}x^5 + a_7^{\circ t}x^7 + \cdots,$$

with each $a_k^{\circ t}$ now viewed as a function of t. This change from n to t is meant to suggest that t might not stay within the natural numbers. Consider, for example, $f^{\circ 0}(x) = x$. Or, letting $t = -1$ and $f(x) = \sin x$, we get

$$\sin^{\circ -1}(x) = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \frac{5}{112}x^7 + \cdots,$$

which is $\arcsin x$, the inverse-sine function, so that

$$x = \sin(\arcsin x) = (\sin^{\circ 1} \circ \sin^{\circ -1})(x).$$

The additive nature of the iterations is preserved even when t is a non-positive integer.

We can also consider $t \notin \mathbb{Z}$. For example, the function

$$h(x) = \sin^{\circ 1/2}(x) = x - \frac{1}{12}x^3 - \frac{1}{160}x^5 - \frac{53}{40320}x^7 + \cdots$$

has the property that

$$(h \circ h)(x) = \sin x.$$

(Can we denote this $(\sqrt{\sin})(x)$? This function is called the *half-iterate* or the *functional square root* of sine.) Further, the functional cube root of sine, $c(x) = \sin^{\circ 1/3}(x)$, has the property that $(c \circ c \circ c)(x) = \sin x$, and so on. (Note that we are not addressing the question of whether these functions have non-trivial domains. We are only considering them as symbolic power series. Indeed, we already know that the domain has been reduced to just the interval $[-1, 1]$ when we consider arcsine.)

Back to the problem

At this point, we return to the problem of understanding the coefficients $a_k^{\circ n}$ of $f^{\circ n}$. Since $f^{\circ n}(x) = f(f^{\circ n-1}(x))$, we consider

$$\begin{aligned} f(f^{\circ n-1}(x)) = x + a_3^{\circ n-1}x^3 + a_4^{\circ n-1}x^4 + \cdots \\ + a_3(x + a_3^{\circ n-1}x^3 + a_4^{\circ n-1}x^4 + \cdots)^3 \\ + a_4(x + a_3^{\circ n-1}x^3 + a_4^{\circ n-1}x^4 + \cdots)^4 \\ + \cdots . \end{aligned}$$

We can easily see that $a_0^{\circ n} = 0$, $a_1^{\circ n} = 1$, and $a_2^{\circ n} = 0$ for all n. Our goal is to look at each $a_k^{\circ n}$ in terms of n.

Without too much more difficulty, we can check that $a_3^{\circ n} = a_3^{\circ n-1} + a_3$. Summing $a_3^{\circ i} - a_3^{\circ i-1}$ from $i = 2$ to n yields a telescoping sum and we get

$$a_3^{\circ n} - a_3 = \sum_{i=2}^{n}(a_3^{\circ i} - a_3^{\circ i-1}) = (n-1)a_3,$$

so

$$a_3^{\circ n} = na_3.$$

And similarly,

$$a_4^{\circ n} = na_4.$$

The formulas for $a_k^{\circ n}$ get complicated quickly. For example,

$$a_5^{\circ n} = na_5 + 3a_3^2 n(n-1)/2,$$

as we will see momentarily.

Faà di Bruno

Here we will use the Faà di Bruno formula for the higher derivatives of the composition of functions. The statement, below, of the Faà di Bruno formula is adapted from [**3**] with the notation changed to match our work thus far. (Other good sources include [**1**] and [**9**].) Given two functions g and h, Faà di Bruno gives us a (not-so-simple) expression for the kth derivative of the composition, $g \circ h$. First we establish a small bit of notation.

DEFINITION. Let M_k be the set of k-tuples of non-negative integers $\vec{m} = (m_1, m_2, \ldots, m_k)$ such that $m_1 + 2m_2 + \cdots + km_k = k$. For each $\vec{m} \in M_k$, let the size of the m be $|\vec{m}| = m_1 + m_2 + \cdots + m_k$.

FAÀ DI BRUNO FORMULA. *If g and h are functions with a sufficient number of derivatives, then*

$$\frac{d^k}{dx^k} g(h(x)) = \sum_{\vec{m} \in M_k} \frac{k!}{m_1!\,m_2!\cdots m_k!} g^{(j)}(h(x)) \left(\frac{h'(x)}{1!}\right)^{m_1} \left(\frac{h''(x)}{2!}\right)^{m_2} \cdots \left(\frac{h^{(k)}(x)}{k!}\right)^{m_k},$$

where we have written j for $|\vec{m}|$ to ease notation.

Ignoring the many constants, we see that each term is a $g^{(j)} \circ h$ multiplied by (powers of) derivatives of h. This formula can be a bit intimidating at first. It may help to start the $k = 4$ case. We see that M_4 consists of five vectors: $(0, 0, 0, 1)$, $(1, 0, 1, 0)$, $(0, 2, 0, 0)$, $(2, 1, 0, 0)$, and $(4, 0, 0, 0)$, with sizes 1, 2, 2, 3, and 4, respectively. So the 4th derivative of $g \circ h$ has five terms with $g^{(2)} \circ h$ appearing twice.

For our purposes, we let $g(x) = f(x)$ and $h(x) = f^{\circ n-1}(x)$. Then Taylor's (even MacLaurin's) formula says that $g^{(j)}(0) = j!\, a_j$ and $h^{(\ell)}(0) = \ell!\, a_\ell^{\circ n-1}$. And, in addition, if we temporarily fix n and let $F(x) = f^{\circ n}(x)$, then the coefficient $F^{(k)}(0)$ is just $k!\, a_k^{\circ n}$. Additionally, we have the hypotheses that $f(0) = f''(0) = 0$ and $f'(0) = 1$. So Faà di Bruno's formula is significantly simplified by the assumptions that $g(0) = h(0) = g''(0) = h''(0) = 0$ and $g'(0) = h'(0) = 1$. As an illustration, we note that if we remove these assumptions (though we retain $f(0) = 0$ so that our series are all centered at $x = 0$), the Faà di Bruno formula would tell us that $6a_3^{\circ n} = 6a_1 a_3^{\circ n-1} + 12a_2 a_1^{\circ n-1} a_2^{\circ n-1} + 6a_3 (a_1^{\circ n-1})^3$. Reimposing our assumptions that $a_1 = a_1^{\circ n-1} = 1$ and $a_2 = 0$ yields our formula for $a_3^{\circ n}$ as above: $a_3^{\circ n} = a_3 + a_3^{\circ n-1}$.

It is convenient to restate Faà de Bruno in terms of the coefficients of the MacLaurin series for an iterated function. In the spirit of MacLaurin, we simply plug in $x = 0$ (and cancel a $k!$ from both sides).

COROLLARY. *Let f be a function defined by a power series of the form $f(x) = a_1 x + a_2 x^2 + a_3 x^3 + \cdots$, and let $a_k^{\circ n}$ be the coefficient of x^k in the power series for the nth iterate of f. Then*

$$a_k^{\circ n} = \sum_{\vec{m} \in M_k} \frac{j!}{m_1!\, m_2! \cdots m_k!} a_j \prod_{\ell=1}^{k} (a_\ell^{\circ n-1})^{m_\ell}$$

where, as before, we have written j for $|\vec{m}| = \sum m_\ell$.

This is a bit more general than the case we are considering, and we won't need this full statement in what follows. (Note that any vector $\vec{m}$ with $j = |\vec{m}| = 2$ will yield a term that has a_2 in it. Similarly, any vector $\vec{m} = (m_1, m_2, \ldots, m_k)$ with $m_2 \neq 0$ will yield a term with $a_2^{\circ n}$ in it. Since, under our assumptions, $a_2 = a_2^{\circ n} = 0$, these terms will vanish. For this reason, we never need to consider these $\vec{m}$'s going forward.)

Details

The only thing we really need to take away from Faà di Bruno is the fact that, for a fixed k, the coefficient $a_k^{\circ n}$ is a linear combination of products of the form

$$a_j \prod_{\ell} (a_\ell^{\circ n-1})^{m_\ell}, \tag{2}$$

with coefficients that are positive. Each m_ℓ is non-negative and each ℓ is a distinct natural number, subject to the condition

$$\sum \ell m_\ell = k. \tag{3}$$

Note that condition (3) implies that we only need to consider terms with $j \leq k$ since $j = \sum m_\ell \leq \sum \ell m_\ell$. One can check that this gives (as noted above)

$$\begin{aligned} a_3^{\circ n} &= a_3^{\circ n-1} + a_3, \\ a_4^{\circ n} &= a_4^{\circ n-1} + a_4, \end{aligned}$$

$$a_5^{\circ n} = a_5^{\circ n-1} + 3a_3 a_3^{\circ n-1} + a_5, \qquad \text{and, additionally,}$$
$$a_6^{\circ n} = a_6^{\circ n-1} + 3a_3 a_4^{\circ n-1} + 4a_4 a_3^{\circ n-1} + a_6.$$

Summing $a_k^{\circ i} - a_k^{\circ i-1}$ from $i = 2$ to n yields a telescoping sum and thus a closed form for $a_k^{\circ n}$, assuming one already has a closed form for $a_s^{\circ n}$ for $s < k$.

We have already seen what happens for $k = 3$ and 4. For $k = 5$, we get

$$\begin{aligned} a_5^{\circ n} &= na_5 + 3a_3 \sum_{i=2}^{n} a_3^{\circ i-1} \\ &= na_5 + 3a_3 \sum_{i=2}^{n} (i-1)a_3 \\ &= na_5 + 3a_3^2 n(n-1)/2. \end{aligned}$$

Condition (3) tells us that the $a_\ell^{\circ n-1}$'s involved in this expression all have $\ell \le k$, and in fact there is exactly one term that involves $a_k^{\circ n-1}$ (when $\vec{m} = (0, \ldots, 0, 1)$, when $|\vec{m}| = 1$). This allows us always to get an expression for $a_k^{\circ n} - a_k^{\circ n-1}$ that only involves $a_i^{\circ n-1}$ for $i < k$, which we can then sum as above. The terms in this expression come from $\vec{m}$ with $j = |\vec{m}| \ne 1, 2$ since $\vec{m}$'s of size 2 all vanish, as we observed earlier.

Note that in each of the expressions so far, the coefficient a_3 appears and it appears in the dominant term (with respect to n) for $a_k^{\circ n}$.

PROPOSITION. *Let f be a function defined by a power series of the form $f(x) = x + a_3x^3 + a_4x^4 + \cdots$ with $a_3 \ne 0$, and let $a_k^{\circ n}$ be the coefficient of x^k in the power series for the nth iterate of f.*

For k odd, the coefficient $a_k^{\circ n}$ is a polynomial in n with leading term of the form

$$\gamma_k (a_3 n)^{(k-1)/2},$$

where $\gamma_k \in \mathbb{Q}$.

For k even, the coefficient $a_k^{\circ n}$ is a polynomial in n of degree at most $(k-2)/2$.

Note that the γ_k do *not* depend on f. They are absolute constants: the same whether $f(x) = \sin x$ or any other function of this form. In fact, we can compute $\gamma_3 = 1$, $\gamma_5 = 3/2$, $\gamma_7 = 5/2$, and so on. (The pattern should not be clear at this point. We will return to this after the proof of the proposition.)

Proof. We will use induction on k to look at the leading term. It is now clear that $a_k^{\circ n}$ is a polynomial in n. We have already seen that $a_3^{\circ n} = na_3$ and $a_4^{\circ n} = na_4$, and that $a_5^{\circ n} = (3/2)(n^2 - n)a_3^2 + na_5$. Assume $k \ge 6$.

Consider the degree with respect to n of a polynomial expression (in n), which we denote by $\deg_n$. We have $\deg_n(a) = 0$ if a is constant with respect to n. Also, $\deg_n(a_1a_2) = \deg_n(a_1) + \deg_n(a_2)$. Lastly, $\deg_n(a_1 + a_2) \le \max(\deg_n(a_1), \deg_n(a_2))$, with equality as long as the highest-degree terms do not cancel. Of course, if all coefficients are positive (and we are in characteristic zero), no cancellation can occur.

We compute the degree of each term $a_j \prod_\ell (a_\ell^{\circ n-1})^{m_\ell}$. Again, note that $j \ge 3$.

$$\begin{aligned} \deg_n\left(a_j \prod_\ell (a_\ell^{\circ n-1})^{m_\ell}\right) &= \sum m_\ell \deg_n(a_\ell^{\circ n-1}) \\ &\le \sum_{\ell \text{ odd}} m_\ell(\ell-1)/2 + \sum_{\ell \text{ even}} m_\ell(\ell-2)/2 \end{aligned}$$

$$= (1/2)\left(\sum m_\ell \ell - \sum_{\ell \text{ odd}} m_\ell - 2\sum_{\ell \text{ even}} m_\ell\right)$$
$$= (1/2)\left(k - j - \sum_{\ell \text{ even}} m_\ell\right)$$
$$\le (k-j)/2 \le (k-3)/2.$$

We still need to sum these terms. The key is that, whether k is odd or even, there are maximum-degree terms with positive coefficients. Note that each product of the form (2) appears with a coefficient that is a combinatorial symbol, and therefore positive. Thus, we only need to show that there is one term with the specified degree (and none with higher degree), as no cancellation can occur.

Case 1 (k odd). The question is whether there is a term of maximal degree, $(k-3)/2$. In other words, for an odd k, is there a term with $j = 3$? (Recall that $j = |\vec{m}| = \sum m_\ell$ for some $\vec{m} \in M_k$.)

If we let $\vec{m} = (2, 0, \ldots, 0, 1, 0, 0)$—in other words, $\ell_1 = 1$, $\ell_2 = k-2$, $m_{\ell_1} = 2$, and $m_{\ell_2} = 1$—then $|\vec{m}| = 3 = j$, and $\sum m_\ell \ell = k$. In addition, $\sum_{\ell \text{ even}} m_\ell = 0$, since $k-2$ and 1 are odd. Thus, $a_k^{\circ n}$ contains a term of the form $a_3 \cdot (a_1^{\circ n-1})^2 a_{k-2}^{\circ n-1} = a_3 a_{k-2}^{\circ n-1}$, which has degree $(k-3)/2$ in n.

Case 2 (k even). For a variety of reasons, there cannot be any terms of degree $(k-3)/2$. This would be a non-integer degree, for one! Thus, we are trying to show that there exists a term of degree $(k-4)/2$ in n. We do this by demonstrating a term with $j = 4$.

If we let $\vec{m} = (3, 0, \ldots, 0, 1, 0, 0, 0)$—in other words $\ell_1 = 1$, $\ell_2 = k-3$, $m_{\ell_1} = 3$, and $m_{\ell_2} = 1$—then $|\vec{m}| = 4 = j$, and $\sum m_\ell \ell = k$. In addition, $\sum_{\ell \text{ even}} m_\ell = 0$, since $k-3$ and 1 are odd. Thus, $a_k^{\circ n}$ contains a term of the form $a_4 \cdot (a_1^{\circ n-1})^3 a_{k-3}^{\circ n-1} = a_4 a_{k-3}^{\circ n-1}$, which has degree $(k-4)/2$ in n. Of course, a_4 could be zero, in which case the degree would be lower. This is the case when $f(x) = \sin x$, for instance.

For odd k, we get some additional information: Since any term that starts with a_j has degree at most $(k-j)/2$, we see that there are no terms with $j > 3$ in the expression for $a_k^{\circ n}$ that have this highest degree. Thus, we get the added fact that (by induction on k) the expression for $a_k^{\circ n}$ is of degree $(k-3)/2 + 1 = (k-1)/2$ in a_3.

Finally, we see that for odd k the expression for $a_k^{\circ n}$ looks like

$$a_k^{\circ n} - a_k^{\circ n-1} = \sum_{\substack{\vec{m} \in M_k \\ |\vec{m}|=3}} C_{\vec{m}} a_3 \prod_\ell (a_\ell^{\circ n-1})^{m_\ell} + \text{lower-order terms}$$

where the sum is over sets of ℓ's that satisfy (3) with $j = |\vec{m}| = 3$, and the $C_{\vec{m}}$'s are combinatorial symbols.

Using what we have just shown, this says that there is some polynomial of degree $(k-3)/2$ in n and degree $(k-1)/2$ in a_3 such that

$$a_k^{\circ n} - a_k^{\circ n-1} = p_k(n) = \beta a_3^{(k-1)/2} n^{(k-3)/2} + \text{lower-order terms},$$

where β is a constant.

Summing from $i = 2$ to n, we get

$$\begin{aligned} a_k^{\circ n} - a_k &= \sum_{i=2}^{n} a_k^{\circ i} - a_k^{\circ i-1} \\ &= \sum_{i=2}^{n} p_k(i) \\ &= \beta a_3^{(k-1)/2} \sum_{i=2}^{n} i^{(k-3)/2} + \text{lower-order terms} \\ &= \beta a_3^{(k-1)/2} (2/(k-3)) n^{(k-1)/2} + \text{l.o.t.} \\ &= \gamma_k (a_3 n)^{(k-1)/2} + \text{l.o.t.} \end{aligned}$$

Here we have used the fact that $\sum_{i=2}^{n} i^c = (1/c)n^{c+1} + \text{l.o.t.}$ Both β and γ_k are constants, not dependent on n or the initial choice of f.

The proof of the even k case is essentially the same. We do not prove any claims about the appearance of a_j's. For example, $a_6^{\circ n} = (7/2)a_3 a_4 (n^2 - n)$. ■

Consequences

We have established that the nth iterate of f is

$$f^{\circ n}(x) = x + a_3^{\circ n} x^3 + a_4^{\circ n} x^4 + \cdots$$

where

$$a_k^{\circ n} = \gamma_k (a_3 n)^{(k-1)/2} + \cdots \quad \text{when } n \text{ is odd,}$$

and $a_k^{\circ n}$ has degree $(k-2)/2$ when n is even.

We are now in a position to isolate the γ's as promised.

Since we are looking at the nth iterate of f, it is natural to wonder what happens as n grows. We have

$$\begin{aligned} f^{\circ n}(x) &= x + a_3^{\circ n} x^3 + a_4^{\circ n} x^4 + \cdots \\ &= x + n a_3 x^3 + n a_4 x^4 + \left[\frac{3}{2} a_3^2 n^2 + \left(\frac{3}{2} a_3^2 + a_5\right) n\right] x^5 + \cdots . \end{aligned}$$

Even the x^3 coefficient grows with n, so simply taking the limit as n goes to infinity of $f^{\circ n}$ would not be very interesting. We can, however, consider the following scaling trick to control that particular growth. We conjugate by the scaling function $x \mapsto x/\sqrt{n}$ and consider $\sqrt{n} f^{\circ n}(x/\sqrt{n})$. The x term is unaffected. The x^3 term is $\sqrt{n}(n a_3)(x/\sqrt{n})^3 = a_3 x^3$; as we intended, the growth has been controlled. But notice that the x^5 term is

$$\sqrt{n} \left[\frac{3}{2} a_3^2 n^2 + \left(\frac{3}{2} a_3^2 + a_5\right) n\right] \left(\frac{x}{\sqrt{n}}\right)^5 = \left[\frac{3}{2} a_3^2 + \left(\frac{3}{2} a_3^2 + a_5\right) \cdot \frac{1}{n}\right] x^5.$$

By controlling for the growth of the x^3 term, we have automatically controlled for the x^5 term.

In fact, one of our main observations is that, due to the work we did in the Proposition, we can now see that *all* the terms are controlled by this one trick; for, taking the limit as $n \to \infty$, we get

$$\begin{aligned}\lim_{n\to\infty} \sqrt{n}\, f^{\circ n}(x/\sqrt{n}) &= x + a_3x^3 + (3/2)a_3^2x^5 + \cdots \\ &= x + \gamma_3 a_3 x^3 + \gamma_5 a_3^2 x^5 + \cdots .\end{aligned}$$

Note the intriguing result that the only dependence that this limiting function has on the original function f is on a_3. In fact, we can completely identify the limiting function in terms of a_3.

THEOREM. *Let f be a power series of the form $f(x) = x + a_3x^3 + a_4x^4 + \cdots$. Then*

$$\lim_{n\to\infty} \sqrt{n}\, f^{\circ n}(x/\sqrt{n}) = \frac{x}{\sqrt{1 - 2a_3x^2}}.$$

Proof. Consider the function

$$g(x) = \frac{x}{\sqrt{1 - 2a_3x^2}}.$$

The power series for g is (suspiciously)

$$g(x) = x + a_3x^3 + (3/2)a_3^2x^5 + (5/2)a_3^3x^7 + \cdots .$$

In addition, we can quickly check that $g^{\circ n}(x) = \frac{x}{\sqrt{1-2a_3nx^2}}$ for all n.

We have shown that *any* power series f of the form

$$f(x) = x + a_3x^3 + \cdots$$

has

$$\lim_{n\to\infty} \sqrt{n}\, f^{\circ n}(x/\sqrt{n}) = x + \gamma_3 a_3x^3 + \gamma_5 a_3^2x^5 + \gamma_7 a_3^3 x^7 + \cdots$$

where the γ's are absolute constants. This limit power series only depends on a_3. In particular, g is one such f. Since

$$g^{\circ n}(x) = \frac{x}{\sqrt{1 - 2a_3nx^2}},$$

we see that $\sqrt{n}\, g^{\circ n}(x/\sqrt{n}) = g(x)$ for all n.

Therefore, for any f,

$$\lim_{n\to\infty} \sqrt{n}\, f^{\circ n}(x/\sqrt{n}) = g(x). \qquad \blacksquare$$

Earlier we promised that we would return to and identify the γ's. Here we see that the values of the γ's are simply the coefficients in the power series for $\frac{x}{\sqrt{1-2x^2}}$.

Lastly, we return to the original question of iterating the sine function.

COROLLARY. *The nth iterate of the sine function is*

$$\sin^{\circ n}(x) = x - \frac{n}{6}x^3 + \left(\frac{n^2}{24} - \textit{l.o.t.}\right)x^5 - \left(\frac{5n^3}{432} - \textit{l.o.t.}\right)x^7 + \cdots ,$$

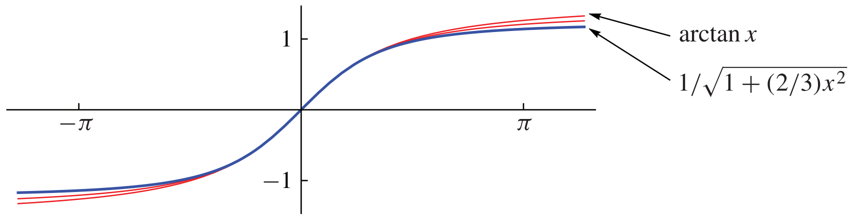


Figure 4 $y = \arctan x$, $y = 2\arctan^{\circ 4}(x/2)$, and the limit function $y = \frac{x}{\sqrt{1+(2/3)x^2}}$

where the coefficient of x^k is a polynomial in n whose leading term is $\gamma_k \left(\frac{n}{6}\right)^{(k-1)/2}$, where γ_k is the coefficient of x^k in the power series for $\frac{x}{\sqrt{1-2x^2}}$.

Consider the function we obtain by scaling (to control the main growth term of the coefficients), $\sqrt{n}\,\sin^{\circ n}(x/\sqrt{n})$. This sequence of functions converges as follows:

$$\lim_{n\to\infty} \sqrt{n}\,\sin^{\circ n}(x/\sqrt{n}) = \frac{x}{\sqrt{1+(1/3)x^2}}$$
$$= x - \frac{1}{6}x^3 + \frac{1}{24}x^5 - \frac{5}{432}x^7 + \cdots.$$

Of course, this same limit applies to any $f(x) = x - \frac{1}{3!}x^3 + \cdots$. Nevertheless, it is interesting to consider other familiar functions as well, such as arctangent (FIGURE 4).

COROLLARY. *The nth iterate of the arctangent function is*

$$\arctan^{\circ n}(x) = x - \frac{n}{3}x^3 + \left(\frac{n^2}{6} - \textit{l.o.t.}\right)x^5 - \left(\frac{5n^3}{54} - \text{l.o.t.}\right)x^7 + \cdots$$

so that, as before

$$\lim_{n\to\infty} \sqrt{n}\,\arctan^{\circ n}(x/\sqrt{n}) = \frac{x}{\sqrt{1+(2/3)x^2}}$$
$$= x - \frac{1}{3}x^3 + \frac{1}{6}x^5 - \frac{5}{54}x^7 + \cdots.$$

Additional comments and further directions

Although our method will give a formal limit to the scaled iterate of the tangent function $\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 - \frac{17}{315}x^7 + \cdots$, the behavior of $\tan^{\circ n}(x)$ is particularly bad. Note that the singularity of tangent at $x = \pi/2$ becomes a singularity of $\tan^{\circ 2}(x)$ at $x = \arctan(\pi/2)$. This becomes a singularity of $\tan^{\circ 3}(x)$ at $x = \arctan^{\circ 2}(\pi/2)$. Thus the singular points of $\tan^{\circ n} x$ converge toward 0 as n grows.

There are, of course, many questions related to our main theorem that have not been answered in this paper. For instance, it may be interesting to note that, for even degree (in x) terms, only a_3's and a_4's appear in the highest-degree (in n) term. Can the degrees to which these constants appear have some significance?

We may certainly wonder what happens if we allow our iterated function to have a non-zero quadratic term: $f(x) = x + a_2x^2 + a_3x^3 + \cdots$. The answer is that a_2 takes over. We can check that $a_k^{\circ n} \approx (na_2)^{k-1}$ (with leading coefficient 1). In order to control for the growth of the coefficients, we need to consider $nf^{\circ n}(x/n)$, and the limit is not nearly as interesting as in our case. Similarly, we have avoided considering the possibility that a_3 might be 0. In this case, our main theorem still holds, but the result just says that $\sqrt{n} f^{\circ n}(x/\sqrt{n}) \to x$. Essentially, $\sqrt{n}$ is not the right conjugate-scaling factor to use.

It has been suggested to the author that the Schwartzian derivative may be playing a role, though this role is not clear to me at this point. For those readers wondering, "What is . . . the Schwarzian derivative?" I would suggest [**6**].

We could also ask about the domain of $f^{\circ n}(x)$, or for the non-integer "iterates" of sine, such as $h(x)$ and $c(x)$, defined earlier. Even stranger, we might consider $t \in \mathbb{R}$ or $t \in \mathbb{C}$. If we think of f as an function over the real (or complex) numbers, we would be lead to the topic of *continuous iteration*. The reader interested in this point of view would do well to see [**4**] for a very thorough introduction via the topic of functional equations. A slightly different approach (perhaps more accessible and more directly related to our investigation) can be found in [**2**], particularly in Chapter 8.

An alternative approach would be to look at the dynamics of formal power series. Although this particular topic is not covered in Silverman's book [**11**], that book has an extensive bibliography that is useful. One common approach in this area is to look at the subject in the context of p-adic dynamics. After all, working within the p-adics is one way to control the growth of the coefficients in the series and not have to worry about convergence issues, at least not in the same way as over the real numbers. We can even consider letting t be in $\mathbb{Z}_p$, the p-adic integers. Typical questions that have been addressed have centered around the periodic points of arithmetic functions, and commuting power series (see [**5**] and [**10**], for example). In [**7**], the coefficients of iterates is addressed somewhat. Lastly, [**8**] is a non-p-adic example that might be of interest.

I would like to thank the referee who directed me to the book *Asymptotic Methods in Analysis* by N. G. de Bruijn [**2**]. As the title of the book suggests, his motivation comes from an entirely different direction, yet much of his chapter on iterated functions featured results related to those presented here. In particular, the function $g(x) = \frac{x}{\sqrt{1-x^2}}$ (which is nearly the $g(x)$ from our theorem) appears as the solution to Abel's equation

$$\phi\{G(x)\} - \phi(x) = 1$$

when we set $\phi(y) = -1/y^2$ [**2**, Sec. 8.7]. It follows (see [**2**] or [**4**]) that $g_n(x) = \frac{x}{\sqrt{1-nx^2}}$ is a solution to $\phi\{G(x)\} - \phi(x) = n$ for the same ϕ. (Abel's equation is an additive version of the more well known Schröder's equation $\psi\{w(x)\} = s\psi(x)$.) The function g in the theorem has a fascinating property that we have alluded to. If we let $g_s(x) = \frac{x}{\sqrt{1-sx^2}}$, then $g_{s+t}(x) = g_s(g_t(x))$, as we can check. Again, see [**4**] to start.

Acknowledgment The author would like to thank Andy Bernoff, Helen Grundman, Jonathan Lubin, Jorge Aarao, Francis Bonohon, and others for their ideas and comments, and for listening to me talk about this topic. The author was supported by a Scripps College Faculty Research Grant.

REFERENCES

1. Alex D. D. Craik, Prehistory of Faà di Bruno's formula, *Amer. Math. Monthly* **112** (2005) 119–130. http://dx.doi.org/10.2307/30037410
2. N. G. de Bruijn, *Asymptotic Methods in Analysis,* 3rd ed., Dover, New York, 1981.
3. Warren P. Johnson, The curious history of Faà di Bruno's formula, *Amer. Math. Monthly* **109** (2002) 217–234. http://dx.doi.org/10.2307/2695352
4. Marek Kuczma, *Functional Equations in a Single Variable,* Monografie Matematyczne, Tom 46, Państwowe Wydawnictwo Naukowe, Warsaw, 1968.
5. Jonathan Lubin, Non-Archimedean dynamical systems, *Compositio Math.* **94** (1994) 321–346.
6. Valentin Ovsienko and Sergei Tabachnikov, What is . . . the Schwarzian derivative? *Notices Amer. Math. Soc.* **56** (2009) 34–36. http://www.ams.org/notices/200901/tx090100034p.pdf

7. Ludwig Reich, Über die Koeffizienten der natürlichen Iterierten einer formalen Potenzreihe, *Math. Pannon.* **7**(1) (1996) 3–22. http://www.emis.de/journals/MP/index_elemei/mp07-1/mp07-1-003-022.pdf
8. Bruce Reznick, When is the iterate of a formal power series odd? *J. Austral. Math. Soc. Ser. A* **28** (1979) 62–66. http://dx.doi.org/10.1017/S1446788700014920
9. Steven Roman, The formula of Faà di Bruno, *Amer. Math. Monthly* **87** (1980) 805–809. http://dx.doi.org/10.2307/2320788
10. Ghassan Sarkis, On lifting commutative dynamical systems, *J. Algebra* **293** (2005) 130–154. http://dx.doi.org/10.1016/j.jalgebra.2005.08.007
11. Joseph H. Silverman, *The Arithmetic of Dynamical Systems*, Vol. 241 of Graduate Texts in Mathematics, Springer, New York, 2007. http://dx.doi.org/10.1007/978-0-387-69904-2

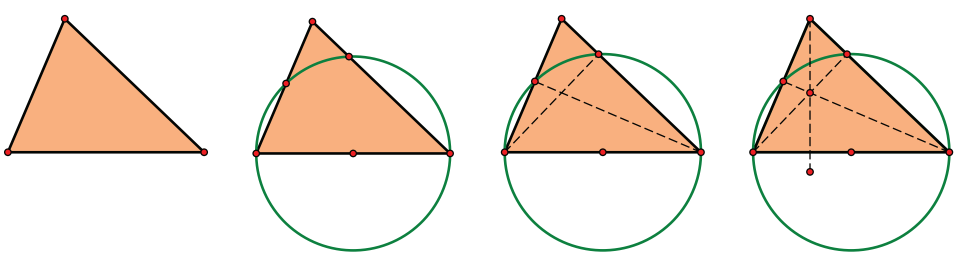
Summary We consider the nth iterate of the sine function and related functions by looking at the growth and form of the coefficients in the resulting power series. By controlling for the growth of the terms, as functions of n, we find a surprising relationship to a family of algebraic functions.

CHRIS TOWSE is an associate professor at Scripps College. He received his Ph.D. from Brown University and held positions at the University of Michigan, Swarthmore College, and Pomona College before arriving at Scripps. He directed the Claremont College Math REU Site from 2009 to 2011. His research is primarily in number theory and arithmetic geometry.

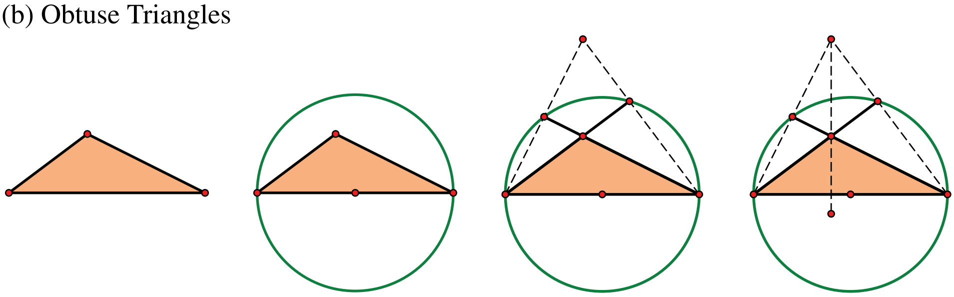
Proof Without Words: Rapid Construction of Altitudes of Triangles

LARRY HOEHN
Austin Peay State University
Clarksville, Tennessee
HoehnL@apsu.edu

(a) Acute Triangles



(b) Obtuse Triangles



Summary The altitudes of triangles are constructed using a circle and three lines.

Math. Mag. **87** (2014) 349. doi:10.4169/math.mag.87.5.349. © Mathematical Association of America

Straightedge and Compass Constructions in Spherical Geometry

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

Mathematicians love constructions. They begin with simple questions that even children can understand, but that often have unexpectedly difficult answers. A complete classification of constructible points, involving a tour through analytic geometry and abstract algebra, is included in many, perhaps most, abstract algebra textbooks.

In the most famous math book of all time, Euclid began with five basic assumptions about geometry and built upon those assumptions [**2**]. The first three of Euclid's assumptions can be seen as instructions in the use of straightedge and compass, two tools employed extensively by Greek mathematicians.

From these three assumptions, a "game" can be made. Given any set of points, we may construct the line determined by any two of them, or we may construct the circle with one of them as the center and another on the circle itself. We may add to the set of points any point arising as an intersection of two constructed lines, two constructed circles, or one constructed line and one constructed circle. Since we must start the game with something, we generally begin with two points, say A and B. The unit of measure being arbitrary in Euclidean geometry, we assume that the distance from A to B is 1. The goal of the game is to decide what, in general, is constructible. Of course, to simplify the game, we may choose specific things that we want to construct, such as a regular pentagon, a 20° angle, or the like, and try to answer those specific questions rather than the general question.

The Greeks were adept at the straightedge and compass game. For example, they were able to show that segments with rational numbers and square roots of positive rational numbers as lengths were constructible. However, they were stymied by some other values, such as $\sqrt[3]{2}$. The straightedge and compass game saw no major breakthroughs for centuries, until Gauss proved in 1796 that a regular 17-gon is constructible by showing that

$$\cos\left(\frac{2\pi}{17}\right) = \frac{-1+\sqrt{17}+\sqrt{34-2\sqrt{17}}+2\sqrt{17+3\sqrt{17}-\sqrt{34-2\sqrt{17}}-2\sqrt{34+2\sqrt{17}}}}{16},$$

together with the knowledge that sums, differences, products, quotients, and square roots of constructible numbers are also constructible. A few years later he extended this result to show that a regular 257-gon and a regular 65537-gon are also constructible [**3**].

Gauss' results imply that a regular n-gon can be constructed if

$$n = 2^{k_0}\, 3^{k_1}\, 5^{k_2}\, 17^{k_3}\, 257^{k_4}\, 65537^{k_5}, \tag{1}$$

where $k_0 \in \{0, 1, 2, \ldots\}$, and each of the other $k_i \in \{0, 1\}$. (In fact, we could throw into that equation any other Fermat prime to the 0 or 1 power, though at this time none are known.)

Math. Mag. **87** (2014) 350–359. doi:10.4169/math.mag.87.5.350. © Mathematical Association of America

The coup de grâce to the (Euclidean) constructibility game was delivered by Wantzel, who was able to classify completely those numbers that are and are not constructible [**11**]. Essentially the idea was to use analytic geometry to turn the geometric problem into an algebraic problem, and to use the tools of abstract algebra—in particular the field structure of the plane considered as the set of complex numbers—to solve the algebraic problem. This result is discussed in detail in many places, including many introductory texts on abstract algebra (such as [**9**]).

Spherical geometry

The infamous fifth postulate of Euclid, however, hogged more attention than the others, in particular more than the three that give rise to the "construction game." Essentially, this postulate says that, given a line and a point not on the line, there is exactly one line through the point and parallel to the given line. Centuries of mathematicians debated whether this postulate should, in fact, be a theorem rather than a postulate. Finally, in the early 19th century, two mathematicians, Bolyai and Lobachevsky, independently showed that the postulate was independent, and that making a different assumption leads to another geometry, currently called *hyperbolic geometry* [**1, 7**]. Years later Riemann noted that another assumption is possible, leading to still other geometries, currently called *elliptic* or *projective geometry*, and *double elliptic* or *spherical geometry* [**8**].

To our knowledge, the first person to investigate constructions in non-Euclidean geometry was al-Buzjani, a Persian mathematician, over 1000 years ago. He studied spherical geometry and trigonometry, and demonstrated numerous constructions on the sphere. A great introduction to his work can be found in [**10**].

In his beautifully written Master's thesis, Kincaid [**5**] showed that constructible segment lengths on the sphere are closed under the operations we will show, together with a "cosine relation." This cosine relation is a spherical analog of the Pythagorean Theorem. If a segment length is constructible, then a cosine relation holds; but the converse does not necessarily hold, so this relation cannot be used to "solve" the spherical constructions game the way Wantzel solved the Euclidean version of the game.

Let A and B represent points in a given geometry. We use notation AB to represent the distance from A to B, $\overleftrightarrow{AB}$ to represent the geodesic (line) determined by A and B (though we shall use the words *geodesic* and *line* interchangeably), and $\circ AB$ to represent the circle with center A and radius AB. We use the notation $\overline{AB}$ to denote the segment defined by A and B, and note that, in the spherical case, there are two such segments; we shall use this notation to refer to the shorter of the two, and $\overline{\overline{AB}}$ to refer to the longer.

Unlike Euclidean planes, two spheres are not necessarily congruent to one another; their radii must be congruent for the spheres themselves to be congruent. Hence, in a strict sense we could discuss the spherical construction game separately for each sphere of radius $0 < r < \infty$. In addition, and again unlike the Euclidean plane, we shall show that the choice of starting conditions affects the outcome of the game. However, we note that, since two points that are distance x apart on a sphere of radius 1 are equivalent to two points that are distance rx apart on a sphere of radius r, we shall assume without loss of generality that spherical geometry is the set of points $S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$. We note that distance between two points on the sphere is measured by arc length rather than chord length, and since we are restricting ourselves to the unit sphere, that is equivalent to the measure of the central angle subtended by those points. Equivalently, the distance is the inverse cosine of the dot product of the two points considered as vectors.

For each point A on the sphere, there is a unique point whose coordinates are just the negative coordinates of A. We call this point the *antipode* of A and denote it A'. We also note that for any two non-antipodal points A and B, there are two points C and C' that are each distance $\pi/2$ from each of A and B. These points are called *poles* of A and B. Since C and C' are also poles for any points on $\overleftrightarrow{AB}$, we sometimes refer to them as *poles* for the line.

We note that, as in the Euclidean case, we may use analytic geometry to transform geometric information into algebraic information. In particular, the equation of a "line" (geodesic) determined by two (non-antipodal) points $A(x_0, y_0, z_0)$ and $B(x_1, y_1, z_1)$ in spherical geometry is

$$(y_0 z_1 - y_1 z_0)\, x + (z_0 x_1 - z_1 x_0)\, y + (x_0 y_1 - x_1 y_0)\, z = 0, \tag{2}$$

which is equivalent to $(\vec{A} \times \vec{B}) \cdot \langle x, y, z\rangle = 0$. The equation of the circle centered at A and containing point B (which may be A') is

$$x_0 x + y_0 y + z_0 z = x_0 x_1 + y_0 y_1 + z_0 z_1, \tag{3}$$

which is equivalent to $\vec{A} \cdot \langle x, y, z\rangle = \vec{A} \cdot \vec{B}$.

In contrast with the Euclidean case, there is no natural field structure on the sphere; in fact, it doesn't even have a Lie group structure. (This means that there isn't a natural way to define something like addition for points on the sphere.) There is, however, a natural group structure on the circle, which we will use. Each point (x, y) on the circle can be written uniquely in the form $(\cos\theta, \sin\theta)$ for some $\theta \in (-\pi, \pi]$. For a group operation, we use angle addition (modulo 2π). In other words, $(\cos\theta_1, \sin\theta_1) + (\cos\theta_2, \sin\theta_2) = (\cos(\theta_1 + \theta_2), \sin(\theta_1 + \theta_2))$.

We let $\{A, B\}$ be the set of *starting conditions* for the construction game. The ultimate goal is to decide what points on the sphere are constructible using only a (spherical) straightedge and compass. If the distance between points A and B is $\pi/4$, then the starting conditions are called *standard starting conditions*. As we will see below, several different (most) starting conditions imply the existence of two constructible points of distance $\pi/4$, so it makes sense to start with something that is in general constructible anyway.

Some special starting conditions

EXAMPLE 1. *Let $AB = \pi$, that is, the starting conditions are two antipodal points. Then the set of constructible points on the sphere is exactly $\{A, B\}$.*

This is easy to verify. We note that since antipodal points do not determine a line, we cannot construct $\overleftrightarrow{AB}$. We then note that $\circ AB = \{B\}$, since the set of points of distance π from A is just the antipode, B. Similarly, $\circ BA = \{A\}$. Hence $\circ AB \cap \circ BA = \{\,\}$, and no new points are constructible. We call a set of starting conditions of this form *trivial of type 1*.

EXAMPLE 2. *Let $AB = \pi/2$. Then there is a point $C \in S^2$ (a pole of $\overleftrightarrow{AB}$) so that $AB = AC = BC = \pi/2$, and the set of constructible points on the sphere is exactly $\{A, A', B, B', C, C'\}$.*

To see this, we note that

$$\circ AB \cap \circ BA = \{C, C'\},$$
$$\circ AC \cap \circ CA = \{B, B'\}, \text{ and}$$
$$\circ BC \cap \circ CB = \{A, A'\}.$$

Now it is elementary to check that $\circ AB = \circ AB' = \circ A'B = \circ A'B' = \circ AC = \circ AC' = \circ A'C = \circ A'C' = \overleftrightarrow{BC} = \overleftrightarrow{BC'} = \overleftrightarrow{B'C} = \overleftrightarrow{B'C'}$, and so on. Although there are $\binom{6}{2} = 15$ pairs of points that we might use to construct lines and circles, still there are only three distinct non-point constructions.

We call a set of starting conditions of this form *trivial of type 2*.

Note that the constructible set for starting conditions that are trivial of type 2 contains three disjoint sets of constructible points that are trivial of type 1. This shows that spherical constructions cannot, in general, be "reverse engineered." In other words, just because we can construct point C given points A and B does *not* mean that we can construct point B if given points A and C.

EXAMPLE 3. *Let $AB = 2\pi/3$. Then there is a unique point $C \in S^2$ so that the set of constructible points is exactly $\{A, B, C\}$.*

To make calculations easier, we take $A = (1, 0, 0)$ and $B = (-1/2, \sqrt{3}/2, 0)$. Then

$$\overleftrightarrow{AB} = \left\{(x, y, z) \in S^2 : z = 0\right\},$$
$$\circ AB = \left\{(x, y, z) \in S^2 : x = -\tfrac{1}{2}\right\}, \text{ and}$$
$$\circ BA = \left\{(x, y, z) \in S^2 : -\tfrac{1}{2}x + \tfrac{\sqrt{3}}{2}y + 0z = -\tfrac{1}{2}\right\}.$$

Then $C = \circ AB \cap \circ BA = (-1/2, -\sqrt{3}/2, 0)$.

We note that $\overleftrightarrow{AB} \cap \circ AB = \{B, C\}$ and $\overleftrightarrow{AB} \cap \circ BA = \{A, C\}$. It only remains to show that the intersection of any lines or circles determined by A, B, and/or C is one or two of A, B, and/or C. This is elementary to verify.

We call a set of starting conditions of this form *trivial of type 3*. Any starting conditions that are not trivial of type 1, 2, or 3 are said to be *non-trivial*.

We note that, in light of Examples 1, 2 and 3, the starting conditions change the outcome. Thus, in the spherical version of the construction game, unlike the Euclidean version, there cannot be said to be a *general* set of starting conditions. However, in the next section we will see that, except for these three examples, constructible points are in fact dense on the sphere.

Main results

In this section, we prove some results in a traditional theorem-proof format. Each proof proceeds by construction, and should be replicable by undergraduate readers by finding the equations involved (using (2) for lines and (3) for circles), and then solving those equations using elimination of variables.

THEOREM 1. *Let A and B be given so that $AB < 2\pi/3$ and $AB \neq \pi/2$. The perpendicular bisector of $\overleftrightarrow{AB}$ is constructible.*

Proof. Essentially, we use Euclid's proof of Proposition 10, noting exceptions along the way.

Without loss of generality, let $A = (1, 0, 0)$ and $B = (x_0, y_0, 0)$. The conditions on AB give us $-1/2 < x_0 \neq 0$. Then

$$\circ AB \cap \circ BA = \left\{ \left(x_0, \frac{x_0 - x_0^2}{y_0}, \pm\sqrt{\frac{(1-x_0)(2x_0+1)}{1+x_0}} \right) \right\}.$$

The condition $-1/2 < x_0$ means that the set contains two distinct points; call them C and D. The condition $x_0 \neq 0$ means that they are not antipodes, so they determine a line. Next we consider

$$\overleftrightarrow{AB} \cap \overleftrightarrow{CD} = \left\{ \pm \left(\sqrt{\frac{1+x_0}{2}}, \sqrt{\frac{1-x_0}{2}}, 0 \right) \right\}.$$

We call these two points M and M'.

We use an analytic argument to see that M bisects $\overline{AB}$. Since $\cos(AB) = x_0$, $\cos(AM) = \sqrt{(1-x_0)/2}$. The latter, solved for x_0, is just $x_0 = 1 - 2\cos^2(AM)$. A standard trigonometric identity now shows that $2AM = AB$. We note also that M' is the bisector of $\overline{\overline{AB}}$.

We could use an analytic argument to show that $\overleftrightarrow{AB} \perp \overleftrightarrow{MM'}$, too, but a geometric argument is simpler. If two triangles have each of their three sides pairwise congruent to one another, then those two triangles are congruent. (This theorem is known to hold in each of Euclidean, hyperbolic, and elliptic geometries [**4**].) Because $AM = BM$ has been shown, $PC = QC$ holds by construction, and because $MC = MC$ holds by reflexivity, we have $\Delta AMC \cong \Delta BMC$, and thus $\angle AMC \cong \angle BMC$. Since, in addition, the measures of $\angle AMC$ and $\angle BMC$ sum to 180°, this shows that $\angle AMC$ must be a right angle, completing the proof. ■

THEOREM 2. *Let A and B be given so that $AB \neq \pi$ and $AB \neq 2\pi/3$. Then the poles for $\overleftrightarrow{AB}$ are constructible.*

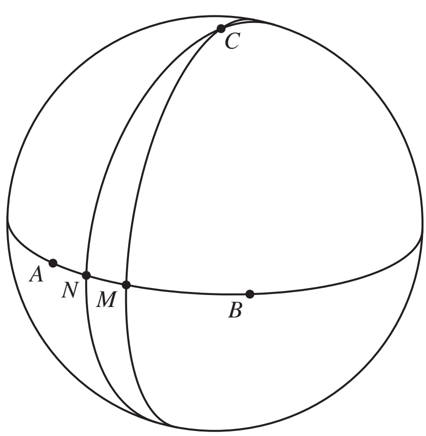


Figure 1 Constructing a pole

Proof. Without loss of generality, we let $A = (1, 0, 0)$ and $B = (x_0, y_0, 0)$. We may in addition take $y_0 > 0$, so that we have $y_0 = \sqrt{1 - x_0^2}$.

We first show that we may assume $AB < 2\pi/3$. Suppose, instead, that $2\pi/3 < AB < \pi$, so $-1 < x_0 < -1/2$. Note that $\overleftrightarrow{AB} \cap \circ BA = \{A, C\}$, where (using the fact that $x_0^2 + y_0^2 = 1$) $C = \left(x_0^2 - y_0^2, 2x_0y_0, 0\right)$. In addition, $AC = \cos^{-1}(\vec{A} \cdot \vec{C}) =$

$\cos^{-1}(x_0^2 - y_0^2)$. Since $x_0^2 + y_0^2 = 1$, we have that $x_0^2 - y_0^2 = 2x_0^2 - 1$. Since $-1 < x_0 < -1/2$, we have that $-1/2 < x_0^2 - y_0^2 < 1$. Together, these imply that $0 < AC < 2\pi/3$. Hence there is a constructible point C on $\overleftrightarrow{AB}$ whose distance is less than $2\pi/3$ from A. Any points constructible from $\{A, C\}$ are thus constructible from $\{A, B\}$.

If $AB = \pi/2$, then Example 2 already showed how to construct the poles for $\overleftrightarrow{AB}$. Hence we assume that $\pi/2 \neq AB < 2\pi/3$.

Now we apply Theorem 1 to $\overleftrightarrow{AB}$, obtaining perpendicular bisector ℓ and midpoint M, with

$$\ell = \left\{(x, y, z) \in S^2 : -2x_0 \frac{1 - x_0}{1 + x_0}\sqrt{2x_0 + 1} \cdot x + 2x_0\sqrt{\frac{(1 - x_0)(2x_0 + 1)}{1 + x_0}} \cdot y + 0 \cdot z = 0\right\},$$

$$\{M, M'\} = \left\{\pm\left(\sqrt{\frac{1 + x_0}{2}}, \sqrt{\frac{1 - x_0}{2}}, 0\right)\right\}.$$

Note that $AM < \pi/3$, so we may apply Theorem 1 to $\overleftrightarrow{AM}$, obtaining perpendicular bisector m and midpoint N. We do so by merely incrementing subscripts, setting $x_1 = \sqrt{(1 + x_0)/2}$, and obtaining

$$m = \left\{(x, y, z) \in S^2 : -2x_1 \frac{1 - x_1}{1 + x_1}\sqrt{2x_1 + 1} \cdot x + 2x_1\sqrt{\frac{(1 - x_1)(2x_1 + 1)}{1 + x_1}} \cdot y + 0 \cdot z = 0\right\}.$$

The two lines ℓ and m intersect in $\{C, C'\} = \{(0, 0, \pm 1)\}$, as in FIGURE 1. It is easy to see that these are the poles for $\overleftrightarrow{AB}$, completing the proof. ■

COROLLARY. *Let distinct points P, Q, and R be given, with $Q \neq P'$ and $PR \neq PQ$. Then the poles of $\overleftrightarrow{PQ}$ are constructible.*

Proof. First we note that we may assume $R \in \overleftrightarrow{PQ}$, since $\circ PR \cap \overleftrightarrow{PQ}$ is a pair of points in $\overleftrightarrow{PQ}$ that are distance PR from P. Then we have either $\{P, Q\}$, $\{P, R\}$, or $\{Q, R\}$ satisfy the hypothesis of Theorem 2, so the poles of $\overleftrightarrow{PQ}$ are constructible, or $PQ = QR = \pi/2$, in which case Example 2 demonstrates how to construct the poles. This completes the proof. ■

Let $\{A, B\}$ be given with $0 < AB < \pi$. Define a bijection ϕ between points of $\overleftrightarrow{AB}$ and the real numbers $(-\pi, \pi]$, so that $\phi(A) = 0$, $\phi(A') = \pi$, and $\phi(X) = AX$ if $X \neq A, A'$ is on the same component of $\overleftrightarrow{AB} \backslash \{A, A'\}$ as B, or $\phi(X) = -AX$ otherwise. In other words, $\phi(X)$ is just the signed distance $\pm AX$, where the sign is positive if X is in the same direction as B and negative otherwise.

The set of numbers $(-\pi, \pi]$ form a group under addition modulo 2π. For points $X, Y \in \overleftrightarrow{AB}$, we define $-X = \phi^{-1}(-\phi(X))$ and $X + Y = \phi^{-1}(\phi(X) + \phi(Y))$. In this sense, the set of points in $\overleftrightarrow{AB}$ form an additive group, which we shall denote G. This is the circle group, in which elements of the circle are "added" via standard angle addition, where the angle is as measured from point A in the (angular) direction of point B.

THEOREM 3. *The constructible numbers in $\overleftrightarrow{AB}$ form a subgroup of G.*

Proof. If the set of constructible numbers is as in Example 1 or 3, the result is straightforward, so we assume that the set of constructible numbers has at least five members. We show that if $X, Y \in \overleftrightarrow{AB}$ are constructible, then $-X = \phi^{-1}(-\phi(X))$ and

$X + Y = \phi^{-1}(\phi(X) + \phi(Y))$ are constructible. The proof then follows from associativity of G and the fact that A acts as the identity element.

Again for ease of computation we assume $A = (1, 0, 0)$ and that $B = (x_0, y_0, 0)$. If $X = A'$, then

$$-X = \phi^{-1}(-\phi(A')) \pmod{2\pi} = \phi^{-1}(\pi) = X,$$

so X is its own inverse.

Let $X \neq A, A'$ be a constructible point on $\overleftrightarrow{AB}$, and write $X = (x_1, y_1, 0)$. Consider $\overleftrightarrow{AB} \cap \circ AX = \{(x_1, y_1, 0), (x_1, -y_1, 0)\}$. The first of these points is X. Notice that $\phi\,((x_1, -y_1, 0)) = -\phi(X)$; in other words, $-X = (x_1, -y_1, 0) = \phi^{-1}(-\phi(X))$. This shows that $-X$ is constructible.

Let $X = (x_1, y_1, 0)$ and $Y = (x_2, y_2, 0)$ be constructible points on $\overleftrightarrow{AB}$. The poles for $\overleftrightarrow{AB}$ are constructible by the Corollary. One of the poles, which we denote C, is $(0, 0, 1)$. Then $\circ AX \cap \overleftrightarrow{AC} = \{(x_1, 0, \pm y_1)\}$. We denote these points D_0 and D_1. Then we obtain

$$(\circ CD_0 \cup \circ CD_1) \cap \overleftrightarrow{YC} = \{(x_1x_2, x_1y_2, \pm y_1), (-x_1x_2, -x_1y_2, \pm y_1)\},$$

which we denote E_i for $i = 0, 1, 2, 3$. Lastly, we consider $\circ YE_i \cap \overleftrightarrow{AX}$, which is the set

$$\left\{ \begin{array}{ll} (x_1x_2 + y_1y_2, -y_1x_2 + x_1y_2, 0), & (-x_1x_2 - y_1y_2, -y_1x_2 - x_1y_2, 0), \\ (x_1x_2 - y_1y_2, y_1x_2 + x_1y_2, 0), & (-x_1x_2 + y_1y_2, y_1x_2 - x_1y_2, 0) \end{array} \right\}.$$

We call these points F_i for $i = 0, 1, 2, 3$; part of their construction process can be seen in FIGURE 2. Recall that $AX = \cos^{-1}(x_1)$ and $AY = \cos^{-1}(x_2)$. This gives us

$$\begin{aligned} AF_i &= \cos^{-1}(\pm x_1x_2 \pm y_1y_2) \\ &= \cos^{-1}\left(\pm \cos(AX)\cos(AY) \pm \sin(AX)\sin(AY)\right) \\ &= \cos^{-1}\left(\cos(\pm AX \pm AY)\right) \\ &= \pm AX \pm AY, \end{aligned}$$

and thus the F_i are the points $\pm X \pm Y$, and in particular one of them is $X + Y$, which is therefore constructible. ■

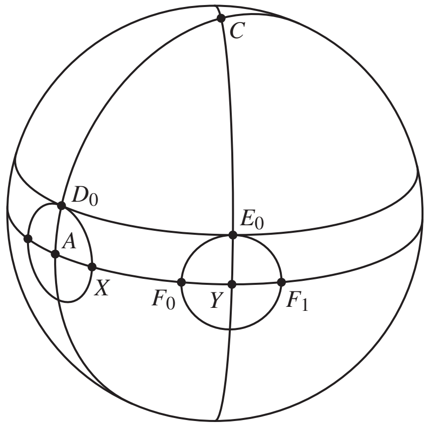


Figure 2 Adding intervals

THEOREM 4. *Let AB be neither $\pi/2$, $2\pi/3$, nor π; that is, assume the starting conditions are non-trivial. Then there exists a constructible point $M \in \overleftrightarrow{AB}$ with $AM = \pi/4$.*

Proof. Following the proof of Theorem 2, we construct the set $\{A, A', C, C', D, D'\}$ so that $AC = \pi/2$, where C is on the same component of $\overleftrightarrow{AB} \setminus \{A, A'\}$ as B, and so that $\{D, D'\} = \circ AC \cap \circ CA$, that is, D and D' are the poles for $\overleftrightarrow{AC}$.

We may assume $AB < \pi/2$; for if not, $E = B - C$ is constructible by Theorem 3, and $AE < \pi/2$. We may also assume that B falls on $\overline{AC}$; for if not, $-B$ is constructible by Theorem 3, and falls on $\overline{AC}$. Now we may also assume that $AB < \pi/4$, since if $AB > \pi/4$ then $E = C - B$ is constructible, where $AE < \pi/4$, and if $AB = \pi/4$, we are done.

Let $F = C - B$. Since $0 < AB < AC = \pi/2$, we have that $0 < AF < \pi/2$, so we can find the perpendicular bisector ℓ of $\overleftrightarrow{BF}$ by Theorem 1, and the points $\{M, M'\} = \ell \cap \overleftrightarrow{BF}$ are the midpoints of $\overline{BF}$ and $\overline{\overline{BF}}$. Assume that $M \in \overline{BF}$ (and $M' \in \overline{\overline{BF}}$). We will show that $AM = CM$, that is, $AM = \pi/4$.

Note that $BM = FM$ and $AB = FC$ by construction. Hence, in the group structure, $M = (M - B) + (B - A)$ and $M = (C - F) + (F - M)$, which implies that $M - A = C - M$; that is, $MA = CM$. Since $AC = \pi/2$, this implies that $AM = \pi/4$. This completes the proof. ∎

Since a segment of length $\pi/4$ is constructible under any non-trivial starting conditions, we shall call starting conditions $\{A, B\}$ with $AB = \pi/4$ *standard starting conditions.*

THEOREM 5. *Assume standard starting conditions. Then the set of constructible points contained in $\overleftrightarrow{AB}$ is dense in $\overleftrightarrow{AB}$.*

Proof. Let $\epsilon > 0$ be given, and let $X \in \overleftrightarrow{AB}$. We will show that there is a constructible point on $\overleftrightarrow{AB}$ whose distance from X is less than ϵ. Without loss of generality, we assume that $\phi(X) > 0$. As we have already seen, $\overleftrightarrow{AB}$ can be bisected by C_0, so that $AC_0 = \pi/8$. Recursively, $\overleftrightarrow{AC}_{i-1}$ can be bisected by C_i such that $AC_i = \pi/(2^{3+i})$, and thus $\phi(C_i) = \pi/(2^{3+i})$. Choose $i \in \mathbb{N}$ such that $\pi/2^{3+i} < \epsilon$. By the Archimedean Property, there is some $n \in \mathbb{N}$ such that $(n-1)\pi/(2^{3+i}) < \phi(X) < n\pi/(2^{3+i})$. Let $W = n \cdot C_i$, which is constructible because C_i is, and because constructible points are closed under (group) addition. Then $WX < \epsilon$, completing the proof. ∎

THEOREM 6. *Assume standard starting conditions. Then the set of constructible points is dense in S^2.*

Proof. Let $\epsilon > 0$ be given, and choose a point $X \in S^2$. If $X \in \overleftrightarrow{AB}$ we are done by Theorem 5, so assume not. Let C be a pole for $\overleftrightarrow{AB}$, and let $\{Y, Y'\} = \overleftrightarrow{CX} \cap \overleftrightarrow{AB}$. Without loss of generality, we assume that $AY > 0$. Repeat the argument of Theorem 5 to find some constructible point D such that $DY < \epsilon/2$. Let E be a pole for $\overleftrightarrow{CD}$, and $\{Z, Z'\} = \overleftrightarrow{EX} \cap \overleftrightarrow{CD}$. Again without loss of generality, we assume that Z is closer to X (and therefore that Z' is closer to X'). Repeat the argument of Theorem 5 to find some constructible point F so that $FZ < \epsilon/2$; see FIGURE 3.

We claim that $FX < \epsilon$. We have that $FX \leq FZ + ZX$ by the triangle inequality, which holds in spherical geometry [**4**]. In addition, using the spherical law of sines we have $\sin(ZX) = \sin(CX)\sin(\angle ZCX)$ and $\sin(DY) = \sin(\angle DCY)$. However, $\angle ZCX = \angle DCY$, and $0 < \sin(CX) < 1$ implies that $\sin(ZX) < \sin(DY)$. Again, since ZX and

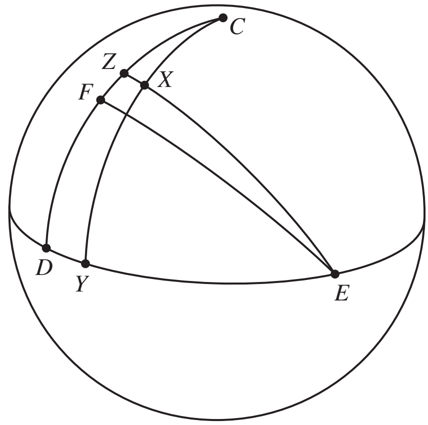


Figure 3 Constructible points are dense

DY are between 0 and $\pi/2$, we have $ZX < DY$. It follows that $FX < \epsilon/2 + \epsilon/2 = \epsilon$, and the proof is complete. ■

More results and even more questions

Not only have we *not* solved the constructibility question for the sphere, we doubt whether the question is answerable in general. (In a forthcoming paper, we give a necessary condition for a point in S^2 to be constructible, allowing us to give examples of non-constructible points, but casting doubt on points that fulfill the condition.) However, we can use the results thus far to explore the question of constructions in the projective plane, which we envision as the set of points satisfying the equation $x^2 + y^2 + z^2 = 1$, together with the equivalence relation $(x, y, z) \sim (-x, -y, -z)$.

We note that the greatest distance between two points in this geometry is $\pi/2$, so Examples 1 and 3 cannot occur here. However, we do have the following, which is the projective plane's analog of Example 2.

EXAMPLE 4. *Let $\{A, B\}$ be the set of starting conditions for the construction game in the projective plane, and assume that $AB = \pi/2$. Then there is a unique point C such that $AB = AC = BC = \pi/2$, and the constructible set is exactly $\{A, B, C\}$.*

We call starting conditions of this form *trivial*, and any other starting conditions *non-trivial*. We can follow the proofs of the theorems of the previous section to show the following.

THEOREM 7. *Let $\{A, B\}$ be a set of non-trivial starting conditions in the projective plane. Then the set of constructible points is dense in the projective plane.*

Note, however, that the results we have found in spherical and projective geometries do not easily generalize to the hyperbolic case. Is it possible to show that different sets of starting conditions lead to different sets of constructible points? Is there a length, like $\pi/4$ in the sphere, that is "always" constructible (i.e., regardless of starting conditions) in hyperbolic geometry? Is the set of constructible points dense in the hyperbolic plane? These seem to be great questions for an undergraduate research project.

Kincaid notes at the end of [**5**] that his results have an analog in the hyperbolic plane. Although his hyperbolic cosine relation (the hyperbolic analog of the Pythagorean Theorem) is transcendental and thus not particularly useful given our current tools, it doesn't suffer from the same geometric limitations as its spherical analog. Thus we can imagine that someday our tools will become sophisticated enough to solve the hyperbolic constructibility question in general.

We would like to close by noting that both [**6**] and [**10**] give examples of constructions of the vertices of Platonic solids on the sphere. In particular, the constructions of the dodecahedron and icosahedron allow us, given standard starting conditions, to construct segments of length $2\pi/3$ and $2\pi/5$ in any geodesic $\overleftrightarrow{PQ}$ in the sphere (by placing a vertex of the construction at a pole of $\overleftrightarrow{PQ}$ and then joining that vertex to the 3 or 5 closest vertices by lines that divide $\overleftrightarrow{PQ}$ into 3 or 5 congruent segments). This means that we know how to construct segments of length $m\pi/n$, where m is an integer, $0 < m < n$, and $n = 2^{k_0} 3^{k_1} 5^{k_2}$ with $k_0 \in \{0, 1, 2, \ldots\}$ and $k_i \in \{0, 1\}$ for $i > 0$. This should hearken back to Equation (1). Hence we conjecture:

CONJECTURE. *Given standard starting conditions, segments of length*

$$\frac{2\pi}{17}, \quad \frac{2\pi}{257}, \quad \frac{2\pi}{65537}$$

are constructible in spherical and projective geometries.

REFERENCES

1. F. W. Bolyai, *Tentamen,* 1832. J. Bolyai's work on non-Euclidean geometry appears as an appendix.
2. Euclid, *The Thirteen Books of the Elements,* translated by Sir T. L. Heath, Dover, New York, 1956.
3. C. F. Gauss, *Disquisitiones Arithmeticae,* translated by A. A. Clarke, Yale University Press, New Haven, CT, 1965.
4. D. C. Kay, *College Geometry: A Discovery Approach*, Addison Wesley Longman, Boston, 2001.
5. J. M. Kincaid, Constructible numbers in double elliptic geometry, Master's thesis, Emporia State University, 1985.
6. I. Lénárt, *Non-Euclidean Adventures on the Lénárt Sphere,* Key Curriculum Press, Emeryville, CA, 1996.
7. N. I. Lobachevsky, *Geometrical Investigations on the Theory of Parallel Lines,* translated by G. N. Halstead, 1891. Reprinted in Bonola, *NonEuclidean Geometry*, 1912; Dover reprint, 1955.
8. B. Riemann, On the hypotheses which lie at the bases of geometry, inaugural lecture transcribed by D. R. Wilkins, translated by W. K. Clifford, *Nature* **8** (183 & 184) (1998) 14–17, 36, 37. http://dx.doi.org/10.1038/008014a0
9. J. J. Rotman, *A First Course in Abstract Algebra,* Prentice-Hall, Upper Saddle River, NJ, 1996.
10. R. Sarhangi, An introduction to medieval spherical geometry for artists and artisans, proceedings of the Bridges Conference, London, 2006, 551–560.
11. P. Wantzel, Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas, *Journal de Mathématiques Pures et Appliquées* **1**(2) (1837) 366–372.

Summary We examine straightedge and compass constructions in spherical geometry. We show via examples that the starting conditions affect the set of constructible points. Although current tools do not allow for a complete solution, we take a tour through group theory and real analysis to show that, in general, the set of constructible points is dense on the sphere. However, we conclude with more questions than answers.

DANIEL J. HEATH received his Ph.D. in mathematics from the University of California at Davis under the direction of Abigail Thompson. He is an associate professor and the chair of mathematics at Pacific Lutheran University. His mathematical interests range from algebra to mathematical origami to topology, and his non-mathematical interests range from bicycling to jazz to mushroom hunting.

ACROSS

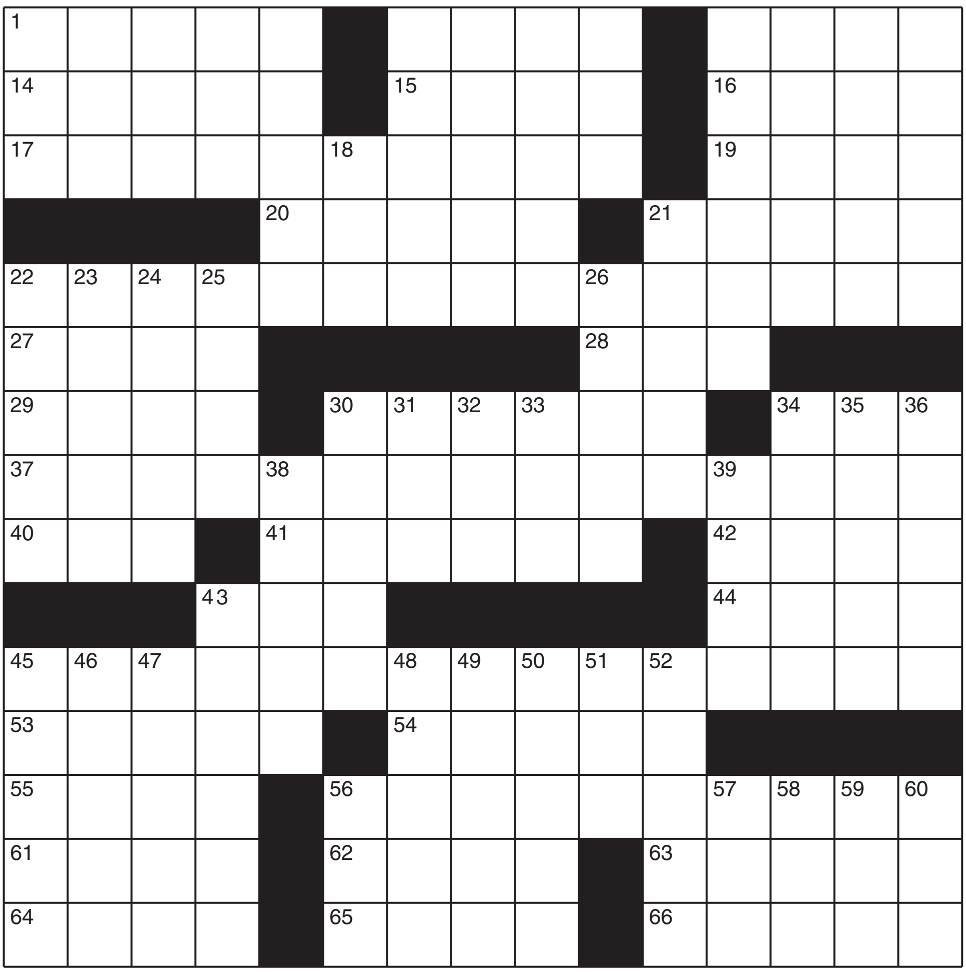
1. Like some avengers
6. Like show horses' feet
10. Marines ranked just above Pvts.
14. Pain reliever brand
15. Source of protein for vegetarians
16. The E in QED
***17.** 2006 (Australian)
19. Constellation whose principal star is Vega
20. Signal used to find shipwrecks
21. Type of plane transformation
***22.** 1950 (French)
27. Gawk at
28. Align the crosshairs
29. Willy of "Free Willy"
30. "Skinny Legs ______": 1990 Tom Robbins novel
34. Common ending of chemical sugars
37. What the answers to all of the starred clues are
40. Birth mo. of Euler
41. Solvent formed when a hydroxyl group is protonated
42. Style of jazz singing
43. Auguste Chevalier, to Evariste Galois
44. Colt's mother
***45.** 1954 (French)
53. Ends of prayers
54. Widely used text: "Abstract Algebra" by Dummit & ______
55. Writer/director Ephron
***56.** 2010 (Vietnamese)
61. Gas or elec., e.g.
62. "Alas!"
63. Computer operators
64. Gets
65. Snake's sound
66. Sri ______

DOWN

1. Schroedinger's pet in a box
2. It comes in Pale and Brown varieties
3. Each
4. WALL-E's love
5. Adjective that describes the set $\mathbb{Q}$ vis-a-vis the set $\mathbb{R}$, e.g.
6. Surgical tube
7. "It's ______ heck in here!"
8. Saint Joan ______
9. Twosome
10. Denzel Washington movie: "The Taking of ______ 123"
11. Appliance with an oil vat and basket
12. Unit of mass and gold purity
13. Premium cable channel
18. Not pro
21. Term for bad, watered down beer
22. Bath sponge
23. "Get ______!"
24. Stomach malady
25. Like some numbers
26. Islamic equivalent of kosher
30. Computer scheme for character-encoding
31. Weekly music magazine in the UK
32. This can be used to identify whether a system of eqns. has a soln.
33. Mathematician and pioneering computer scientist Lovelace, credited with writing the first ever computer program
34. Algebraic geometer Zariski, 1981 Wolf Prize winner
35. Drummer with Lennon and McCartney
36. Cosmetician Lauder
38. Reduces the amplitude of an oscillating system
39. "Woe ______!"
43. ______ of Mathematics: bimonthly journal published in Princeton
45. Two-faced Roman god who is an eponym for a month
46. Express dismay or elation, for instance
47. Eagle's home
48. Run of letters towards the beginning of the alphabet
49. You'll find infinitely many of these in the Hilbert Hotel
50. Judges' wear
51. Greek letter used to denote the Dirichlet function that is the alternating version of the series expression for the Riemann zeta function
52. Asian capital, and the location of the 2014 ceremony to recognize more 37-Across
56. Slangy "no"
57. Jefferson Davis was its Pres.
58. Rooster's partner
59. Noah built one in the Bible
60. Home country of eleven 37-Across, to date (abbr.)

Award Winners

BRENDAN W. SULLIVAN
Emmanuel College
Boston, MA 02115
sullivanb@emmanuel.edu



Clues are at left, on page 360. The solution is on page 402.

Extra copies of the puzzle, in both .pdf and .puz (AcrossLite) formats, can be found at the MAGAZINE's website, or (temporarily) at http://www.mathematicsmagazine.org.

Math. Mag. **87** (2014) 360–361,402. doi:10.4169/math.mag.87.5.360. © Mathematical Association of America

Mathematics, Models, and Magz, Part II: Investigating Patterns in Pascal's Simplex

VICTOR GARCIA
Santa Clara University
Santa Clara, CA 95053-0290
wizgarcia@gmail.com

JEAN PEDERSEN
Santa Clara University
Santa Clara, CA 95053-0290
jpedersen@scu.edu

Illustrations by Sylvie Donmoyer

Dedicated to the memory of Peter Hilton

This paper began when one of the authors, JP, gave a colloquium talk at Santa Clara University on the content of "Mathematics, Models, and Magz, Part I" describing results later reported in this MAGAZINE [**5**]. She described the "Star of David" theorems, which involve two-dimensional arrays of binomial coefficients, and their three-dimensional analogues involving trinomial coefficients. The other author, VG, then a sophomore at Santa Clara, was in the audience and became interested in seeing what would happen in four dimensions. The authors applied for, and received, a grant from Santa Clara University's Student Research Initiative Program that allowed them to spend the summer of 2009 studying extensions of the results in two and three dimensions.

To get a handle on what results to expect in four dimensions we again relied on a large set of magnetic toys called Magz. We were continually asking ourselves: What results in dimension four would naturally extend the two- and three-dimensional results? In the lower dimensions the Magz gave us a very clear picture of what was happening. It was only a shortage of time and lack of four-dimensional Magz that held us back, and we realized that we could just as easily have asked for analogous theorems in higher dimensions. We will pose some possible questions for the reader to explore at the end of this article.

Assumptions and definitions

We use the notations for binomial and trinomial coefficients from [**4**] and [**5**], and extend them to four dimensions. These are multinomial, or "m-nomial" coefficients, in the cases of $m = 2, 3, 4$.

If n, r, s, t, u are all non-negative integers, we define

$$\begin{pmatrix} & n & \\ r & & s \end{pmatrix} = \frac{n!}{r!\,s!} \qquad \text{when } r + s = n;$$

$$\begin{pmatrix} & n & \\ r & s & t \end{pmatrix} = \frac{n!}{r!\,s!\,t!} \qquad \text{when } r + s + t = n;$$

$$\begin{pmatrix} & & n & \\ r & s & t & u \end{pmatrix} = \frac{n!}{r!\,s!\,t!\,u!} \qquad \text{when } r + s + t + u = n.$$

Math. Mag. **87** (2014) 362–376. doi:10.4169/math.mag.87.5.362. © Mathematical Association of America

The binomial coefficient $\binom{n}{r\ s}$ is usually written as $\binom{n}{r}$ or $\binom{n}{s}$. We use the expanded notation to emphasize the analogy to higher dimensions.

We will, as before, sometimes refer to $\binom{n}{r\ s}$, $\binom{n}{r\ s\ t}$, and $\binom{n}{r\ s\ t\ u}$ as the "addresses" of the binomial, trinomial, or tetranomial coefficients, respectively. That is, we will use n, r, s, and so on as a kind of coordinates for points in two-, three-, or four-dimensional space. As was illustrated in Figure 3 of [**5**], the binomial addresses naturally arrange themselves in the triangular array known as the Pascal Triangle, while the trinomial addresses naturally arrange themselves in a tetrahedral array that we called the Pascal Tetrahedron. Because the tetranomial coefficients exist in four dimensions, they are harder to visualize. We introduce the idea of a *simplex* (as in [**1**]) in order to be able to draw figures of them as two-dimensional projections. The four-dimensional regular simplex may be thought of as the natural analogue of the triangle and tetrahedron, as shown in FIGURE 1.

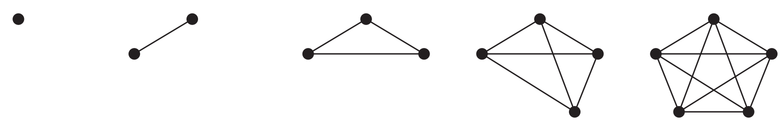


Figure 1 Simplices in dimensions 0 through 4. The 3-simplex is easily visualized as a tetrahedron, but visualizing the 4-simplex is more challenging.

In each case we are concerned with a discrete set of points. We will be particularly interested in those coefficients that lie on the boundary of certain special geometric configurations that we will refer to as *homologues*, as defined in [**5**].

Computing distances If we think of the shortest distance between two binomial (or trinomial, or tetranomial) coefficients in Pascal's triangle as being a *Pascal unit*, then we may compute the Pascal distance, d, between two tetranomial coefficients $\binom{n_1}{r_1\ s_1\ t_1\ u_1}$ and $\binom{n_2}{r_2\ s_2\ t_2\ u_2}$ using the formula

$$d = \frac{1}{\sqrt{2}}\sqrt{(n_2-n_1)^2+(r_2-r_1)^2+(s_2-s_1)^2+(t_2-t_1)^2+(u_2-u_1)^2}.$$

For trinomial coefficients we omit the u component, and for binomial coefficients we also omit the t component. This means, for example, that the Pascal distance between $\binom{3}{1\ 2}$ and $\binom{3}{2\ 1}$ would be

$$\frac{1}{\sqrt{2}}\sqrt{(3-3)^2+(2-1)^2+(1-2)^2} = \frac{\sqrt{0+1+1}}{\sqrt{2}} = 1,$$

just as we would wish. In the rest of this paper, when we refer to "distance" we always mean the Pascal distance.

The Star of David Theorem and the Hoggatt-Alexanderson Theorem

Recall from [**4**] or [**5**] that the original Star of David Theorem involved the six nearest neighbors to $\bigstar = \binom{n}{r\ s}$, as illustrated here in FIGURE 2 (and also in Figure 5 in [**5**]). The product of all six coefficients nearest to $\bigstar$ is a perfect square, say S^2, where

$$\left\{\frac{(n-1)!\,n!\,(n+1)!}{(r-1)!\,(s-1)!\,r!\,s!\,(r+1)!\,(s+1)!}\right\}^2 = S^2.$$

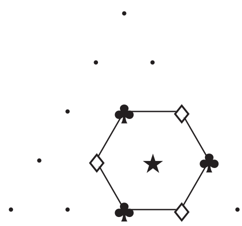


Figure 2 Original Star of David theorem

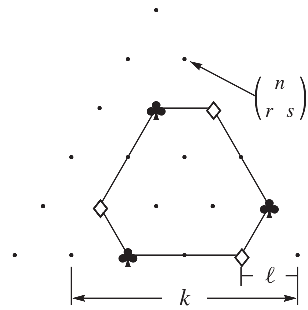


Figure 3 Generalizing to a semi-regular hexagon

Furthermore,

$$\prod\clubsuit=\binom{n-1}{r-1\ \ s}\binom{n}{r+1\ \ s-1}\binom{n+1}{r\ \ s+1}$$

and

$$\prod\diamondsuit=\binom{n-1}{r\ \ s-1}\binom{n}{r-1\ \ s+1}\binom{n+1}{r+1\ \ s},$$

from which we can immediately conclude that

$$\prod\clubsuit=\prod\diamondsuit=\frac{(n-1)!\,n!\,(n+1)!}{(r-1)!\,(s-1)!\,r!\,s!\,(r+1)!\,(s+1)!}=S.$$

We leave the proof that $\prod\clubsuit=\prod\diamondsuit$ in FIGURE 3 (where $k=4$ and $\ell=1$) as an exercise for the reader. A slick geometric proof of the more general case appears in [**5**].

Moving on to multinomial coefficients, we take note of a paper by Hoggatt and Alexanderson [**6**]. That paper begins with the following abstract:

> The multinomial coefficients "surrounding" a given multinomial coefficient in a generalized Pascal pyramid are partitioned into subsets such that the product of the coefficients in each subset is a constant N and such that the product of the coefficients "surrounding" a given m-nomial coefficient is N^m....

This result is an m-dimensional analogue of the Star of David Theorem, and we call it the *Hoggatt-Alexanderson Theorem* (HAT).

We recall the form taken by the HAT in trinomial case $m=3$. The polyhedron that sits inside the Pascal tetrahedron having the 12 nearest neighbors to $\binom{n}{r\ s\ t}$ is a cuboctahedron, and the product of the 12 coefficients is

$$\left\{\frac{(n-1)!\,[n!]^2\,(n+1)!}{(r-1)!\,(s-1)!\,(t-1)!\,[r!]^2\,[s!]^2\,[t!]^2\,(r+1)!\,(s+1)!\,(t+1)!}\right\}^3=T^3.$$

The 12 points can be partitioned into three sets of four points each, the homologues, which are congruent to each other as geometric objects, and which have the property described in the HAT. It turned out that the three homologues are not in the shape of tetrahedra as we had anticipated. With the model in hand, we see that the four vertices connecting either the $\clubsuit$s, $\diamondsuit$s, or $\heartsuit$s form a *square*! A proof can be provided either by ordinary vector analysis, or, more simply, by computing the relevant distances. For example, using our distance formula for the ♣ homologue, shown in Figure 7 of [**5**], we see that the four line segments that would be on the boundary of the model have a length of $\sqrt{2}$, while the two line segments going through the interior of the model (joining opposite vertices of the quadrilateral) have a length of 2. Thus, the shape of this homologue *must* be a square, which, of course, lives not only in three dimensions, but also in a *plane*.

Known two- and three-dimensional analogues

In [**5**], the authors presented a two-dimensional generalization of the Star of David Theorem, which involved six binomial coefficients arranged in a hexagon. As illustrated in FIGURE 3, the hexagon can be seen as a vertex-truncated triangle. Start with an equilateral triangle with side k, and remove from each corner a smaller triangle of side ℓ (with $0 < \ell < k/2$). The six vertices at the corners can be partitioned into two homologues. Each homologue is itself a triangle, and the coefficients satisfy

$$\prod \clubsuit = \prod \diamond = \frac{(n+k)!\,(n+\ell)!\,(n+k-\ell)!}{r!\,s!\,(r+\ell)!\,(s+\ell)!\,(r+k-\ell)!\,(s+k-\ell)!} = J,$$

with the product of all six coefficients being J^2.

The authors found a similar generalization for trinomial coefficients by beginning with a regular tetrahedron within the Pascal tetrahedron from which they truncated a smaller tetrahedron from each vertex, obtaining a vertex-truncated tetrahedron, as shown in Figure 11 of [**5**]. The top point of the tetrahedron is labeled $\binom{n}{r\ s\ t}$. We describe it more fully below.

It turned out that, although this generalization produced Generalized Star of David Theorems—in the sense that the 12 vertices could be partitioned into 3 homologues, each of which was a semi-regular tetrahedron—it didn't explain the HAT. So at that point the authors of [**5**] considered what happens when you begin with a regular tetrahedron of edge length k, and truncate ℓ units from each edge, with $0 < \ell < k/2$. (It turned out that allowing $\ell \geq k/3$ led to some non-typical degenerate cases; so in this paper, we will consider only the case of $\ell < k/3$.) An example of this edge-truncated tetrahedron is shown in Figure 13 of [**5**], with $k = 5$ and $\ell = 1$.

When the edge truncation results in genuine rectangles (beneath the original edges of the original tetrahedron) the partition results in homologues that form interpenetrating *tetrahedra*. However, when the dimensions of the truncation with relation to the original tetrahedron are such that the rectangles become squares, then the partitioned sets become the vertices of *squares*; and, at last, we see that the HAT may be regarded as a simple case of the edge truncation.

Homologues in the four-dimensional simplex

We now turn to the study of some *tetranomial* analogues. We will examine two types. The first type is a vertex-truncated simplex, which will naturally be the easier case because we have the binomial and trinomial coefficient results from which to glean possible analogues. The second type, an edge-truncated simplex, will be more difficult because truncating an edge from a triangle in the display of binomial coefficients doesn't produce the desired theorems. So we only have the case of the edge-truncated tetrahedron from which to guess the result when we truncate the edge of a simplex. This is harder! Nevertheless, we were successful at getting some results in this case, and the results are, at least to us, quite surprising and more than a little vexing.

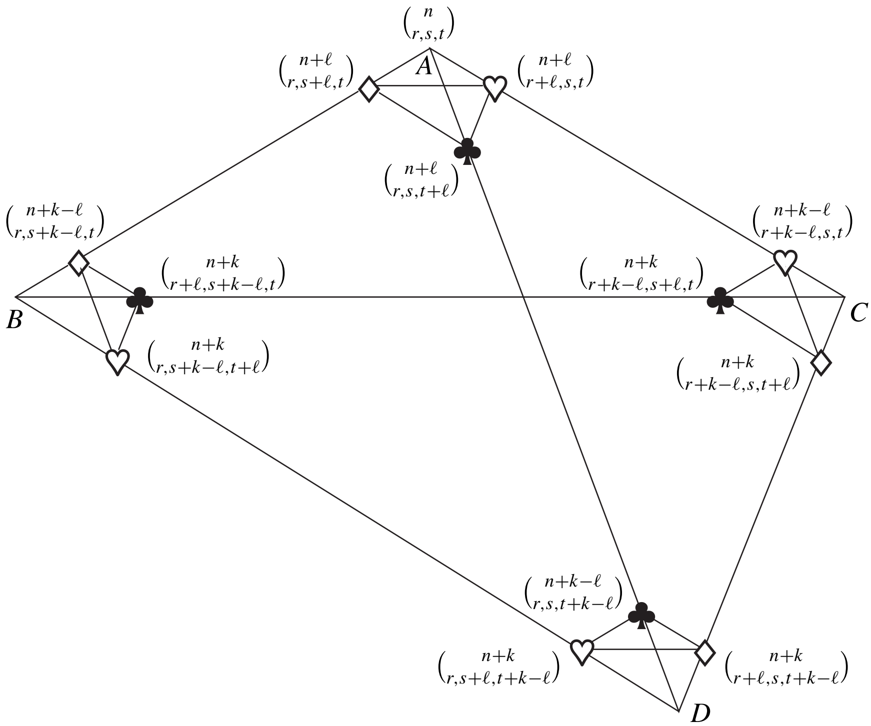


Figure 4 The vertex-truncated tetrahedron: $\prod\clubsuit = \prod\diamondsuit = \prod\heartsuit$

The vertex-truncated simplex

We now look more closely at the three-dimensional case of the vertex-truncated tetrahedron. FIGURE 4 represents a perspective view of the vertex-truncated tetrahedron shown in Figure 11 of [**5**], where the edge length of the original tetrahedron is k and the edge length of the tetrahedron cut off from each vertex is ℓ $(0 < \ell < k/2)$. The vertices in FIGURE 4 are suggestively labeled not only with their appropriate trinomial coefficients, but also with $\clubsuit$s, $\diamondsuit$s, and $\heartsuit$s, identifying the homologues. We have

$$\prod\clubsuit = \prod\diamondsuit = \prod\heartsuit$$
$$= \frac{[(n+k)!]^2(n+\ell)!\,(n+k-\ell)!}{[r!]^2[s!]^2[t!]^2(r+\ell)!\,(s+\ell)!\,(t+\ell)!\,(r+k-\ell)!\,(s+k-\ell)!\,(t+k-\ell)!}$$
$$= K,$$

so that the product of all 12 coefficients is indeed a perfect cube, K^3.

Armed with this information we made what for us was a *mighty guess*; namely, that if we were to truncate from a regular simplex, with $\binom{n}{r\ s\ t\ u}$ at its top vertex, we would get four homologues for which, in each case, the product of the five relevant tetranomials would be

$$\frac{[(n+k)!]^3(n+\ell)!\,(n+k-\ell)!}{\left(\begin{array}{c}[r!]^3[s!]^3[t!]^3[u!]^3(r+\ell)!\,(s+\ell)!\,(t+\ell)!\,(u+\ell)! \\ (r+k-\ell)!\,(s+k-\ell)!\,(t+k-\ell)!\,(u+k-\ell)!\end{array}\right)} = L$$

so that the product of all 20 tetranomials would be L^4.

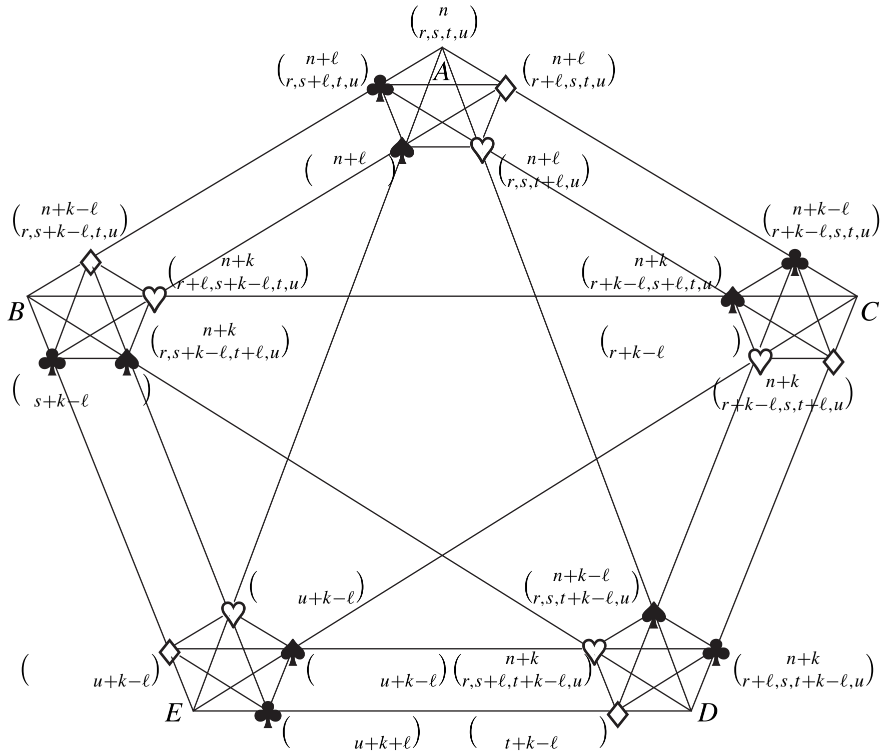


Figure 5 Aspiring to be a vertex-truncated simplex

Now, because FIGURE 4 depicts how the tetrahedron sits inside of a regular simplex of edge length k, the task remains, after truncating a regular tetrahedron of edge length ℓ from each vertex, to complete the simplex shown in FIGURE 5. What clues do we have to help us?

First, notice that in the simplex,

- near vertex A all the values in the n position are $n + \ell$;
- near vertex B all the values in the s position are $s + k - \ell$;
- near vertex C all the values in the r position are $r + k - \ell$; and
- near vertex D all the values in the t position are $t + k - \ell$.

Second, carrying out our investigations thus far made us strongly aware of the fact that you

Ignore symmetry at your Peril!

So, adhering to this belief we made another guess; namely, that at vertex E all the values in the u position would be $u + k - \ell$.

Third, we knew that the values in the u position at vertices A, B, C, and D would have to be u (otherwise the homologues wouldn't be correct).

Assuming our guesses were correct, all we had to do was complete the diagram shown in FIGURE 5 so that $\prod \clubsuit = \prod \diamondsuit = \prod \heartsuit = \prod \spadesuit = L$. You might like to try it yourself before going on. (At this point it can be viewed merely as a puzzle.)

There are now several things to investigate: What do the homologues of the vertex-truncated simplex look like? Do those homologues live in four dimensions, or could

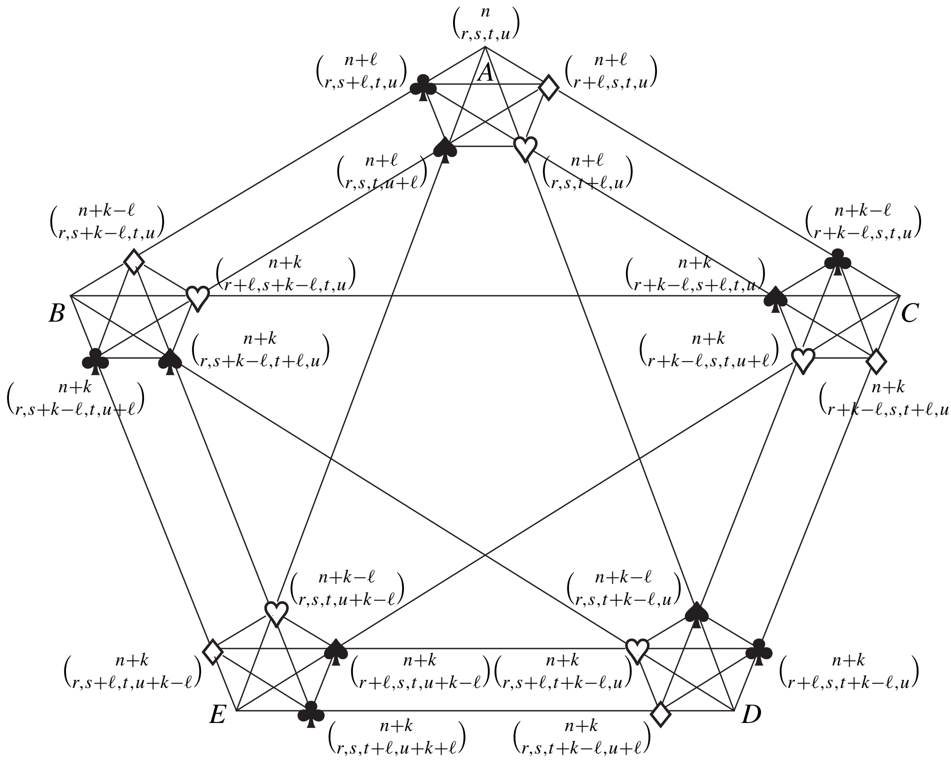


Figure 6 The vertex-truncated simplex

they possibly also live in three dimensions (which would be analogous to the square of the HAT for the 12 nearest neighbors of $\binom{n}{r\ s\ t}$)? And, *if* the homologues live in three dimensions, then what are their shapes?

Before stating the results we give some interesting constants that will appear in the rest of this section. We like to list them together because we notice some nice patterns among them. The first three appear in the vertex-truncated simplex (with $0 < \ell < k/2$), while the last three will appear later in the edge-truncated simplex (with $0 < \ell < k/3$).

$$
\begin{aligned}
a &= \sqrt{k^2 - 2k\ell + 2\ell^2} \\
b &= \sqrt{k^2 - 3k\ell + 3\ell^2} \\
c &= \sqrt{k^2 - 4k\ell + 4\ell^2} \quad (= k - 2\ell) \\
d &= \sqrt{k^2 - 5k\ell + 7\ell^2} \\
e &= \sqrt{k^2 - 6k\ell + 10\ell^2}
\end{aligned}
$$

Using ordinary vector analysis, we find that the homologues of FIGURE 6, as seen in FIGURE 7, really do live in four dimensions. Since the symmetry group for the simplex, S_5, contains all possible permutations of the five vertices of the simplex, the homologue shown in FIGURE 7(a) is actually the same as that shown in FIGURE 7(b).

We should check to make sure there are five vertex-truncated tetrahedra in the vertex-truncated simplex of FIGURE 6. It turns out that, in FIGURE 6, for $x = r$, s, t, or u, there is a vertex of the original simplex where the four vertices of the tetrahe-

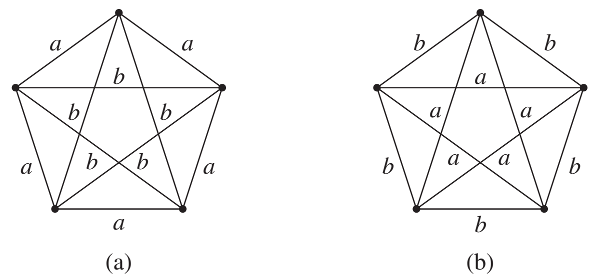


Figure 7 Homologues of the vertex-truncated simplex

dron near that vertex all contain $x + k - \ell$ in the x position. If you eliminate those four vertices and take the remaining 12 entries in which a single x appears, what remains is a vertex-truncated tetrahedron. In particular, if at vertex E, where $x = u + k - \ell$ for all four entries, we eliminate those four vertices (and let the remaining u's equal 0), we then obtain the vertex-truncated tetrahedron of FIGURE 4. Making $u = 0$ isn't necessary, since for any fixed u this is still a vertex-truncated tetrahedron in a different location of the Pascal pyramid. And the homologues remain the same, as can be seen by calculating the distances between constituent parts of each homologue.

The case where the values in the position n are fixed (at vertex A) appear to be quite different, But, in fact, if you eliminate the four vertices at vertex A, where the value in the n position is $n + \ell$, then the only remaining vertices that make sense to take are the 12 that have $n + k$ in the n position. At first, this may look dreadful because all of these seem to be tetranomial coefficients; but it can be shown by vector analysis that this configuration does, in fact, live in three dimensions. However, it is simpler to use our Pascal distance formula to show that all the edge lengths for each of the homologues are exactly the same as for the other four cases mentioned above.

In each of the five cases the homologues of the vertex-truncated tetrahedra are either the (a), (b), or (c) part of FIGURE 8 (which is not oriented, nor to scale). In each case, the homologue is an isosceles tetrahedron with each face having side lengths a, a, and c.

We should perhaps mention that the symbolic representation of the homologues looks very different in the last case where all the coefficients have $n + k$ in the n position; but in view of the fact that the n position plays a special role in the multinomial coefficients, this isn't too surprising. What is very pleasing to the authors is how the mathematics connected with the geometry works out so well even though the symbolic representation of the homologues, in this case, looks so utterly different from the other four cases.

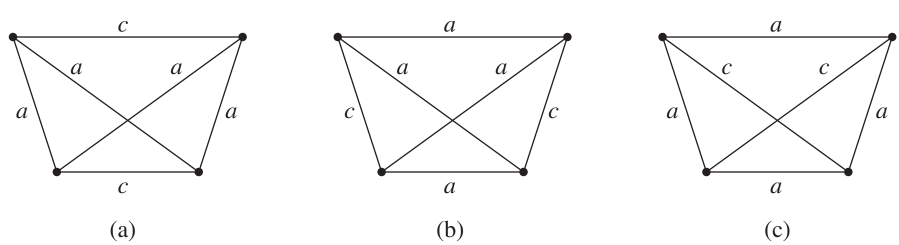


Figure 8 Homologues of the vertex-truncated tetrahedra that live in the vertex-truncated simplex

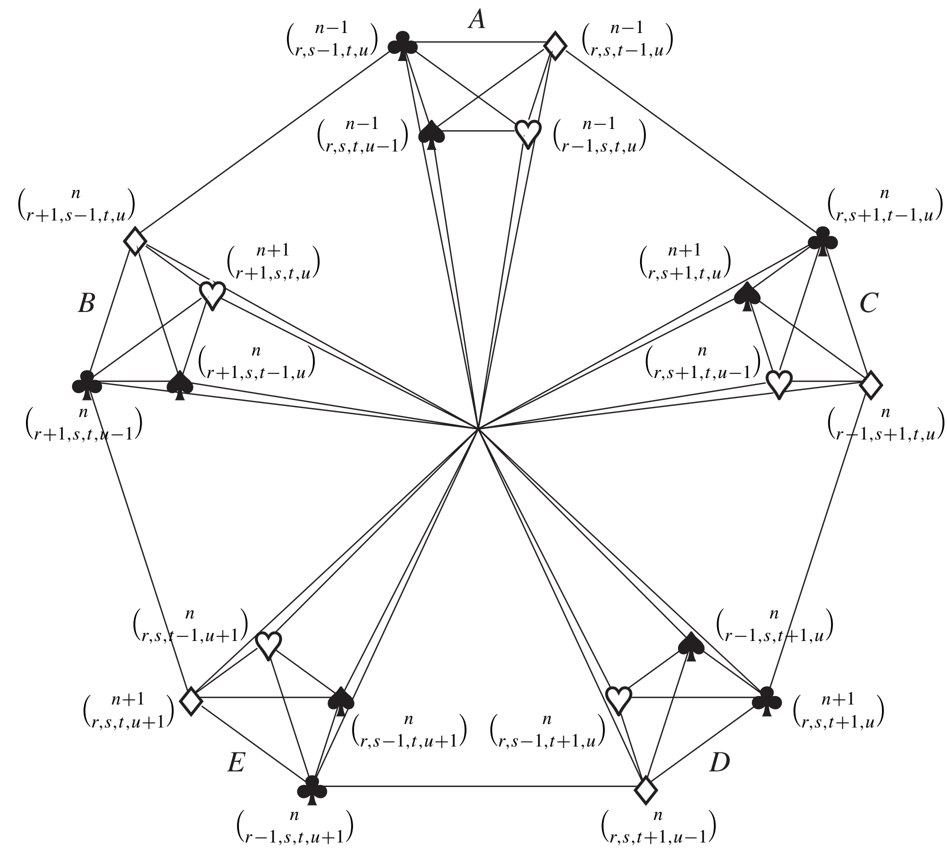


Figure 9 The 20 nearest neighbors to $\binom{n}{r\,s\,t\,u}$

The edge-truncated simplex

Since the 12 nearest neighbors to $\binom{n}{r\ s\ t}$ produce a cuboctahedron, which is a special case of the edge-truncated tetrahedron, we are tempted to look first at the HAT for tetranomial coefficients. FIGURE 9 shows the 20 nearest neighbors to $\binom{n}{r\ s\ t\ u}$, which is assumed to be at the center of the diagram.

We used the distance formula to ascertain the edge lengths, in Pascal units, for the constituent parts of the four homologues. The diagrams in FIGURE 10 show the two presentations of the homologues with the edge lengths indicated. In fact, the two parts of FIGURE 10 are congruent, since they have the same connectivity and the group structure for the original simplex is S_5.

What is striking about this homologue is that there exists a three-dimensional model with these dimensions. It is, in fact, the triangular dipyramid shown in FIGURE 11. So here we have homologues in the four-dimensional simplex that live in three dimensions (just as the homologues for the HAT in three dimension actually lived in two dimensions).

However, we would like to get more general results, so we seek an edge-truncated simplex with one of the vertices of the original simplex labeled $\binom{n}{r\ s\ t\ u}$, as we did for the vertex-truncated case.

We were unable to draw a satisfactory diagram of the actual edge-truncated tetrahedron as it sits in the simplex. So we tried a different approach, remembering to take into account *symmetry*. We first looked at the edge-truncated tetrahedron and noticed that

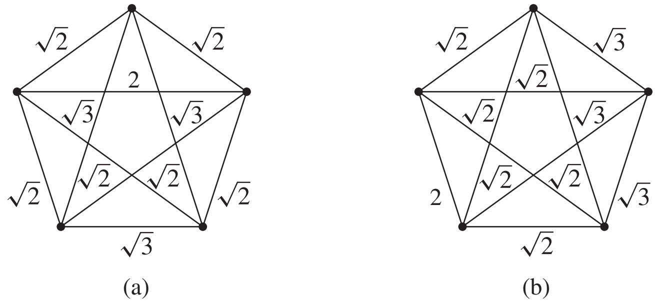


Figure 10 Homologues with edge lengths shown (the numbers inside the pentagon indicate the length of the diagonals)

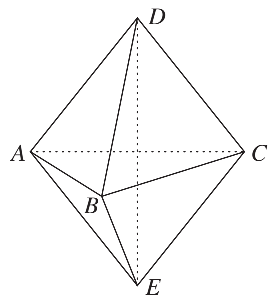


Figure 11 The triangular dipyramid. The triangle *ABC* is equilateral with side length $\sqrt{3}$, *DE* is an interior diagonal with length 2, and the remaining edges have length $\sqrt{2}$.

the four triangles, symmetrically placed on a sphere with their centers lying beneath the vertices of a regular inscribed tetrahedron, completely determined the edge-truncated tetrahedron. We then studied the topographical diagram of FIGURE 12, which shows how it looks as viewed looking down from the top vertex $\binom{n}{r\ s\ t}$. Finally, we inserted the addresses of the topographical diagram into the simplex at vertices A, B, C, and D so that they had the right connectivity and yielded the proper homologues, obtaining FIGURE 13.

Reasoning that the analogous situation for the simplex should involve four tetrahedra symmetrically placed on the boundary of a hypersphere, we tried to complete the configuration for a simplex of edge length k where we truncate from each edge ℓ units (k and ℓ being positive integers, with $0 < \ell < k/3$). The information we had to work from was sketchy. We know that when we take a tetrahedron of edge-length k and truncate from each edge ℓ units, the homologue for the resulting edge-truncated tetrahedron, represented in trinomial coefficients $\binom{n}{r\ s\ t}$, is

$$\frac{(n+2\ell)!\,[(n+k+\ell)!]^2\,(n+k)!}{r!\,s!\,t!\,[(r+\ell)!\,(s+\ell)!\,(t+\ell)!]^2\,(r+k-2\ell)!\,(s+k-2\ell)!\,(t+k-2\ell)!}$$

and that

- near vertex A all the values in the n position are $n+2\ell$;
- near vertex B all the values in the s position are $s+k-2\ell$;
- near vertex C all the values in the r position are $r+k-2\ell$; and
- near vertex D all the values in the t position are $t+k-2\ell$.

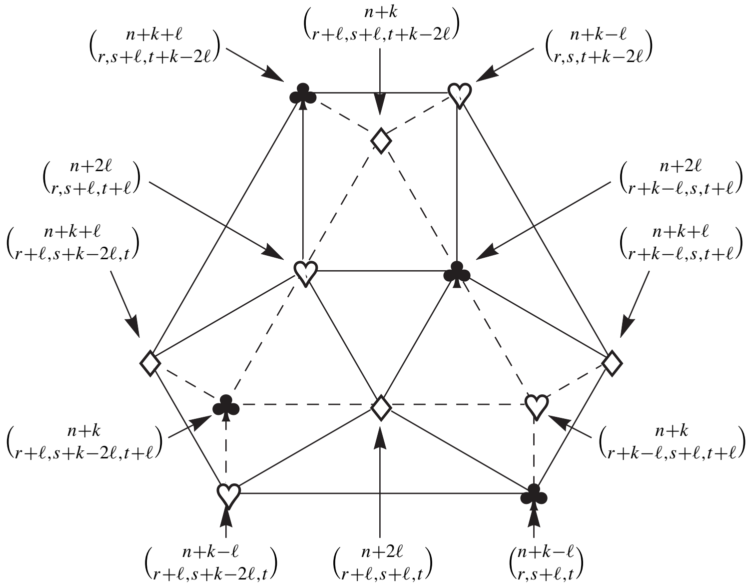


Figure 12 A topographical diagram of the edge-truncated tetrahedron of edge length k after truncating ℓ units along each edge

We also know that in the u position for the existing coefficients in FIGURE 13 there should be just u. By exhaustive trial and error, and using what information we had, including that the product for each of the sets of ♣s, ♢s, ♡s, and ♠s had to be the same, we found a representation of the edge-truncated simplex as shown in FIGURE 14.

However, the coefficients did not *in any way* look like what we had anticipated. The biggest surprise was the presence of $u + k - 3\ell$ at vertex E. We were filled with

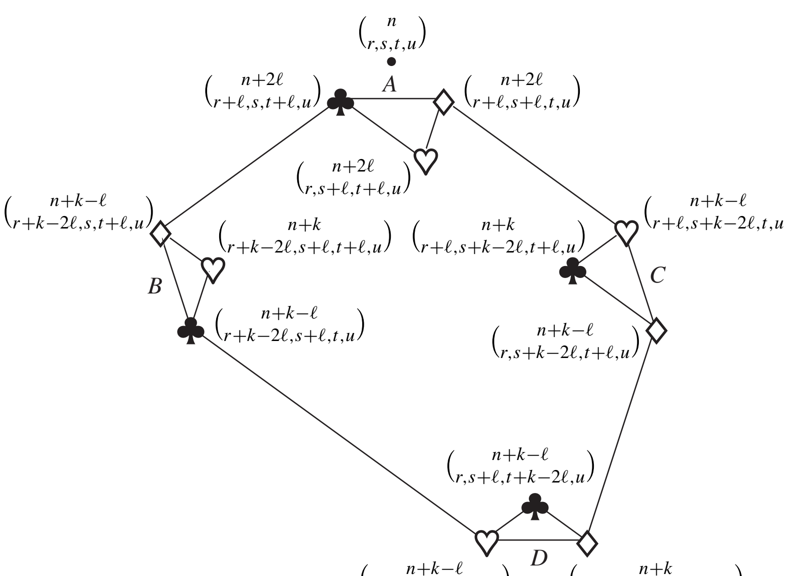


Figure 13 Edge-truncated tetrahedron, where $\prod ♣ = \prod \diamondsuit = \prod \heartsuit$

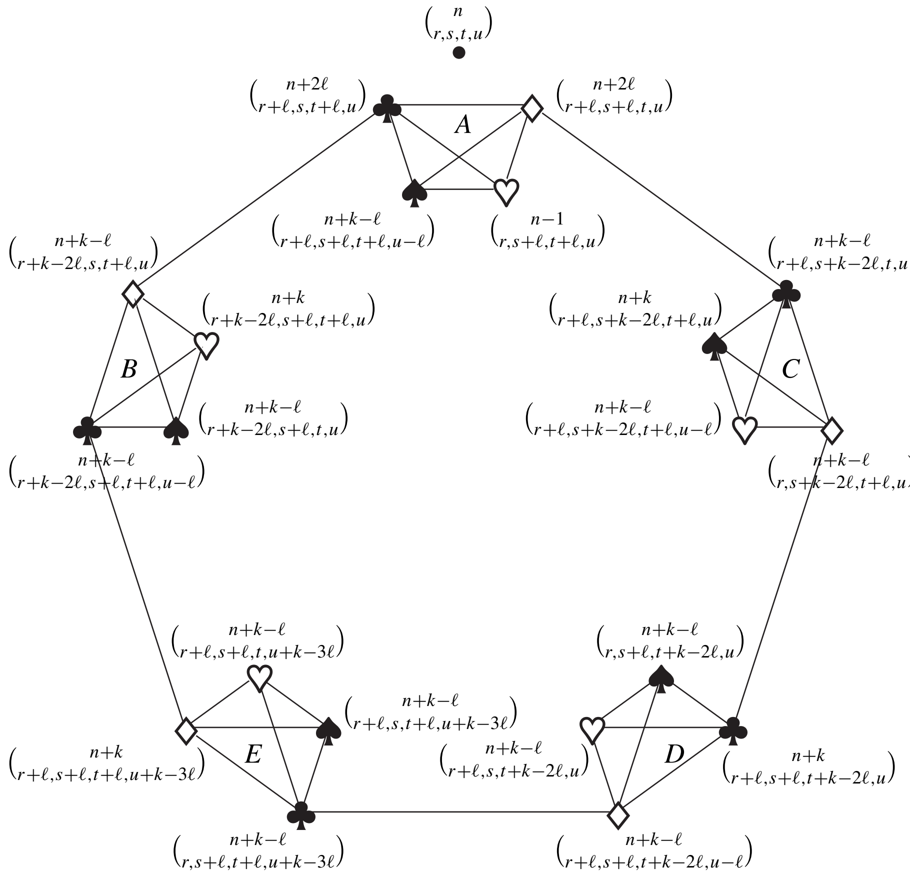


Figure 14 The edge-truncated simplex: $\prod\clubsuit = \prod\diamondsuit = \prod\heartsuit = \prod\spadesuit$

trepidation that we had made some error in our calculations—or, even worse, that this just could not be done. Then, to our surprise and delight, we saw that

$$\prod\clubsuit = \prod\diamondsuit = \prod\heartsuit = \prod\spadesuit$$
$$= \frac{(n+2\ell)!\,[(n+k+\ell)!]^3\,(n+k)!}{\begin{pmatrix} r!\,s!\,t!\,(u-\ell)!\,[(r+\ell)!\,(s+\ell)!\,(t+\ell)!\,u!]^3 \\ (r+k-2\ell)!\,(s+k-2\ell)!\,(t+k-2\ell)!\,(u+k-3\ell)! \end{pmatrix}}.$$

Even more satisfying was when we checked the distances between the coefficients in each of the homologues, as suggested by the labeling of FIGURE 14. We found that they were indeed all the same geometrically, though presented in two different ways, as shown in FIGURE 15. We were especially encouraged by this; and since these geometric objects were consistent with those shown in FIGURE 10, we had to accept that, however strange the coefficients in the diagram may look to us, the mathematics was trying to tell us that this is indeed a representation of an edge-truncated simplex.

By using ordinary vector analysis (assuming $\ell = 1$, since it is the ratio of $k:\ell$ that determines the final result), we can determine that the homologues for the edge-truncated simplex live in four dimensions, except for the cases when $k = 2$ or $k = 4$. If $k = 2$, then the relevant distances for the homologues of the simplex, shown in FIGURE 15, are $c = 0$, $d = 1$, and $e = \sqrt{2}$. So, in this case, the simplex itself degenerates to a tetrahedron in three dimensions with four vertices and six edges.

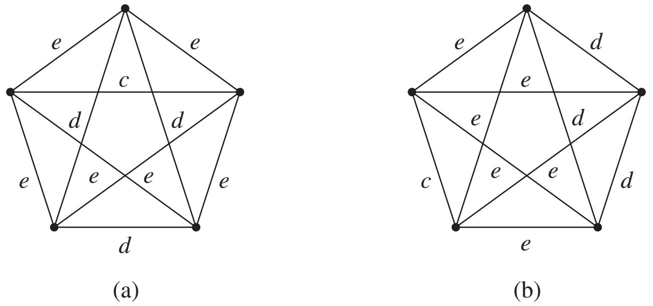


Figure 15 Homologues of the edge-truncated simplex, with $0 < \ell < k/3$, $c = \sqrt{k^2 - 4k\ell + 4\ell^2}$ $(= k - 2\ell)$, $d = \sqrt{k^2 - 5k\ell + 7\ell^2}$, $e = \sqrt{k^2 - 6k\ell + 10\ell^2}$

We wanted more confirmation that this is a valid representation, so we tried to identify all five edge-truncated tetrahedra that should be in this simplex. From FIGURE 14, it is possible to verify that the edge lengths for each of these edge-truncated tetrahedra are as shown in FIGURE 16.

If $k = 4$, then the homologue actually lives in three dimensions with $c = 2$, $d = \sqrt{3}$, and $e = \sqrt{2}$, so that, in this case, the homologue within the simplex lives in three dimensions, and it is the triangular dipyramid shown in FIGURE 11.

To see how we obtained this diagram, first observe in FIGURE 14 that

- near vertex A each tetranomial has the value of $n - 2\ell$ in the n position;
- near vertex B each tetranomial has the value of $r + k - 2\ell$ in the s position;
- near vertex C each tetranomial has the value of $s + k - 2\ell$ in the r position;
- near vertex D each tetranomial has the value of $t + k - 2\ell$ in the t position; and,
- near vertex E each tetranomial has the value of $u + k - 3\ell$ in the u position.

To obtain the diagrams in FIGURE 17 we picked one of the five vertices and eliminated the four vertices at a particular vertex. Next, we considered the remaining 12 vertices in the simplex for which one of the five positions in the tetranomial was filled with the same variable. For example, if we eliminated the four vertices near vertex A of FIGURE 14, of the 16 vertices remaining there were 12 that had a value of $n + k - \ell$ in the n position. We then shaded the four triangles with $n + k - \ell$ in the n position and focused on the explicit values for the tetranomials at the vertices of these four triangles in FIGURE 14. We then had to find the appropriate labeling for the homologues. This was done by calculating distances so that the edges of the homologues matched one of the diagrams in FIGURE 16. As you might expect, this turned out to be a fascinating

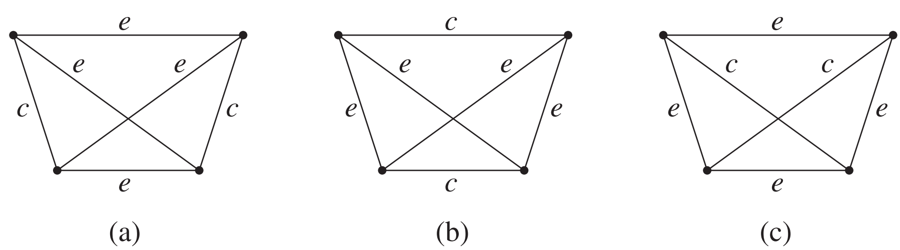


Figure 16 Homologues of the edge-truncated tetrahedra that live in the edge-truncated simplex

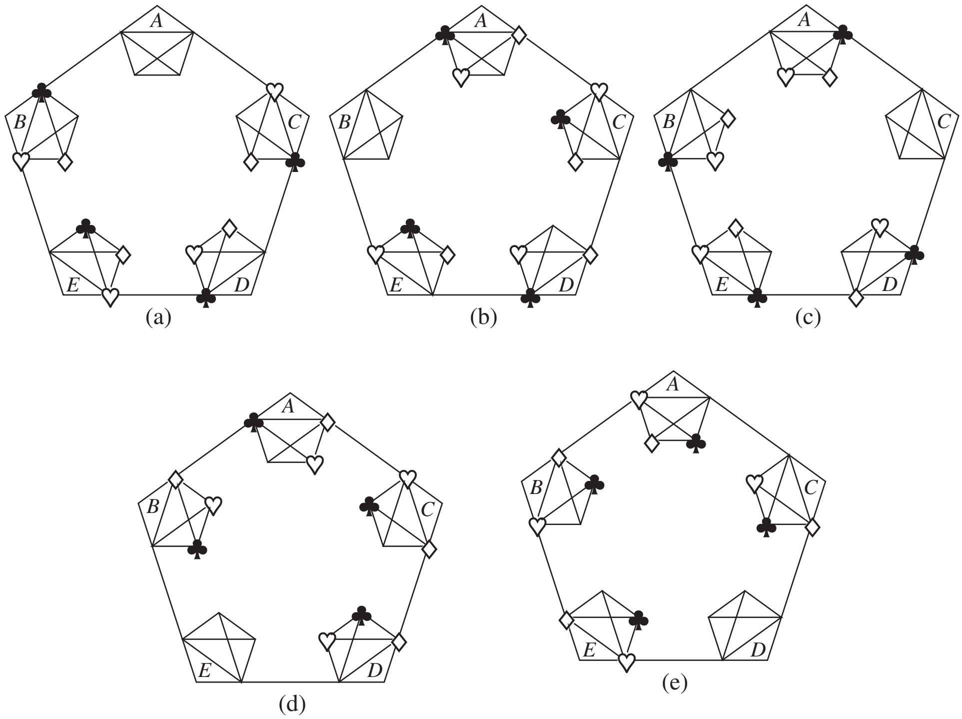


Figure 17 Highlighting edge-truncated tetrahedra within the edge-truncated simplex. Part (a), (b), (c), (d), and (e), is what results when you begin by eliminating the tetrahedron near vertex A, B, C, D, and E, respectively, in FIGURE 14.

puzzle. It was particularly satisfying to fill in the last set of homologues and find that the distances worked out just right.

So now we see the HAT and its analogue through the perspective of a larger universe. But it seems the more we learn about this, the more questions we have. Here are some of our questions.

Questions for further study

1. George Pólya (1887–1985) said, "Geometry is the science of correct reasoning on incorrect figures." We think this may be how we obtained the values for the coordinates in FIGURE 14. Is there a better, more symmetric-looking way to represent the diagram of the edge-truncated simplex?
2. Does there exist another, more symmetric labeling of the edge-truncated simplex?
3. Do the analogues of the vertex-truncated simplex extend to higher dimensions?
4. Do the analogues of the edge-truncated simplex extend to higher dimensions?
5. When considering dimensions greater than four, does the idea of a face truncation make sense, and what would be the analogues of these results for those types of truncations?
6. Will there always be homologues of the HAT in n dimensions that live in dimension $n - 1$?

Several other papers ([**2**] and [**7–12**]) might be of interest to readers who want more background on this topic.

Acknowledgment The authors would like to thank Richard Scott for his help with the vector analysis involved and Frank Farris for his discussions with us about polyhedra in higher dimensions. We are also grateful to Gerald Alexanderson and Geoffrey Shephard for reading an earlier version of this manuscript and offering valuable suggestions for its improvement. We thank the Student Research Initiative Program at Santa Clara University for funding this research during the summer 2009.

REFERENCES

1. H. S. M. Coxeter, *Regular Polytopes,* 2nd ed., Macmillan, New York, 1963.
2. H. W. Gould, Equal products of generalized binomial coefficients, *Fibonacci Quarterly* **9** (1971) 337–346. http://www.fq.math.ca/Scanned/9-4/gould.pdf
3. Peter Hilton, Derek Holton, and Jean Pedersen, *Mathematical Reflections—In a Room with Many Mirrors,* 2nd printing, Springer Verlag, New York, 1998, 211–216.
4. Peter Hilton, Derek Holton, and Jean Pedersen, *Mathematical Vistas—From a Room with Many Windows,* Springer Verlag, New York, 2002, 172–193. http://dx.doi.org/10.1007/978-1-4757-3681-6
5. Peter Hilton and Jean Pedersen, Mathematics, models, and Magz, part 1: Investigating patterns in Pascal's triangle and tetrahedron, *Math. Mag.* **85** (2012) 97–109. http://dx.doi.org/10.4169/math.mag.85.2.097
6. V. E. Hoggatt, Jr. and G. L. Alexanderson, A property of multinomial coefficients, *Fibonacci Quarterly* **9** (1971) 351–356, 420. http://www.fq.math.ca/Scanned/9-4/hoggatt1-a.pdf and http://www.fq.math.ca/Scanned/9-4/hoggatt1-b.pdf
7. V. E. Hoggatt and W. Hansell, The hidden hexagon squares, *Fibonacci Quarterly* **9** (1971) 120. http://www.fq.math.ca/Scanned/9-2/hoggatt1-a.pdf
8. Calvin T. Long, Arrays of binomial coefficients whose products are squares, *Fibonacci Quarterly* **11** (1973) 449–456. http://www.fq.math.ca/Scanned/11-5/long.pdf
9. Calvin T. Long and Verner E. Hoggatt, Jr., Sets of binomial coefficients with equal products, *Fibonacci Quarterly* **12** (1974) 71–79. http://www.fq.math.ca/Scanned/12-1/long.pdf
10. Calvin T. Long and Shiro Ando, Another generalization of Gould's Star of David theorem, *Fibonacci Quarterly* **30** (1992) 251–255. http://www.fq.math.ca/Scanned/30-3/long.pdf
11. John F. Putz, The Pascal polytope: An extension of Pascal's triangle to *N* dimensions, *College Math. J.* **17** (1986) 144–155. http://dx.doi.org/10.2307/2686833
12. Zalman Usiskin, Perfect square patterns in the Pascal triangle, *Math. Mag.* **46** (1973) 203–208. http://dx.doi.org/10.2307/2688302

Summary This paper is a sequel to *Mathematics, models, and Magz, Part 1: Investigating patterns in Pascal's triangle and tetrahedron.* In the current paper the authors use a set of magnetic toys to extend results of Part I about patterns concerning binomial and tetranomial coefficients. In Pascal's tetrahedron they study vertex truncations and edge truncations. Then they extend those ideas to obtain analogous results in Pascal's simplex.

JEAN PEDERSEN was the first woman to teach in the mathematics department at Brigham Young University, where she obtained a B.S. degree. She then attended the University of Utah, where she taught for 10 years after receiving a M.S. degree. After that she was the first woman to teach in the mathematics department at Santa Clara University. She is completing her 50th year of teaching at SCU, during which time she has, among her articles and books, co-authored papers with at least six of her students.

VICTOR GARCIA graduated from Santa Clara University with a B.S. in mathematics and went on to pursue a higher-education degree from Claremont Graduate University. Aside from combinatorics, he enjoys mathematical programming and video games as pastimes. As a future educator, his main goal is to spread a love of mathematics and to encourage students to see the mathematics in everything they do.

NOTES

Cardano, *Casus Irreducibilis*, and Finite Fields

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

Undoubtedly, the mathematical gem of the 16th century is Gerolamo Cardano's *Ars Magna*, a work that contains the formula for the solution of the general cubic equation. Every cubic, through a simple transformation, can be "depressed" to a cubic without a quadratic term. In modern notation, Cardano's formula for the three roots of the depressed cubic $x^3 + qx + r$ (with $q \neq 0$) is given by $\alpha + \beta$, $\omega\alpha + \omega^2\beta$, and $\omega^2\alpha + \omega\beta$, where

$$\alpha^3 = \frac{-r + \sqrt{R}}{2},$$
$$R = r^2 + \frac{4q^3}{27},$$
$$\beta = -\frac{q}{3\alpha},$$

and where $\omega = \left(-1 + \sqrt{-3}\right)/2$ is a primitive cube root of unity.

To demonstrate this classic formula, consider the depressed cubic equation $x^3 - 6x + 1 = 0$, which from its graph has three real roots. A little calculation reveals that

$$\alpha^3 = \frac{-1 + \sqrt{-31}}{2} \qquad \text{and} \qquad \beta^3 = \left(\frac{2}{\alpha}\right)^3 = \frac{-1 - \sqrt{-31}}{2}.$$

Then, one of the roots is given by

$$\alpha + \beta = \sqrt[3]{\frac{-1 + \sqrt{-31}}{2}} + \sqrt[3]{\frac{-1 - \sqrt{-31}}{2}},$$

where the cube roots are taken to preserve the condition $\beta = (2/\alpha)$.

Note how Cardano's formula expresses this real root as a sum of cube roots of complex numbers. For centuries, mathematicians were puzzled by this fact: Cardano's formula requires an excursion into the complex numbers to express in radicals the roots of an irreducible cubic polynomial with rational coefficients having only real roots. This phenomenon, known as the *casus irreducibilis*, is a famous theorem in the theory of polynomial equations. Before stating the result, we recall a few definitions from field theory.

Let F be a field. A polynomial $f(x) \in F[x]$ is *irreducible* over F if it cannot be written as the product of two polynomials in $F[x]$ of lower degree. A field K is a

Math. Mag. **87** (2014) 377–380. doi:10.4169/math.mag.87.5.377. © Mathematical Association of America

radical extension of F if $K = F(a)$ and $a^m \in F$ for some positive integer m. A chain of fields $F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t$ is a *radical tower* over F if each successive field extension is a radical extension of the preceding field. And finally, a field E is called a *splitting field* of the polynomial $f(x) \in F[x]$ if $F \subseteq E$, $f(x)$ splits into linear factors over E, and E is generated by F and the roots of $f(x)$.

To illustrate these concepts, consider the cubic polynomial $x^3 - 6x + 1$ above. It is irreducible over the field $\mathbb{Q}$ of rational numbers since it has no rational root. If we let

$$a = \sqrt{-31} \qquad \text{and} \qquad b = \sqrt[3]{-1 + \sqrt{-31}},$$

then $\mathbb{Q} \subseteq \mathbb{Q}(a) \subseteq \mathbb{Q}(a, b) \subseteq \mathbb{Q}(a, b, \omega)$ is a radical tower over $\mathbb{Q}$. Finally, $E = \mathbb{Q}(u_1, u_2, u_3)$ is a splitting field of $x^3 - 6x + 1$, where u_1, u_2, u_3 denote the three real roots of $x^3 - 6x + 1 = 0$. We observe, by Cardano's formula, that $E \subseteq \mathbb{Q}(a, b, \omega)$.

We can now state the famous theorem proved by Otto Hölder in the late 19th century.

CASUS IRREDUCIBILIS [**2**, p. 217]. *Let $f(x)$ be an irreducible cubic over the field $\mathbb{Q}$ of rational numbers having three real roots with splitting field E, and let $\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t$ be a radical tower with $E \subseteq K_t$. Then $K_t \not\subseteq \mathbb{R}$, where $\mathbb{R}$ is the field of real numbers.*

The remarkable consequence of this result is the necessity of complex numbers to express by radicals the three real roots of the irreducible cubic. More specifically, the proof of the theorem establishes that a primitive cube root of unity must lie in the field K_t.

In this paper, we seek an analogous result for finite fields. To do so, we need a generalized version of the *casus irreducibilis* that requires a slight modification to the definition of radical extension. To motivate this adaptation, suppose we wish to adjoin a primitive eighth root of unity γ to the field $\mathbb{Q}$ of rational numbers. An obvious way would be the single radical extension $\mathbb{Q}(\gamma)$. However, with this rather naive approach, it is not obvious which new elements have been added to $\mathbb{Q}$. Alternatively, we could choose to work exclusively with irreducible polynomials in our radical extensions. For instance, observe that the polynomial $x^8 - 1$, of which γ is a root, is reducible over $\mathbb{Q}$,

$$x^8 - 1 = (x - 1)(x + 1)(x^2 + 1)(x^4 + 1).$$

The quadratic factor $x^2 + 1$, of which the imaginary unit i is a root, is irreducible over $\mathbb{Q}$. Furthermore, the quartic factor $x^4 + 1$ is reducible over the radical extension field $\mathbb{Q}(i)$ of $\mathbb{Q}$ and factors into two irreducible quadratic polynomials,

$$x^4 + 1 = (x^2 + i)(x^2 - i).$$

If we take the root $\sqrt{i}$ of the second irreducible factor $(x^2 - i)$, then the radical tower

$$\mathbb{Q} \subseteq \mathbb{Q}(i) \subseteq \mathbb{Q}\big(i, \sqrt{i}\big)$$

also adjoins γ to $\mathbb{Q}$ (since $\mathbb{Q}(\gamma) = \mathbb{Q}(i, \sqrt{i})$), but now the field $\mathbb{Q}$ has been extended in a systematic way. With this in mind, we give the refined definition.

A field $K = F(a)$ is an *irreducible radical extension* of F if it is a radical extension (i.e., $a^m \in F$) and the polynomial $x^m - a^m$ is irreducible over F. Similarly, a chain of fields $F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t$ is an *irreducible radical tower* over F if each successive field extension is an irreducible radical extension of the preceding field. Finally, an element is expressible in terms of *irreducible radicals* if it lies in some irreducible radical tower.

With this new concept, we can state a generalized form of the *casus irreducibilis* due to Mann [**1**]: Let $f(x)$ be an irreducible cubic over the field F of characteristic not equal to 2 or 3 with splitting field E, and $F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t$ be an irreducible radical tower with $E \subseteq K_t$. Then K_t contains a primitive cube root of unity.

Mann's result clarifies the situation in the *casus irreducibilis*: If the radical extensions used by Cardano's formula to solve the irreducible cubic with three real roots can be expressed as irreducible radical extensions (which they can), then the ultimate field extension in the radical tower must contain a primitive cube root of unity and therefore cannot be a subfield of the real numbers. The requirement that the characteristic of F be different from 3 is a necessary restriction; otherwise primitive cube roots of unity do not exist in F. (Mann also addresses the characteristic 2 case in the same note.)

We consider an analogous result for irreducible cubic polynomials over finite fields. Let F_p denote the prime finite field of characteristic prime p and assume $p > 3$. Also, let $GF(p^n)$ denote the Galois field with p^n elements, which is a radical extension of degree n over F_p. In this simpler setting, we explore whether the roots of $f(x) \in F_p[x]$ are expressible in terms of *irreducible* radicals present in the splitting field of $f(x)$.

For example, since $x^2 - 2$ is irreducible over F_5, as it does not have a root in F_5, adjoining a square root of 2 to the field F_5 results in an irreducible radical extension $GF(5^2)$ of F_5. In contrast, $x^3 - 2$ is reducible over F_5,

$$x^3 - 2 = (x + 2)(x^2 + 3x + 4),$$

and the same is true for each binomial cubic equation $x^3 - c$ over F_5. Thus it is not clear whether $GF(5^3)$ is an irreducible radical extension of F_5. In fact, it is not, as we shall see shortly.

For the remainder of this paper, $f(x) \in F_p[x]$ will be an irreducible cubic over F_p and $E = GF(p^3)$ will denote the splitting field of $f(x)$ over F_p. We begin with a lemma about the existence of primitive cube roots of unity in the splitting field E.

LEMMA. *$GF(p^3)$ contains a primitive cube root of unity if and only if $p \equiv 1$ (mod 3).*

Proof. Let $E = GF(p^3)$. Then, $E^* = E - \{0\}$ is a cyclic group under multiplication with order $p^3 - 1$. Therefore, applying both Lagrange's theorem and Cauchy's theorem for finite groups, E^* has an element of order 3—and thus E has a primitive cube root of unity—if and only if $3 \mid (p^3 - 1)$. However, as $p^3 \equiv p \pmod 3$ by Fermat's little theorem, this last condition is easily seen to be equivalent to $p \equiv 1 \pmod 3$. ■

In fact, $GF(p^3)$ contains a primitive cube root of unity if and only if F_p does, but we will not need this fact. We are now ready to prove the main result of this paper.

THEOREM. *E is an irreducible radical extension of F_p if and only if $p \equiv 1$ (mod 3).*

Proof. Suppose that 3 is not a divisor of $p - 1$. By the result of Mann, every irreducible radical tower $F_p = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t$ containing E must contain a primitive cube root of unity. By the lemma above, E does not contain a primitive cube root of unity. Hence, E is not an irreducible radical extension of F_p. Conversely, suppose that E is not an irreducible radical extension of F_p. Consider the polynomials $x^3 - c \in F_p[x]$ for each $c \in F_p$. It must be the case that each of these polynomials is reducible over F_p, otherwise E is the splitting field of the irreducible polynomial $x^3 - c \in F_p[x]$ for some $c \in F_p$, and thus E is an irreducible radical extension of F_p, contrary to the hypothesis. But this can happen only if the mapping $c \mapsto c^3$ is a bijection on F_p. So the group $F_p^* = F_p - \{0\}$ does not have an element of order 3. It follows from Cauchy's theorem that 3 is not a divisor of $p-1$. Consequently, $p \not\equiv 1 \pmod 3$. ■

The theorem gives a simple divisibility condition to determine whether the roots of $f(x)$ are expressible in terms of irreducible radicals present in the splitting field E. In particular, the roots cannot be expressed in terms of irreducible radicals in E exactly when $p \not\equiv 1 \pmod 3$. This situation mirrors the historic *casus irreducibilis* in that both cases require the presence of elements not in the splitting field in order to express the roots of an irreducible cubic in terms of irreducible radicals.

To demonstrate, consider the polynomial x^3+3x+2, which is readily seen to be irreducible over both F_5 and F_7. If $p=5$, then the roots of this polynomial are not expressible in terms of irreducible radicals in $GF(5^3)$. In particular, a cube root of $-1+\sqrt{2}$, as given by Cardano's formula, is not present in the splitting field. This element, however, does lie in the extension field $GF(5^6)$. If, on the other hand, $p=7$, then $F_7 \subseteq GF(7^3)$ is an irreducible radical extension, and a cube root of $-1+\sqrt{2}$ is present in the splitting field. Hence, the roots of this same polynomial are expressible in terms of irreducible radicals in $GF(7^3)$, namely,

$$\begin{aligned}
\alpha_1 &= \sqrt[3]{-1+\sqrt{2}} + \sqrt[3]{-1-\sqrt{2}}, \\
\alpha_2 &= \omega\left(\sqrt[3]{-1+\sqrt{2}}\right) + \omega^2\left(\sqrt[3]{-1-\sqrt{2}}\right), \\
\alpha_3 &= \omega^2\left(\sqrt[3]{-1+\sqrt{2}}\right) + \omega\left(\sqrt[3]{-1-\sqrt{2}}\right),
\end{aligned}$$

where $\omega=2$ and $\omega^2=4$ are the primitive cube roots of unity.

REFERENCES

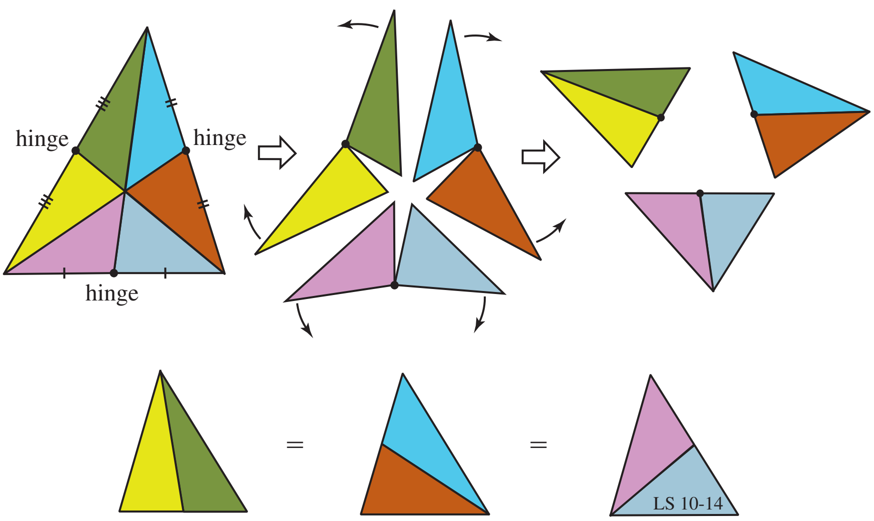
1. Henry B. Mann, On the *Casus Irreducibilis*, *Amer. Math. Monthly* **71** (1964) 288–290. http://dx.doi.org/10.2307/2312188
2. Joseph J. Rotman, *Advanced Modern Algebra*, American Mathematical Society, Providence, RI, 2010.

Summary For centuries, mathematicians were puzzled by the fact that Cardano's formula requires an excursion into the complex numbers to express in radicals the roots of an irreducible cubic polynomial with rational coefficients having only real roots. In this paper, we consider an analogous case for irreducible cubics over finite fields. In this simpler setting, we explore whether the roots of an irreducible cubic polynomial are expressible in terms of irreducible radicals present in its splitting field. The key theorem gives a simple divisibility condition that answers this question completely. Thus there is a situation that mirrors the historic "casus irreducibilis," in that both cases require the presence of elements not in the splitting field in order to express the roots of an irreducible cubic in terms of irreducible radicals.

A Triangle Theorem

LEE SALLOWS
Johannaweg 12, 6523 MA, Nijmegen
The Netherlands
lee.sal@inter.nl.net

For any triangle dissected along its medians, rotating the smaller triangles about three hinges placed at the center of each side yields three congruent triangles.



The medians of any triangle T dissect it into six smaller triangles, three adjacent pairs of which share a common vertex point that is the midpoint of a side of T. If the two triangles in each pair are rotated about their common vertex until they meet so as to share a common side, then the three new triangles formed by the union of each pair are congruent. The proof is elementary.

If T is isosceles (or equilateral) then the three triangles are also isosceles (or equilateral).

Summary The medians of a triangle divide it into six parts, which can be reassembled into three congruent triangles.

Math. Mag. **87** (2014) 381. doi:10.4169/math.mag.87.5.381. © Mathematical Association of America

Multiplicative Subgroups of $\mathbb{C}$ that Contain Regular Jordan Curves

WIJARN SODSIRI
Khon Kaen University
Khon Kaen 40002, Thailand
wijsod@kku.ac.th

Let X be a subset of a group G and $g \in G$. By the *left translate* gX (respectively, the *right translate* Xg) of X by g, we mean the set $\{gx : x \in X\}$ (respectively, $\{xg : x \in X\}$). If H is a subgroup of G, the translates gH and Hg are called, respectively, the *left* and *right cosets* of H in G. We know from elementary abstract algebra that the left (respectively, right) cosets of a subgroup *partition* the group; that is, any two left (respectively, right) cosets of the subgroup are either equal or disjoint, and the union of all the left (respectively, right) cosets is the whole group. This yields the following useful fact:

LEMMA 1. *Let H be a subgroup of a group G, $g \in G$, and let J be any subset of H. If the left translate gJ intersects with J, then H contains gJ.*

Proof. If $gJ \cap J \neq \emptyset$, then $gH \cap H \neq \emptyset$, so $gH = H$. This implies that H contains gJ. ■

In an article in this MAGAZINE [**6**], Tugger raised the following question: *Are there subsets of G other than subgroups of G whose left translates partition G?* She proved the following theorem:

TUGGER'S THEOREM. *Let X be a subset of a group G. Then the left translates of X partition G if and only if X is a (left or right) coset of a subgroup of G.*

Tugger gave an illustration and an application of the above theorem by considering the group $\mathbb{C}^\times$ of nonzero complex numbers under multiplication. If $X \subseteq \mathbb{C}^\times$ and $r > 0$, then the left translate $re^{i\theta}X$ is obtained geometrically from X by rotating around 0 through the angle θ and dilating (or "stretching") by r.

Tugger defined a closed curve in $\mathbb{C}^\times$ to be a continuous image of a circle in $\mathbb{C}^\times$ that contains at least two points. Note that a closed curve in this sense can degenerate into a simple arc. In general, a continuous image of a circle (or any connected, closed, bounded set) must be connected, closed, and bounded. So a closed curve must contain at least a segment of a curve with positive length.

For each $r \in \mathbb{R}^+$, we denote the circle in the complex plane $\mathbb{C}$ with center 0 and radius r by $S(r)$.

Tugger's very informative discussion mentioned without proof that "closed curves in $\mathbb{C}^\times$ apart from the circles centered at 0 are not cosets of any subgroup of $\mathbb{C}^\times$." Her statement is correct. This follows since, if a coset C of a subgroup H of $\mathbb{C}^\times$ is a closed curve, then H itself must be a closed curve. Hence H is bounded. If H contained an element z with $|z| \neq 1$, then H would contain the unbounded set of powers of z (or z^{-1}). So H must lie on the unit circle $S(1)$, and therefore H contains an arc J_1, where $J_1 := \{e^{it} : t \in [a, a+d]\}$ for some $a, d \in \mathbb{R}$ with $d > 0$, of $S(1)$.

Note that $e^{i(a+d)} \in e^{id}J_1 \cap J_1$. It follows from LEMMA 1 that H contains the arc $J_2 := e^{id}J_1 \cup J_1 = \{e^{it} : t \in [a, a+2d]\}$ of $S(1)$. Continuing this process inductively,

Math. Mag. **87** (2014) 382–385. doi:10.4169/math.mag.87.5.382. © Mathematical Association of America

we can eventually conclude that H must be $S(1)$. This implies that the coset C of H is actually a circle centered at 0.

On the other hand, the left translates of a circle centered at 0 partition $\mathbb{C}^\times$ and are cosets of the subgroup $S(1)$ of $\mathbb{C}^\times$. From these facts, Tugger came to the following incorrect result:

Let H be a subgroup of $\mathbb{C}^\times$ containing a closed curve. Then H is either the whole group $\mathbb{C}^\times$ or the unit circle $S(1)$.

Unfortunately, it is false as stated, as the closed curve may be a simple arc, or a circle centered at 0 apart from $S(1)$. In fact, there are many subgroups of $\mathbb{C}^\times$, other than $S(1)$ and $\mathbb{C}^\times$, that contain a simple arc or $S(1)$: The subgroups $\mathbb{R}^+$ and, for $n \in \mathbb{N}$, $\{rz : r \in \mathbb{R}^+$ and $z \in \mathbb{C}^\times$ is an nth root of 1$\}$ as well as $\{\sqrt[n]{r}e^{i\frac{\ln r}{n}} : r \in \mathbb{R}^+\}$ of $\mathbb{C}^\times$, mentioned in [**6**], all contain simple arcs; and if P is any multiplicative subgroup of $\mathbb{R}^+$, then the subgroup $\{z \in \mathbb{C}^\times : |z| \in P\}$ of $\mathbb{C}^\times$ consists of all circles with center 0 and radii in P.

Before giving a correct version of the result, we note that the properties of the absolute value show that the mapping $z \mapsto |z|$ is a homomorphism from $\mathbb{C}^\times$ to $\mathbb{R}^+$. We thus have the following lemma:

LEMMA 2. *Let H be any subgroup of $\mathbb{C}^\times$. Then $\{|z| : z \in H\}$ is a subgroup of $\mathbb{R}^+$.*

As usual, we identify the complex plane $\mathbb{C}$ with $\mathbb{R}^2$. Given nondegenerate closed intervals $[a_1, b_1]$ and $[a_2, b_2]$ in $\mathbb{R}$, the Cartesian product $R := [a_1, b_1] \times [a_2, b_2]$ is called a *rectangle* in $\mathbb{R}^2$ with *area* $|R| = (b_1 - a_1)(b_2 - a_2)$. Let $P_1 := (x_0, x_1, \ldots, x_m)$ with $a_1 = x_0 < x_1 < \cdots < x_m = b_1$ and $P_2 := (y_0, y_1, \ldots, y_n)$ with $a_2 = y_0 < y_1 < \cdots < y_n = b_2$ be subdivisions of $[a_1, b_1]$ and $[a_2, b_2]$, respectively. A rectangle R^* is said to be *obtained by P_1 and P_2* provided that there exist $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$ such that $R^* = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$. A *grid* on R is a collection $\mathcal{G}$ of all rectangles obtained by some subdivisions of $[a_1, b_1]$ and $[a_2, b_2]$.

Let E be a subset of $\mathbb{R}^2$ contained in a rectangle R, and let $\mathcal{G}$ be a grid on R. We define the *outer sum* of E with respect to $\mathcal{G}$ by $\mathcal{A}(E; \mathcal{G}) := \sum\{|R^*| : R^* \in \mathcal{G}$ and R^* intersects with the closure of $E\}$, where the *empty* sum is by definition zero. And E is said to be *of area zero* if given $\varepsilon > 0$, there exist a rectangle R with $R \supseteq E$ and a grid $\mathcal{G}$ on R such that $\mathcal{A}(E; \mathcal{G}) < \varepsilon$.

A *Jordan curve* in $\mathbb{C}$ is the range of a continuous function γ from a nondegenerate closed interval $[a, b]$ to $\mathbb{C}$ such that γ is injective on $[a, b)$ and $\gamma(a) = \gamma(b)$. We first state a main theorem about Jordan curves.

JORDAN CURVE THEOREM. *If J is a Jordan curve in $\mathbb{C}$, then $\mathbb{C} \setminus J$ consists of two nonempty disjoint open connected regions. One, denoted by $\mathcal{I}(J)$, is the interior of J and is bounded, and the other, denoted by $\mathcal{E}(J)$, is the exterior of J and is unbounded. Moreover, J is the boundary of both $\mathcal{I}(J)$ and $\mathcal{E}(J)$, and the closure of $\mathcal{I}(J)$ is $\mathcal{I}(J) \cup J$.*

Note that if two Jordan curves J_1 and J_2 do not intersect, then either their interiors have disjoint closures or the closure of one's interior is properly contained in the other's interior.

Osgood [**3**] discovered that a Jordan curve is not necessarily of area zero. In this article, the Jordan curves of area zero are called *regular Jordan curves*. For instance, rectangles and circles in $\mathbb{R}^2$ are regular Jordan curves. If J is a regular Jordan curve, then the areas of $\mathcal{I}(J)$ and its closure can be defined formally; these areas are indeed equal (consult [**7**] for more details).

If we restrict ourselves to regular Jordan curves (as Tugger may have intended), then we can characterize all subgroups of $\mathbb{C}^\times$ that contain such regular Jordan curves.

LEMMA 3. *For any regular Jordan curve J and its image $e^{i\theta}J$ under a rotation around* 0 *through the angle θ, if the interiors of J and $e^{i\theta}J$ intersect, then so do J and $e^{i\theta}J$ themselves.*

Proof. Note that rotations around 0 are rigid motions that preserve area. If J and $e^{i\theta}J$ did not intersect, we would have from the previous paragraph that either $\mathcal{I}(J) \cup J \subsetneq \mathcal{I}(e^{i\theta}J)$ or $\mathcal{I}(e^{i\theta}J) \cup e^{i\theta}J \subsetneq \mathcal{I}(J)$. This is impossible since the areas of $\mathcal{I}(J) \cup J$, $\mathcal{I}(e^{i\theta}J)$, $\mathcal{I}(e^{i\theta}J) \cup e^{i\theta}J$, and $\mathcal{I}(J)$ are all identical. ■

We now have all tools to establish our main result.

CHARACTERIZATION OF SUBGROUPS OF $\mathbb{C}^\times$ CONTAINING REGULAR JORDAN CURVES. *Let H be any subgroup of $\mathbb{C}^\times$. Then the following are equivalent:*

(a) H contains a regular Jordan curve.
(b) H contains an arc of a circle centered at 0.
(c) H contains a circle centered at 0.
(d) H is a union of circles centered at 0, *whose radii form a subgroup of the multiplicative group $\mathbb{R}^+$.*

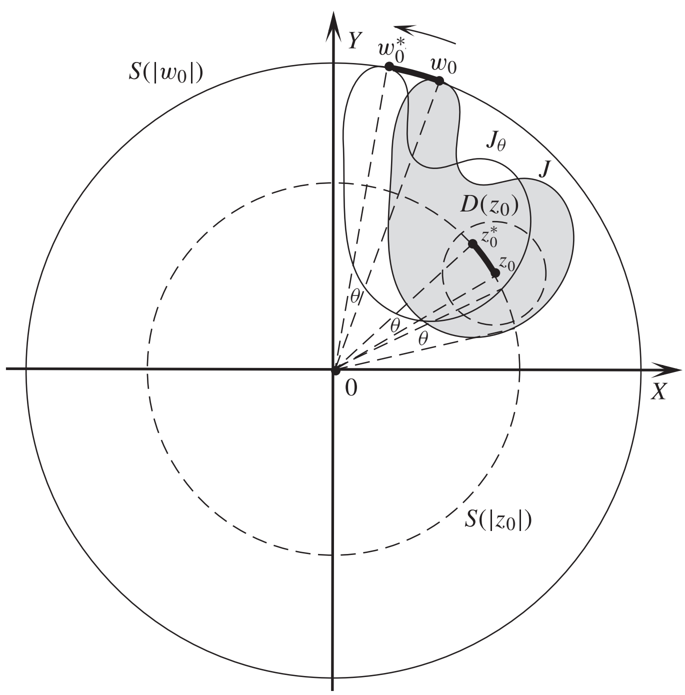


Figure 1 The Jordan curve J and its rotational image J_θ.

Proof. (a) ⇒ (b). Assume that H contains a regular Jordan curve, say J. If J itself is a circle centered at 0, we are done. Assume that J is not a circle centered at 0. Let w_0 be an arbitrary point on J. We wish to show that H contains an arc of $S(|w_0|)$. Let z_0 be a point in the interior of J; then there is an open disk $D(z_0)$ with center z_0 that is contained in the interior of J. Now, let $\theta \in \mathbb{R}^+$ be small enough so that the arc $\widehat{z_0^* z_0} := \{e^{it}z_0 : t \in [0, \theta]\}$ of $S(|z_0|)$ lies in $D(z_0)$. (See FIGURE 1.) It is readily seen that, for each $t \in [0, \theta]$, the point $e^{it}z_0$ belongs to the interior of J and the interior of the image $J_t := e^{it}J = \{e^{it}z : z \in J\}$ of J under the counterclockwise rotation around 0 through the angle t. It follows from LEMMA 3 that, for each $t \in [0, \theta]$, J and its image J_t intersect; by LEMMA 1, H also contains J_t. We deduce that H contains the arc $\widehat{w_0^* w_0} := \{e^{it}w_0 : t \in [0, \theta]\}$ of $S(|w_0|)$, as desired.

(b) $\Rightarrow$ (c). Assume that H contains an arc of a circle C centered at 0. Rotations of this arc around 0 with application of LEMMA 1 show that H contains the whole circle C.

(c) $\Rightarrow$ (d). Assume that H contains a circle C centered at 0. For each $\theta \in \mathbb{R}$, since $e^{i\theta}C = C$, the left coset $e^{i\theta}H$ also contains C. Thus, $e^{i\theta}H = H$ for all $\theta \in \mathbb{R}$. We infer that $H = \bigcup_{z \in H} S(|z|)$, which is a union of circles centered at 0. Moreover, it follows from LEMMA 2 that the set of radii of these circles is a subgroup of the multiplicative group $\mathbb{R}^+$.

(d) $\Rightarrow$ (a). Clear.

This completes the proof. ■

Acknowledgment The author would like to express his gratitude to the referees for their suggestions to improve the results of this article.

REFERENCES

1. James Ward Brown and Ruel V. Churchill, *Complex Variables and Applications*, 8th ed., McGraw-Hill, New York, 2009.
2. John B. Fraleigh, *A First Course in Abstract Algebra*, 5th ed., Addison-Wesley, Reading, MA, 1994.
3. William F. Osgood, A Jordan curve of positive area, *Trans. Amer. Math. Soc.* **4** (1903) 107–112. http://dx.doi.org/10.1090/S0002-9947-1903-1500628-5
4. Bruce P. Palka, *An Introduction to Complex Function Theory*, Spriger-Verlag, New York, 1991.
5. Richard A. Silverman, *Complex Analysis with Applications*, Dover, Mineola, NY, 1984.
6. Roberta T. Tugger, Subgroups and the partitioning property, *Math. Mag.* **66** (1993) 114–115. http://dx.doi.org/10.2307/2691120
7. William R. Wade, *An Introduction to Analysis*, 4th ed., Pearson, Upper Saddle River, NJ, 2010.

Summary In this paper, we characterize subgroups of the multiplicative group $\mathbb{C}^\times$ of nonzero complex numbers that contain Jordan curves of area zero.

Coming Soon in *The College Mathematics Journal*

Weak Contractions and Fixed Points, *by Daniel Acosta and Terry Lawson*

Sequences of Power Lines, *by Ricardo Alfaro*

The Origin of Quaternions, *by Thomas J. Bannon*

To Be (a Circle) or Not To Be? *by Hassan Boualem and Robert Brouzet*

An Intrinsic Formula for Cross Ratio in Spherical and Hyperbolic Geometries, *by Robert L. Foote and Xidian Sun*

Adjusting Child Support Payments in Michigan, *by Michael A. Jones and Jennifer Wilson*

What Distributes Over Exponentiation? *by Sherman Stein*

One Picture, All the Conics

ÓSCAR CIAURRI
Universidad de La Rioja
26004 Logroño, Spain
oscar.ciaurri@unirioja.es

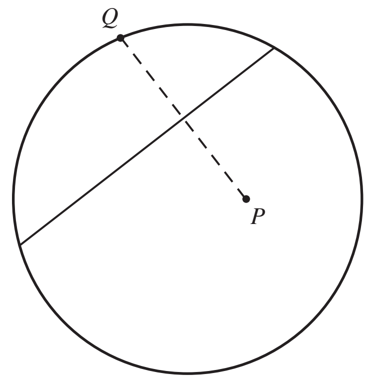


Figure 1 The crease obtained by folding Q onto P

Take a piece of paper, draw a circle of radius 1 on it, and fix a point P in its interior. Take a point Q on the circumference of the circle and fold the paper over until Q meets P, as in FIGURE 1. Draw a line along the crease in the paper. Do this for each point on the circumference. These lines enclose a region on the paper; that is, they form the envelope of tangent lines to a curve in the plane. What is the shape of this region?

This is a problem that could be proposed in a first-year course in calculus. The shape of the region limited by the creases is an ellipse, and it can be obtained as the envelope of the family of lines containing the creases. Indeed, each crease belongs to the line perpendicular to the segment PQ through its midpoint. To determinate each one of these lines we take a coordinate system in which the given circle has center $(0, 0)$ and radius 1, and in which $P = (a, 0)$ for some $-1 < a < 1$. Each point Q can be written by $Q = (\cos t, \sin t)$ for some $t \in [0, 2\pi)$, and the lines containing the creases are $c_a(t, x, y) = 0$, where

$$c_a(t, x, y) = x(a - \cos t) - y \sin t + \frac{1 - a^2}{2}.$$

According to the method described by E. Goursat in [**1**, chap. X], the envelope of this family of lines is a solution of the system

$$c_a(t, x, y) = 0 \qquad \text{and} \qquad \frac{\partial c_a(t, x, y)}{\partial t} = 0.$$

From this fact, we deduce that the parametric coordinates of the envelope are

$$x_a(t) = \frac{1 - a^2}{2(1 - a\cos t)} \cos t \qquad \text{and} \qquad y_a(t) = \frac{1 - a^2}{2(1 - a\cos t)} \sin t,$$

for $t \in [0, 2\pi)$, and this is a parametrization of the ellipse, denoted by E_a,

$$4\left(x - \frac{a}{2}\right)^2 + \frac{4y^2}{1 - a^2} = 1.$$

Each of these ellipses has foci $(0, 0)$ and $(a, 0)$ and major axis 1.

Taking the point P outside the circle, the situation is exactly the same, because we can reproduce all the previous computations considering $|a| > 1$. But in this case the envelope E_a is a hyperbola. The family of lines $c_a(t, x, y) = 0$ and its envelope E_a are shown in FIGURE 2 for the case $|a| < 1$, and in FIGURE 3 for the case $|a| > 1$.

Math. Mag. **87** (2014) 386–387. doi:10.4169/math.mag.87.5.386. © Mathematical Association of America

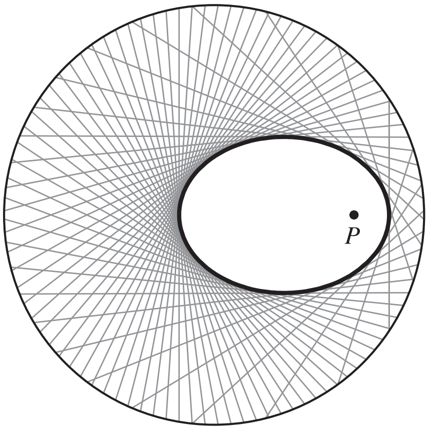


Figure 2 The lines $c_a(t, x, y) = 0$ and their elliptical envelope E_a for the case $|a| < 1$

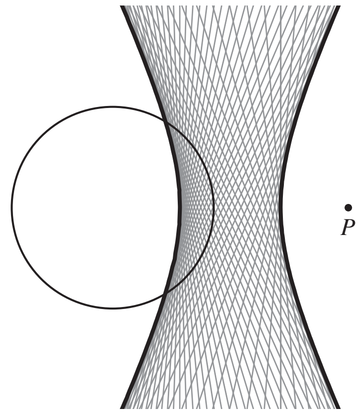


Figure 3 The lines $c_a(t, x, y) = 0$ and their hyperbolic envelope E_a for the case $|a| > 1$

But the family of conics E_a contains a final surprise: The envelope of the family of curves E_a, with $a \in \mathbb{R}$, is the curve $(-1 + x^2 - 2y)(-1 + x^2 + 2y) = 0$, that is, the pair of parabolas $y = \pm(1 - x^2)/2$. This is shown in FIGURE 4: One picture, all the conics.

We have to mention that in FIGURE 3, for the case $|a| > 1$, we have shown both branches of the hyperbola, but in FIGURE 4 we only show the branch of each hyperbola that is closer to the point $(a, 0)$.

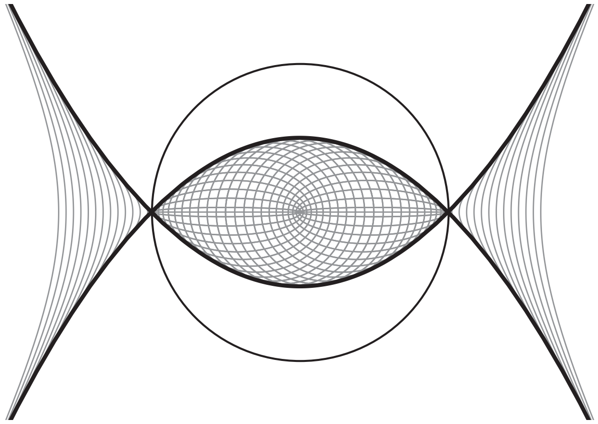


Figure 4 The family of conics E_a and its envelope: One picture, all the conics

Acknowledgment The author's research was supported by grant MTM2012-36732-C03-02 from the Spanish government.

REFERENCE

1. E. Goursat, *A Course in Mathematical Analysis, Vol. I: Derivatives and Differentials, Definite Integrals, Expansion in Series, Applications to Geometry,* trans. E. R. Hedrick, Dover, New York, 1959.

Summary In this note we obtain the envelopes of some families of lines and the envelope of the family of envelopes. All the conics appear.

The Pill Problem, Lattice Paths, and Catalan Numbers

MARGARET BAYER
University of Kansas
Lawrence, KS 66045-7594
bayer@ku.edu

KEITH BRANDT
Rockhurst University
Kansas City, MO 64110
Keith.Brandt@Rockhurst.edu

In 1991, Knuth and McCarthy [**5**] posed the following problem in the *American Mathematical Monthly*:

> A certain pill bottle contains m large pills and n small pills initially, where each large pill is equivalent to two small ones. Each day the patient chooses a pill at random, if a small pill is selected, (s)he eats it; otherwise (s)he breaks the selected pill and eats one half, replacing the other half, which thenceforth is considered to be a small pill.
>
> (a) What is the expected number of small pills remaining when the last large pill is selected?
>
> (b) On which day can we expect the last large pill to be selected?

When the solution was published [**11**], the *Monthly* editors commented that the origins of the problem were not clear. It turns out that several of those who submitted solutions had seen the problem before (at MIT and Michigan State, for example). In 2003, Brennan and Prodinger [**3**] studied the problem further and considered some variations, such as breaking the whole pills into more than two pieces.

In this paper we study a related question. The different sequences of pill selections are represented as paths in a binary tree that we call the *pill tree*. We count the vertices in the pill tree as a function of the initial numbers of large and small pills. In what follows we shall use the terms *whole* and *half* pills for large and small terms, to emphasize the size relationship. We observe some connections among the pill tree, lattice paths, and Catalan numbers.

The pill tree arises naturally in the work of Brandt and Waite [**2**] on the following probability question. What is the probability $P_k(w, h)$ that a whole pill is selected on the kth day, given that the bottle starts with w whole pills and h half pills? In [**2**] a recurrence relation is given for $P_k(w, h)$. The size of the pill tree shows the difficulty of implementing the recursion efficiently. Brandt and Waite study various storage techniques (arrays, trees, etc.) to eliminate redundant calculations and thereby improve the performance of their implementation.

The pill tree

For nonnegative integers w and h, the *pill tree* $PT(w, h)$ is a labeled, binary rooted tree with root labeled $\langle w, h\rangle$. A node labeled $\langle u, v\rangle$ has *left child* $\langle u-1, v+1\rangle$ (if $u > 0$) and *right child* $\langle u, v-1\rangle$ (if $v > 0$).

Math. Mag. **87** (2014) 388–394. doi:10.4169/math.mag.87.5.388. © Mathematical Association of America

A node labeled $\langle u, w\rangle$ represents a bottle containing u whole pills and v half pills. The root $\langle w, h\rangle$ represents the starting configuration of w whole pills and h half pills. The leaves all represent the empty configuration $\langle 0, 0\rangle$. The paths from root to leaves describe all possible sequences of configurations of pills in the bottle. A step down to the left represents choosing a whole pill; a step down to the right represents choosing a half pill. For example, $PT(2, 1)$ is given in FIGURE 1.

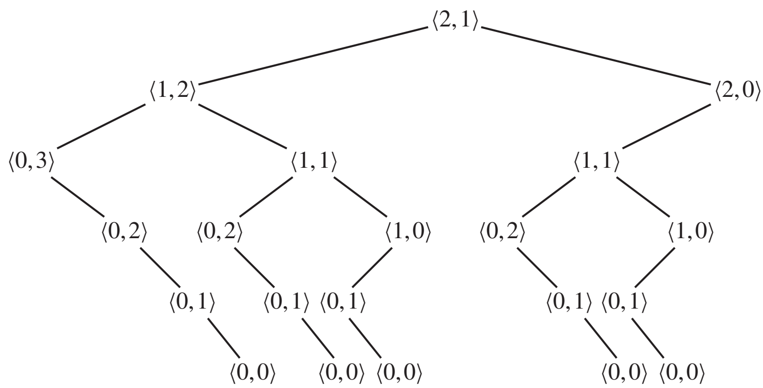


Figure 1 The Pill Tree *PT*(2, 1)

Let $T(w, h)$ be the number of nodes in the pill tree with initial configuration $\langle w, h\rangle$. The function T has an initial condition and recurrence relation similar to the recurrence of Brandt and Waite for $P_k(w, h)$. For $w, h \geq 0$,

$$T(w,h) = \begin{cases} h+1 & \text{if } w = 0; \\ 1 + T(w-1, 1) & \text{if } h = 0,\ w > 0; \\ 1 + T(w-1, h+1) + T(w, h-1) & \text{otherwise.} \end{cases} \tag{1}$$

TABLE 1 gives $T(w, h)$ for some small values of w and h.

TABLE 1: Number of nodes in the pill tree *PT*(*w*, *h*)

$w\backslash h$	0	1	2	3	4	5
0	1	2	3	4	5	6
1	3	7	12	18	25	33
2	8	21	40	66	100	143
3	22	63	130	231	375	572
4	64	195	427	803	1376	2210
5	196	624	1428	2805	5016	8398
6	625	2054	4860	9877	18276	31654

What are these numbers?

How can mathematicians, faced with a table like this, find a formula for, or at least better understand, the numbers? We can turn to the *On-Line Encyclopedia of Integer*

Sequences (OEIS) [**8**]. There we find that the numbers in the $h = 0$ column are partial sums of Catalan numbers. The Catalan numbers C_n are defined by $C_0 = 1$ and

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

for $n \geq 1$. The first few terms in the sequence are 1, 2, 5, 14, 42. An equivalent and commonly used formula is $C_n = \binom{2n}{n} - \binom{2n}{n-1}$. In the 1750s Euler gave these as the numbers of triangulations of convex polygons. Eugene Catalan rediscovered them in 1838, using them to count well-formed sequences of parentheses. Since then, these numbers have turned up in a vast number of settings. Stanley [**9**] stresses this in his exercise 6.19 where he says, "Show that the Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n}$ count the number of elements of the 66 sets S_i... given below." Stanley's web page addendum [**10**] extends this list to over 200. A brief history with citations for the original references can be found in [**9**, p. 212]. For a comprehensive introduction to Catalan numbers see Koshy [**6**].

Returning to our table, we notice that each entry in the $h = 1$ column is one less than the entry in the $h = 0$ column one row down. This is easily confirmed in general by the middle line of equation (1). The entries of the $h = 2$ column are Catalan numbers minus 2. Now it seems that we run out of luck: Neither we nor the OEIS recognizes the sequence of numbers in the $h = 3$ column. But what if these numbers are the partial sums of some sequence, as were the entries in the $h = 0$ column? We ask OEIS about the sequence of differences from the $h = 3$ column: 4, 14, 48, 165, 572, 2002, 7072. Again these numbers are related to Catalan numbers. The OEIS does not help to identify the sequences in the rows of the table, but it might be interesting to investigate these further. The connections between the finite sequences in the columns and the Catalan numbers lead to the following theorem.

THEOREM 1. *For $w \geq 1$ the number of nodes in the pill tree is*

(a) $T(w, 0) = \sum_{i=1}^{w+1} C_i$

(b) $T(w, 1) = \sum_{i=2}^{w+2} C_i$

(c) $T(w, 2) = C_{w+3} - 2$

Because of the recursion (1), each part of the theorem implies the others.

Catalan numbers and lattice paths

To understand the role of the Catalan numbers in the pill tree, we focus on the application of Catalan numbers to counting lattice paths.

The *integer lattice* is the set of points in the Cartesian plane having integer coordinates. We think of this as a grid with length-one vertical and horizontal line segments connecting adjacent lattice points. A *lattice path* is a sequence of connected rightward and upward segments going from $(0, 0)$ to a point (c, d). Moving along a rightward edge means adding $(1, 0)$ to the lattice point; moving along an upward edge means adding $(0, 1)$. The whole path must contain exactly $c + d$ edges, c horizontal (rightward) edges and d vertical (upward), and these c and d edges can be in any order. The number of lattice paths from $(0, 0)$ to (c, d) is $\binom{c+d}{d}$, because the d horizontal edges can be chosen to be any of the $c + d$ edges of the path.

Catalan numbers count a restricted class of lattice paths. First, assume the final point lies on the line $y = x$. Second, assume the lattice path never goes above the line $y = x$; that is, each lattice point in the path is of the form (s, t), with $s \geq t$. Such a path is called a *ballot path*, sometimes referred to as a *Dyck path*. The number of ballot paths from (0, 0) to (n, n) is the nth Catalan number C_n. (For a proof, see [**6**, p. 259] or your favorite combinatorics textbook.) For example, $C_3 = 5$, and FIGURE 2 shows the five ballot paths from (0, 0) to (3, 3).

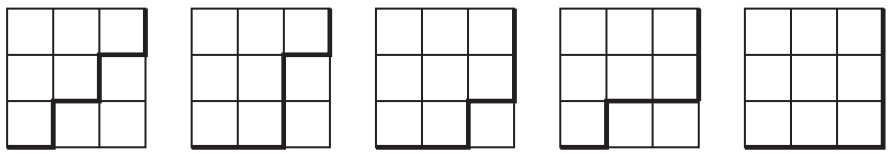


Figure 2 Ballot paths

Counting lattice paths with various restrictions goes back a long way. For a detailed history, see Humphreys [**4**]. Before the 1960s interest was primarily among statisticians. Indeed, the book *Lattice Path Counting and Applications* by S. G. Mohanty [**7**] was published in the series Probability and Mathematical Statistics. In the past thirty years lattice paths have become a standard topic in combinatorics.

Why the term "ballot path" for a lattice path that does not go above the line $y = x$? The following "ballot problem" dates from the 1880s [**1**]:

> Candidate A wins an election with a votes over opponent B, who receives b votes. The votes are counted one at a time. What is the probability that throughout the count, A stays ahead of B?

The vote count can be represented by a lattice path, where a rightward edge is drawn each time a vote for A is counted, and an upward edge is drawn each time a vote for B is counted. To compute the probability that A stays ahead of B we need to count a set of restricted lattice paths: those lattice paths from (0, 0) to (a, b) that stay strictly below the line $y = x$. (What we have called a "ballot path" is slightly different, but it is not too hard to make the conversion: Appending a horizontal segment at the beginning and a vertical segment at the end changes a path that does not go above the diagonal to one that stays strictly below the diagonal.)

It is time to return to the pill problem! It turns out that in the special case of $h = 0$, the pill tree $PT(w, 0)$ represents ballot paths from (0, 0) to (w, w). More generally, we show a connection between pill sequences and lattice paths.

In the pill tree $PT(w, h)$, a node labeled $\langle u, v\rangle$ represents a bottle having u whole pills and v half pills. We reach $\langle u, v\rangle$ by selecting whole pills $s = w - u$ times and selecting half pills $t = (h + w - u) - v$ times. Note that t can also be interpreted as the total reduction in the number of whole or half pills. The pill configuration labeled $\langle u, v\rangle$ in the pill tree can alternatively be identified by the pair (s, t). To empty the bottle, whole pills are selected w times and half pills are selected $w + h$ times. Thus, when the nodes of the pill tree are relabeled, each path in the tree from the root (now labeled (0, 0)) to a leaf (now labeled $(w, w + h)$) represents a lattice path from (0, 0) to $(w, w + h)$. See FIGURE 3 for the pill tree $PT(2, 1)$ relabeled as the lattice path tree.

When $h = 0$, the only half pills available in the process are those that came from whole pills, so $s \geq t$. This is also clear from the form $(s, t) = (w - u, w - (u + v))$. In this case the lattice paths corresponding to pill sequences are ballot paths.

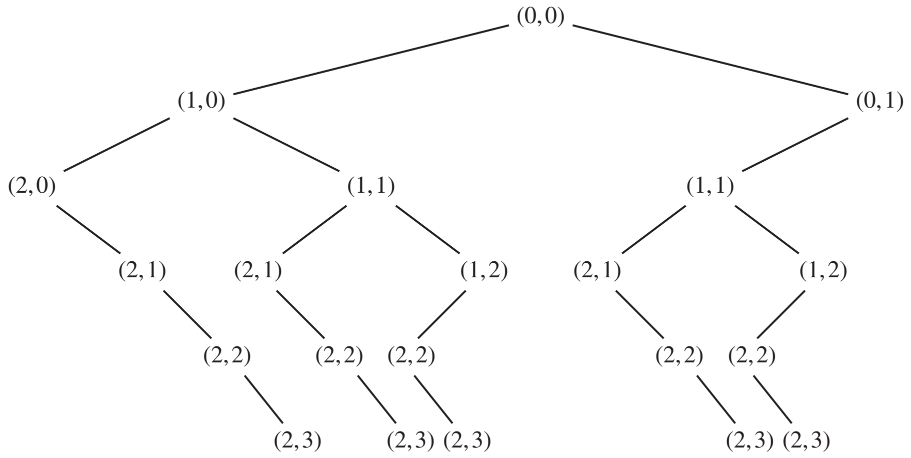


Figure 3 The lattice path tree

THEOREM 2. *The number of pill sequences that start with w whole pills and no half pills is the Catalan number* C_w.

Counting nodes in the pill tree

In the case $h = 0$, we found the number of lattice paths, or, equivalently, the number of leaves in the lattice path tree. But we want the total number of nodes in the pill tree (or lattice path tree). For the moment we still restrict ourselves to the case $h = 0$, and we work with the lattice path version of the pill tree. To count all the nodes in the lattice path tree, we count how many times a fixed label (s, t) (with $0 \le t \le s \le w$) occurs in the tree. For each node labeled (s, t), there is a unique path from the root to that node—representing a lattice path from $(0, 0)$ to (s, t) that does not go above the line $y = x$—and all such lattice paths are represented by paths in the tree. (Such a path can be extended in one or more ways to form a ballot path from $(0, 0)$ to (w, w).) Write $C(s, t)$ for the number of such lattice paths. A formula for $C(s, t)$ comes later, but we do not need it now, because our goal is a formula for the sum of such numbers for all pairs (s, t) with $0 \le t \le s \le w$.

LEMMA. *For all* $s \ge 0$, $\sum_{t=0}^{s} C(s, t) = C_{s+1}$.

Proof. Divide all allowable lattice paths from $(0, 0)$ to $(s + 1, s + 1)$ into $s + 1$ categories, depending on which lattice point with x-coordinate $s + 1$ they reach first. (Here "allowable" means not going above the line $y = x$.) Any path first reaching an x-coordinate of $s + 1$ at the point $(s + 1, t)$, where $0 \le t \le s$, must have previously passed through the point (s, t). The number of such lattice paths is $C(s, t)$. Summing these numbers $C(s, t)$ gives the total number of lattice paths from $(0, 0)$ to $(s + 1, s + 1)$ that do not go above the line $y = x$, that is, C_{s+1}. ∎

Proof of Theorem 1. The number of nodes in the pill tree $PT(w, 0)$ is the sum of $C(s, t)$ for all pairs (s, t) with $0 \le t \le s \le w$. Thus, by the lemma,

$$T(w, 0) = \sum_{s=0}^{w} C_{s+1} = \sum_{i=1}^{w+1} C_i.$$

The recursion (1) gives $T(w,1) = T(w+1,0) - 1$ and $T(w,2) = T(w+1,1) - T(w+1,0) - 1$. Thus

$$T(w,1) = \sum_{i=1}^{w+2} C_i - 1 = \sum_{i=2}^{w+2} C_i$$

and

$$T(w,2) = \sum_{i=2}^{w+3} C_i - \sum_{i=1}^{w+2} C_i - 1 = C_{w+3} - 2. \qquad \blacksquare$$

Can we use lattice paths to compute $T(w,h)$ for $h > 2$? The lattice points (s,t) no longer stay at or below the line $y = x$. For $s = w - u$ and $t = (w+h) - (u+v)$, we know only that $s + h \geq t$; that is, that the lattice points (s,t) do not go above the line $y = x + h$.

These and other variations of ballot paths also have a long history (see [**4**]). The following result was rediscovered by various people. We take it from [**7**, p. 3]. Write $C_h(s,t)$ for the number of lattice paths from $(0,0)$ to (s,t) that do not go above the line $y = x + h$. ($C_0(s,t)$ is what we called $C(s,t)$ before.)

THEOREM 3. *For $s \geq 0$, $h \geq 0$, and $t \leq s + h$,*

$$C_h(s,t) = \binom{s+t}{s} - \binom{s+t}{s+h+1}.$$

The number of nodes in the pill tree $P(w,h)$ is then the sum of the numbers $C_h(s,t)$ over all pairs (s,t) satisfying $0 \leq s \leq w$ and $0 \leq t \leq s + h$.

THEOREM 4. *The number of nodes in the pill tree $PT(w,h)$ is*

$$T(w,h) = \sum_{s=0}^{w} \left[\binom{2s+h+2}{s+1} - \binom{2s+h+2}{s} \right].$$

Proof. The number of nodes in the pill tree $PT(w,h)$ is

$$\sum_{s=0}^{w} \sum_{t=0}^{s+h} C_h(s,t).$$

The inner sum works the same way as the sum in the Lemma. That is,

$$\sum_{t=0}^{s+h} C_h(s,t) = C_h(s+1, s+h+1) = \binom{2s+h+2}{s+1} - \binom{2s+h+2}{s}. \qquad \blacksquare$$

The expression $\binom{2s+h+2}{s+1} - \binom{2s+h+2}{s}$ in the last equation can also be written as

$$\frac{h+1}{s+h+2} \binom{2s+h+2}{s+1}.$$

This theorem gives part (a) of Theorem 1 directly. For $h = 1$, note that

$$\binom{2s+3}{s+1} - \binom{2s+3}{s} = \binom{2s+4}{s+2} - \binom{2s+4}{s+1} = C_{s+2},$$

so we get part (b) of Theorem 1.

Conclusion

Lattice paths come up in a variety of contexts, so the lattice path tree can be interpreted in various ways. For example, a lattice path from (0, 0) to (n, m) gives a binary sequence of n minus ones and m plus ones, when an upward step is represented by $+1$ and a rightward step by -1. In the lattice path tree, moving to the left child appends a -1 to the sequence, moving to the right child appends a $+1$. The number of nodes in the pill tree $PT(w, h)$ is thus the number of sequences of plus and minus ones containing at most w minus ones and at most $w + h$ plus ones, and having all partial sums (of initial sequences) at most h.

Yet again, the ubiquitous Catalan numbers have shown up in an unexpected place. Theorem 2 does not surprise us, as one can easily associate a pill sequence with a lattice path or with a well-formed sequence of parentheses. But we found quite striking the values of T we observed in TABLE 1, hence this paper. Theorem 1 reveals surprising connections between Catalan numbers and the lattice paths that are often used to define them.

Dedication This article is dedicated by Margaret Bayer to the memory of her late husband, Ralph Byers, and by Keith Brandt to his father, John C. Brandt, in celebration of his 80th birthday.

REFERENCES

1. J. Bertrand, Calcul des probabilités. Solution d'un problème, *C. R. Math. Acad. Sci. Paris* **105** (1887) 369.
2. K. Brandt and K. Waite, Using recursion to solve the pill problem, *J. Computing Sciences in College* **24** (2009) 136–141. http://dl.acm.org/citation.cfm?id=1516595.1516624
3. Charlotte A. C. Brennan and Helmut Prodinger, The pills problem revisited, *Quaest. Math.* **26**(4) (2003) 427–439. http://www.ajol.info/index.php/qm/article/view/21711
4. Katherine Humphreys, A history and a survey of lattice path enumeration, *J. Statist. Plann. Inference* **140** (2010) 2237–2254. http://dx.doi.org/doi:10.1016/j.jspi.2010.01.020
5. Donald E. Knuth and John McCarthy, Problem E3429, *Amer. Math. Monthly* **98** (1991) 263. http://dx.doi.org/10.2307/2325034
6. Thomas Koshy, *Catalan Numbers with Applications*, Oxford University Press, Oxford, 2009.
7. Sri Gopal Mohanty, *Lattice Path Counting and Applications*, Academic Press, New York, 1979.
8. The on-line encyclopedia of integer sequences. http://oeis.org, accessed 2011.
9. Richard P. Stanley, *Enumerative combinatorics*, vol. 2, volume 62 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, 1999.
10. Richard P. Stanley, Catalan addendum, 2013. http://www-math.mit.edu/~rstan/ec/catadd.pdf
11. Walter Stromquist and Tim Hesterberg, Solution to E3429, *Amer. Math. Monthly* **99** (1992) 684. http://dx.doi.org/10.2307/2325015

Summary We define the pill tree, which is a rooted, binary tree consisting of different sequences of whole and half pills from a problem posed by Knuth and McCarthy. We observe some connections among the pill tree, lattice paths, and Catalan numbers, and give an explicit formula for the number of nodes in the tree.

PROBLEMS

BERNARDO M. ÁBREGO, *Editor*
California State University, Northridge

Assistant Editors: SILVIA FERNÁNDEZ-MERCHANT, California State University, Northridge; JOSÉ A. GÓMEZ, Facultad de Ciencias, UNAM, México; EUGEN J. IONASCU, Columbus State University; ROGELIO VALDEZ, Facultad de Ciencias, UAEM, México; WILLIAM WATKINS, California State University, Northridge

PROPOSALS

To be considered for publication, solutions should be received by May 1, 2015.

1956. *Proposed by Valery Karachik, South Ural State University, Chelyabinsk, Russia.*

Let a be a real number and k a positive integer. Let $\{a_n\}$ be a sequence of real numbers such that $\lim_{n\to\infty} a_n = a$. Evaluate

$$\lim_{n\to\infty}\left(\frac{a_0}{n+k}+\frac{a_1}{n+k+1}+\cdots+\frac{a_{(n-1)k}}{n+nk}\right).$$

1957. *Proposed by Wong Fook Sung, Temasek Polytechnic, Singapore.*

Evaluate

$$\sum_{n=1}^{\infty}\arccos\left(\frac{1+\sqrt{4n^2-1}}{2\sqrt{n(n+1)}}\right).$$

1958. *Proposed by Marcel Chiriţă, Bucharest, Romania.*

Determine all functions $f:\mathbb{N}\to[1,\infty)$ that satisfy the following conditions:

(i) $f(2)=4$,
(ii) $(n+1)f(n)\le nf(n+1)$ for every $n\in\mathbb{N}$, and
(iii) $f(nm)=f(n)f(m)$ for every $n,m\in\mathbb{N}$.

Math. Mag. **87** (2014) 395–402. doi:10.4169/math.mag.87.5.395. © Mathematical Association of America

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. Submitted problems should not be under consideration for publication elsewhere.

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

Solutions and new proposals should be mailed to Bernardo M. Ábrego, Problems Editor, Department of Mathematics, California State University, Northridge, 18111 Nordhoff St, Northridge, CA 91330-8313, or mailed electronically (ideally as a LaTeX or pdf file) to mathmagproblems@csun.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.

1959. *Proposed by Eddie Cheng and Jerrold W. Grossman, Oakland University, Rochester, MI.*

Let n be a positive integer, and let S be the set of subsets of $\{1, 2, \ldots, 2n\}$ of size n. Form an undirected graph with vertex set S by putting an edge between A and B if $B = \{1, 2, \ldots, 2n\} \setminus A$ or $B = \{2n + 1 - i : i \in A\}$. How many connected components does this graph have?

1960. *Proposed by Michael Goldenberg, The Ingenuity Project, Baltimore Polytechnic Institute, Baltimore, MD, and Mark Kaplan, Towson University, Towson, MD.*

Given a nonisosceles triangle ABC, let λ be a real number different from 1 and -1. Let X, Y, and Z be points on the lines BC, AC, and AB, respectively, such that

$$\frac{\overline{AZ}}{\overline{ZB}} = \frac{\overline{BX}}{\overline{XC}} = \frac{\overline{CY}}{\overline{YA}} = \lambda.$$

Let A', B', and C' be the intersections of the pairs of lines (BY, CZ), (CZ, AX), and (AX, BY), respectively, and call $\triangle A'B'C'$ a λ-triangle of $\triangle ABC$. Find the values of λ for which the triangles ABC and $A'B'C'$ are similar. (We call two triangles similar if, for some one-to-one correspondence among their vertices, the corresponding angles of the triangles are equal.)

Quickies

Answers to the Quickies are on page 401.

Q1045. *Proposed by Timothy Hall, PQI Consulting, Cambridge, MA.*

Evaluate

$$\sum_{k=1}^{\infty} \frac{2^{-k} + 1}{4^{-k} - 2^{-k+1} + 2^{k+1} - 1}.$$

Q1046. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let f be a Riemann integrable real-valued function on the closed interval $[a, b]$. Suppose that $\int_a^b f(x)\,dx = 0$ and $|f(x)| \le M$ for every x in $[a, b]$. Prove that

$$\left| \int_a^b x f(x)\,dx \right| \le \frac{M(b-a)^2}{4}.$$

Solutions

The smallest disk containing a unit length closed curve **December 2013**

1931. *Proposed by John E. Wetzel, University of Illinois at Urbana-Champaign, Urbana, IL.*

Find with proof the radius of the smallest disk that can cover every simple closed curve of unit length.

Solution by Edward Schmeichel, San Jose State University, San Jose, CA.
The smallest radius is $1/4$. An isosceles triangle with base $1/2 - \epsilon$ and perimeter 1 shows the radius cannot be smaller than $1/4$. To prove that $1/4$ is large enough, let Γ be a simple close curve with length 1, and let A, B, C be three distinct points on Γ. Triangle ABC has perimeter

$$|AB| + |BC| + |CA| \leq |\Gamma(A, B)| + |\Gamma(B, C)| + |\Gamma(C, A)| = 1.$$

We claim that there exists a point X such that $\max\{|XA|, |XB|, |XC|\} \leq 1/4$. In other words, the disks with centers at A, B, and C and radius $1/4$ have a point in common. Indeed, if $\triangle ABC$ is acute, let X be its circumcenter (in an acute triangle the circumradius is less than one-fourth the perimeter). If $\triangle ABC$ is obtuse, let X be the midpoint of its longest side.

Center a closed disk with radius $1/4$ at each point of Γ, denoting the resulting collection of disks by D. By the previous claim, any three disks in D intersect. So by Helly's Theorem, all the disks in D intersect. If Y is a point in the intersection of the disks in D, the disk with radius $1/4$ centered at Y covers Γ.

Also solved by Robert Calcaterra, Natacha Fontes-Merz, Omran Kouba (Syria), Thomas L. McCoy, and the proposer.

Rotations as compositions of varied reflections December 2013

1932. *Proposed by Michel Bataille, Rouen, France.*

In 3-space, let R_O, R_ℓ, and R_P be the reflections by a point O, a line ℓ, and a plane P, respectively. Characterize the configurations for which $R_O \circ R_\ell \circ R_P$ is a rotation; if such is the case, show that $R_\ell \circ R_O \circ R_P$ and $R_O \circ R_P \circ R_\ell$ are rotations whose axes are parallel to the axis of $R_O \circ R_\ell \circ R_P$.

Solution by the proposer.
We show that $R_O \circ R_\ell \circ R_P$ is a rotation if and only if ℓ is perpendicular to P at O, or ℓ is not orthogonal to P and the point O is in the plane containing ℓ and perpendicular to P.

In what follows, $T_{\overrightarrow{w}}$ denotes the translation with vector $\overrightarrow{w}$. If ℓ is orthogonal to P, then $R_\ell \circ R_P = R_{O'}$ where O' is the point of intersection of ℓ and P, and so $R_O \circ R_\ell \circ R_P = R_O \circ R_{O'}$ is a translation, which is not a rotation unless $O = O'$.

Conversely, suppose that ℓ is not orthogonal to P. Then, we first have

$$R_O \circ R_\ell \circ R_P = R_O \circ R_P \circ (R_P \circ R_\ell \circ R_P) = R_O \circ R_P \circ R_{\ell'},$$

where $\ell' = R_P(\ell)$. Note that ℓ' is contained in the plane Q containing ℓ and perpendicular to P. Let P' be the plane parallel to P through O, and let m be the line through O orthogonal to P' (and P). Because $R_O = R_m \circ R_{P'}$, it follows that

$$R_O \circ R_P = R_m \circ R_{P'} \circ R_P = R_m \circ T_{\overrightarrow{u}} = T_{\overrightarrow{u}} \circ R_m,$$

where $\overrightarrow{u}$ is some vector orthogonal to P. We deduce that

$$R_O \circ R_\ell \circ R_P = T_{\overrightarrow{u}} \circ R_m \circ R_{\ell'} = T_{\overrightarrow{u}} \circ \rho \circ T_{\overrightarrow{v}},$$

where ρ is a rotation around the common perpendicular n of the lines m and ℓ' (which are not parallel since ℓ' is not orthogonal to P), and $\overrightarrow{v}$ is parallel to n or $\overrightarrow{v} = \overrightarrow{0}$. Since $\overrightarrow{u}$ is orthogonal to n, $T_{\overrightarrow{u}} \circ \rho$ is a rotation ρ' with axis n' parallel to n, and finally $R_O \circ R_\ell \circ R_P = \rho' \circ T_{\overrightarrow{v}}$. Thus $R_O \circ R_\ell \circ R_P$ is a rotation if and only if $\overrightarrow{v} = \overrightarrow{0}$, that is, if and only if m and ℓ' intersect, which occurs if and only if O is in the plane Q.

If ℓ is orthogonal to P and O is the intersection of ℓ and P, then $R_O \circ R_\ell \circ R_P = I$ is the identity. Thus $R_P = R_O \circ R_\ell$ and $R_O = R_\ell \circ R_P$. It follows that

$$R_\ell \circ R_O \circ R_P = R_\ell \circ R_O \circ (R_O \circ R_\ell) = I \text{ and}$$
$$R_O \circ R_P \circ R_\ell = (R_\ell \circ R_P) \circ R_P \circ R_\ell = I.$$

If ℓ is not orthogonal to P and the point O is in the plane Q containing ℓ and perpendicular to P, then from what we have just seen, $R_O \circ R_\ell \circ R_P$ is a rotation whose axis n' is orthogonal to Q. Each of $R_\ell \circ R_O \circ R_P$ and $R_O \circ R_P \circ R_\ell$ is a direct isometry, hence each is a rotation if and only if the set of its fixed points is a line (and this line is then its axis). Note that $R_P \circ R_\ell \circ R_O = (R_O \circ R_\ell \circ R_P)^{-1}$ has axis n'. Now, for any point M,

$$\begin{aligned} R_\ell \circ R_O \circ R_P(M) = M &\Leftrightarrow R_P \circ R_\ell \circ R_O(R_P(M)) = R_P(M) \\ &\Leftrightarrow R_P(M) \in n' \Leftrightarrow M \in R_P(n'). \end{aligned}$$

Thus, $R_\ell \circ R_O \circ R_P$ has a line of fixed points, namely $R_P(n')$, which is parallel to n' (since n' is parallel to P). Similarly, for any point M,

$$\begin{aligned} R_O \circ R_P \circ R_\ell(M) = M &\Leftrightarrow R_\ell \circ R_O \circ R_P(R_\ell(M)) = R_\ell(M) \\ &\Leftrightarrow R_\ell(M) \in R_P(n') \Leftrightarrow M \in R_\ell(R_P(n')), \end{aligned}$$

and $R_\ell(R_P(n'))$ is a line parallel to n' since $R_P(n')$ is orthogonal to ℓ and parallel to n'.

Also solved by Sara Armstrong, Kinley Crump, Dawn Hendrix, and Justin Mills; Stephen Broughton; Shelby Davis, Pamela King, Melissa Borrelli, and Zachary Matlock; Heather Evans, Brandon Fenwick, and Jeffory Hogan; Eugene A. Herman; Peter McPolin (Northern Ireland); Vedaste Mutambuka, Claude Riche, Paul Scott, and Jerome Sims; and Xiaoshen Wang.

A multiple logarithmic integral — December 2013

1933. *Proposed by Ovidiu Furdui, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.*

Let $n \geq 2$ be a natural number. Calculate

$$\int_0^1 \int_0^1 \cdots \int_0^1 \ln(1 - x_1 x_2 \cdots x_n)\, dx_1\, dx_2 \cdots dx_n.$$

Solution by The Iowa State University Student Problem Solving Group, Iowa State University, Ames, IA.

Using the McLauren expansion for $\ln(1-x)$,

$$\begin{aligned} I &= \int_0^1 \int_0^1 \cdots \int_0^1 \ln(1 - x_1 x_2 \cdots x_n)\, dx_1\, dx_2 \cdots dx_n \\ &= \lim_{a \to 1^-} \int_0^a \int_0^1 \cdots \int_0^1 \ln(1 - x_1 x_2 \cdots x_n)\, dx_1\, dx_2 \cdots dx_n \\ &= \lim_{a \to 1^-} \int_0^a \int_0^1 \cdots \int_0^1 \left(-\sum_{k=1}^{\infty} \frac{(x_1 x_2 \cdots x_n)^k}{k} \right) dx_1\, dx_2 \cdots dx_n. \end{aligned}$$

Because the series converges uniformly on the set of the integration, we can integrate term by term to get

$$I = -\lim_{a \to 1^-} \sum_{k=1}^{\infty} \frac{a^{k+1}}{k(k+1)^n} = -\sum_{k=1}^{\infty} \frac{1}{k(k+1)^n},$$

the latter because the sum converges uniformly for $a \in [0, 1]$. Finally,

$$I = -\sum_{k=1}^{\infty} \frac{1}{k(k+1)^n} = \sum_{k=1}^{\infty} \left(-\frac{1}{k} + \frac{1}{k+1} + \frac{1}{(k+1)^2} + \cdots + \frac{1}{(k+1)^n} \right)$$

$$= -1 + \sum_{j=2}^{n} (\zeta(j) - 1) = \sum_{j=2}^{n} \zeta(j) - n.$$

Also solved by Robert A. Agnew, Michel Bataille (France), Khristo Boyadzhie, Brian Bradie, Bruce S. Burdick, Robert Calcaterra, Kennard Callender, Minh Can, Chip Curtis, Bruce E. Davis, John N. Fitch, Dmitry Fleischman, J. A. Grzesik, GWstat Problem Solving Group, Eugene A. Herman, Maung Soe Htet, the Iowa State University Student Problem Solving Group, Parviz Khalili, John C. Kieffer, Omran Kouba (Syria), Elias Lampakis (Greece), Kee-Wai Lau (China), Ang Li, Rituraj Nandan, Northwestern University Math Problem Solving Group, Jennifer O'Day, Moubinool Omarjee (France), Paolo Perfetti (Italy), Tomas Persson and Mikael P. Sundqvist (Sweden), the Pittsburg State University Problem Solving Group, Ángel Plaza (Spain), Nicholas C. Singer, Jayson Smith, Dave Trautman, Thomas P. Turiel, Traian Viteam (South Africa), Michael Vowe (Switzerland), Haohao Wang and Yangping Xia, and the proposer.

Quotient subrings of the same size **December 2013**

1934. *Proposed by Greg Oman, University of Colorado, Colorado Springs, CO.*

Let R be an infinite ring (not assumed commutative or to contain an identity). Consider the following two conditions:

(a) R has zero divisors; that is, there exist nonzero elements $s, t \in R$ such that $st = 0$.

(b) There exist nonzero elements $s, t \in R$ such that $st = 0$. Further, $Rs \neq \{0\}$ and $Rt \neq \{0\}$ (here $Rs := \{rs : r \in R\}$).

Does (a) imply that R possesses a nonzero left ideal I such that R and R/I have the same cardinality? Does the answer change if we assume (b) instead?

Solution by Andrew Samer, Cheektowaga, NY.

Let R be the ring $\mathbb{Z}$ with the usual addition and where all products in R are 0. Clearly R has zero divisors. As in the ring $\mathbb{Z}$ with the usual addition and multiplication, the only ideals are principal ideals. Thus any nonzero ideal I is generated by some element k, where $k \in \mathbb{Z}$ and $k \neq 0$. The derived quotient ring $\mathbb{Z}/k\mathbb{Z}$ has an additive structure isomorphic to the ring of integers modulo k. Thus, $\mathbb{Z}/k\mathbb{Z}$ is finite, as are all quotient groups. Thus R and R/I never have the same cardinality.

Condition (b) does imply the desired property. Consider the following two cases:

Case 1. $|Rt| = |R|$. Let $\phi : R \to R$ such that $\phi(r) = rt$. By the First Isomorphism Theorem, $R/\ker(\phi) \cong Rt$. As $st = 0$, $\ker(\phi)$ is nonempty. This implies that $|R/\ker(\phi)| = |Rt| = |R|$. Therefore, $\ker(\phi)$ is a nonzero left ideal that produces a quotient ring the same size as R.

Case 2. $|Rt| < |R|$. By our assumption Rt is nonzero. Note that $|R| = |R/Rt \times Rt|$. The cardinality of $R/Rt \times Rt$ is equal to $\max\{|R/Rt|, |Rt|\}$. As $|R| = |R/Rt \times Rt|$ and $|Rt| < |R|$, $|R/Rt|$ must equal $|R|$. Therefore, Rt is a nonzero left ideal that yields a quotient ring the same size as R.

Also solved by Robert Calcaterra and the proposer.

Royalty is mostly imperfect **December 2013**

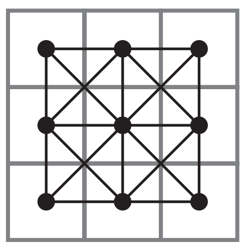
1935. *Proposed by Stan Wagon, Macalester College, St. Paul, MN.*

Let $K_{m,n}$ be the graph on the vertex set $\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$, where the vertex (m_1, n_1) is connected to the vertex (m_2, n_2) if the king piece in chess can move from the square (m_1, n_1) to the square (m_2, n_2). Define $Q_{m,n}$ accordingly for the possible queen moves on an $m \times n$ chessboard.

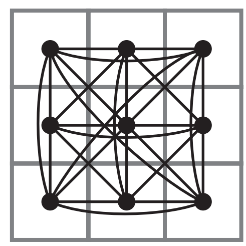
(a) For which pairs (m, n) is $K_{m,n}$ perfect?

(b) For which pairs (m, n) is $Q_{m,n}$ perfect?

Note: A graph is *perfect* if neither the graph nor its complement has a chordless odd cycle of length 5 or more.



$K_{3,3}$



$Q_{3,3}$

Solution by the proposer.

Assume in all graphs that $m \leq n$, and orient the graphs so that m is the number of rows and n the number of columns. Let G^c be the graph complement of G. Two easy general facts are useful:

Fact 1. The complement of a chordless 5-cycle is a chordless 5-cycle.

Fact 2. If C is a chordless cycle of length at least 7 in G^c, then every vertex in C has degree ≥ 4 in the subgraph of G induced by C (because the non-chords in G^c are edges in G).

(a) $K_{m,n}$ is perfect if and only if $m \leq 3$.

If $m \geq 4$, then $K_{m,n}$ is imperfect, via the chordless 7-cycle in FIGURE 4 (left).

There are no cycles in $G = K_{1,n}$ and since the the maximum degree of G is 2, the two facts above show that G^c has no chordless odd cycles of length 5 or more. Thus $K_{1,n}$ is perfect.

For $G = K_{2,n}$, let A be a leftmost vertex of C, a cycle of length 5 or more; the two C-edges leaving A go to two ends of an edge, so C has a chord. There is no chordless 5-cycle in G^c by Fact 1. Let C be a chordless cycle in G^c of size 7 or larger. Again, let A be the leftmost vertex in C. Fact 2 yields a contradiction since A has at most 3 neighbors that are not to its left. Thus $K_{2,n}$ is perfect.

Finally, let $G = K_{3,n}$, and again let A be a leftmost vertex of C, a cycle of length 5 or more. As in the $K_{2,n}$ case, A cannot lie on the first or last row of the graph; so A is the central vertex of a column. The two C-edges leaving A must be two diagonals in order to avoid a chord. Let AB and AC be the two diagonal edges leaving A. The edges leaving B and C can include no diagonals, since including one would lead, because of the lack of a chord, to a closing of the cycle at length 4. Similarly, a vertical edge is impossible, since it too would close the cycle at length 4. So the edges leaving B and C are horizontal. This argument can be repeated as we move to the right, so the cycle, if it is to avoid a chord, will have to close up with even length.

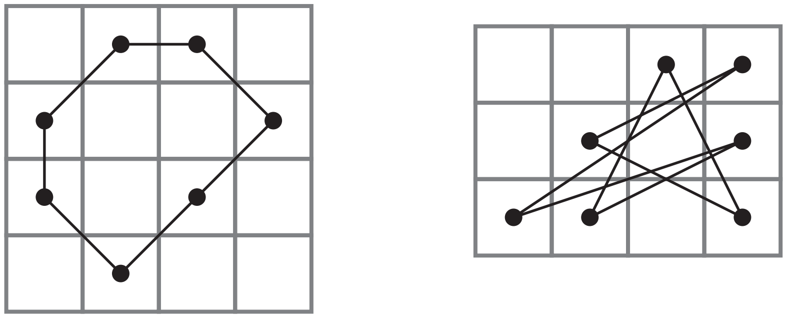


Figure 4 Chordless 7-cycles in $K_{4,4}$ (left) and in $(Q_{3,4})^c$ (right)

There is no chordless 5-cycle in G^c by Fact 1. Let C be a chordless cycle in G^c of size 7 or larger. Again, let A be a leftmost vertex in C. By Fact 2, A has at least four neighbors in C. Vertex A cannot be on the first or last row because it would have at most three neighbors in C. So it is in the central row. But for C to have four G-neighbors of A that are not to the left of A requires using a vertical neighbor, and this puts a leftmost vertex of C in the first or last row, contradicting an earlier observation.
(b) $Q_{m,n}$ is perfect if and only if $m \le 2$ or $m = n = 3$.

First, we have that $G = Q_{3,4}$ is imperfect by the chordless 7-cycle in G^c shown in FIGURE 4 (right). It remains to settle the $Q_{1,n}$, $Q_{2,n}$, and $Q_{3,3}$ cases. For $G = Q_{3,3}$, G^c is an 8-cycle and an isolated point. Fact 1 then tells us that G^c has no chordless 5-cycle. A chordless cycle in G^c can have at most two vertices in each row, and hence can have at most six vertices. So G is perfect. The graphs $Q_{1,n}$ are complete, and so perfect. Let $G = Q_{2,n}$. An odd cycle of length 5 or more in G must have three vertices in the same row, yielding a chord. A cycle $v_1, v_2, \ldots, v_s$ in G^c must alternate rows, but, if s is odd, that means v_s is a G-neighbor of v_1, which is a contradiction. Thus $Q_{2,n}$ is perfect.

Answers

Solutions to the Quickies from page 396.

A1045. For $n \ge 1$,

$$\begin{aligned}\sum_{k=1}^{n} \frac{2^{-k}+1}{4^{-k}-2^{-k+1}+2^{k+1}-1} &= \sum_{k=1}^{n} \frac{2^{-k}+1}{(2^{-k}+1)(2^{-k}-3+2^{k+1})} \\ &= \sum_{k=1}^{n} \frac{1}{2^{-k}-3+2^{k+1}} = \sum_{k=1}^{n} \frac{2^k}{(2^k-1)(2^{k+1}-1)} \\ &= \sum_{k=1}^{n} \left(\frac{1}{2^k-1} - \frac{1}{2^{k+1}-1}\right) = 1 - \frac{1}{2^{n+1}-1}.\end{aligned}$$

Thus

$$\sum_{k=1}^{\infty} \frac{2^{-k}+1}{4^{-k}-2^{-k+1}+2^{k+1}-1} = \lim_{n\to\infty} \left(1 - \frac{1}{2^{n+1}-1}\right) = 1.$$

A1046. Because $\int_a^b f(x)\,dx = 0$, it follows that

$$\int_a^b xf(x)\,dx = \int_a^b \left(x - \frac{a+b}{2}\right) f(x)\,dx.$$

By the Triangle Inequality for Integrals and the bound $|f(x)| \leq M$,

$$\begin{aligned}\left|\int_a^b xf(x)\,dx\right| &= \left|\int_a^b \left(x - \frac{a+b}{2}\right) f(x)\,dx\right| \\ &\leq \int_a^b \left|x - \frac{a+b}{2}\right| |f(x)|\,dx \leq M\int_a^b \left|x - \frac{a+b}{2}\right|\,dx \\ &= 2M\int_a^{(a+b)/2} \left(\frac{a+b}{2} - x\right)\,dx = \frac{M(b-a)^2}{4}.\end{aligned}$$

Solution to puzzle on page 360

C	A	P	E	D		S	H	O	D		P	F	C	S
A	L	E	V	E		T	O	F	U		E	R	A	T
T	E	R	E	N	C	E	T	A	O		L	Y	R	A
				S	O	N	A	R		S	H	E	A	R
L	A	U	R	E	N	T	S	C	H	W	A	R	T	Z
O	G	L	E						A	I	M			
O	R	C	A		A	N	D	A	L	L		O	S	E
F	I	E	L	D	S	M	E	D	A	L	I	S	T	S
A	P	R		A	C	E	T	A	L		S	C	A	T
			A	M	I						M	A	R	E
J	E	A	N	P	I	E	R	R	E	S	E	R	R	E
A	M	E	N	S		F	O	O	T	E				
N	O	R	A		N	G	O	B	A	O	C	H	A	U
U	T	I	L		A	H	M	E		U	S	E	R	S
S	E	E	S		H	I	S	S		L	A	N	K	A

Problems Section Index for Volume 87

The letters P, Q, and S refer to Proposals, Quickies, and Solutions, respectively; page numbers appear in parentheses. For example, P1938 (62) refers to Proposal 1938, which appears on page 62.

Math. Mag. **87** (2014) 403. doi:10.4169/math.mag.87.5.403. © Mathematical Association of America

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.

Alcock, Laura, *How to Study as a Mathematics Major*, Oxford University Press, xvi + 272 pp, $24.95(P). ISBN 978-0-199-66131-2. Available also under the title *How to Study for a Mathematics Degree* (with some localization to the British university system). ISBN 978-0-199-66132-9.

This book starts with the proven premise that "a student who expects mathematics to come in the form of procedures to copy will not know how to interact with material presented via definitions, theorems, and proofs." The first part of the book is about mathematical content and the second is about how to learn. The first part treats calculation procedures, abstract objects, definitions, theorems, proof, proof types, and reading and writing mathematics. The second part is about study skills and is less specific to mathematics (getting the most out of lectures, time management, the psychology of not being the best). Although the author is now based in England, she previously was a professor of mathematics education in the U.S., hence knows both the British and the American university systems. The writing style is pleasant and informal, and the advice is "spot on" (British) and "dead right" (U.S.). Author Alcock also has a newer book, *How to Think About Analysis*, due out at the end of 2014.

Hand, David J., *The Improbability Principle: Why Coincidences, Miracles, and Rare Events Happen Every Day*, Scientific American / Farrar, Straus and Giroux, 2014; $27, $15(P), $12.99 (ebook), $39.99 (audio CD). ISBNs 978-0-374-17534-4, 978-0-374-53500-1, 978-0-374-71139-9, 978-1-42725882-3.

Watkins, Phillip, "Things which ought to be expected can seem quite extraordinary if you've got the wrong model," *Significance* 11 (4) (October 2014) 36–39.

The weight of expectations: Interview, Jasper Fforde, *Significance* 11 (4) (October 2014) 6.

"Rare events happen all the time," I tell my students. And author Hand, former president of the Royal Statistical Society, explains in popular style (no equations) why this "Improbability Principle" holds. He bases it on Borel's law (if it's *too* unlikely, it won't happen), plus his own five laws: inevitability (*something* must happen), truly large numbers (with enough chances, "any outrageous thing is likely to happen"), selection ("postdiction" can be sure), the probability lever (slight changes can magnify probabilities), and near enough (call it a match even if it's not quite). The book is full of concrete examples from contemporary events. Author Watkins interviews Hand, starting from the question: "What is the probability of experiencing more than one once-in-a-million event in your lifetime?" Hand responds in terms of his Improbability Principle and also with a conventional (Poisson) model. Nevertheless, "things which ought to be expected can seem quite extraordinary if you've got the wrong model." He goes on to say that the Improbability Principle has not cost him his "sense of wonder" but still lets him "live as though everything is a miracle." Meanwhile, fantasy author Fforde asserts his own laws of "expectation-influenced probability" (if you expect something to happen, it probably will) and its inverse (most "random bad events" won't happen because you don't expect them to).

Math. Mag. **87** (2014) 404–405. doi:10.4169/math.mag.87.5.404. © Mathematical Association of America

Polster, Burkard, and Marty Ross, *Math Goes to the Movies*, Johns Hopkins Press, 2012; xiv + 286 pp, $70, $35(P), $35 (ebook). ISBNs 978-1-4214-0483-7, 978-1-4214-0484-4, 978-1-4214-0608-4.

Sklar, Jessica K. and Elizabeth S. Sklar (eds.), *Mathematics in Popular Culture: Essays on Appearances in Film, Fiction, Games, Television and Other Media*, Jefferson, NC: McFarland & Company, 2012; ix + 345 pp, $45(P), $45(ebook). ISBNs 978-0-7864-4978-1, 978-0-7864-8994-7.

Polster and Ross distill their collection of more than 800 "math movies" that they have seen. Their book has the attraction of including many movie stills and mathematical figures, which draw the reader in; and their website http://www.qedcat.com contains a database with summaries, dialog, and links to clips. Some chapters are built around a single movie, while others are thematic (infinity, the fourth dimension, problems and puzzles). Sklar and Sklar's book features many movies but also goes beyond movies, with two dozen essays by various authors that broach TV (brief discussion of an episode of *Numb3rs*), cartoons (http://www.xkcd.com), role-playing games (Dungeons and Dragons), the *New York Times* crossword puzzle (!), the plays *Arcadia* and *Proof*, and even Tolstoy's integration metaphor in *War and Peace*. Music—even Tom Lehrer—does not appear at all. Alex Kasman's essay provides a much-needed guide to the varieties of mathematics available in fiction: "plausible," "bogus," "verisimilar," "applied," "surreal," and "meta-mathematical," with some categories more likely than others to attract readers to become interested in mathematics itself (e.g., forget "bogus").

Conway, John H., On unsettleable arithmetical problems, *American Mathematical Monthly* 120 (3) (March 2013), 192–198. Reprinted in *The Best Writing on Mathematics 2014*, edited by Mircea Pitici, 39–48. Princeton University Press, 2014; $24.95(P). ISBN 978-0-691-16417-5.

Did you too miss this gem from almost two years ago? I must have seen it but overlooked it, hence am glad for its reprinting in Pitici's annual collection. Conway explores the possibility that the Collatz $3n + 1$ Conjecture and its relatives may be "un*set*tleable," a term he uses for statements that are true but unprovable from *set* theory. He then metatheorizes: Maybe it is true that the assertion of true-but-unprovable is itself not provable. He cites a 40-year-old example of his of a "Collatzian game" that *is* unsettleable, and he considers in detail an even simpler problem that he *believes* to be unsettleable. As for the Collatz Conjecture itself, Conway is convinced that it is *very likely* to be unsettleable. We are thus yet further "unsettled" and rebuffed from Hilbert's optimistic rallying cry, "Wir müssen wissen—wir werden wissen!" ("We must know—we shall know!"); that is the epitaph on his gravestone, proclaimed in his retirement address in 1930—one day after Gödel announced his incompleteness theorem.

Bliss, Karen M., Kathleen R. Fowler, and Benjamin J. Galluzzo, *Math Modeling: Getting Started & Getting Solutions* and *Math Modeling: Getting Started & Getting Solutions: Connections to Common Core State Standards*, with *Math Modeling Reference Cards*, Society for Industrial and Applied Mathematics, 2014; 68 pp + 15 pp + 6 pp, $15. Free PDF download at http://m3challenge.siam.org/about/mm/.

This guide for students and teachers of mathematical modeling features modeling an amount of plastic recycled, the growth of an epidemic, and the comparative thrills of roller coasters. The modeling process is elaborated into steps: defining the problem statement, making assumptions, defining variables, building solutions, analysis and model assessment, and reporting results, with a reference card for each step with explicit advice. The derivative enters briefly in connection with a varying-rate disease model (with two pages on Euler's method). Also included is the text of the 2013 problem of Moody's Mega Math Challenge and a winning solution paper (but the authors and their school are not identified). (The Moody's Foundation provided funding for the booklets.) The smaller booklet quotes specific Common Core standards in math, English, and science for each of the steps in modeling. The guide is suitable at both high school and college levels and would be useful for both instructors and students. The booklets are informative, very attractively designed, and should be disseminated widely—so send the link for free PDF download to a mathematics teacher whom you know.

ACKNOWLEDGMENTS

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

Adkins, William A., *Louisiana State University, Baton Rouge, LA*
Allgower, Eugene, *Colorado State University, Fort Collins, CO*
Atanasov, Risto, *Western Carolina University, Cullowhee, NC*
Barbeau, Edward J. Jr, *University of Toronto, Toronto, Ontario, Canada*
Beals, Richard, *Yale University, New Haven, CT*
Beezer, Robert A., *University of Puget Sound, Tacoma, WA*
Beier, Julie, *Earlham College, Richmond, IN*
Beineke, Lowell, *Indiana University–Purdue University Fort Wayne, Fort Wayne, IN*
Berkove, Ethan, *Lafayette College, Easton, PA*
Boardman, Michael, *Pacific University, Forest Grove, OR*
Brookfield, Gary, *California State University, Los Angeles, Los Angeles, CA*
Brown, Jason I., *Dalhousie University, Halifax, Nova Scotia, Canada*
Bryant, Lance, *Shippensburg University, Shippensburg, PA*
Bush, Michael, *Washington and Lee University, Lexingon, VA*
Butler, Steve, *Iowa State University, Ames, IA*
Callan, C. David, *University of Wisconsin–Madison, Madison, WI*
Case, Jeremy S., *Taylor University, Upland, IN*
Chakerian, G. Don, *University of California, Davis, Davis, CA*
Chinn, Phyllis, *Humboldt State University, Arcata, CA*
Daly, Dan, *Southeast Missouri State University, Cape Girardeau, MO*
Davidson, Mark G., *Lousiana State University, Baton Rouge, LA*
Dodge, Clayton W., *University of Maine, Orono, ME*
Dorée, Suzanne, *Augsburg College, Minneapolis, MN*
Egge, Eric, *Carleton College, Northfield, MN*
Fisher, Evan, *Lafayette College, Easton, PA*
Fisher, J. Chris, *University of Regina, Regina, Saskatchewan, Canada*
Frederickson, Greg N., *Purdue University, W. Lafayette, IN*
Gelman, Andrew, *Columbia University, New York, NY*
Glickman, Mark, *Boston University School of Public Health, Boston, MA*
Grosshans, Frank D., *West Chester University of Pennsylvania, West Chester, PA*
Guenther, Ronald B., *Oregon State University, Corvallis, OR*
Hall, Rachel Wells, *Saint Joseph's University, Philadelphia, PA*
Harper, James D., *Central Washington University, Ellensburg, WA*
Hartmann, Frederick, *Villanova University, Villanova, PA*
Henle, James, *Smith College, Northampton, MA*
Hill, Theodore P., *California Polytechnic State University, San Luis Obispo, CA*
Isaksen, Dan, *Wayne State University, Detroit, MI*
Janke, Steven, *Colorado College, Colorado Springs, CO*
Johnson, Stewart D., *Williams College, Williamstown, MA*
Kerckhove, Michael G., *Univeristy of Richmond, Richmond, VA*
Knapp, Michael P., *Loyola University Maryland, Baltimore, MD*
Koch, Richard, *University of Oregon, Eugene, OR*
Larson, Loren C., *Saint Olaf College, Northfield, MN*
Levi, Mark, *Pensylvania State University, State College, PA*
Lynch, Mark, *Millsaps College, Jackson, MS*
Markov, Lubomir P., *Barry University, Miami Shores, FL*

Math. Mag. **87** (2014) 406–407. doi:10.4169/math.mag.87.5.406. © Mathematical Association of America

McCleary, John, *Vassar College, Poughkeepsie, NY*
Miller, Steven J., *Williams College, Williamstown, MA*
Mills, Mark A., *Central College, Pella, IA*
Morrison, Kent E., *California Polytechnic State University, San Luis Obispo, CA*
Murray, Will, *California State University, Long Beach, Long Beach, CA*
Oxman, Victor, *The Western Galilee College, Acre, Israel*
Pak, Igor, *University of California, Los Angeles, Los Angeles, CA*
Pasles, Paul C., *Villanova University, Villanova, PA*
Pence, Dennis, *Western Michigan University, Kalamazoo, MI*
Reid, Michael, *University of Central Florida, Orlando, Fl*
Reiner, Vic, *University of Minnesota, Minneapolis, MN*
Rickey, V. Frederick, *United States Military Academy, West Point, NY*
Robins, Sinai, *Nanyang Technical University, Singapore*
Rogers, Edwin L., *Siena College, Loudonville, NY*
Rogers, Michael, *Emory University, Oxford College, Atlanta, GA*
Sauerberg, Jim, *Saint Mary's College of California, Moraga, CA*
Seggev, Itai, *Wolfram Research, Champaign, IL*
Sellers, James A., *The Pennsylvania State University, State College, PA*
Spivey, Michael Z., *University of Puget Sound, Tacoma, WA*
Tangora, Martin, *University of Illinois at Chicago, Chicago, IL*
Trevino, Enrique, *Lake Forest College, Lake Forest, IL*
Warrington, Gregory, *University of Vermont, Burlington, VT*

We thank also the referees whose participation has been entirely via the Editorial Manager system, and who are not listed above. They will be recognized more fully on another occasion.

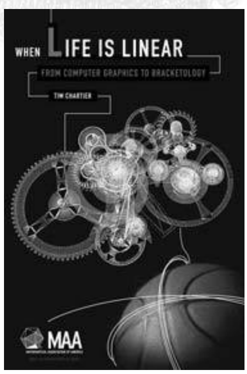
This is the last issue of the MAGAZINE *prepared with the assistance of Don DeLand, who has been the compositor for the MAA print journals since 2001. He has been copyeditor, typesetter, artist, and manager; whatever you have liked about the appearance of the journals has been due to his work. We appreciate his expertise and his extraordinary dedication to the* MAGAZINE.

In view of the wider availability of search tools and databases, the MAA journals are no longer printing annual indexes of articles and notes. The annual index for Volume 87 is available as a supplement, in the standard format, at the MAGAZINE'S website and (temporarily) at http://www.mathematicsmagazine.org.

The annual index to the Problems Section appears on page 403.

New from the MAA

When Life is Linear: From Computer Graphics to Bracketology
By Tim Chartier



Catalog Code: NML-45
ISBN: 978-0-88385-649-9
140 pp., Paperbound, 2015
List Price: $50.00
Member Price: $40.00
Series: Anneli Lax NML

Tim Chartier has written the perfect supplement to a linear algebra course. Every major topic is driven by applications, such as computer graphics, cryptography, webpage ranking, sports ranking and data mining. Anyone reading this book will have a clear understanding of the power and scope of linear algebra.

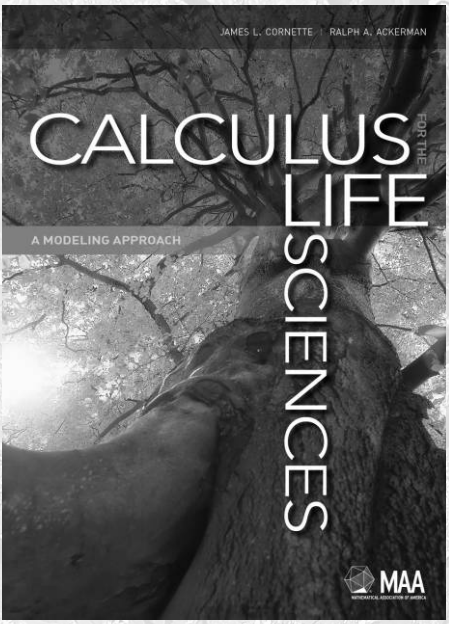
—Arthur Benjamin, Harvey Mudd College

From simulating complex phenomenon on supercomputers to storing the coordinates needed in modern 3D printing, data is a huge and growing part of our world. A major tool to manipulate and study this data is linear algebra. This book introduces concepts of matrix algebra with an emphasis on application, particularly in the fields of computer graphics and data mining. Readers will learn to make an image transparent, compress an image and rotate a 3D wireframe model. In data mining, readers will use linear algebra to read zip codes on envelopes and encrypt sensitive information. The books details methods behind web search, utilized by such companies as Google, and algorithms for sports ranking which have been applied to creating brackets for March Madness and predict outcomes in FIFA World Cup soccer. The book can serve as its own resource or to supplement a course on linear algebra.

To order, visit maa-store.hostedbywebstore.com or call 800-331-1622.

MAA 100
MATHEMATICAL ASSOCIATION OF AMERICA
CELEBRATING A CENTURY OF ADVANCING MATHEMATICS

Save on textbooks
in the MAA eBooks Store



Calculus
for the Life Sciences
A Modeling Approach

James L. Cornette and
Ralph A. Ackerman

MAA Textbooks

Freshman and sophomore life sciences students respond well to the modeling approach to calculus, difference equations, and differential equations presented in this book. Examples of population dynamics, pharmacokinetics, and biologically relevant physical processes are introduced in Chapter 1, and these and other life sciences topics are developed throughout the text.

The ultimate goal of calculus for many life sciences students primarily involves modeling living systems with difference and differential equations. Understanding the concepts of derivative and integral is crucial, but the ability to compute a large array of derivatives and integrals is of secondary importance.

The students should have studied algebra, geometry and trigonometry, but may be life sciences students because they have not enjoyed their previous mathematics courses. This text can help them understand the relevance and importance of mathematics to their world. It is not a simplistic approach, however, and indeed is written with the belief that the mathematical depth of a course in calculus for the life sciences should be comparable to that of the traditional course for physics and engineering students.

eISBN: 978-1-61444-615-6
2015, 731 pp

Price: $35.00

To order, visit maa.org/ebooks/CLS

MAA 100
MATHEMATICAL ASSOCIATION OF AMERICA
CELEBRATING A CENTURY OF ADVANCING MATHEMATICS

MAA
MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth St., NW
Washington, DC 20036

CONTENTS