
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:

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).

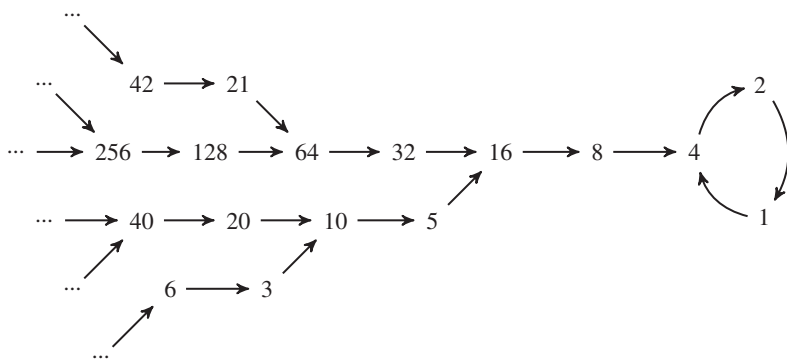


Figure 1 Digraph generated by the $3x + 1$ sequences

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, \dots . This is a *trivial cycle*. Sequences that end in trivial cycles are *trivial sequences*, such as 5, 15, 10, 5, 5, 5, \dots , or $-143, 39, -52, -13, -13, -13, \dots$. 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., $-17, 7, -5, 2, -3, -1, -2, \dots$, 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, \dots ; 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 EOOEEOEOEOEO... Our first subprime Fibonacci sequence had shape EOOEEOEOEOEO... 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 OOOOEOOOEOOOEOOE 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 ...

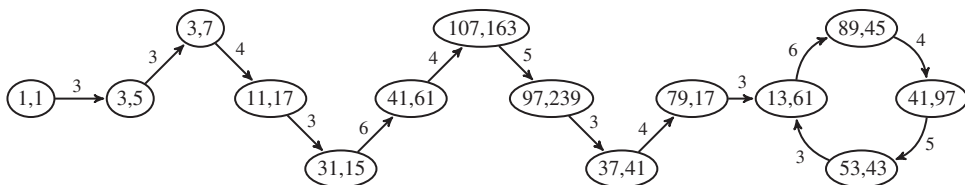


Figure 2 Path generated by the 0, 1 sequence

Taking the sequence’s shape, we can treat each substring of the form $O \dots OE$ 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 $\dots, -83, 109, 13, 61, 37, 49, 43, 46$. 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.

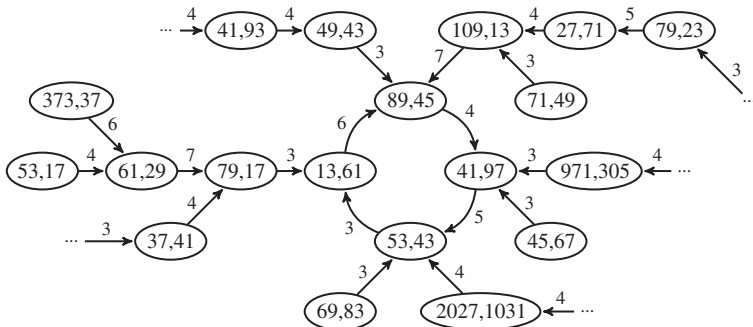


Figure 3 Some paths leading to the 18-cycle

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

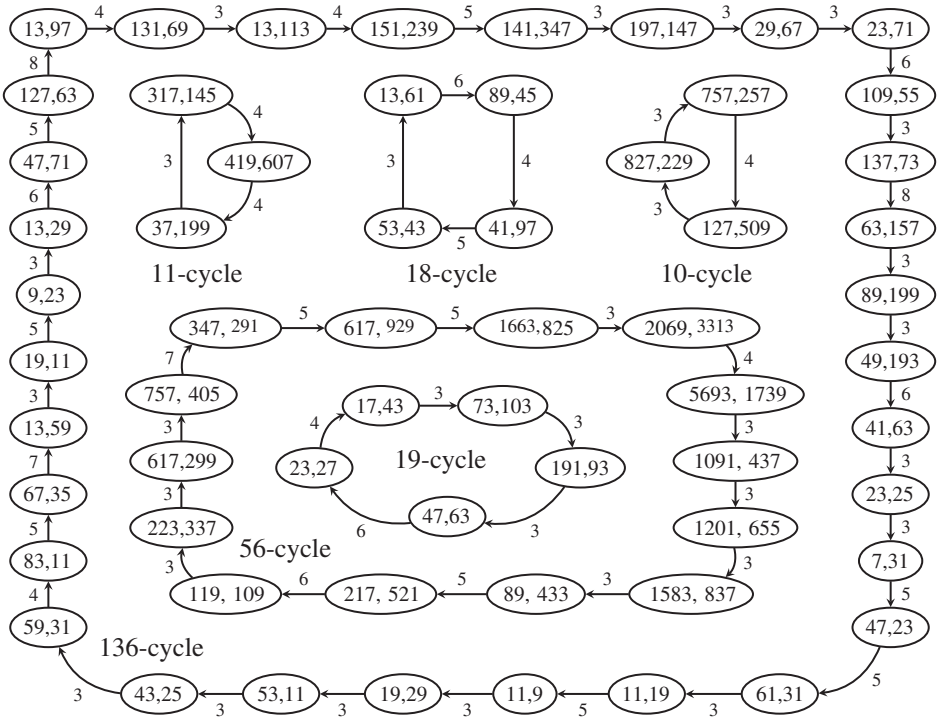


Figure 4 Digraphs of the six known non-trivial cycles

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 $a, a \dots$ 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 (OOEOEOOOE) 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,

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

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

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

By definition, $s_3, \dots, s_m = 2$, 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 OOEEOE.*

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, \dots, 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

$$a + 2^{k-1}J_0d = a, \quad a + 2^{k-2}J_1d = a + 2^{k-2}d, \quad \dots, \quad a + 2^1J_{k-2}d \quad (2)$$

are all odd since a and d are odd by definition, while the last (k th) term $a + 2^0J_{k-1}d$ 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):

$$\begin{aligned} a_1, a_1 + 2^{k_1-2}d_1, a_1 + 2^{k_1-3}d_1, \dots, a_1 + 2J_{k_1-2}d_1, a_1 + J_{k_1-1}d_1 \\ a_2, a_2 + 2^{k_2-2}d_2, a_2 + 2^{k_2-3}d_2, \dots, a_2 + 2J_{k_2-2}d_2, a_2 + J_{k_2-1}d_2 \end{aligned}$$

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

$$\frac{2a_1 + J_{k_1}d_1}{p_2} \quad \text{and} \quad \frac{(p_2 + 2)a_1 + (J_{k_1-1}p_2 + J_{k_1})d_1}{p_2q_2},$$

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

$$\begin{aligned} p_2a_2 &= 2a_1 + J_{k_1}d_1 \\ p_2q_2(a_2 + 2^{k_1-2}d_2) &= (p_2 + 2)a_1 + (J_{k_1-1}p_2 + J_{k_1})d_1 \\ p_1a_1 &= 2a_2 + J_{k_2}d_2 \\ p_1q_1(a_1 + 2^{k_2-2}d_1) &= (p_1 + 2)a_2 + (J_{k_2-1}p_1 + J_{k_2})d_2 \end{aligned}$$

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

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

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

$$\begin{aligned} 2^{k_1+k_2-4}p_1q_1p_2q_2 &= J_{k_1-1}J_{k_2-1}p_1p_2 + J_{k_1}J_{k_2-2}p_1q_2 + J_{k_1-2}J_{k_2}q_1p_2 \\ &+ J_{k_1-1}J_{k_2-1}q_1q_2 + J_{k_1}J_{k_2-1}p_1 + J_{k_1-1}J_{k_2}q_1 \\ &+ J_{k_1-1}J_{k_2}p_2 + J_{k_1}J_{k_2-1}q_2 + J_{k_1}J_{k_2} - (-1)^{k_1+k_2-4}. \end{aligned}$$

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 OOEEOEOOE 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

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

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

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

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 $p_1, q_1, \dots, p_n, q_n$ is 1, and that $p_1, q_1, \dots, p_n, q_n$ 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:

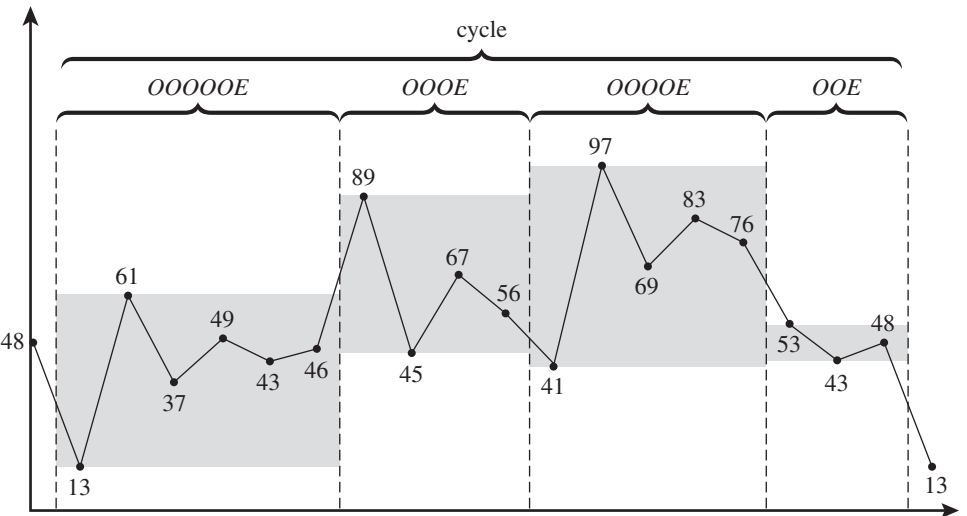


Figure 5 18-cycle with runs labeled and local run bounds shaded

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 $p_1, q_1, \dots, p_n, q_n$ do not equal 1.*

PROPOSITION 6. *At least two of $p_1, q_1, \dots, p_n, q_n$ 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, \dots, 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 OOEEOE 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, \dots, 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

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

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. Andalaro, 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 unshiftable 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.

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

*Contributed by Joseph Tonien,
University of Wollongong, Australia*