Conway's subprime Fibonacci sequences

Richard K. Guy, Tanya Khovanova, Julian Salazar

May 2, 2014

In memory of Martin Gardner

Abstract

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 behaviour 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.

1 Introduction

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 $3x + 1$ problem (if a number is odd, triple it and add one; if even, halve it), with which they seem to have something in common, and of which Erdős has said, "Mathematics is not yet ready for such problems."

Start with 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 a **cycle** of 18 terms. If we start with $1, 1$ or $1, 2$ we clearly 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 $13, 11$, or ....

One might suspect that every such sequence enters this 18-cycle, just as it is conjectured that every $3x + 1$ sequence enters the 3-cycle 1, 4, 2, 1, 4, 2, 1, ... After all, since our sequences are bounded or unbounded they must either end in a cycle or increase indefinitely.

We do not believe the latter happens, and though we have no proof we provide a heuristic argument in §3. 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**; e.g., 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 are of opposite sign or include a zero, then they bound further terms until two consecutive terms of the same sign appear, e.g., −17, 7, −5, 2, −3, −1, −2, …, after which the sign remains constant.

Next, note that two terms of opposite parity are followed by an odd term, and that two odd terms may be followed by an even or an odd term, depending on whether their sum is a multiple of 4 or an odd multiple of 2. Finally, one can have an arbitrarily long string of even terms, but it must terminate since the power of two in consecutive terms must decrease steadily, e.g., 128, 160, 144, 152, 148, 150, 149, …, and once we have an odd term (unless this is a trivial sequence), the even terms are isolated with each followed by at least two odd terms. Therefore, *we need only concern ourselves with sequences of positive terms, comprised of these '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 $EOOEOOEOOEOO\ldots$. Our sample subprime Fibonacci sequence has shape $EOOEOOEOOOEOO\ldots$; the 'extra' odd term here arises when 2 is a composite's smallest prime factor such that dividing by it gives an odd number. In the example, starting at 13, 61 inclusive gives the shape $OOOOEOOOEOOOOEOOE$ that repeats with the cycle (though for cycles, the starting point doesn't matter; $OOOEOOOOEOOEOOOOOE$ is just as valid as a shape for the cycle).

2 Nodes and other cycles

To get a 'sense' for the original sequence, one could plot its trajectories on a directed graph. This approach is quite favored for the $3x + 1$ problem [2], but because of the recurrence we must be purposeful in defining vertices for this sequence. We introduce two important terms:

- The **nodes** of a sequence are the ordered pairs of two positive coprime odd integers not preceded by an odd term.
- **Runs** are the strings of odd terms and a terminating even generated by a node, such that each node corresponds to the first two terms of a run.

In §3 we will see that at some point a non-trivial sequence becomes composed of runs. Here is the same 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 …

As already discussed, we can treat each substring of $O \ldots OE$ as a unit, starting when the first two terms of such a substring are coprime and are not preceded by an odd. The corresponding terms comprise a run, and the first two terms of the run comprise a node. Let us now construct our first sequence graph (Figure 1). For notational convenience we weight the digraph by assigning to each arc the length of the run corresponding to the node at the arc's tail.

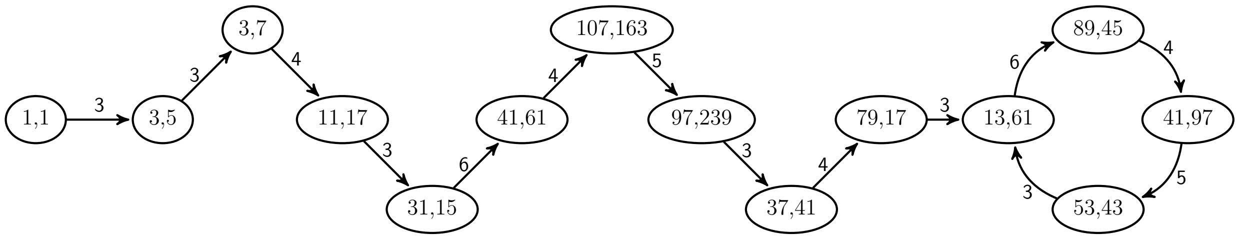

Figure 1: Digraph generated by the 0,1 sequence

One reason a digraph is a useful representation is that many nodes can be **direct predecessors** to a single node. If a node is a predecessor (not necessarily direct) to another node or cycle, we say it is **tributary** to the node or cycle. Some examples are shown in Figure 2.

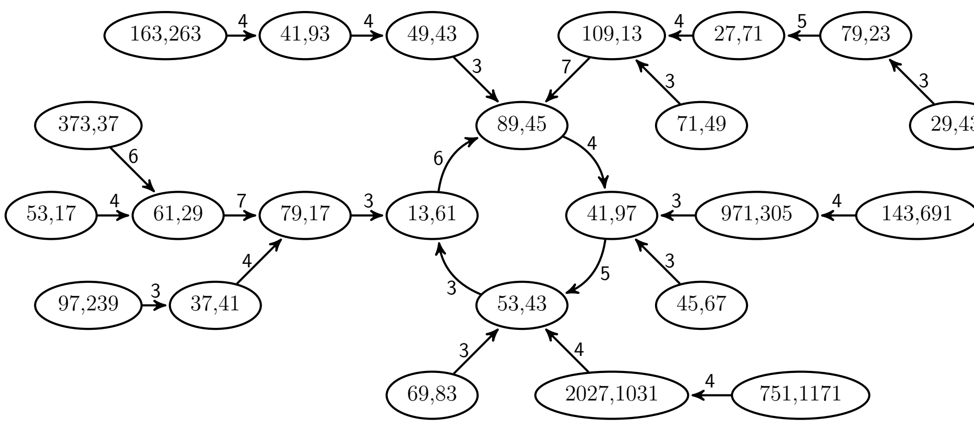

Figure 2: Some tributaries to the 18-cycle

Can we enumerate a node's direct predecessors? Yes, but with some work. For example, take the node $(89, 45)$ of the 18-cycle. The preceding even term t must satisfy $t + 89 = 45q$, where q is 1 or 3 ($q = 2$ makes t odd, prime $q > 3$ would follow 89 with $15q$ instead), which gives $t = -44$ or 46, so the node must always be proceeded by 46. Let the odd term (by definition) before t 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 $r = 49$. In fact, none of the other possible values of s work since they would make $r \leq 0$. In fact, we can keep assuming each prior step used a division of two, which gives $\ldots$, -83, 109, 13, 61, 37, 49, 43, 46. Thus $(109, 13)$, $(13, 61)$, $(61, 37)$, $(37, 49)$, and $(49, 43)$ are exactly the direct predecessors. This process for constructively generating the graph has thus far discouraged a graph theory approach to analysis.

Do sequences all finish up in the 18-cycle we have already seen, similar to how we suspect all $3x + 1$ sequences end in $1, 4, 2, 1, 4, 2, 1, \ldots$? 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 would then be predecessors to the node $(23, 27)$.

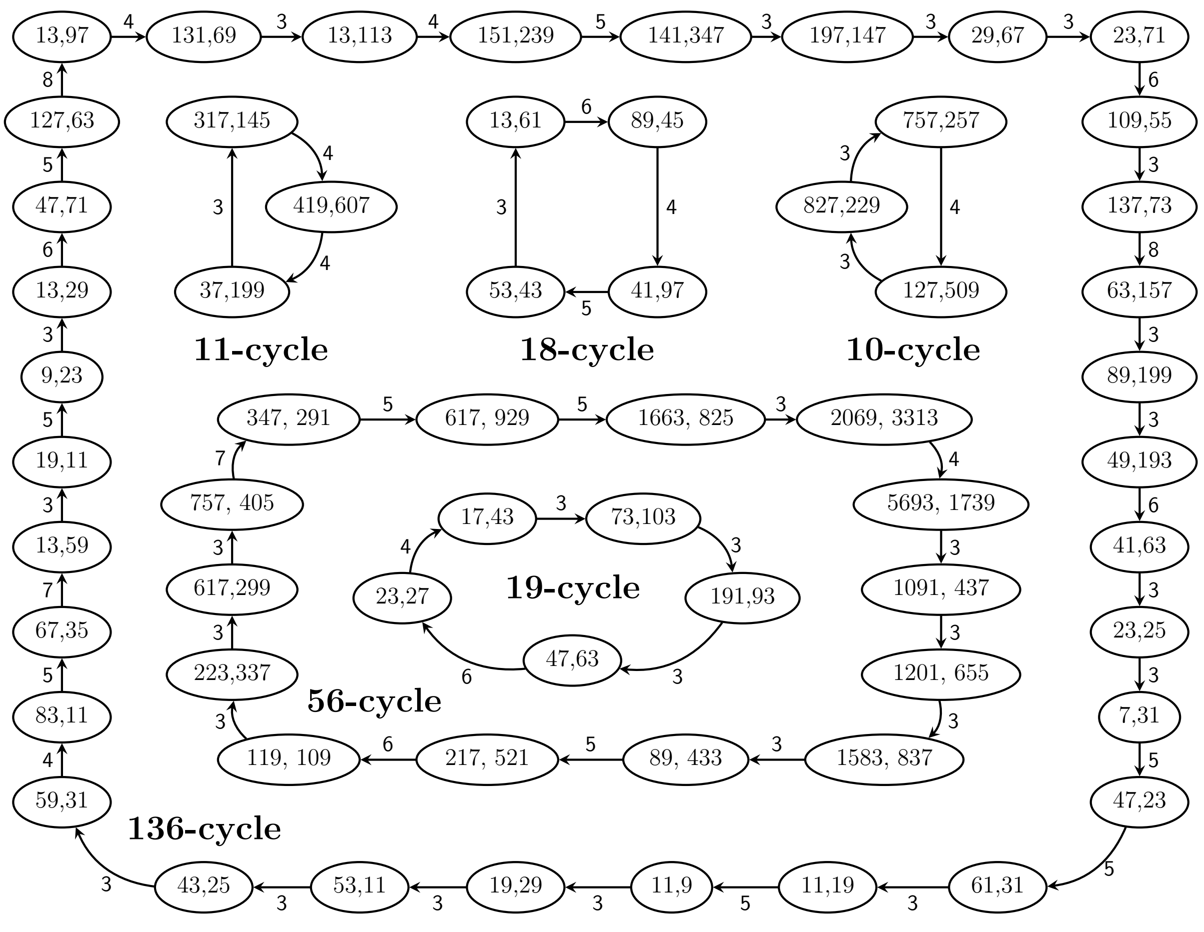

Figure 3: Digraphs of the six known non-trivial cycles

If you start with $5, 13$ you will enter a 136-cycle through node $(47, 23)$ (though simpler starting terms like $1, 4$ will suffice). If you start with $5, 23$ you will enter a 56-cycle through node $(119, 109)$ with 5693 as its largest term. The four cycles we have seen so far were found by hand, and it was not until we used a computer that we found two shorter cycles: an 11-cycle generated by the node $(37, 199)$ and a 10-cycle generated by the node $(127, 509)$. We checked sequences that start with two numbers 1,000,000 or below and found no other non-trivial cycles.

In Table 1 the headings indicate the range for the first two terms of the sequence and the values are how often each cycle occurs. The distribution of non-trivial cycles is quite stable with regard to the starting conditions, suggesting that the number of predecessors to a node is fairly arbitrary and not determined by the cycle it is tributary to, such that the number of starting conditions tributary to a non-trivial cycle increases at a roughly equal rate for each cycle. 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.

Cycle length	$1..10, 1..10$	$1..10^2, 1..10^2$	$1..10^3, 1..10^3$	$1..10^4, 1..10^4$	$1..10^5, 1..10^5$
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

Table 1: Distribution of cycles based on ranges for first two terms

It should be clear why non-trivial sequences exhibit such pseudo-random behavior, in their terms and their digraphs, as their definition involves prime factorization. However, another source of apparent randomness seems to be the conditional iteration itself, as with $3x + 1$ 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 arise. This makes certain questions seem intractable, specifically those regarding the growth and end conditions of a sequence given its starting conditions.

3 End conditions

A sequence must either end in a trivial cycle, 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:

The Fibonacci sequence has shape $EOOEOOEOOEOO\ldots$, but for subprime sequences, when we add two odds there is an even chance that after dividing by 2 the result is odd, a quarter of a chance the next one is odd as well, and so on. We would expect a typical sequence to have three times as many odds as evens, so that in twelve consecutive terms we would divide by 2 six times on average and that among the other six steps the odd sum is divisible by 3 around one-third of the time. We expect to have divided by $2^6 \cdot 3^2 = 576$ on average, while the usual Fibonacci growth would have been only by a factor $\phi^{12} \approx 322$, where $\phi = (1 + \sqrt{5})/2$ is the golden ratio (the limit of the ratios of two successive Fibonacci terms). With this argument we have not considered larger primes than 3 or the rarity of primes among large numbers.

Considering that starting terms $5, 23$ produce a term as large as 5693, the difficulty of proving the non-existence of divergent sequences seems comparable in difficulty to the analogous problem for the $3x + 1$ sequences. However, this does not prevent the deduction of results like the following:

Proposition 1. *A non-trivial sequence contains infinitely many primes (not necessarily distinct), each greater than their two preceding terms.*

Proof. If after some point the sequence does not contain primes, 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 2. *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. Indeed, $\gcd(a, b) = \gcd(b, a + b)$, and so $\gcd(b, \frac{a+b}{p}) \leq \gcd(a, b)$. 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 3. *If the two starting terms of a sequence are coprime, the sequence is non-trivial.*

These results 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).

Another consequence of Proposition 2 and Corollary 3 is a simplification of the search for non-trivial cycles. Since all starting conditions a, b with $\gcd(a, b) > 1$ become trivial or reach $\gcd(a, b) = 1$ for consecutive terms, it is sufficient to study starting conditions with $\gcd(a, b) = 1$ to enumerate all non-trivial end conditions. This, combined with the memorization of nodes and a lookup table for primes could give a more efficient search method than testing all positive integer ordered pairs.

4 The general system

We devote the rest of our paper to the cycles that non-trivial sequences generate. By the definition of a run, we can conclude that a non-trivial cycle must consist of a concatenation of runs. It follows from Proposition 2 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 **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.

It is possible that signature terms are not bounded, as 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 above. Using Proposition 1, this result on signatures follows:

Corollary 4. *Let a, b, c be consecutive terms of a sequence of a cycle. Then $c > \max(a, b)$ if and only if $c = a + b$, i.e., the corresponding signature value of c must be 1 (and c is thus prime). Thus, the largest term of a cycle must be prime with a signature value of 1.*

With the terms and the signature of a cycle, we can establish a homogeneous linear system. Let $t_1, \ldots, t_m$ be the terms of the cycle and $s_1, \ldots, s_m$ be the corresponding signature. Then $\ldots,\ t_{m-1} + t_m = s_1 t_1,\ t_m + t_1 = s_2 t_2,\ t_1 + t_2 = s_3 t_3,\ \ldots,\ t_{i-2} + t_{i-1} = s_i t_i,\ \ldots.$ In matrix form,

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

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

Theorem 5. *No two cycles have the same signature.*

Proof. This is equivalent to showing that a potential signature $s_1, \ldots, s_m$ defines at most one cycle. Given a potential signature, consider solutions for $t_1, \ldots, t_m$ 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 (Corollary 4). □

Theorem 6. *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 equations of (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 $2(t_1 + t_2) = s_1 t_1 + s_2 t_2$. Let t_1 be 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 7. *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 Proposition 6 does not generalize to helping find cycles of more than one run. In this regard, we instead look to Theorem 5, because it means that results on signatures are necessarily results on cycles, which makes it desirable to relate signature terms to one another 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 is to find 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 method where the only signature terms of interest are those corresponding to the node of each run (divisors not equal to 2).

5 The run-centric system

We begin by relating a run's terms with its node:

Theorem 8. *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 and J_n is the n-th Jacobsthal number. 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 but the two node terms. Thus k denotes run length. The Jacobsthal numbers are defined by the recurrence $J_n = J_{n-1} + 2J_{n-2}$, where $J_0 = 0$, $J_1 = 1$ (see A001045 in 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.

We conclude that the first $k - 1$ members

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

are all odd since a and d are odd by definition, while the last (k-th) 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 all our terms are positive. □

Corollary 9. *The terms of the run defined by (a, b) are bounded by a, b, and terms after the node are bounded by $b, \frac{a+b}{2}$. In general, two consecutive run terms bound the rest of the run.*

We refer to Theorem 8 for a more run-based system for a cycle. Write out two runs in the style of (2):

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

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.

When we bend the two runs into a cycle, the divisors of the first run will be denoted by p_1 and q_1. Like with (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 in the four unknowns 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

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

When writing out this expression or expressions for more runs, since entries of q_i appear twice there will be 2^n terms for each possible product of p_is and q_is (excluding the product of all p_i, q_i) where n is the number of q_i. For example, there are 2 terms of q_1, 2 other terms of $q_1 p_2$, and $2^2 = 4$ terms of $q_1 q_2$. Fortunately, the identity $2^{k-1} - J_k = J_{k-1}$ lets us combine these multiple terms into one in each instance. In the $n = 2$ case, the 4 terms involving $q_1 q_2$ simplify to one term:

$$\begin{aligned}
&- J_{k_1}J_{k_2}q_1q_2 + 2^{k_2-1}J_{k_1}q_1q_2 + 2^{k_1-1}J_{k_1}q_1q_2 - 2^{k_1+k_2-2}q_1q_2 \\
&\qquad = -\left(2^{k_1-1} - J_{k_1}\right)\left(2^{k_2-1} - J_{k_2}\right)q_1q_2 = -J_{k_1-1}J_{k_2-1}q_1q_2.
\end{aligned}$$

Using the identity means (3) reduces to:

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

Applying similar manipulation for 3-run cycles gives:

$$\begin{gathered}
2^{k_1+k_2+k_3-6}p_1q_1p_2q_2p_3q_3 = \\
J_{k_1-1}J_{k_2-1}J_{k_3-1}p_1p_2p_3 + J_{k_1-1}J_{k_2}J_{k_3-2}p_1p_2q_3 + J_{k_1}J_{k_2-2}J_{k_3-1}p_1q_2p_3 + J_{k_1}J_{k_2-1}J_{k_3-2}p_1q_2q_3 + \\
J_{k_1-2}J_{k_2-1}J_{k_3}q_1p_2p_3 + J_{k_1-2}J_{k_2}J_{k_3-1}q_1p_2q_3 + J_{k_1-1}J_{k_2-2}J_{k_3}q_1q_2p_3 + J_{k_1-1}J_{k_2-1}J_{k_3-1}q_1q_2q_3 + \\
J_{k_1-1}J_{k_2-1}J_{k_3}p_2p_3 + J_{k_1-1}J_{k_2}J_{k_3-1}p_2q_3 + J_{k_1}J_{k_2-2}J_{k_3}q_2p_3 + J_{k_1}J_{k_2-1}J_{k_3-1}q_2q_3 + \\
J_{k_1}J_{k_2-1}J_{k_3-1}p_3p_1 + J_{k_1-1}J_{k_2-1}J_{k_3}p_3q_1 + J_{k_1}J_{k_2}J_{k_3-2}q_3p_1 + J_{k_1-1}J_{k_2}J_{k_3-1}q_3q_1 + \\
J_{k_1-1}J_{k_2}J_{k_3-1}p_1p_2 + J_{k_1}J_{k_2-1}J_{k_3-1}p_1q_2 + J_{k_1-2}J_{k_2}J_{k_3}q_1p_2 + J_{k_1-1}J_{k_2-1}J_{k_3}q_1q_2 + \\
J_{k_1}J_{k_2}J_{k_3-1}p_1 + J_{k_1-1}J_{k_2}J_{k_3}p_2 + J_{k_1}J_{k_2-1}J_{k_3}p_3 + \\
J_{k_1-1}J_{k_2}J_{k_3}q_1 + J_{k_1}J_{k_2-1}J_{k_3}q_2 + J_{k_1}J_{k_2}J_{k_3-1}q_3 + \\
J_{k_1}J_{k_2}J_{k_3} - (-1)^{k_1+k_2+k_3-6}.
\end{gathered}$$

Is there a method to the madness? Yes. Let's move back the subscripts of p cyclically ($p_1 \to p_3$, $p_2 \to p_1$, $p_3 \to p_2$) to see the pattern. Here are some terms after this shifting:

$$\begin{aligned}
\cdots + J_{k_1-2}J_{k_2}J_{k_3-1}q_1p_1q_2 + J_{k_1-1}J_{k_2-2}J_{k_3}q_1q_2p_2 + J_{k_1-1}J_{k_2-1}J_{k_3-1}q_1q_2q_3 + \\
J_{k_1-1}J_{k_2-1}J_{k_3}p_1p_2 + J_{k_1-1}J_{k_2}J_{k_3-1}p_1q_3 + J_{k_1}J_{k_2-2}J_{k_3}q_2p_2 + J_{k_1}J_{k_2-1}J_{k_3-1}q_2q_3 + \cdots .
\end{aligned}$$

The subscripts of the Jacobsthal numbers are now k_i, k_i-1 or k_i-2, corresponding to 0, 1 or 2 occurrences of the subscript i on the shifted p_i and q_i in the term. Thus, for unshifted p this holds for p_{i-1}, q_i, where p_0 represents p_n.

We can now generalize of this expression for n-run cycles. Here is the determinant for 3-run cycles (with original p_i):

$$\begin{vmatrix}
2 & J_{k_1} & -p_2 & 0 & 0 & 0 \\
1 & J_{k_1-1} & -q_2+1 & -2^{k_2-2}q_2 & 0 & 0 \\
0 & 0 & 2 & J_{k_2} & -p_3 & 0 \\
0 & 0 & 1 & J_{k_2-1} & -q_3+1 & -2^{k_3-2}q_3 \\
-p_1 & 0 & 0 & 0 & 2 & J_{k_3} \\
-q_1+1 & -2^{k_1-2}q_1 & 0 & 0 & 1 & J_{k_3-1}
\end{vmatrix} = 0. \tag{4}$$

We can see that in the $2n \times 2n$ generalization of (4) there are just $7n$ nonzero entries, three in each odd-numbered row, four in each even-numbered row. Every two rows repeat, but with the items shifted right two columns (for the last two rows, we 'wrap around' back to the first two columns) and the relevant subscripts incremented.

Notice that the p_i occur only in the odd-numbered columns (or rows), whereas the q_i occur in all the columns (i.e., twice in the even-numbered rows). In the expansion there is just one term of degree $2n$ in p_i, q_i; it is

$$2^{k_1+k_2+\cdots+k_n-2n}p_1p_2\cdots p_nq_1q_2\cdots q_n.$$

This is the only term that takes the product of consecutive divisors, e.g., p_1q_1, p_2q_2. To see why, without loss of generality suppose we have a term with the product p_2q_2. Let the matrix in (4) be (a_{ij}). Then the product $a_{13}a_{24} = (-p_2)(-2^{k_2-2}q_2)$ was taken. The term taken from the next row is necessarily $a_{35} = -p_3$, and so on until we get the above term.

The remaining terms contain the product

$$p_1^{\delta_1}p_2^{\delta_2}\cdots p_n^{\delta_n}q_1^{\delta_{n+1}}q_2^{\delta_{n+2}}\cdots q_n^{\delta_{2n}}$$

where the δ_i are 0 or 1, with their nonzero values ranging over all subsets of $\{1, 2, \ldots, 2n\}$ with cardinality less than or equal to n. A further restriction is that δ_i and δ_{i+n} cannot both be 1 for $i \leq n$ since terms with $p_i q_i$ are excluded.

With the relationship between the Jacobsthal and divisor subscripts, and letting p_0 refer to p_n, the equation for n-run cycles may be written as

$$\prod_{i=1}^{n} 2^{\sum(k_i-2)} p_{i-1}q_i = \sum_{\substack{\delta_1,\ldots,\delta_{2n}\in\{0,1\},\\ \sum\delta_j\leq n,\\ \forall j\leq n:\ \delta_j+\delta_{j+n}<2}} \prod_{i=1}^{n} J_{k_i-\delta_i-\delta_{i+n}} p_{i-1}^{\delta_i} q_i^{\delta_{i+n}} - (-1)^{\sum(k_i-2)} \tag{5}$$

where the absolute term ($\sum\delta_j = 0$) arises from 2^n products of n Jacobsthal numbers with n 2 s or 1 s, which combine into a single term $(-1)^{\sum(k_i-2)}$ in the same fashion as already illustrated, together with the product $-J_{k_1}J_{k_2}\cdots J_{k_n}$ through the $+1$ s in the $q_i + 1$ entries. Note that the equation holds for $n \geq 2$.

The input for this formula only requires the number of runs in the cycle 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 $m = \sum k_i = 10$. 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 is preserved.

6 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 $p_1, q_1, 2, 2, \ldots, p_n, q_n, 2, 2, \ldots$ 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 (the cycle's divisors). As already established, $n \geq 2$.

We were able to disprove one-run cycles and relate the divisors of n-run cycles to one another given the run configuration. But even with a given two-run configuration, we are left with a relation in 4 unknown variables p_1, q_1, p_2, q_2. 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 term vs. index for the 10-cycle (Figure 4), which is composed of three runs:

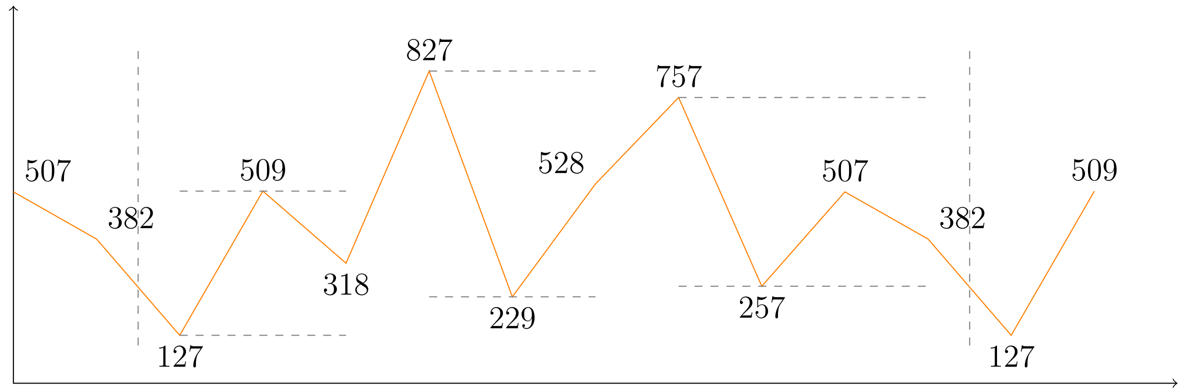

Figure 4: 10-cycle; vertical dashes define the cycle, horizontal dashes illustrate run bounds

The bounds are a consequence of Corollary 9. Remembering Corollary 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, we provide three stronger results:

Proposition 10. *At least two of $p_1, q_1, \ldots, p_n, q_n$ do not equal 1.*

Proof. Without loss of generality, let all but p_2 or q_2 be 1.

Case 1. $p_2 \neq 1$. Then b_1 and the rest of the first run are $> a_1$, so $a_2 > 0$ and $b_2 > a_1$. Since all other $p_i, q_i = 1$, all terms before a_1 are $> \frac{1}{2}a_1$. Then $a_1 > a_1$, a contradiction.

Case 2. $q_2 \neq 1$. Then b_1 and the rest of the first run are $> a_1$, so $a_2 > 2a_1$ and $b_2 > 0$, and $\frac{a_2+b_2}{2} > a_1$. Since all other $p_i, q_i = 1$, all terms before a_1 are $> \frac{1}{2}a_1$. Then $a_1 > a_1$, a contradiction. □

Proposition 11. *At least two of $p_1, q_1, \ldots, p_n, q_n$ equal 1.*

Proof. Without loss of generality, let exactly one of p_1, q_1 be 1 (the rest are ≥ 3). Then the corresponding term a_1 or b_1 is prime and the largest term of the cycle.

Case 1. $p_1 = 1$. Then $b_1 < \frac{2}{3}a_1$ and the rest of the run is $< \frac{5}{6}a_1$. Then $a_2 < \frac{5}{9}a_1$, $b_2 < \frac{25}{54}a_1$, and the rest of the second run is $< \frac{55}{108}a_1$. Since all further p_i,q_i (if any) are also ≥ 3, they do not increase the maximum. We iterate our bounding to get $b_1 < \frac{163}{324}a_1 < \frac{5}{9}a_1$ and the rest of the run is $< \frac{7}{9}a_1$. Then $a_2 < \frac{14}{27}a_1$, $b_2 < \frac{35}{81}a_1$, and the rest of the terms before a_1 are $< \frac{77}{162}a_1 < \frac{1}{2}a_1$. Then $a_1 < a_1$, a contradiction.

Case 2. $q_1 = 1$. The first run is $< b_1$ except for b_1 itself. Then $a_2 < \frac{2}{3}b_1$, $b_2 < \frac{5}{9}b_1$, and the rest of the second run is $< \frac{11}{18}b_1$. Since all further p_i,q_i (if any) are also ≥ 3, they do not increase the maximum. Then $a_1 < \frac{11}{27}b_1 < \frac{1}{2}b_1$. We iterate our bounding to see that $\frac{a_1+b_1}{2} < \frac{3}{4}b_1$ and that the rest of the run is $< \frac{7}{8}b_1$. Then $a_2 < \frac{7}{12}b_1$, $b_2 < \frac{35}{72}b_1$, and the rest of the terms before a_1 are $< \frac{77}{144}b_1$. Then $a_1 < \frac{77}{216}b_1$ and $b_1 < \frac{385}{432}b_1 < b_1$, a contradiction. □

Proposition 12. *If there are only two $p_1, q_1, \ldots, p_n, q_n$ that equal 1, the two cannot be of the form p_i, q_i unless $(p_{i+1}, q_{i+1}) = (3, 3)$, $(3, 5)$, or $(5, 3)$.*

Proof. Without loss of generality, let $p_1, q_1 = 1$ and all further $p_i, q_i \geq 3$. Then b_1 is the largest term, with $b_1 < 2a_1$, $\frac{a_1+b_1}{2} < \frac{3}{2}a_1$, and the rest of the run is $< \frac{7}{4}a_1$. Then $a_2 < \frac{7}{4p_2}a_1$, $b_2 < \frac{7(p_2+1)}{4p_2q_2}a_1$. Since all further p_i, q_i (if any) are ≥ 3 then all subsequent terms before

a_1 are $< \max\left(\frac{7}{4p_2}a_1, \frac{7(p_2+1)}{4p_2q_2}a_1\right)$. Hence $a_1 < \max\left(\frac{7}{2p_2}a_1, \frac{7(r+1)}{2p_2q_2}a_1\right)$, a contradiction unless $(p_2, q_2) = (3, 3)$, $(3, 5)$, or $(5, 3)$. ☐

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 that

Conjecture 13. *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, and eliminating the three exceptions of Proposition 12. 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 5 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.

7 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 ((1) and (5))
- Each signature corresponds to a unique potential cycle (Theorem 5)
- Non-existence of one-run cycles (Theorem 6)
- Restrictions on possible signatures (Propositions 10 to 12)

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

Theorem 14. *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 either (3) with run configuration $(k_1, k_2) = (3, 3)$ or the more general (1) for $n = 6$, 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. There are only four cases:

Case 1. $p_1, q_1 = 1$ (equivalent to $p_2, q_2 = 1$). Then $4p_2q_2 = 7(p_2 + q_2 + 2)$. Since we only want solutions over the odd primes, then exactly one of p_2 and q_2 is 7 and $(p_2, q_2) = (3, 7), (7, 3)$. These solutions fail by Proposition 12.

Case 2. $p_1, p_2 = 1$. Then $q_1q_2 = 2q_1 + 2q_2 + 5 \Rightarrow (q_1 - 2)(q_2 - 2) = 9$. The solutions over the odd primes are $(q_1, q_2) = (3, 11), (5, 5), (11, 3)$, though reordering the runs shows $(3, 11)$ and $(11, 3)$ are equivalent.

Case 3. $q_1, q_2 = 1$. Then $p_1p_2 = 2p_1 + 2p_2 + 5 \Rightarrow (p_1, p_2) = (3, 11), (5, 5), (11, 3)$, though reordering the runs shows $(3, 11)$ and $(11, 3)$ are equivalent.

Case 4. $p_1, q_2 = 1$ (equivalent to $p_2, q_1 = 1$). Then $q_1p_2 = 4q_1 + 4p_2 + 17 \Rightarrow (q_1 - 4)(p_2 - 4) = 33$. The solutions over the odd primes are $(q_1, p_2) = (5, 37), (37, 5)$.

Substitute the signature values into (1) and solve the system. Since the solution space is 1-dimensional, we can express all the cycle terms $t_1, \ldots, t_n$ in terms of t_1 (even better: let $t_1 = 1$), and then multiply by the common denominator to get the unique cycle candidate:

(p_1, q_1, p_2, q_2)	t_2/t_1	t_3/t_1	t_4/t_1	t_5/t_1	t_6/t_1	Cycle candidate	It should be...
$(1, 3, 1, 11)$	3/5	4/5	7/5	1/5	4/5	$5, 3, 4, 7, 1, 4$	$4, 7, 11$, not $4, 7, 1$
$(1, 5, 1, 5)$	1/3	2/3	1	1/3	2/3	$3, 1, 2, 3, 1, 2$	$2, 3, 5$, not $2, 3, 1$
$(3, 1, 11, 1)$	19/9	14/9	1/3	17/9	10/9	$9, 19, 14, 3, 17, 10$	$19, 14, 11$, not $19, 14, 3$
$(5, 1, 5, 1)$	3	2	1	3	2	$1, 3, 2, 1, 3, 2$	$3, 2, 5$, not $3, 2, 1$
$(1, 5, 37, 1)$	11/41	26/41	1/41	27/41	14/41	$41, 11, 26, 1, 27, 14$	$11, 26, 37$, not $11, 26, 1$
$(1, 37, 5, 1)$	1/27	14/27	1/9	17/27	10/27	$27, 1, 14, 3, 17, 10$	$10, 27, 37$, not $10, 27, 1$

Table 2: Candidates for a 6-cycle

The last entry also fails because $1, 14$ should be followed by 5, not 3, and because the largest term is not prime. Since all candidates fail, the theorem is proved. □

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.

There are also two lemmas that can simplify things:

Lemma 15. *The smallest term in a non-trivial cycle must be a node term (and thus odd), and at least 7.*

Proof. Since node terms bound a run's terms, the smallest number of the cycle must also be one of its node terms, which are odd by definition.

Dividing a composite number by its smallest prime factor never produces 1. If 3 is the smallest cycle term, the previous members a, b must add to 3, 6 or 9. Since the same integer cannot be separated by only one term (the sequence a, b, a continues into trivial cycle $a, a, \ldots$), it follows that the smallest number in the cycle is less than the two preceding members. Hence, a and b are greater than 3, giving the two cases $(a, b) = (5, 4), (4, 5)$ which are tributary to, but not part of, non-trivial cycles.

If 5 occurs in a sequence, the previous members a, b must add to 5, 10, 15, or 25, and if they are to be greater than 5, $(a, b) = (6, 9), (7, 8), (8, 7), (9, 6), (6, 19), \ldots (19, 6)$ (this list would be deduced by the first half of the 'direct predecessor' method of §2). Calculation shows these are tributary to, but not part of, non-trivial cycles.

7 is the smallest member of the 136-cycle, completing the proof. □

Note that this result immediately disqualifies all the 6-cycle candidates. Also, using lower bound arguments omitted here, we can eliminate the three exceptions of Proposition 12 for all two-run cycles:

Lemma 16. *In a two-run cycle, p_1, q_1 cannot both be 1 and p_2, q_2 cannot both be 1.*

Thus for cycles of longer length, we can apply Theorem 14's method of generating candidates and easily show why they fail. For 7-cycles, the only possible run configuration is $(3, 4)$; for 8-cycles, the two possible configurations are $(3, 5)$ and $(4, 4)$. At the time of writing, we have confirmed that:

Theorem 17. *There are no cycles of length 8 or less. There are no two-run cycles of length 11 or less.*

8 Open questions

1. Give an algorithm for finding solutions $\{p_i, q_i\}$ to (5) given a run configuration $(k_1, \ldots, k_n)$, at least for $n = 3$.
2. Determine if and how divisors and terms are bounded based on the number of runs.
3. Find more non-trivial cycles, or prove that no more exist.
4. Prove that subprime Fibonacci sequences do not increase indefinitely.
5. Explore the subprime variants of other linear recurrences.

References

[1] *The On-Line Encyclopedia of Integer Sequences*, published electronically at `http://oeis.org`, 2012, Sequence A001045

[2] Jeffrey C. Lagarias, *The $3x + 1$ problem and its generalizations*, Amer. Math. Monthly **92** (1985), 3–23.