ON CONWAY'S SUBPRIME FUNCTION, A COVERING OF $\mathbb{N}$ AND AN UNEXPECTED APPEARANCE OF THE GOLDEN RATIO

MIHAI CARAGIU, PAUL A. VICOL, AND MOHAMMAD ZAKI

Abstract. The *subprime* function, introduced by Conway ($s(n) = n$ if n is prime, otherwise $s(n) = n/p$ where p is the least prime factor of n) was used to design interesting analogues of the Fibonacci sequence that are conjectured to be ultimately periodic. In the present article we prove that the binary operation $\circ$ on $\mathbb{N}$ defined by $x \circ y = s(x + y)$ induces a magma structure that is cyclic with generator 1, i.e., $\mathbb{N} = \langle 1 \rangle$. Moreover, if we consider, in this context, the sequence of subsets of natural numbers $\{C_n\}_{n \geq 0}$ defined by $C_0 = \{1\}$ and $C_{n+1} = C_n \cup (C_n \circ C_n)$ thereafter, we provide computational evidence to the effect that $\lim_{n\to\infty} \frac{|C_{n+1}|}{|C_n|} = \frac{1+\sqrt{5}}{2}$, thus providing an unexpected appearance of the Golden ratio φ.

1. Introduction

Conway's *subprime* function s is defined [9, 8] as follows. If n is a prime number, we set $s(n) = n$. If n is composite, define $s(n) = n/p$ where p is the least prime factor of n. We may set $s(1) = 1$ to complete the definition of an arithmetic function $s : \mathbb{N} \to \mathbb{N}$.

The function s is instrumental in the definition of *subprime Fibonacci* sequences, i.e. integer sequences $\{x_k\}_{k\geq 0}$ satisfying the recursion $x_k = s(x_{k-1} + x_{k-2})$. Sequences in this class are conjectured to be ultimately periodic regardless of the choice of the initial *seed* (x_0, x_1). Among the cycle lengths discovered so far [5] are 1, 10, 11, 18, 19, 56, 136. The interesting phenomenon of ultimate periodicity suggests an analogy between the subprime Fibonacci recursion and the perennial $3x + 1$ conjecture [4, 6, 11]. Sequences similar to the subprime Fibonacci ones that involve the greatest prime factor function gpf instead of s were proved to be ultimately periodic with the (unique) limit cycle $(7, 3, 5, 2)$ [1]. A conjecture of ultimate periodicity for general prime sequences of the form $x_k = P(c_1x_{k-1} + c_2x_{k-2} + \cdots + c_dx_{k-d})$ was also formulated [1].

In the present paper we will consider a shift in the usage of Conway's subprime function s, from a study of the sequences arising from s to the algebraic structures arising from s. Towards this end we will introduce a binary operation $\circ$ on the set of positive integers $\mathbb{N} = \{1, 2, 3, \ldots\}$ defined as follows:

$$x \circ y := s(x + y), \tag{1}$$

for any $x, y \in \mathbb{N}$.

The operation (1) displays no immediate, *textbook* properties (it is commutative, not associative, and has no identity element), which makes $(\mathbb{N}, \circ)$ just a seemingly amorphous *magma* [3]. However a closer look suggests an interesting property stated in Theorem 1 below.

Let C_0 be a set of integers. Consider the sequence $\{C_n\}_{n\geq 0}$ defined by

$$C_{n+1} = C_n \cup (C_n \circ C_n), \tag{2}$$

where we used the notation $A \circ B := \{a \circ b | a \in A \text{ and } b \in B\}$ for $A, B \subseteq \mathbb{N}$.

We define

$$\langle C_0 \rangle := \bigcup_{n=0}^{\infty} C_n. \tag{3}$$

We have the following interesting result.

Theorem 1. $(\mathbb{N}, \circ)$ *is a cyclic structure generated by 1, i.e.,*

$$\mathbb{N} = \langle 1 \rangle. \tag{4}$$

Corollary 2. *Let $C_0 = \{1\}$. Then using (3) and (4) we have*

$$\bigcup_{n=0}^{\infty} C_n = \mathbb{N}. \tag{5}$$

We will prove Theorem 1 in Section 3. In the next section we discuss a surprising conjecture that emerged from Theorem 1.

2. A Conjecture Involving the Golden Ratio

An analysis of the cardinalities of the sets C_n satisfying (2) starting with $C_0 = \{1\}$ is presented in the Table 1.

Table 1: *Consecutive quotients for the sequence $\{C_n\}_{n\geq 0}$*

n	$\lvert C_n\rvert$	$\lvert C_n\rvert/\lvert C_{n-1}\rvert$	n	$\lvert C_n\rvert$	$\lvert C_n\rvert/\lvert C_{n-1}\rvert$
1	2	2.000000000	17	1429	1.620181406
2	3	1.500000000	18	2299	1.608817355
3	4	1.333333333	19	3705	1.611570248
4	6	1.500000000	20	5961	1.608906883
5	8	1.333333333	21	9615	1.612984399
6	12	1.500000000	22	15524	1.614560582
7	18	1.500000000	23	25057	1.614081422
8	25	1.388888889	24	40442	1.614000080
9	38	1.520000000	25	65247	1.613347510
10	56	1.473684211	26	105412	1.615583858
11	89	1.589285714	27	170224	1.614844610
12	138	1.550561798	28	274963	1.615301015
13	218	1.579710145	29	444156	1.615330063
14	342	1.568807339	30	717551	1.615538234
15	547	1.599415205	31	1159406	1.615782014
16	882	1.612431444	32	1873356	1.615789465

The computations were done by using a Julia program running in a Google cloud environment and were fairly intensive [2]. Table 1 suggests an unexpected relationship between the subprime function s, defined in terms of primes and the golden ratio $\varphi = \frac{1+\sqrt{5}}{2}$. This is summarized in the following conjecture.

Conjecture 3. $\varphi = \lim_{n\to\infty} \frac{|C_{n+1}|}{|C_n|}$.

The last computed quotient satisfies:

$$\frac{|C_{32}|}{|C_{31}|} - \varphi = -0.00224453\ldots.$$

A few words about the programming method that was used: first, preliminary calculations done with MATLAB were used to calculate (fairly quickly) the sets C_n for $n \leq 17$, together with their maximum elements. The maximum element of C_{17} turns out to be 2143. Since at every step of the iteration the maximal element of C_{n+1} is less than twice the maximal element of C_n we estimated that the elements of C_{32} (our target at the moment) are necessarily less than $2143 \cdot 2^{15} = 70221824$.

Moving on to the Julia program, for the purpose of computing the necessary subprime functions, we first generated and stored a list of all primes up to 200 million (the last one being 199999991), so that, if possible, we could advance past C_{32} (we didn't, though, due to the large computing time for C_{32}).

The Julia programming strategy was fairly brute: at each level $n > 1$ of the computation, we used a split of the set C_n into *old* elements (those in C_{n-1}) and *new* elements (those in $C_n \setminus C_{n-1}$), so that instead of applying (2), the calculation of the subsequent set C_{n+1} was done by appending to C_n all possible products of *new* elements, as well as all possible products between an *old* element and a *new* element (in either order). This is because the products of *old* elements are already present in C_n. Afterward, C_{n+1} splits itself into its own *old* and *new* elements and the computation proceeds as shown before. This is only a slight improvement compared to a brute-force application of (2). We believe it would be nice to have a parallel version of this algorithm: it would run faster, but it would still be exponential.

If the conjecture is true, it would provide an unexpected computation of φ with sets defined by primes. The conjecture would also add meaningful context to a problem (Conway's subprime Fibonacci recursion [5, 9]) of a $3x + 1$-type [1, 11].

3. Proof of $\mathbb{N} = \langle 1 \rangle$

For the proof of Theorem 1 we will need an effective result concerning the distribution of primes in small intervals. For our immediate purposes, the following proposition, due to Nagura [7] (who follows Ramanujan's method of proving Bertrand's postulate) would be sufficient.

Proposition 4. *For $x \geq 8$ there exists at least one prime p with*

$$x < p < 3x/2.$$

Alternatively, this ensures the existence of a prime in the interval $(2y/3, y)$ for every $y \geq 12$.

The proof of Theorem 1, presented in Section 2, is based on a basic result involving the distribution of primes in short intervals.

Proof of Theorem 1. Let $\{C_n\}_{n\geq 0}$ be the ascending sequence of subsets of $\mathbb{N}$ satisfying the recursion (3) starting with $C_0 = \{1\}$. Since $1 \in C_n$ for all $n \geq 0$, we may define $A_n \subseteq C_n$ to be the maximal initial segment of $\mathbb{N} = \{1, 2, 3, \ldots\}$ included in C_n. Let $c_n := |C_n|$ and $a_n := |A_n|$. Clearly $a_n \leq c_n$ and $A_n \subseteq A_{n+1}$ for $n \geq 0$ for all n. Theorem 1 would follow once we establish that

$$\lim_{n\to\infty} a_n = \infty. \tag{6}$$

A preliminary calculation shows the values

$$(a_0, a_1, a_2, a_3, a_4, a_5, a_6) = (1, 2, 3, 3, 5, 7, 9) \tag{7}$$

for $a_0, a_1, \ldots, a_6$, respectively. In order to prove (6), we will estimate the rate of growth of the sequence $\{a_n\}_{n\geq 0}$.

From the form of the recurrence (3), and since $A_n = \{1, 2, \ldots, a_n\} \subseteq C_n$, it follows that the set of all possible products of elements in the initial segment A_n,

$$A_n \circ A_n = \{x \circ y | 1 \leq x, y \leq a_n\} \tag{8}$$

is a subset of C_{n+1}. We will estimate the size of the maximum initial segment of $\mathbb{N}$ included in the set of products (8).

Lemma 5. *If p is any prime with*
$$p < 2a_n, \tag{9}$$
then
$$p \in A_n \circ A_n. \tag{10}$$
Proof of Lemma 5. Clearly (10) is vacuous for $n = 0$ (we refer to the preliminary list (7)). Let $n \geq 1$. If $a_n < p < 2a_n$, p can be written as a sum $p = x + y$ with $x, y \in A_n = \{1, 2, \ldots, a_n\}$. Then, from (1) we have
$$x \circ y = s(x + y) = s(p) = p \in A_n \circ A_n.$$
If $p \leq a_n$ then $p \in A_n$ and $p = s(p+p) = p \circ p \in A_n \circ A_n$. This concludes the proof of Lemma 5.

Let q_n be the largest prime p satisfying (9). If $n \geq 5$ (which, according to the preliminary calculation (7), would imply $2a_n > 12$) setting $y := 2a_n$ in Proposition 4 demonstrates that that the largest prime q_n less than $2a_n$ (necessarily odd) must satisfy the estimate
$$4a_n/3 < q_n < 2a_n. \tag{11}$$
Note that $q_n \in A_n \circ A_n \subseteq C_{n+1}$. Then any product of the form
$$(2k + 1) \circ q_n \tag{12}$$
where $1 \leq 2k+1 \leq a_n$ (i.e., $2k+1$ is an odd element of A_n) must necessarily be in $A_n \circ C_{n+1} \subseteq C_n \circ C_{n+1} \subseteq C_{n+1} \circ C_{n+1} \subseteq C_{n+2}$, according to (3). From the form of the subprime function s, since $2k + 1 + q_n$ is even, an element in the list (12) can be written as follows:
$$(2k + 1) \circ q_n = s(2k + 1 + q_n) = \frac{2k + 1 + q_n}{2} = k + \frac{q_n + 1}{2}. \tag{13}$$
We thus found a series (13) of $\left\lceil a_n/2 \right\rceil$ consecutive elements of C_{n+2} starting from $(q_n + 1)/2$ upwards.

Since $q_n < 2a_n$ or, equivalently, $(q_n + 1)/2 \leq a_n$, (13) represents a extension of the initial segment $A_n = \{1, 2, \ldots, a_n\}$ up to
$$\frac{q_n + a_n - \varepsilon_n}{2},$$
where $\varepsilon_n = 1-(a_n \bmod 2)$. Using the estimate (11) for the prime q_n, we find that (13) actually represents an initial segment of C_{n+2} of size bounded from below by
$$\frac{4a_n/3 + a_n - 1}{2} = \frac{7a_n - 3}{6},$$
and consequently, starting from $n = 5$ (corresponding to $a_5 = 7$) we have
$$a_{n+2} \geq \frac{7a_n - 3}{6}. \tag{14}$$
From (14) and the fact that the sequence $\{a_n\}_{n\geq 0}$ is non-decreasing, (6) follows. This concludes the proof of Theorem 1 and consequently the validity of (5).

Acknowledgement

The authors would like to thank the anonymous referee for very helpful, insightful and constructive comments.

References

[1] G. Back and M. Caragiu, *The greatest prime factor and recurrent sequences*, The Fibonacci Quarterly, **48.4** (2010), 358–362.
[2] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman, *Julia: a fast dynamic language for technical computing*, (2012) arXiv:1209.5145v1.
[3] M. Caragiu and G. Back, *The greatest prime factor and related magmas*, JP J. of Algebra, Number Theory and Appl., **15** (2009), 127–136.
[4] J. H. Conway, *On unsettleable arithmetical problems*, Amer. Math. Monthly, **120.3** (2013), 192–198.
[5] R. K. Guy, T. Khovanova, and J. Salazar, *Conway's subprime Fibonacci sequences*, Mathematics Magazine, **87.5** (2014), 323–337.
[6] J. C. Lagarias, editor, *The Ultimate Challenge: The 3x+1 Problem*, American Mathematical Society, Providence, RI, 2010.
[7] J. Nagura, *On the interval containing at least one prime number*, Proceedings of the Japan Academy, **28** (1952), 177–181.
[8] OEIS Foundation Inc. (2011), The On-Line Encyclopedia of Integer Sequences, `http://oeis.org/A117818`.
[9] S. Roberts, *Genius At Play: The Curious Mind of John Horton Conway*, Audiobook, Jennifer Van Dyck (Reader), Bloomsbury Press, 2015.
[10] N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, Ann. Math. Inform., **41** (2013), 219–234.
[11] D. Zeilberger, *Teaching the computer how to discover(!) and then prove(!!) (all by itself(!!!)) analogues of Collatz's notorious 3x+1 conjecture*, J. Difference Equ. Appl., **17.3** (2011), 375–386.

MSC2010: 11B39, 11A41, 11N05

Department of Mathematics and Statistics, Ohio Northern University, Ada, OH 45810
E-mail address: `m-caragiu.1@onu.edu`

School of Computing Science, Simon Fraser University, Burnaby BC, V5A 1S6, Canada
E-mail address: `pvicol@sfu.ca`

Department of Mathematics and Statistics, Ohio Northern University, Ada, OH 45810
E-mail address: `m-zaki@onu.edu`