On Subprime Recurrences

(author names suppressed for review)

Abstract. Mathematical legend John Conway invented a sequence that combines elements of Fibonacci numbers and the Collatz problem: after adding the previous two elements, we divide that sum by its least prime factor (but only if that sum is composite). This leads to some very interesting structure, which has yet to receive the widespread attention it deserves. We generalize from Fibonacci numbers to general second order recurrences, and prove a variety of results for such sequences.

The Fibonacci sequence $F_n = F_{n-1} + F_{n-2}$ (with $F_0 = 0, F_1 = 1$) is everybody's favorite recurrence (see, e.g., [**5, 6, 7**]). John Conway invented an interesting variation, the subprime Fibonacci sequence. We include an operation after adding the previous two terms: if the sum is composite, it is divided by its least (positive) prime factor. That is,

$$a_n = \frac{a_{n-1} + a_{n-2}}{B(a_{n-1} + a_{n-2})},$$

where

$$B(x) = \begin{cases} \text{lpf}(x) & x \in \mathbb{X} \\ 1 & \text{otherwise} \end{cases}.$$

Here $\text{lpf}(x)$ denotes the least positive prime factor of x, $\mathbb{X}$ denotes the set of composites $\{n \in \mathbb{Z} : |n| \geq 2 \text{ and } \exists a, b \in \mathbb{P},\ ab|n\}$, and $\mathbb{P}$ denotes the set of positive primes. Conventionally we set $\text{lpf}(x) = 1$ for $|x| \leq 1$, so $B(x) = \text{lpf}(x)$ unless x is prime.

More generally, we can define the Conway subprime function on $\mathbb{Z}$ via

$$C(n) = \frac{n}{B(n)},$$

with which the Conway subprime Fibonacci sequence is $a_n = C(a_{n-1} + a_{n-2})$.

This family of subprime sequences (with various initial conditions) have interesting structure, combining elements of Fibonacci sequences with elements of the famous Collatz $(3x + 1)$ problem. Like with the Collatz problem, long-term behavior remains unknown in general, although substantial progress was made in [**4**]. Also, a group of undergraduate students generalized this to the subprime Tribonacci recurrence in [**2**]. Very few other papers have appeared even close to this area, e.g. [**1**] and [**3**].

We propose to generalize to arbitrary first- and second-order recurrences on $\mathbb{Z}$. This leads to many beautiful sequences and a wealth of interesting and accessible questions.

In particular, we consider the problem of fixing $a_0, a_1, c_1, c_2 \in \mathbb{Z}$, and looking at the subprime sequence $a_n = C(c_1 a_{n-1} + c_2 a_{n-2})$ (for $n \geq 2$). Here the special case of $c_1 = c_2 = 1$ gives the original subprime Fibonacci sequence.

An important step to understanding these sequences is looking at end conditions, such as cycles. For fixed c_1, c_2, we call $[x_0, x_1, \ldots, x_{n-1}]$ a *cycle* or *n-cycle* if the sequence beginning with $a_0 = x_0, a_1 = x_1$ satisfies $a_i = x_{i \pmod n}$ for all $i \geq 0$. In particular, we say that $[x_0]$ is a 1-cycle if $a_0 = a_1 = x_0$ leads to the constant sequence

$a_i = x_0$ for all $i \geq 0$. Note that $[0]$ (or $[0, 0]$, etc.) is always a 1-cycle, regardless of c_1, c_2; we call any such cycle trivial. For convenience, let $\mathbb{Z}^\star$ denote the set of nonzero integers.

We begin by considering cycles of the form $[x]$ and $[x, -x]$.

Theorem 1. *The subprime sequence for any $c_1, c_2 \in \mathbb{Z}$ and any $x \in \mathbb{Z}^\star$ has a 1-cycle $[x]$ exactly when:*

1. $c_1 + c_2 \in \mathbb{P}$ *and lpf*$(x) \geq c_1 + c_2$*; or*
2. $c_1 + c_2 = 1$ *and* $x \notin \mathbb{X}$.

Proof. A nontrivial 1-cycle $[x]$ arises exactly when $x = C(c_1x + c_2x) = x\frac{c_1+c_2}{B(x(c_1+c_2))}$, i.e. exactly when $c_1 + c_2 = B(x(c_1 + c_2))$. In particular, $c_1 + c_2 \in \mathbb{P} \cup \{1\}$. If $c_1 + c_2 \in \mathbb{P}$, then x must have a prime factor (else $B(x(c_1 + c_2)) = 1$), and in fact $\text{lpf}(x) \geq c_1 + c_2$ (else $B(x(c_1 + c_2)) = \text{lpf}(x)$). If instead $c_1 + c_2 = 1$, then $B(x) = 1$, so $x \notin \mathbb{X}$. ■

A similar result holds for the class of 2-cycles $[x, -x]$.

Theorem 2. *The subprime sequence for any $c_1, c_2 \in \mathbb{Z}$ and any $x \in \mathbb{Z}^\star$ has a nontrivial 2-cycle $[x, -x]$ exactly when:*

1. $c_2 - c_1 \in \mathbb{P}$ *and lpf*$(x) \geq c_2 - c_1$*; or*
2. $c_2 - c_1 = 1$ *and* $x \notin \mathbb{X}$.

Proof. A nontrivial 2-cycle $[x, -x]$ arises exactly when $x = C(-c_1x + c_2x) = x\frac{-c_1+c_2}{B(x(-c_1+c_2))}$ (and also $-x = C(c_1x - c_2x) = -x\frac{c_1-c_2}{B(x(c_1-c_2))}$, which is algebraically equivalent). This occurs exactly when $c_2 - c_1 = B(x(c_2 - c_1))$. In particular, $c_2 - c_1 \in \mathbb{P} \cup \{1\}$. If $c_2 - c_1 \in \mathbb{P}$, then x must have a prime factor (else $B(x(-c_1 + c_2)) = 1$), and in fact $\text{lpf}(x) \geq c_2 - c_1$ (else $B(x(-c_1 + c_2)) = \text{lpf}(x)$). If instead $c_2 - c_1 = 1$, then $B(x) = 1$, so $x \notin \mathbb{X}$. ■

It turns out that the first order special case (i.e. $a_n = C(c_1a_{n-1})$ for $n \geq 1$) is completely characterized by these simple cycles. Note that if $c_1 = 0$ we immediately fall into the trivial cycle $[0]$.

Theorem 3. *Consider the first order subprime sequence for any $c_1, a_0 \in \mathbb{Z}^\star$. Then its long-term behavior is determined by c_1, as follows:*

1. *If $c_1 \notin \mathbb{X}$ and $c_1 > 0$, then the sequence will fall into a 1-cycle.*
2. *If $c_1 \notin \mathbb{X}$ and $c_1 < 0$, then the sequence will fall into some $[x, -x]$ cycle.*
3. *If $c_1 \in \mathbb{X}$, then the sequence will diverge.*

Proof. First consider $c_1 = 1$. If $a_0 \notin \mathbb{X}$, then the 1-cycle is immediate by Theorem 1 (taking $c_2 = 0$). Otherwise, $a_n = \frac{a_{n-1}}{B(a_{n-1})}$, so each successive term in the sequence loses the smallest prime factor of the preceding term. This continues until all that is left is ($\pm$) the largest prime factor of a_0, and then the 1-cycle begins.

Now consider $c_1 \in \mathbb{P}$. If $\text{lpf}(a_0) \geq c_1$, then again the 1-cycle is immediate. Otherwise, suppose a_0 has j prime factors (counted with multiplicity) strictly less than c_1. Now $a_1 = \frac{c_1a_0}{B(c_1a_0)} = c_1\frac{a_0}{\text{lpf}(a_0)}$, so a_1 now has $j - 1$ prime factors strictly less than c_1. Continuing in this way we see that $a_j = c_1^j a_0'$, where a_0' is just a_0 with its j smallest prime factors removed. Now $\text{lpf}(a_0') \geq c_1$, so the 1-cycle begins at this point.

Next consider $c_1 = -1$. If $a_0 \notin \mathbb{X}$, then the 2-cycle is immediate by Theorem 2 (taking $c_2 = 0$). Otherwise, $a_n = \frac{-a_{n-1}}{B(-a_{n-1})}$, so each successive term in the sequence alternates sign, and loses the smallest prime factor of the preceding term. This continues until all that is left is ($\pm$) the largest prime factor of a_0, and then the 2-cycle begins.

Now consider $-c_1 \in \mathbb{P}$. If $\text{lpf}(a_0) \geq -c_1$, then again the 2-cycle is immediate. Otherwise, suppose a_0 has j prime factors (counted with multiplicity) strictly less than c_1. Now $a_1 = \frac{c_1 a_0}{B(c_1 a_0)} = c_1 \frac{a_0}{\text{lpf}(a_0)}$, so a_1 now has $j-1$ prime factors strictly less than c_1. Continuing in this way we see that $a_j = c_1^j a_0'$, where a_0' is just a_0 with its j smallest prime factors removed. Now $\text{lpf}(a_0') \geq -c_1$, so the 2-cycle begins at this point.

Lastly, suppose $c_1 \in \mathbb{X}$. Now we have $|a_n| = \left|\frac{c_1 a_{n-1}}{B(c_1 a_{n-1})}\right| \geq \left|\frac{c_1}{\text{lpf}(c_1)} a_{n-1}\right| \geq 2|a_{n-1}|$. Hence $|a_n| \geq 2^n |a_0| \geq 2^n$. ∎

Note that if $c_1 = 0$ the second order sequence is really two interwoven first order sequences: $a_{2n} = C(c_2 a_{2n-2})$ and $a_{2n+1} = C(c_2 a_{2n-1})$, which are each fully explained by Theorem 3. Henceforth we restrict to the case of $c_1, c_2 \in \mathbb{Z}^\star$. We now continue our study of 2-cycles $[x, y]$ by restricting to $x, y \in \mathbb{Z}^\star$ via the following.

Lemma 4. *Suppose the subprime sequence for some $c_1, c_2 \in \mathbb{Z}^\star$ and some $x, y \in \mathbb{Z}$ has a nontrivial 2-cycle $[x, y]$. Then $x, y \in \mathbb{Z}^\star$.*

Proof. Suppose to the contrary that $y = 0$. Then $0 = y = C(c_1 x + c_2 y) = \frac{c_1 x}{B(c_1 x)}$ and hence $c_1 x = 0$. Since $c_1 \neq 0$ by hypothesis, we must have $x = 0$, so the 2-cycle is trivial. The case $x = 0$ is similar. ∎

We now offer a characterization of nontrivial 2-cycles.

Theorem 5. *The subprime sequence for some $c_1, c_2, x, y \in \mathbb{Z}^\star$ has a nontrivial $[x, y]$ 2-cycle, if and only if*

1. $B(c_1 y + c_2 x) = c_1 \frac{y}{x} + c_2$*; and*
2. $B(c_1 x + c_2 y) = c_1 \frac{x}{y} + c_2$.

Proof. Set $b_1 = B(c_1 x + c_2 y)$, $b_2 = B(c_1 y + c_2 x)$ for convenience. Suppose now that we have a $[x, y]$ cycle. Then $x = C(c_1 y + c_2 x) = \frac{c_1 y + c_2 x}{b_2}$ and $y = C(c_1 x + c_2 y) = \frac{c_1 x + c_2 y}{b_1}$. These rearrange to $b_2 = \frac{c_1 y + c_2 x}{x}$ and $b_1 = \frac{c_1 x + c_2 y}{y}$, respectively. ∎

Theorem 5 admits various corollaries.

Corollary 6. *If the subprime sequence for some $c_1, c_2 \in \mathbb{Z}^\star$ admits a nontrivial $[x, y]$ cycle, then it also admits the cycle $[kx, ky]$, for any $k \in \mathbb{Z}$ with*

$$\textit{lpf}(k) \geq \max(c_1 \frac{y}{x} + c_2, c_1 \frac{x}{y} + c_2).$$

These 2-cycles have some interesting number theoretic properties.

Corollary 7. *Suppose the subprime sequence for some $c_1, c_2 \in \mathbb{Z}^\star$ admits a nontrivial $[x, y]$ cycle with $y \neq -x$. Set $x' = \frac{x}{\gcd(x,y)}, y' = \frac{y}{\gcd(x,y)}$. The following must hold:*

1. *Both $c_1 \frac{y}{x} + c_2$ and $c_1 \frac{x}{y} + c_2$ lie in $\mathbb{P} \cup \{1\}$, and are not equal; and*
2. *If $c_1 \frac{y}{x} + c_2 \in \mathbb{P}$, then lpf$(x) \geq c_1 \frac{y}{x} + c_2$ (otherwise $x \notin \mathbb{X}$); and*
3. *If $c_1 \frac{x}{y} + c_2 \in \mathbb{P}$, then lpf$(y) \geq c_1 \frac{x}{y} + c_2$ (otherwise $y \notin \mathbb{X}$); and*

4. $\frac{xy}{\gcd(x,y)^2} = \frac{\text{lcm}(x,y)}{\gcd(x,y)} = x'y'$ *divides* c_1*; and*
5. $|x'| \le |c_1|$ *and* $|y'| \le |c_1|$*; and*
6. $c_2 > 0$*; and*
7. $c_1^2 = (c_2 - B(c_1y + c_2x))(c_2 - B(c_1x + c_2y))$.

Proof. (1) If $c_1\frac{y}{x} + c_2 = c_1\frac{x}{y} + c_2$, then $\frac{y}{x} = \frac{x}{y}$. Hence $x = \pm y$.
(6) $c_2 \ge 0$ follows from combining $\frac{1}{c_1} \le \frac{x}{y} \le c_1$ with $c_1\frac{y}{x} + c_2 \ge 1$. ■

Corollary 8. *Let* $c_1, c_2 \in \mathbb{Z}^\star$*, and set* $S = \{c_2 - n : n \in \mathbb{P} \cup \{1\}\}$*. If* S *does not contain two distinct elements whose product is* c_1^2*, then the subprime sequence with* c_1, c_2 *admits no nontrivial 2-cycles.*

Proof. Corollary 7(7). ■

For example, with $c_1 = 2, c_2 = 3$, we have $S = \{2, 1, 0, -2, -4, \ldots\}$. No two distinct elements of S have product $4 = c_1^2$. Indeed, nontrivial 2-cycles seem rather rare due to these many conditions, but once one is found in a sequence there are infinitely many others due to Corollary 6.

We now demonstrate infinitely many examples of 2-cycles that are neither trivial nor simple.

Proposition 9. *Let* $a, b \in \mathbb{Z}^\star$ *with* $a + b \in \mathbb{P}$*. Let* $c, x \in \mathbb{Z}^\star$ *with* $ac^2 + b \in \mathbb{P}$*, lpf*$(c) \ge a + b$*, and lpf*$(x) \ge \max(a + b, ac^2 + b)$*. Then the subprime sequence with* $c_1 = ac, c_2 = b$ *admits 2-cycle* $[x, cx]$*.*

Proof. We calculate $B(c_1y + c_2x) = B(ac(cx) + bx) = B(x(ac^2 + b)) = ac^2 + b = c_1\frac{y}{x} + c_2$ and $B(c_1x + c_2y) = B(acx + b(cx)) = B(xc(a + b)) = a + b = c_1\frac{x}{y} + c_2$. ■

For example, we could take $a = b = 1, c = x = 2$ in Proposition 9, which gives cycle $[2, 4]$ in subprime sequence $c_1 = 2, c_2 = 1$. However, all of the 2-cycles we found have this structure, and we believe there are no others.

Conjecture 10. *If the subprime sequence for some* $c_1, c_2, x, y \in \mathbb{Z}$ *has a 2-cycle* $[x, y]$*, then one of* x, y *must divide the other.*

Our analysis of the first order case shows that if a cycle arises in a subprime sequence, it need not do so immediately. However, this is not true for the trivial cycle $[0]$. We show that if it does not begin immediately with $a_0 = a_1 = 0$, it never does.

Proposition 11. *Let* $c_1, c_2 \in \mathbb{Z}^\star$*, and let* $a_0, a_1 \in \mathbb{Z}$*, not both zero. Then this subprime sequence will not fall into the trivial cycle* $[0]$*.*

Proof. Suppose otherwise; let n be minimal with $a_n = a_{n+1} = 0$, so $a_{n-1} \ne 0$. Now $a_{n+1} = C(c_1a_n + c_2a_{n-1}) = C(c_2a_{n-1}) \ne 0$, a contradiction. ■

There is much more to learn about cycles. For example, Figure 1 contains a plot of the subprime sequence with $c_1 = 2, c_2 = 5, a_0 = 0, a_1 = 1$. It gets as big as 10932575866748112593 (which is a prime and approximately 1.09×10^{19}), but at $n = 133581$ it falls into a cycle of length 37790.

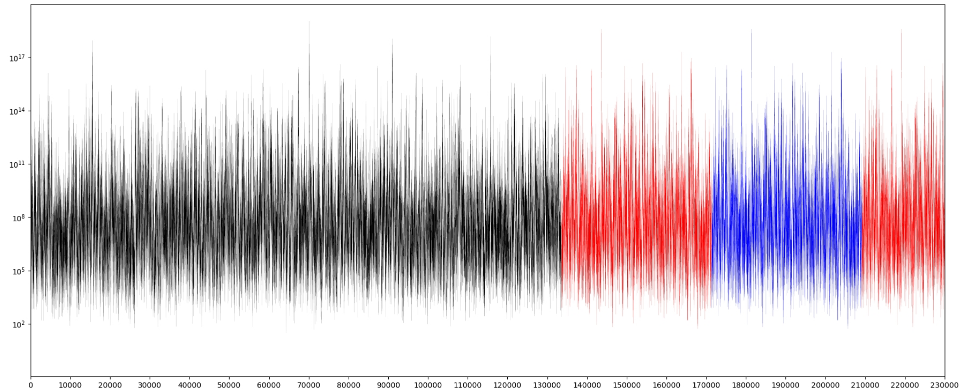

Figure 1. Subprime sequence with $c_1 = 2, c_2 = 5, a_0 = 0, a_1 = 1$

We close with some families of sequences without cycles.

Theorem 12. *Let $c_1, c_2, a_0, a_1 \in \mathbb{N}$ with $\gcd(c_1, c_2) > 1$. Then this subprime sequence diverges.*

Proof. Set $k = \text{lpf}(\gcd(c_1, c_2)), c_1' = \frac{c_1}{\gcd(c_1,c_2)}, c_2' = \frac{c_2}{\gcd(c_1,c_2)}$. We have $a_n = \frac{c_1 a_{n-1} + c_2 a_{n-2}}{B(c_1 a_{n-1} + c_2 a_{n-2})} \geq \frac{\gcd(c_1,c_2)}{k}(c_1' a_{n-1} + c_2' a_{n-2}) \geq c_1' a_{n-1} + c_2' a_{n-2}$. Now the non-subprime recurrence with c_1', c_2', a_0, a_1 is bigger than the Fibonacci sequence F_n (easily proved by induction). Hence $a_n \geq F_n$, which diverges. ■

Theorem 12 can be immediately generalized to allow $c_1, c_2, a_0, a_1 \in \mathbb{Z}^\star$ with $c_1 c_2 > 0$ and $a_0 a_1 > 0$. More complicated are the cases with mixed signs, which sometimes gives sequences that alternate sign (for a while?). There is more to learn here. Our final result shows a familiar sequence that is also a subprime sequence.

Theorem 13. *Let $c_1, a_0, k \in \mathbb{Z}^\star$, and set $a_1 = a_0 k$. Choose $p \in \mathbb{P}$ with $p \leq lpf(a_1)$, and set $c_2 = pk^2 - c_1 k$. This gives a subprime sequence that is also geometric, namely $a_n = k^n a_0$.*

Proof. By strong induction, assume $a_{n-1} = k^{n-1} a_0$ and $a_{n-2} = k^{n-2} a_0$. Now $a_n = C(c_1 a_{n-1} + c_2 a_{n-2}) = C(pk^n a_0) = \frac{pk^n a_0}{B(pk^n a_0)} = \frac{pk^n a_0}{p} = k^n a_0$. ■

REFERENCES

1. Jeremy F. Alm and Taylor Herald. A note on prime Fibonacci sequences. *Fibonacci Quart.*, 54(1):55–58, 2016.
2. Sara Barrows, Emily Noye, Sarah Uttormark, and Matthew Wright. Three's a crowd: an exploration of subprime tribonacci sequences. *College Math. J.*, 54(5):464–475, 2023.
3. Mihai Caragiu, Paul A. Vicol, and Mohammad Zaki. On Conway's subprime function, a covering of $\mathbb{N}$ and an unexpected appearance of the golden ratio. *Fibonacci Quart.*, 55(4):327–331, 2017.
4. Richard K. Guy, Tanya Khovanova, and Julian Salazar. Conway's subprime Fibonacci sequences. *Math. Mag.*, 87(5):323–337, 2014.
5. Mohammad Javaheri and Stephanie Cambrea. The distribution of Fibonacci numbers modulo primes. *Amer. Math. Monthly*, 129(1):75–79, 2022.
6. E. P. Miles, Jr. Generalized Fibonacci numbers and associated matrices. *Amer. Math. Monthly*, 67:745–752, 1960.
7. D. D. Wall. Fibonacci series modulo m. *Amer. Math. Monthly*, 67:525–532, 1960.