A SIMPLE PRIME-GENERATING RECURRENCE

ERIC S. ROWLAND
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
PISCATAWAY, NJ 08854, USA

Abstract. We provide a proof of the conjecture that the sequence of first differences of the solution to $f(n) = f(n-1) + \gcd(n, f(n-1))$ with $f(1) = 7$ consists only of 1s and primes. The limiting behavior of $f(n)/n$ is also studied.

1. Introduction

There is substantial literature on formulas that generate primes. These formulas fall into two general categories:

(1) formulas that were discovered to sometimes generate primes (for example, the Mersenne formula $2^n - 1$). In practice these provide new prime candidates.
(2) formulas that were engineered to always generate primes (for example, Mills' formula and its relatives [2] and prime-valued polynomials [5]). In practice these do not generate any primes at all.

It seems to be quite rare for a formula to always generate primes and yet to be "naturally occurring" in the sense that it was not constructed for this purpose but simply found to do so.

The subject of this article is such a formula — a recurrence, in fact — that was discovered in 2003 at the NKS Summer School[1] [6], at which I was a participant. Primary interest at the summer school is in systems with simple definitions that exhibit complex behavior. In one of Stephen Wolfram's live computer experiments[2], we pursued just that in a class of nested recurrence equation (which turned up some interesting behavior but is another subject entirely). Afterward, Matt Frank (who was one of the instructors) and a few participants performed some additional experiments, somewhat simplifying the structure of the equation and introducing different components. One of the recurrences they looked at was

$$f(n) = f(n-1) + \gcd(n, f(n-1)). \tag{1}$$

They observed that with (for example) the initial condition $f(1) = 7$, the sequence of differences $f(n) - f(n-1) = \gcd(n, f(n-1))$ has an unpredictable character to it [4]. When they presented this result to the rest of the participants, it was realized that, additionally, this difference sequence appears to be composed entirely of 1s and primes:

Date: September 29, 2007.

[1] The NKS Summer School is a three-week program in which participants conduct original research informed by *A New Kind of Science* [8].

[2] This is just what it sounds like: an experiment conducted with a live audience.

1, 1, 1, 5, 3, 1, 1, 1, 1, 11, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 23, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 47, 3, 1, 5, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 101, 3, 1, 1, 7, 1, 1, 1, 1, 11, 3, 1, 1, 1,
1, 1, 13, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 233, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 467, 3, 1, 5, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, . . .

While $f(n)$ certainly has something to do with factorization (due to the gcd), it was not clear why $\gcd(n, f(n-1))$ should never be composite. In the following few days several of us spent some time trying to find a reason or a counterexample, but nothing emerged.

I was sufficiently interested in this phenomenon to come back to it a few times over the years, and in the experiments of one such trip I noticed some structure that led to the contents of this paper, the main result being that (for $f(1) = 7$ anyway) $\gcd(n, f(n-1))$ is always 1 or prime. The proof is elementary; our most useful tool is the fact that $\gcd(n, m)$ divides the linear combination $rn + sm$ for all integers r and s.

At this point the reader may object that the 1s produced by $\gcd(n, f(n-1))$ contradict the previous claim that the recurrence *always* generates primes. If these 1s are deemed too inconvenient, one can use a shortcut at any step to jump directly to the next non-1 gcd: Hindsight reveals that, perhaps rather surprisingly, there is actually some local structure to $f(n)$.

It certainly seems to be the case that, as one iteratively applies this shortcut, one obtains larger and larger primes appearing every so often (after each large gap). However, executing the shortcut requires finding the smallest prime factor of a large integer, so it is not a practical generator of large primes because we must know a prime when we see one in the first place.[3]

Equation 1 is therefore like category 2 above in that it does not magically produce primes, but it remains to be seen whether, like category 1, it can be coerced into a practical producer of candidates.

2. Initial observations

Before presenting the main proof, I think it is important to reveal several features that were discovered experimentally.

For brevity, let $g(n) = \gcd(n, f(n-1))$. Table 1 lists $f(n)$ and $g(n)$ as well as the quantities $\Delta(n) = f(n-1) - n$ and $f(n)/n$, which we now motivate.

[3] As proven in theorem 1, without the shortcut generally a string of $\frac{p-3}{2}$ 1s precedes the prime p in the sequence of gcds (which is essentially performing trial division). Since it is faster to determine the primality of p by other methods (and also faster to factor it), the shortcut does in fact reduce the amount of computation required.

n	$\Delta(n)$	$g(n)$	$f(n)$	$f(n)/n$
1			7	7
2	5	1	8	4
3	5	1	9	3
4	5	1	10	2.5
5	5	5	15	3
6	9	3	18	3
7	11	1	19	2.71429
8	11	1	20	2.5
9	11	1	21	2.33333
10	11	1	22	2.2
11	11	11	33	3
12	21	3	36	3
13	23	1	37	2.84615
14	23	1	38	2.71429
15	23	1	39	2.6
16	23	1	40	2.5
17	23	1	41	2.41176
18	23	1	42	2.33333
19	23	1	43	2.26316
20	23	1	44	2.2
21	23	1	45	2.14286
22	23	1	46	2.09091
23	23	23	69	3
24	45	3	72	3
25	47	1	73	2.92
26	47	1	74	2.84615
27	47	1	75	2.77778
28	47	1	76	2.71429
29	47	1	77	2.65517
30	47	1	78	2.6
31	47	1	79	2.54839
32	47	1	80	2.5
33	47	1	81	2.45455
34	47	1	82	2.41176
35	47	1	83	2.37143
36	47	1	84	2.33333
37	47	1	85	2.2973
38	47	1	86	2.26316
39	47	1	87	2.23077
40	47	1	88	2.2
41	47	1	89	2.17073
42	47	1	90	2.14286
43	47	1	91	2.11628
44	47	1	92	2.09091
45	47	1	93	2.06667
46	47	1	94	2.04348
47	47	47	141	3
48	93	3	144	3
49	95	1	145	2.95918
50	95	5	150	3
51	99	3	153	3
52	101	1	154	2.96154
53	101	1	155	2.92453
54	101	1	156	2.88889
55	101	1	157	2.85455
56	101	1	158	2.82143
57	101	1	159	2.78947
58	101	1	160	2.75862
59	101	1	161	2.72881
60	101	1	162	2.7
61	101	1	163	2.67213
62	101	1	164	2.64516
63	101	1	165	2.61905
64	101	1	166	2.59375
65	101	1	167	2.56923
66	101	1	168	2.54545
67	101	1	169	2.52239
68	101	1	170	2.5
69	101	1	171	2.47826
70	101	1	172	2.45714
71	101	1	173	2.43662
72	101	1	174	2.41667
73	101	1	175	2.39726
74	101	1	176	2.37838
75	101	1	177	2.36
76	101	1	178	2.34211
77	101	1	179	2.32468
78	101	1	180	2.30769
79	101	1	181	2.29114
80	101	1	182	2.275
81	101	1	183	2.25926
82	101	1	184	2.2439
83	101	1	185	2.22892
84	101	1	186	2.21429
85	101	1	187	2.2
86	101	1	188	2.18605
87	101	1	189	2.17241
88	101	1	190	2.15909
89	101	1	191	2.14607
90	101	1	192	2.13333
91	101	1	193	2.12088
92	101	1	194	2.1087
93	101	1	195	2.09677
94	101	1	196	2.08511
95	101	1	197	2.07368
96	101	1	198	2.0625
97	101	1	199	2.05155
98	101	1	200	2.04082
99	101	1	201	2.0303
100	101	1	202	2.02
101	101	101	303	3
102	201	3	306	3
103	203	1	307	2.98058
104	203	1	308	2.96154
105	203	7	315	3
106	209	1	316	2.98113

TABLE 1. The first few terms for $f(1) = 7$, where $\Delta(n) = f(n-1) - n$.

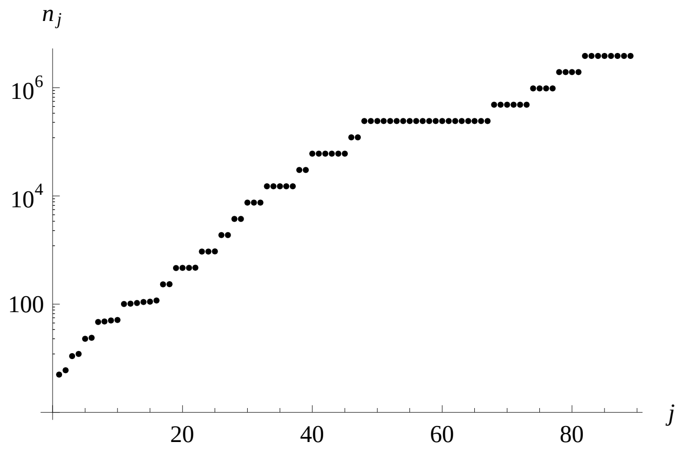

FIGURE 1. Logarithmic plot of n_j, the jth value of n for which $\gcd(n, f(n-1)) \neq 1$. Initially, the regularity of the vertical spacing between clumps is quite unexpected.

In general, consider n_1 and $f(n_1)$. As long as n and $f(n-1)$ are relatively prime (say for $n_1 < n < n_1 + k$), then $g(n) = 1$, and so

$$f(n) = f(n_1) + \sum_{i=1}^{n-n_1} g(n_1 + i) = f(n_1) + (n - n_1). \tag{2}$$

Therefore $f(n) - n = f(n_1) - n_1$ is invariant in this range.

It turns out that a slight modification is significantly more useful:

$$\Delta(n) = f(n-1) - n = f(n_1) - 1 - n_1$$

is invariant on $n_1 < n \leq n_1 + k$, and as table 1 suggests $\Delta(n)$ is always divisible by the next non-1 gcd. This observation (which is easy to show) is a first hint of the shortcut mentioned in section 1.

In studying $f(n)$ experimentally one also notices that $f(n) = 3n$ when $g(n) \neq 1$. This observation is a key ingredient in the proof of theorem 1, and it suggests that $f(n)/n$ may be of interest in general. We study the behavior of $f(n)/n$ in section 4.

Figure 1 plots the values of n for which $g(n) \neq 1$. Clearly they occur in clumps, and the length of each "large gap" between the end of one clump and the beginning of the next is very nearly a power of 2. Upon further examination one finds that when $2n_j - 1 = p$ is prime, we obtain a large gap and $n_{j+1} = p$. This then seriously directs one's attention to the quantity $2n - 1$ (which is $\Delta(n + 1)$ when $f(n) = 3n$).

These observations guide one to an outline of the proof of theorem 1 below.

3. LOCAL STRUCTURE

We now turn to the main result. Recall that $g(n) = \gcd(n, f(n-1))$. Also, note that we no longer assume $f(1) = 7$, and accordingly we may broaden the result: In section 2 we saw that $f(n)/n = 3$ is a significant event; this is one of two cases addressed by the following theorem, which identifies 2 and 3 as recurring values of $f(n)/n$.

Theorem 1. *Let $1 \le n_1 \le f(n_1) - 3$ such that $f(n_1)/n_1$ is 2 or 3. For $n > n_1$ let*

$$f(n) = f(n-1) + \gcd(n, f(n-1)).$$

Let $n_2 > n_1$ be minimal such that $g(n_2) \neq 1$. Then $f(n_2)/n_2 = f(n_1)/n_1$, and moreover $g(n_2)$ is prime.

(We stipulate $f(n_1) \neq n_1 + 2$ because otherwise n_2 does not exist; note however that this excludes only two cases, $n_1 = 2, f(n_1) = 4$ and $n_1 = 1, f(n_1) = 3$. A third case, $n_1 = 1, f(n_1) = 2$, is eliminated by the (strict) inequality; although the conclusion holds in this case (since $n_2 = 2$, $f(n_2)/n_2 = 2$, and $g(n_2) = 2$ is prime), it is not covered by the following proof.)

Proof. Let $r = f(n_1)/n_1$, let p be the smallest prime divisor of $f(n_1) - 1 - n_1 = (r-1)n_1 - 1$, and let $k = n_2 - n_1$. (Since $f(n_1) - 1 - n_1 \ge 2$, p exists.) We show that $g(n_2) = p$ and $k = \frac{p-1}{r-1}$.

For $1 \le i \le k$ we have by assumption $g(n_1 + i) = \gcd(n_1 + i, rn_1 - 1 + i)$. Therefore, $g(n_1 + i)$ divides both $n_1 + i$ and $rn_1 - 1 + i$, so $g(n_1 + i)$ also divides both their difference

$$(rn_1 - 1 + i) - (n_1 + i) = (r-1)n_1 - 1$$

and the linear combination

$$r \cdot (n_1 + i) - (rn_1 - 1 + i) = (r-1)i + 1.$$

Since $g(n_1 + k)$ divides $(r-1)n_1 - 1$ and by assumption $g(n_1 + k) \neq 1$, we have $g(n_1 + k) \ge p$. Since $g(n_1 + k)$ also divides $(r-1)k + 1$, we have $p \le g(n_1 + k) \le (r-1)k + 1$, so $\frac{p-1}{r-1} \le k$.

To show that $k \le \frac{p-1}{r-1}$, assume that $g(n_1 + i) = 1$ for $1 \le i < \frac{p-1}{r-1}$. Then

$$\begin{aligned} g\left(n_1 + \tfrac{p-1}{r-1}\right) &= \gcd\left(n_1 + \tfrac{p-1}{r-1}, rn_1 - 1 + \tfrac{p-1}{r-1}\right) \\ &= \gcd\left(p \cdot \frac{\frac{(r-1)n_1 - 1}{p} + 1}{r-1}, p \cdot \frac{r \cdot \frac{(r-1)n_1 - 1}{p} + 1}{r-1}\right) \\ &= p \neq 1 \end{aligned}$$

since p divides both arguments of the gcd but $g(n_1 + \frac{p-1}{r-1})$ divides $(r-1) \cdot \frac{p-1}{r-1} + 1 = p$. (Note that each quotient here is an integer.)

Therefore $k = \frac{p-1}{r-1}$, and in this case $g(n_2) = g(n_1 + k) = p = (r-1)k + 1$. It follows that

$$\begin{aligned} f(n_2) &= f(n_2 - 1) + g(n_2) \\ &= (rn_1 - 1 + k) + ((r-1)k + 1) \\ &= r(n_1 + k) \\ &= rn_2. \end{aligned}$$

□

We immediately obtain the following result for $f(1) = 7$; one simply computes $g(2) = g(3) = 1$, and $f(3)/3 = 3$ so theorem 1 applies inductively thereafter.

Corollary. *Let $f(1) = 7$. For each $n \ge 2$, the only positive divisors of $g(n)$ are itself and 1.*

(Similar results can be obtained for many other initial conditions. However, the statement is false in general; see section 4.)

Another corollary of theorem 1 is that for $r = 3$, the case $p = 2$ never occurs since $f(n_1) - 1 - n_1 = 2n_1 - 1$ is odd. Furthermore, for $r = 2$, the case $p = 2$ can only occur once for a given initial condition; a simple checking of cases shows that n_2 is even, so applying the theorem to n_2 we find $f(n_2) - 1 - n_2 = n_2 - 1$ is odd.

When $r = 3$ and $2n_1 - 1 = p$ is prime (leading to a large gap, as in figure 1), then $g(n_2) = p \equiv 5 \mod 6$ and $g(n_2 + 1) = 3$. The reason is that eventually we have $f(n) \equiv n \mod 6$ with exceptions only when $g(n) \equiv 5 \mod 6$ (in which case $f(n) \equiv n + 4 \mod 6$). Therefore $p = 2n_1 - 1 = \Delta(n) = f(n-1) - n \equiv 5 \mod 6$, so

$$\begin{aligned} g(n_2 + 1) &= \gcd(n_2 + 1, f(n_2)) \\ &= \gcd(p + 1, 3p) \\ &= 3. \end{aligned}$$

An analogous result holds for $r = 2$ and $n_1 - 1 = p$ prime: $g(n_2) = p \equiv 5 \mod 6$, $g(n_2 + 1) = 1$, and $g(n_2 + 2) = 3$.

Although theorem 1 is stated only for $r = 2$ and $r = 3$, the only distinguishing feature of these values is the guarantee that $\frac{p-1}{r-1}$ is an integer, where p is again the smallest prime divisor of $(r-1)n_1 - 1$. If $r \geq 4$ is an integer and $p - 1$ is divisible by $r - 1$, then the proof goes through.

However, we should also say something about the case when theorem 1 does not apply.

In general one can interpret the evolution of equation 1 as repeatedly computing the minimal $k \geq 1$ such that $\gcd(n+k, f(n-1)+k) \neq 1$ for various n and $f(n-1)$, so let us explore this question in isolation. Let $f(n-1) = n + d$ (with $d \geq 1$); we seek k. (Theorem 1 determines k for the special cases $d = n - 1$ and $d = 2n - 1$.)

Clearly $\gcd(n + k, n + d + k)$ divides d.

Suppose $d = p$ is prime; then we must have $\gcd(n + k, n + p + k) = p$. This is equivalent to $k \equiv -n \mod p$. Since $k \geq 1$ is minimal, we have $k = \operatorname{mod}_1(-n, p)$, where $\operatorname{mod}_j(a, b)$ is the unique number $x \equiv a \mod b$ such that $j \leq x < j + b$.

Now consider a general d. A prime p divides $\gcd(n + i, n + d + i)$ if and only if it divides both $n + i$ and d. Therefore

$$\{\, i : \gcd(n + i, n + d + i) \neq 1 \,\} = \bigcup_{p \mid d} (-n + p\,\mathbb{Z}).$$

Calling this set I, we have

$$k = \min \{\, i \in I : i \geq 1 \,\} = \min \{\, \operatorname{mod}_1(-n, p) : p \mid d \,\}.$$

Therefore (as we record in slightly more generality) k is the minimum of $\operatorname{mod}_1(-n, p)$ over all primes dividing d.

Proposition 1. *Let $n \geq 0$, $d \geq 2$, and j be integers. Let $k \geq j$ be minimal such that* $\gcd(n + k, n + d + k) \neq 1$. *Then*

$$k = \min \{\, \operatorname{mod}_j(-n, p) : p \text{ is a prime dividing } d \,\}.$$

This result generalizes the shortcut of theorem 1 for computing $f(n)$ by skipping all 1s (at the cost of factoring d). One can use this shortcut to feasibly track the evolution from a given initial condition up to large values of n and thereby estimate the number of initial conditions within a certain range whose evolutions do not eventually coincide. For instance, in the range $2^2 \le f(1) \le 2^{13}$ one finds that there are only 203 distinct equivalence classes established below $n = 2^{23}$, and no two of these classes converge below $n = 2^{60}$.

4. Global behavior

One naturally wonders whether $f(1) = 7$ is the only initial condition for which $g(n)$ is always 1 or prime. It turns out that not all initial conditions have this property: $g(18) = 9$ for $f(1) = 532$, and $g(21) = 21$ for $f(1) = 801$. However, with additional experimentation one comes to suspect that it is eventually true for every initial condition.

Conjecture. *Let $n_1 \ge 1$ and $f(n_1) \ge 1$. For $n > n_1$ let*

$$f(n) = f(n-1) + \gcd(n, f(n-1)).$$

Then there exists an N such that for each $n > N$ the only positive divisors of $\gcd(n, f(n-1))$ *are itself and* 1.

A proof of this conjecture (which I do not have) would show that the transience is in fact transient — that if $f(n_1) \ne n_1 + 2$ then $f(N)/N$ is 1, 2, or 3 for some N. (If $f(N) = N + 2$ or $f(N)/N = 1$, then $g(n) = 1$ for $n > N$.) Thus we should try to understand the long-term behavior of $f(n)/n$.

In general we observe that when $f(n)/n$ is large, it tends to decrease. In fact, $f(n)/n$ can never rise above an integer that was previously attained.

Proposition 2. *If $n_1 \ge 1$ and $f(n_1)/n_1$ is a positive integer, then $f(n)/n \le f(n_1)/n_1$ for all $n \ge n_1$.*

Proof. Let $r = f(n_1)/n_1$. We proceed inductively; assume that $f(n-1)/(n-1) \le r$. Then

$$rn - f(n-1) \ge r \ge 1.$$

Since $g(n)$ divides the linear combination $r \cdot n - f(n-1) \ge 1$, we have

$$g(n) \le rn - f(n-1);$$

thus

$$f(n) = f(n-1) + g(n) \le rn.$$

□

There seem to be arbitrarily long repetitions of an integer $r \ge 4$. Searching in the range $1 \le n_1 \le 1000$, $4 \le r \le 10$, one finds the example $n_1 = 757$, $r = 7$, $f(n_1) = rn_1 = 5299$, in which 7 reoccurs nine times (the last at $n = 824$). This suggests the possible difficulty of a sharper result.

From equation 2 in section 2 we see that $g(n_1 + i) = 1$ for $1 \le i < k$ implies that $f(n_1 + i)/(n_1 + i) = (f(n_1) + i)/(n_1 + i)$, and so $f(n)/n$ is strictly decreasing in this range if $f(n_1) > n_1$. Moreover, if the non-1 gcds are overall sufficiently few and sufficiently small, then we would expect $f(n)/n \to 1$ as n gets large. However, in practice we rarely see this occurring. Rather, $f(n_1)/n_1 > 2$ seems to (almost always) imply that $f(n)/n > 2$ for all $n \ge n_1$. Why is this the case?

Suppose the sequence crosses 2 for some n: $f(n)/n > 2 \ge f(n+1)/(n+1)$. Then

$$2 \ge \frac{f(n+1)}{n+1} = \frac{f(n) + \gcd(n+1, f(n))}{n+1} \ge \frac{f(n)+1}{n+1},$$

so $f(n) \leq 2n+1$. Since $f(n) > 2n$, we are left with $f(n) = 2n+1$; and indeed in this case we have

$$\frac{f(n+1)}{n+1} = \frac{2n+1+\gcd(n+1, 2n+1)}{n+1} = \frac{2n+2}{n+1} = 2.$$

The task, then, is to determine whether $f(n) = 2n+1$ can happen in practice. That is, if $f(n_1) > 2n_1 + 1$, is there ever an $n > n_1$ such that $f(n) = 2n+1$?

Let's work backward from $f(n) = 2n+1$. What could $f(n-1)$ have been?

Assume $f(n-1) = 2n$; then

$$2n+1 = f(n) = 2n + \gcd(n, 2n) = 3n,$$

so $n = 1$. In fact $f(1) = 3$ has an infinite history but is a moot case if we restrict to positive initial conditions.

Alternatively, assume $f(n-1) = 2n - j$ for some $j \geq 1$. Then

$$2n+1 = f(n) = 2n - j + \gcd(n, 2n-j),$$

so $j+1 = \gcd(n, 2n-j)$ divides $2 \cdot n - (2n-j) = j$. This is a contradiction.

Thus the state $f(n) = 2n+1$ has no predecessor for $n > 1$, and we have proved the following.

Proposition 3. *If $n_1 \geq 1$ and $f(n_1) > 2n_1 + 1$, then $f(n)/n > 2$ for all $n \geq n_1$.*

References

[1] Stephen Brown, 'Prime' pedagogical schemes, *American Mathematical Monthly* **75** (1968) 660–664.

[2] Underwood Dudley, History of a formula for primes, *American Mathematical Monthly* **76** (1969) 23–28.

[3] Graham Everest, Shaun Stevens, Duncan Tamsett, and Tom Ward, Primes generated by recurrence sequences, *American Mathematical Monthly* **114** (2007) 417–431.

[4] Matthew Frank, personal communication, July 15, 2003.

[5] James Jones, Daihachiro Sato, Hideo Wada, and Douglas Wiens, Diophantine representation of the set of prime numbers, *American Mathematical Monthly* **83** (1976) 449-464.

[6] NKS Summer School, `http://www.wolframscience.com/summerschool`.

[7] Neil Sloane, The On-Line Encyclopedia of Integer Sequences, `http://www.research.att.com/~njas/sequences`, sequences A084662 and A084663.

[8] Stephen Wolfram, *A New Kind of Science*, Wolfram Media, Inc., Champaign, IL, 2002.