Collatz Cycles and $3n + c$ Cycles

Darrell Cox, Sourangshu Ghosh, Eldar Sultanow

January 13, 2021

Abstract

Halbeisen and Hungerb"uhler [1] determined optimal bounds for the length of rational Collatz cycles. Their methods are extended to $3n + c$ cycles. Another sequence having properties similar to those of Riemann zeta function zeros is introduced.

1 Introduction

Their main results are reproduced here since they are directly applicable to $3n + c$ cycles where $c$ is an odd integer not divisible by 3. Proofs of all but one lemma are omitted.

2 Halbeisen and Hungerb"uhler’s Results for Collatz Cycles

Let $S_{l,n}$ denote the set of all 0-1 sequences of length $l$ containing exactly $n$ ones, $S_l = \cup_{n=0}^{l} S_{l,n}$ and $S = \cup_{l \in \mathbb{N}} S_l$. With every $s = (s_1, ..., s_l)$ we associate the affine function $\phi_s : \mathbb{R} \to \mathbb{R}$, $\phi_s = g_{s_l} \circ ... \circ g_{s_2} \circ g_{s_1}$. A sequence $(x_0, ... x_l)$ of real numbers $x_i$ is called a pseudo-cycle of length $l$ if there exists $s = (s_1, ..., s_l) \in S_l$ such that (1) $\phi_s(x_0) = x_0 \in \mathbb{Q}$ and (2) $g_{s_{i+1}}(x_i) = x_{i+1}$ for $i = 0, \ldots, l-1$.

Notice that if $p/q \in \mathbb{Q}$ with $2^r | q$ then $2^r | \tilde{q}$ where $\tilde{q}$ denotes the denominator of $g_i(p/q)$ ($i = 0, 1$). Hence every element of a pseudo-cycle is in $\mathbb{Q}(2)$. Thus, if $p/q$ and $g_i(p/q) = \tilde{p}/\tilde{q}$ are consecutive elements of a pseudo-cycle, then $i = 0$ if $p$ is even (since else $\tilde{p}/\tilde{q} \notin \mathbb{Q}(2)$) or $i = 1$ if $p$ is odd (since else $\tilde{p}/\tilde{q} \notin \mathbb{Q}(2)$).

The conclusion of this observation is

LEMA 1. The set of pseudo-cycles coincides with the set of Collatz cycles in $\mathbb{Q}(2)$. Cycles are either positive or negative.

The function $\varphi : S \to \mathbb{N}$ will be defined recursively by $\varphi(\emptyset) = 0$, $\varphi(s_0) = \varphi(s)$, $\varphi(s_1) = 3\varphi(s) + 2^{l(s)}$, where $s$ denotes an arbitrary element of $S$ and $l(s)$ the
length of $s$. $\varphi$ is computed explicitly by $\varphi(s) = \sum_{j=1}^{\ell(s)} s_j 3^{s_{j+1} + \ldots + s_{(j-1)}} 2^{j-1}$.

A consequence of the above definition is the decomposition formula $\varphi(\bar{s}\bar{t}) = 3^{\ell(\bar{s})} \varphi(s) + 2^{\ell(\bar{t})} \varphi(\bar{t})$. Here $\bar{s}\bar{t}$ is the concatenation of $s$, $\bar{s} \in S$, and $n(s)$ denotes the number of 1’s in the sequence $s$.

The next lemma shows how $\phi$ is used to explicitly compute the function $\phi_s$.

**Lemma 2** (Lagarias [2]). For arbitrary $s \in S$, $\phi_s(x) = \frac{3^{\ell(s)} x + \varphi(s)}{2^{\ell(s)}}$ and hence for every $s \in S$ there exists a unique $x_0 \in \mathbb{Q}[\{2\}]$ which generates a Collatz cycle in $\mathbb{Q}[\{2\}]$ of length $l(s)$ and which coincides with the pseudo-cycle generated by $s$. $x_0$ is given by $x_0 = \frac{\varphi(s)}{2^{\ell(s)} - \varphi(s)}$.

Proof: The proof is by induction with respect to $l(s)$. (1) $l(s) = 1$ is checked from the definition. (2) $l(s) > 1$: if $s = \bar{s}_0$ then $\phi_{s_0}(x) = \frac{\phi_s(x)}{2} = \frac{3^{\ell(s)} x + \varphi(s)}{2^{\ell(s)}} = \frac{3^{\ell(s)} x + \varphi(s)}{2^{\ell(s)}}$. (3) The case $s = \bar{s}_1$ is analogous.

For $s \in S_l$ let $\sigma(s)$ denote the orbit of $s$ in $S_l$ generated by the left-shift permutation $\lambda_l : (s_1, \ldots, s_l) \rightarrow (s_2, \ldots, s_l, s_1)$, i.e. $\sigma(s) := \{\lambda_l^k(s) : k = 1, \ldots, l\}$. Furthermore, let $M_{l,n}$ denote $\max_{s \in S_l} \{\min_{t \in \sigma(s)} \varphi(t)\}$.

Now suppose the Collatz conjecture is verified for all initial values $x_0 \leq m$. If one can then show that $\forall n, l < L : \frac{M_{l,n}}{2^{l(n)} - \varphi(n)} \leq m$, it follows that the length of a Collatz cycle in $\mathbb{N}$ which does not contain 1 is at least $L$.

A lemma that is relevant in the following is

**Lemma 3.** Let $f_i$ be a non-increasing sequence of non-negative real numbers, and $c_i$ and $d_i$ be sequences such that for all $n \in \mathbb{N}$, $\sum_{i=1}^{n} c_i \leq \sum_{i=1}^{n} d_i$. Then for all $n \in \mathbb{N}$, $\sum_{i=1}^{n} c_i f_i \leq \sum_{i=1}^{n} d_i f_i$.

The next two lemmas pertain to the minimum in a Collatz cycle of length $l$. A monotonicity property of the function $\varphi$ is

**Lemma 4.** Let $s = (s_1, \ldots, s_l)$ and $t = (t_1, \ldots, t_l)$ be two distinct elements of $S_{l,n}$. If $\sum_{i=1}^{k} s_i \leq \sum_{i=1}^{k} t_i$ for all $k \in \{1, \ldots, l\}$, then $\varphi(s) > \varphi(t)$.

In the next lemma, the sequence $\hat{s}$ for which $\varphi$ attains the value $M_{l,n}$ is determined.

**Lemma 5:** Let $n \leq l$ be natural numbers. Let $\hat{s}_i := \lfloor in/l \rfloor - \lfloor (i-1)n/l \rfloor$ (for $1 \leq i \leq l$). Then $\varphi(\hat{s}) = \min_{t \in \sigma(\hat{s})} \varphi(t) = M_{l,n}$.

Now it must be shown that $\varphi(\bar{s}) = \min_{s \in \sigma(\bar{s})} \{\varphi(s)\}$. Let $\bar{s}' := \lambda_l^n(\bar{s})$ for an
arbitrary m. Then by the construction of \( \bar{s} \), \( \sum_{i=1}^{k} \bar{s}_i' \leq \sum_{i=1}^{k} \bar{s}_i \) (1 \( \leq k \leq l \)) and therefore \( \varphi(\bar{s}) \leq \varphi(\bar{s}') \) (by Lemma 4).

From Lemma(4) and Lemma(5), the conclusion is

**Corollary 1:** For every \( l \) and \( n \leq l \),

\[
M_{l,n} = \sum_{j=1}^{l} ([jn/l] - [(j-1)n/l])2^{j-1}3^{n-[jn/l]}
\]

### 3 The Minimum Element in a 3n+c Cycle

Setting \( c \) to \( 2^{l(s)} - 3^n(s) \) in Lemma (2) gives integer 3n + c cycles. A staircase for \( \bar{s} \) where 27 and \( n = 17 \) along with a staircase representing the partial sums of \( [jn/l] - [(i-1)n/l] \) is:

![Staircase Diagram](image)

The staircase using the floor function can be viewed as being an upside-down staircase where Halbeisen and Hungerbühler’s logic can be used to find a lower bound of the maximum odd element a 3n+c cycle. Let \( t_j = [jn/l] - [(j-1)n/l] \), \( j = 1, 2, 3, ..., l \). (This "parity vector" is an element of \( S_{l,n} \)). Let \( r \) denote \( \gcd(l, n) \). The parity vector \( [jn/l] - [(j-1)n/l] \), \( j = 1, 2, 3, ..., l \), consists of \( r \) identical sub-vectors. Similarly, the parity vector \( t_j \) consists of \( r \) identical sub-vectors and each of these subvectors is the same as the corresponding sub-vector of \( [jn/l] - [(j-1)n/l] \), \( j = 1, 2, 3, ..., l \), except for the first and last elements. First suppose that \( l \) and \( n \) are relatively prime. When the parity
vector \( \lfloor jn/l \rfloor - \lfloor (j - 1)n/l \rfloor, j = 1, 2, 3, \ldots \) is right-rotated by one position (corresponding to a multiplication by 2), it matches \( t_j \) except for the first two elements of each sub-vector. The first mismatch corresponds to a loss of \( 3^{n-1} \) and the second mismatch corresponds to a gain of \( 2 \cdot 3^{n-1} \). In general, the loss is \( \sum_{i=0}^{r-1} 2^{i(l/r)} 3^{n-1-i(n/r)} \). Let \( N_{t,n} \) denote \( 2M_{t,n} - \sum_{i=0}^{r-1} 2^{i(l/r)} 3^{n-1-i(n/r)} \). A primitive \( 3n + c \) cycle doesn’t have any common divisors of its elements. A generalization of Halbeisen and Hungerbühler’s result is:

Corollary 2: If \( c = 2^l - 3^n \), \( M_{t,n} \) is greater than or equal to the minimum elements in the \( 3n + c \) cycles corresponding to \( s \in S_l, n \) (not necessarily primitive) and \( N_{t,n} \) is less than or equal to the maximum odd elements in the cycles. (The elements of the \( 3n + c \) cycles are \( \varphi(t)_{t \in \sigma(s)} \) where \( s \in S_{l, n} \).) From the definition of \( N_{t,n} \), it is not apparent that it is in a cycle, but it appears to be in the same cycle as \( M_{t,n} \). Also, the distance between \( M_{t,n} \) and \( N_{t,n} \) on the parity vector modulo \( l \) (denoted by \( d \)) satisfies the relation \( l | nd + 1 \) (the distance is measured in the clock-wise direction). This will be investigated in the next section.

For example, for \( (l, n) = (6, 2) \), the parity vector for \( M_{l,n} \) is \( \{1, 1, 0, 1, 1, 0\} \), \( c = -17 \), and the odd elements of the cycle containing \( M_{l,n} \) and \( N_{t,n} \) are \( \{85, 119, 85, 119\} \). There is only one more element in \( S_l, n \) and its odd elements are \( \{65, 89, 125, 179\} \). 85 is greater than 65 and 119 is less than 125. The cycle with odd elements of \( \{85, 119\} \) is not primitive and reduces to a cycle with odd elements of \( \{5, 7\} \) for \( c = -1 \). When \( l = 11 \) and \( n = 7 \), \( M_{l,n} = 3767, N_{l,m} = 6805, \) and \( 2^4 - 3^7 = -139. \) 3767/139 (approximately equal to 27) is greater than the minimum element in the \( c = -1 \) cycle \( \{34, 17, 25, 37, 55, 82, 41, 61, 91, 136, 68\} \) and 6805/139 (approximately equal to 49) is less than the maximum odd element. For the \( c = -1 \) cycle of \( \{5, 7, 10\} \), \( M_{5,2} = 5 \) and \( N_{3,2} = 7 \) (-1 equals \( 2^3 - 3^2 \)). Note the relevance of the Catalan conjecture (proved by Mihăilescu [3]) states that the only natural number solutions of \( x^a - y^b \) are \( x = 3, a = 2, y = 2, \) and \( b = 3 \). For \( c = -17 \), the cycles are \( \{85, 119, 170, 85, 119, 170\}, \{103, 146, 73, 101, 143, 206\}, \) and \( \{65, 89, 125, 179, 260, 130\} \). -17 equals \( 2^6 - 3^4 \), \( M_{6,4} = 85 \) and \( N_{6,4} = 119. \) 85 is greater than 73 and 65. 119 is less than 143 and 179. The cycle \( \{85, 119, 170\} \) is not primitive and reduces to the \( c = -1 \) cycle. Note that 17 must divide \( 3^3 + 2^3 \cdot 3 \) for this to be possible. For the \( c = 1 \) cycle of \( \{4, 1\} \), 1 equals \( 2^2 - 3^1 \) and \( M_{2,1} = N_{2,1} = 1 \). There can be no other such \( c = 1 \) cycles. This leaves the possibility of \( 3n + c \) cycles where \( s \in S_{l, n} \) that are not primitive and reduce to \( c = 1 \) cycles.

All the parity vectors in \( S \) are used up by the \( 3n + c \) cycles where \( c = 2^l - 3^n \). Two \( 3n + c \) cycles with different \( c \) values can’t have the same parity vector. For example, the elements of a \( c = 5 \) cycle are \( \{19, 31, 49, 76, 38\} \) and a \( c = 7 \) sequence having the same parity vector is \( \{65, 101, 155, 236, 118, \ldots\} \). The ratios of the odd elements are 0.2923, 0.3069, and 0.3161 and would have to keep increasing to match the iterations of the \( 3n + 5 \) cycle. So the unreduced \( 3n + c \)
cycles where \( c = 2^l - 3^n \) account for all possible primitive \( 3n + c \) cycles.

When \( l \) and \( n \) are not relatively prime and \( c = 2^l - 3^n \), the cycles generated from \( M_{l,n} \) are not primitive (this is true in general for \( 3n + c \) cycles where \( c = 2^l - 3^n \)). This is due to the duplicated sub-vectors in the parity vector forming a geometric progression. This geometric progression is the same as in the expansion of \( (ax - bx)/(a-b) \). For example, when \( c = 2^9 - 3^6 \), reducing the cycle generated from \( M_{9,6} \) effectively divides \( 2^9 - 3^6 \) by \( (2^9 - 3^6)/(2^3 - 3^2) \).

For \( l = 6, 9, 12, 15, 18, \ldots \), there are 0, 0, 1, 2, 3, \ldots, odd cycle elements between \( M_{l,n} \) and \( N_{l,n} \). In this case, \( c, M_{l,n} \) and \( N_{l,n} \) have at least one prime factor in common. The respective prime factors are 37, 7 \cdot 31, 13 \cdot 109, 10177, 78697, \ldots. The respective \( c \) values (after reduction) are 1, -1, -11, -49, -179, \ldots, and the respective cycles are \( \{1, 2\} \), \( \{5, 7, 10\} \), \( \{19, 23, 29, 38\} \), \( \{65, 73, 85, 130\} \), and \( \{211, 227, 251, 287, 341, 422\} \).

After reduction of \( 3n + c \) cycles where \( c = 2^l - 3^n \), all \(|c|\) values appear to be covered. A cycle will be said to have an attachment point if there are two adjacent 0’s in the parity vector (taking into account rotation of the vector).

Conjecture: There is at least one reduced \( 3n + c \) cycle with an attachment point for every \(|c|\) value.

4 On Proving that \( M_{l,n} \) and \( N_{l,n} \) are in the Same Cycle

That \( l| dn + 1 \) remains unproven. In the following, \( l \) is restricted to being prime (for simplicity). In general, only \( l \) and \( n \) that are relatively prime need be considered (since there are duplicated sub-vectors in the parity vector otherwise). For prime \( l \) and the \( n \) values less than \( l \), the \( d \) values appear to be in one-to-one correspondence with the natural numbers less than \( l \). Let \( r \) denote a natural number less than \( l \). For a fixed \( r \) and \( n = 1, 2, 3, \ldots, l-1 \) there exist \( x \) (\( 0 < x < l \)) such that \( l \) divides \( rn + x \) and these \( x \) values are a permutation of the integers \( 1, 2, 3, \ldots, l-1 \). For \( l = 17 \) and \( r = 5 \) this gives \( 17|5 \cdot 10 + 1 \). For \( l = 17 \) and \( r = 7 \), this gives \( 17|7 \cdot 12 + 1 \). The \( n \) values (10 and 12) have to be distinct, otherwise 17 would divide 10 - 12. Continuing in this way for different \( r \) values gives a system of congruences that turns out to be the same as for \( d \). Other \( x \) values besides 1 could have been selected, so another question is why it is not 2 for example.

A similar congruence is \( d(n - x) - (l - d)x \equiv -1 \mod l \). This congruence was derived using the staircases and can be solved given the \( d \) value (so that the values of \( (n - x, x) \) are not specific to properties of \( 3n + c \) cycles).
5 \((n - x, x)\) Values and Uniform Distributions

For a real number \(x\), let \([x]\) denote the integral part of \(x\) and \(\{x\}\) the fractional part. The sequence \(\omega = \{x_n\}, n=1,2,3,...\) of real numbers is said to be uniformly distributed modulo 1 (abbreviated u.d. mod 1) if for every pair \(a, b\) of real numbers with \(0 \leq a < b \leq 1\) we have \(\lim_{N \to \infty} \frac{A([a,b];N,\omega)}{N} = b - a\). The formal definition of u.d. mod 1 was given by Weyl [4] [5]. Let \(\Delta : 0 = z_0 < z_1 < z_2 < \ldots\) be a subdivision of the interval \([0, \infty)\) with \(\lim_{k \to \infty} z_k = \infty\). For \(z_{k-1} \leq x < z_k\) put \([x]_\Delta = z_{k-1}\) and \(\{x\}_\Delta = \frac{x - z_{k-1}}{z_k - z_{k-1}}\) so that \(0 \leq \{x\}_\Delta < 1\). The sequence of \((x_n)\), \(n = 1,2,3,...\) of non-negative real numbers is said to be uniformly distributed modulo \(\Delta\) (abbreviated u.d. mod \(\Delta\)) if the sequence \((\{x_n\}_\Delta)\), \(n = 1,2,3,...\) is u.d. mod 1. The notion of u.d. mod \(\Delta\) was introduced by Levaque [6].

If \(f\) is a function having a Riemann integral in the interval \([a,b]\), then its integral is the limit of Riemann sums taken by sampling the function \(f\) in a set of points chosen from a fine partition of the interval. This is then a criterion for determining if a sequence is uniformly distributed. A sequence of real numbers is uniformly distributed (mod 1) if and only if for every Riemann-integrable function \(f\) on \([0,1]\) one has \(\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(\{x_n\}) = \int_{0}^{1} f(x)dx\). In the following, evidence that \(n - x\) and \(x\) are u. d. mod \(\Delta\) is presented using this criterion and Weyl’s criterion [4] [5].

Weyl’s criterion is that \((\gamma_n)\) is u.d. mod 1 if and only if \(\lim_{N \to \infty} 1/N \sum_{n=1}^{N} e^{2\pi im\gamma_n} = 0\) for every integer \(m \neq 0\). In the following, the \(z\) increments in the subdivision are set to \(\sqrt{2}\) to avoid any aliasing with the integer \(n - x\) and \(x\) values. A plot of the resulting sequence generated from the sorted \(n - x\) values for \(n = 1,2,3,...,l - 1\) and \(l = 97\) (excluding 12 values of 0) is
Setting \((\gamma_n)\) to this sequence in Weyl's formula and computing the moduli of the complex-valued results for \(l = 97, n - x, \) and \(m = 1\) gives
The values are decreasing but don’t reach zero. The moduli for $l = 97$, $x$, and $m = 1$ are
The values become closer to zero than for $n - x$. The moduli for $l = 997$, $n - x$, and $m = 2$ (excluding twelve values of 0 in the input sequence) are
There is one hump in the curve before it decreases to almost zero. The moduli for $l = 997$, $x$, and $m = 2$ are
The results are similar to those for $N - x$ except that the hump is more well-defined. The moduli for $l = 4999$, $n - x$, and $m = 3$ (excluding 24 values of 0 in the input sequence) are
There are two humps in the curve before it decreases toward zero. The moduli for $l = 4999$, $x$, and $m = 3$ are
The results are similar to those for $N - x$ except that the humps are more well-defined and the last point comes closer to zero. The moduli for $l = 4999$, $n - x$, and $m = 1$ along with a cubic least-squares fit of the curve are
The moduli for $l = 4999$, $x$, and $m = 1$ along with a cubic least-squares fit of the curve are
The results are similar to those for \( n - x \) except that the cubic least-squares fit is better.

The functions \( f(x) \) to be considered are \( x, x^2, x^3, x^4, \sqrt{x}, \sqrt[3]{x}, \log(x), e^x, \sin(x), \cos(x), \tan(x), \) and \( \frac{1}{a + 2} \). The values of \( \int_0^1 f(x)dx \) are \( 1/2, 1/3, 1/4, 1/5, 2/3, 4/5, -1, 2.72, 0.84, 0.54, 1.56, \) and \( \frac{1}{a \tan^{-1} \frac{x}{a}} \) (equal to 0.23 for \( a = 2 \) and 0.11 for \( a = 3 \)) respectively. For \( l = 997 \) and the sequence generated from \( n - x \), the results are \( 0.49, 0.31, 0.22, 0.17, 0.66, 0.80, -1.00, 2.69, 0.84, 0.56, 1.56, 0.23 \) (for \( a = 2 \)), and 0.11 (for \( a = 3 \)) respectively.

A plot of the integral for \( f(x) = x^2 \) is
A plot of the integral for $f(x) = e^x$ is
A plot of the integral for \( f(x) = \sin(x) \) is

\[
\int f(x) = \int \sin(x) dx
\]

A plot of the integral for \( f(x) = \frac{1}{ax^2} \), \( a = 3 \) is
The trigonometric functions require a fixed amount to be added to the sequence values (apparently to change the phase). The exponential function also requires a fixed amount to be added to the sequence values - the same as for the cosine function. Apparently, this is due to Euler’s formula \( e^{ix} = \cos(x) + i \cdot \sin(x) \). The amount required for the tangent function is \( \tan^{-1}(1)/2 \). Denote the other amounts by \( j \) and \( k \). These values satisfy the equation \( j^2 + k^2 = \cos(1) \), similar to the formula \( \sin(x)^2 + \cos(x)^2 = 1 \). They also satisfy the equation \( j/k = \sqrt{\tan(1)} \), similar to the formula \( \sin(x)/\cos(x) = \tan(x) \). The amount required for the sine function is \( \sin^{-1}(\cos(1)) \). The amount required for the cosine function can be determined by using the formula \( j^2 + k^2 = \cos(1) \).

6 Discrete Uniform Distributions and the Möbius Function

Cox \[7\] investigated convolving the zeta function zeros with the Möbius function. This method is applicable to any uniformly distributed sequence. Let \( p \) and \( q \) denote distinct primes. A histogram of \( n - x \) for randomly generated \( l \) values of the form \( pq \) and \( n \) set to 47 is

![Graph showing the distribution of values for \( f(x) = 1/(x^2 + t^2) \) with \( t = 397 \), sequence generated from \( N - x \).]
The corresponding histogram of the $x$ values is
A histogram of \( n - x \) for randomly generated \( l \) values of the form \( p^2 q \) and \( n \) set to 11 is

![Histogram of \( n - x \) values for 447 random \( l \) values of the form \( p^2 q \) and \( N=11 \)](image)

The corresponding histogram of the \( x \) values is
Henceforth, $n - x$ and $x$ are ordered in increasing value. A plot of $n - x$ convolved with the Möbius function for the 1228 prime $l$ values less than 10000 is
The convolution consists of many curves. The bottom curve corresponds to the primes and the middle curve corresponds to the primes squared. There are 200 points in the curve corresponding to \( p \). The interpretation of this is that the 201\(^{th} \) prime in the sequence \( \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, \ldots \} \) is greater than 1228, that is, the convolution maps the primes to their indices. There are 11 points in the curve corresponding to \( p^2 \). The interpretation of this is that the square of the 12\(^{th} \) prime in the sequence of primes (37) is greater than 1228. A plot of the curve corresponding to \( pq \) is
A plot of $n - x$ convolved with the Möbius function for $l = 2, 3, ..., 2000$ is
A plot of $x$ convolved with the Möbius function for $l = 2, 3, 4, ..., 2000$ is
In these convolutions, the elements in the curve corresponding to $p$ are \{2, 3, 5, \ldots\}. There are 303 primes less than 2000 and there are that many elements in the curve.

The convolution of the differences between the adjacent $n - x$ values with the Möbius function for a particular curve is normally distributed. For the above $pq$ curve, the mean is -0.1421 with a 95% confidence interval of (-0.2032, -0.0809) and the standard deviation is 0.7387 with a 95% confidence interval of (0.6979, 0.7845). A plot of this distribution along with the corresponding probit function is
The probit function is the inverse cumulative distribution function of the standard normal probability distribution. The more general function \( F^{-1}(p) = \mu + \sigma \Phi^{-1}(p) \) where \( \mu \) and \( \sigma \) are the mean and standard deviation of the distribution is applicable here. Considering the discrete values of the distribution, the fit is fairly good.

7 Empirical Results

In this section, \( l \) is restricted to being prime. Since \( l \) cannot divide \( nd + 1 \) and \( n'd + 1 \) for equal \( d \), the mapping of the rotated floor parity vector to the ceiling parity vector for all possible \( n \) values is one-to-one. The values of \( nd+1 \) modulo \( l \) determines a "basis". For example, the basis for \( l = 31 \) is

\[
\text{least residue} = \begin{array}{c}
1 : & 30, 15, 10, 6, 5, 3, 2, 1 \\
3 : & 23, 4 \\
5 : & 22, 14, 11, 7 \\
7 : & 27, 24, 18, 12, 9, 8 \\
8 : & 19, 13 \\
11 : & 20, 17 \\
15 : & 29, 16 \\
19 : & 28, 21 \\
21 : & 26, 25 \\
\end{array}
\]
The number of elements in this basis is 9. The number of distinct prime factors in the respective elements is \{3, 2, 3, 2, 3, 2, 3, 3\}. The maximum number of distinct prime factors is 3.

A plot of the number of elements in a basis versus the primes less than 10000 is

![Plot](image)

For a quadratic least-squares fit of the curve, \( p_1 = -9.91 \cdot 10^{-7} \) with a 95% confidence interval of \((-1.074 \cdot 10^{-7}, -9.076 \cdot 10^{-7})\), \( p_2 = 0.2332 \) with a 95% confidence interval of \((0.2324, 0.2341)\), \( p_3 = 12.96 \) with a 95% confidence interval of \((11.26, 14.66)\), \( \text{SSE}=1.596 \cdot 10^5 \), \( \text{R-squared}=0.9997 \), and \( \text{RMSE}=11.41 \).

A plot of the maximum number of distinct prime factors of the elements of a basis versus the square roots of the primes less than 10000 is
For a linear least-squares fit of the curve, $p_1 = 0.2636$ with a 95% confidence interval of $(0.2588, 0.2685)$, $p_2 = 1.123$ with a 95% confidence interval of $(0.7893, 1.456)$, $\text{SSE}=5875$, $R^2=0.9017$, and $\text{RMSE}=2.189$.

A plot of the logarithm of the histogram of the number of elements in a basis is
For a quadratic least-squares fit of the curve, $p_1 = 0.01388$ with a 95% confidence interval of (0.01025, 0.01751), $p_2 = -0.6509$ with a 95% confidence interval of (-0.759, -0.5428), $p_3 = 7.759$ with a 95% confidence of (7.7073, 8.445), SSE=4.648, R-squared=0.9568, and RMSE=0.9496.

References

[1] L. Halbeisen and N. Hungerbühler, *Optimal bounds for the length of rational Collatz cycles*, Acta Arith., LXXVIII.3 (1997), pgs. 227-239

[2] J.C. Lagarias, *The set of rational cycles for the 3x+1 problem*, Acta Arith. 56 (1990), 33-53

[3] Mihăilescu, P., *Primary Cyclotomic Units and a Proof of Catalan’s Conjecture*, J. Reine angew, Math. 572: 167-195

[4] Weyl, H., Über ein Problem aus dem Gebiete der diophantischen Approximationen, *Nachr. Ges. Wiss. Göttingen*, Math.-phys. Kl., 1914, 234-244

[5] Weyl, H., Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.*, 77, 313-352 (1916)

[6] LeVeque, W.J., On uniform distribution modulo a subdivision, *Pacific J. Math.*, 3, 757-771 (1953)
[7] Cox, Darrell, (2019), Zeta Function Zeros, the Möbius Function, and Dirichlet Products, 10.13140/RG.2.2.14588.97923