SUPPLEMENT TO THE PAPER "DIVISIONS BY TWO IN COLLATZ SEQUENCES: A DATA SCIENCE APPROACH"

Christian Koch and First Last

Abstract. Discourse in communities constantly contributes to the sharing of findings and knowledge, constructive criticism of scientific work, and improvement of results. In this supplementary short paper, we address a major potential area of improvement in our published article "Divisions by Two in Collatz Sequences: A Data Science Approach" [1] that was raised on StackExchange Mathematics [2].

2010 *Mathematics Subject Classification.* 37P99.
Key words and phrases. Collatz sequence, binary system.

1. Wrapping up the main results

We stated that for a Collatz sequence $v_1, v_2, \ldots, v_{n+1}$ and the corresponding product $\beta = \beta_1\beta_2\cdots\beta_n = (1 + {}^1\!/_{3v_1})(1 + {}^1\!/_{3v_2})\cdots(1 + {}^1\!/_{3v_n})$ the following equation holds:

$$v_{n+1} = \frac{3^n v_1 \beta}{2^\alpha} \tag{1}$$

Note that $\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_n$ is the total number of divisions by two that have been performed within this sequence starting from v_1 and ending with v_{n+1}.

Moreover we stated that the maximum possible number of division by two in such a sequence is given by equation 2.

$$\hat{\alpha} = \lfloor n \log_2 3 + \log_2 v_1 \rfloor + 1 \tag{2}$$

The binary growth Λ of a Collatz sequence we stated to be upper bounded as follows:

$$\Lambda \leq \lfloor n \log_2 3 \rfloor + 2 \tag{3}$$

In the following section we explain this limit more comprehensible.

2. Proving the binary growth's upper limit

Lemma 1. *Let us apply n times a multiplication by 3 followed with an addition by 1 to an odd integer v_1:*

$$\underbrace{3(3(3(3v_1 + 1) + 1) + 1) + 1 \cdots}_{n \text{ times}}$$

The above given repetition leads to a binary growth of v_1 that we denote with Λ and it is upper bounded as follows:

$$\Lambda \leq \lfloor n \log_2 3 \rfloor + 2$$

Proof. We start with an odd integer v_1 and apply n times the growth operation $x \mapsto 3x + 1$, which yields $3^n v_1 + 3^{n-1} + 3^{n-2} + \ldots + 3^0$. This result is a geometric series and can be compressed as follows:

$$3^n v_1 + 3^{n-1} + 3^{n-2} + \ldots + 3^0 = \frac{2 \cdot 3^n v_1 + 3^n - 1}{2}$$

The binary length of this extremal case is:

$$\operatorname{len}\left(\frac{2\cdot 3^n v_1 + 3^n - 1}{2}\right) = \lfloor \log_2(2\cdot 3^n v_1 + 3^n - 1) - 1\rfloor + 1$$

When we substract from this length the binary length of v_1 given by $\operatorname{len}(v_1) = \lfloor \log_2 v_1 \rfloor + 1$, then we obtain the binary growth of this extremal case of an Collatz sequence that is constantly growing:

$$\begin{aligned}\Lambda &= \lfloor \log_2(2\cdot 3^n v_1 + 3^n - 1) - 1\rfloor + 1 - \lfloor \log_2 v_1 \rfloor - 1\\ &= \lfloor \log_2(2\cdot 3^n v_1 + 3^n - 1) - 1\rfloor - \lfloor \log_2 v_1 \rfloor\end{aligned}$$

According to Lemma 1 we claim that $\Lambda \leq \lfloor n \log_2 3 \rfloor + 2$, which leads to:

$$\begin{array}{ccccc}\lfloor \log_2(2\cdot 3^n v_1 + 3^n - 1) - 1\rfloor & - & \lfloor \log_2 v_1 \rfloor & \leq & \lfloor n \log_2 3 \rfloor + 2\\ \lfloor \log_2(2\cdot 3^n v_1 + 3^n - 1) - 1\rfloor & \leq & \lfloor \log_2 v_1 \rfloor & + & \lfloor n \log_2 3 \rfloor + 2\end{array}$$

In the worst case the left side of this inequality is a whole number and for this reason there is nothing to round down (the floor operation has no effect):

$$\begin{array}{lcccc}\log_2(2\cdot 3^n v_1 + 3^n - 1) - 1 & \leq & \lfloor \log_2 v_1 \rfloor & + & \lfloor n \log_2 3 \rfloor + 2\\ \log_2(2\cdot 3^n v_1 + 3^n - 1) - \lfloor \log_2 v_1 \rfloor & \leq & \lfloor n \log_2 3 \rfloor & + & 3\end{array}$$

The worstcase now is $v_1 = 1$ which maximizes the left side:

$$\log_2(3\cdot 3^n - 1) \leq \lfloor n \log_2 3 \rfloor + 3$$

And even if I increase the left side by removing the *minus one* operation inside the Logarithmization, then it still remains below the limit:

$$\begin{aligned}\log_2(3\cdot 3^n) &\leq \lfloor n \log_2 3 \rfloor + 3\\ n\log_2 3 + \log_2 3 &\leq \lfloor n \log_2 3 \rfloor + 3\\ n\log_2 3 - 1.415 &\leq \lfloor n \log_2 3 \rfloor\end{aligned}$$

□

Theorem 1. *Let us consider the binary growth of an odd integer v_1 to which we apply the Collatz function n times:*

$$\underbrace{\frac{3\dfrac{3\dfrac{3\dfrac{3v_1+1}{2^{\alpha_1}}+1}{2^{\alpha_2}}+1}{2^{\alpha_3}}+1}{2^{\alpha_4}}\cdots}_{n\ times}$$

Note that $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the largest possible exponents for which 2^{α_i}, $i = 1, 2, \ldots, n$ exactly divide the numerator. The binary growth Λ of v_1 in this case is as well upper bounded as follows:

$$\Lambda \leq \lfloor n \log_2 3 \rfloor + 2$$

Proof. We start again with an odd number v_1 and let it grow steadily due to applying the $x \mapsto 3x + 1$ operation n times.

When after these n steps we arrive at an even number v_n, we divide by 2^{α_n} which means that we delete n zeros from the binary representation.

In this case we arrive at a new odd number. According to Lemma 1 we are again at a point, where we cannot exceed the binary growth.

□

3. Problem Statement

Now the following argument was raised: Let v_i be a member of a Collatz sequence, for example $v_i = 17$. An *overflow point* is the next (nearest) power of two above v_i, in this case the overflow point is 32. Theoretically, the overflow point can move higher than the next power of two above 17 (due to multiplication by 3).

When considering a Collatz sequence starting at v_1 and ending with $v_{n+1} = 1$ and introducing a variable δ that represents the accumulation of "+1" we would obtain from equation 1:

$$1 = v_{n+1} = \frac{3^n v_1 \beta}{2^\alpha} = \frac{3^n v_1 + \delta}{2^\alpha}$$

The raised concern is now that nothing may prevent δ to grow larger than $3^n v_1$ possibly leading to $\beta > 2$, since $3^n v_1 \beta = 3^n v_1 + \delta$. Having $\beta > 2$, a beta larger than two would imply for $2^\alpha = 3^n v_1 \beta$ and thus for $\alpha = n \log_2 3 + \log_2 v_1 + \log_2 \beta$ that $\log_2 \beta > 1$ violating the inequality given by equation 2.

We can calculate δ directly using the following sum, see equation A.2 in appendinx of [3, p. 36]:

$$\delta = \sum_{j=1}^{n} 3^{j-1} 2^{\alpha_1 + \ldots + \alpha_n - \sum_{l > n-j} \alpha_l} \tag{4}$$

An example is the sequence $(v_1, v_2, v_3, v_4, v_5) = (37, 7, 11, 17, 13)$ where $v_1 = 37$, $n = 4$ and $v_{n+1} = v_5 = 13$. The beta is $\beta = (1 + {}^1\!/_{111})(1 + {}^1\!/_{21})(1 + {}^1\!/_{33})(1 + {}^1\!/_{51}) = {}^{3328}\!/_{2997}$. The alpha is $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 4 + 1 + 1 + 2 = 8$ and finally the delta is $\delta = 3^0 \cdot 2^{\alpha_1+\alpha_2+\alpha_3} + 3^1 \cdot 2^{\alpha_1+\alpha_2} + 3^2 \cdot 2^{\alpha_1} + 3^3 \cdot 2^0 = 3^0 \cdot 2^6 + 3^1 \cdot 2^5 + 3^2 \cdot 2^4 + 3^3 \cdot 2^0 = 331$.

Indeed it appies

$$v_{n+1} = v_5 = \frac{3^4 \cdot 37 \cdot {}^{3328}\!/_{2997}}{2^8} = \frac{3328}{2^8} = \frac{3^4 \cdot 37 + 331}{2^8} = \frac{2997 + 331}{2^8} = 13$$

Halbeisen and Hungerbühler [4] introduced a function φ, which we can use to describe the δ. This function φ takes a binary number (binary string) s of length $l(s)$ as input and produces an integer output as follows:

$$\varphi(s) = \sum_{j=1}^{l(s)} s_j 3^{s_{j+1} + \ldots + s_{l(s)}} 2^{j-1} \tag{5}$$

Let us take for example the binary string $s = s_1 s_2 s_3 s_4 s_5 s_6 s_7 = 1000111 = 71$ as input for the function φ, which will yield the delta from our example $\delta = \varphi(1000111) = 331$:

$$
\begin{array}{llllll}
 & s_1 \cdot 3^{s_2+s_3+s_4+s_5+s_6+s_7}2^0 & & 1 \cdot 3^{0+0+0+1+1+1}2^0 & & 1 \cdot 3^3 2^0 \\
+ & s_2 \cdot 3^{s_3+s_4+s_5+s_6+s_7}2^1 & + & 0 \cdot 3^{0+0+1+1+1}2^1 & + & 0 \cdot 3^3 2^1 \\
+ & s_3 \cdot 3^{s_4+s_5+s_6+s_7}2^2 & + & 0 \cdot 3^{0+1+1+1}2^2 & + & 0 \cdot 3^3 2^2 \\
+ & s_4 \cdot 3^{s_5+s_6+s_7}2^3 & + & 0 \cdot 3^{1+1+1}2^3 & + & 0 \cdot 3^3 2^3 \\
+ & s_5 \cdot 3^{s_6+s_7}2^4 & + & 1 \cdot 3^{1+1}2^4 & + & 1 \cdot 3^2 2^4 \\
+ & s_6 \cdot 3^{s_7}2^5 & + & 1 \cdot 3^{1}2^5 & + & 1 \cdot 3^1 2^5 \\
+ & s_7 \cdot 3^{0}2^6 & + & 1 \cdot 3^{0}2^6 & + & 1 \cdot 3^0 2^6 \\
= & 331 & = & 331 & = & 331
\end{array}
$$

We can calculate 71 directly using the following procedure:

- Determining the length $l(s)$ of our binary number s, which is $\alpha - 1 = l(s)$. In our case the length is $l(s) = 8 - 1 = 7$.
- Determining the Hamming weight of our binary number s, which simply is n. In our case the Hamming weight is $n = 4$.
- Construct the binary number s by expanding the sequence $(v_1, \ldots, v_n)$ using the function
$$f(v) = \begin{cases} {}^{3v+1}\!/_{2} & 2 \nmid v \\ {}^{v}\!/_{2} & \text{otherwise} \end{cases}$$
In our case we get the expanded sequence $(37, 56, 28, 14, 7, 11, 17)$. By replaycing an odd member of this expanded sequence with 1 and an even member with 0 we obtain the binary number $s = 1000111 = 71$.

We have to proove that δ cannot exceed $3^n v_1$.

Halbeisen and Hungerbühler proved that for two distinct binary strings $s = s_1 s_2 \ldots s_l$ and $t = t_1 t_2 \ldots t_l$, which have the same Hamming weight, it applies [4]:

Theorem 2. *If* $\sum_{i=1}^{k} s_i \leq \sum_{i=1}^{k} t_i$ *for all* $k \in \{1, \ldots, l\}$ *then* $\varphi(s) > \varphi(t)$.

References

[1] C. Koch, E. Sultanow, and S. Cox. Divisions by two in collatz sequences: A data science approach. *International Journal of Pure Mathematical Sciences*, 21, 2020.

[2] Collag3n. Comment on answer to "a possible way to prove non-cyclicity of eventual counterexamples of the collatz conjecture?", December 2020.

[3] E. Sultanow, C. Koch, and S. Cox. Collatz sequences in the light of graph theory. Technical report, University of Potsdam, 2020.

[4] L. Halbeisen and N. Hungerbühler. Optimal bounds for the length of rational collatz cycles. *Acta Arithmetica*, 78(3):227–239, 1997.

Christian Koch, Technische Hochschule Georg Simon Ohm, Kessler square 12, 90489 Nuremberg, Germany

Email address: `christian.koch@th-nuernberg.de`

First Lastname, Graduate School of Mathematics, XYZ University, City, Adresszusatz, ZIP, Germany

Email address: `first.last@university.de`