



HAL
open science

Are the Collatz and abc conjectures related?

Olivier Rozier

► **To cite this version:**

| Olivier Rozier. Are the Collatz and abc conjectures related?. 2023. hal-04142385v1

HAL Id: hal-04142385

<https://hal.science/hal-04142385v1>

Preprint submitted on 26 Jun 2023 (v1), last revised 21 Oct 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Are the Collatz and abc conjectures related?

Olivier Rozier

Abstract

The Collatz and *abc* conjectures, both well known and thoroughly studied, appear to be largely unrelated at first sight. We show that assuming the *abc* conjecture true is helpful to improve the lower bound of integers initiating a particular type of Collatz sequences, namely finite sequences of a given length where all terms but one are odd with the usual “shortcut” form. To obtain sharper bounds in this context, we are led to consider a small subset of the *abc*-hits. Then, it turns out that Collatz iterations as well as Wieferich primes may be used to find large triples in this subset.

1 Introduction

The Collatz conjecture deals with positive integer sequences arising when repeatedly applying the function

$$T(n) = \begin{cases} \frac{3n+1}{2}, & \text{if } n \text{ is odd,} \\ \frac{n}{2}, & \text{otherwise.} \end{cases} \quad (1)$$

Originally, Lothar Collatz introduced those sequences with the transformation $n \mapsto 3n + 1$ for odd n , without dividing by 2 in the same step [9]. We say that T has a shortcut, since the number of steps is smaller for the same final result. It is expected that whatever the first term of the sequence, it always ends up reaching the trivial cycle $(1, 2)$. The latter assumption, called Collatz conjecture, is very popular under various names like the $3x + 1$ or Syracuse problem, even outside the mathematical community.

On the other hand, the *abc* conjecture of J. Oesterlé and D. Masser is considered as particularly important for its numerous implications in number theory [13, 21]. It poses a limitation on the presence of high powers of primes in Diophantine equations over three terms. There are various ways, more or less equivalent, to formulate this conjecture, which can be stated as follows:

Conjecture 1.1. (*abc* conjecture) *For every $\varepsilon > 0$, there exist only finitely many triples (a, b, c) of coprime positive integers satisfying $a + b = c$ and such that*

$$c > \text{rad}(abc)^{1+\varepsilon} \quad (2)$$

where $\text{rad}(abc)$ denotes the radical of abc , that is the product of its distinct prime factors.

The purpose of the present article is to highlight a possible relationship between the *abc* and Collatz conjectures. Apparently, no direct relationship is mentioned in the literature, despite a previous attempt¹ by M. Kaneda briefly related at the end of [8]. Let us recall that the *abc* conjecture originates in the theory of elliptic curves, an active field of research in number theory far from the Collatz conjecture [13].

Both conjectures are nevertheless connected to linear forms in logarithms and, thus, to Baker theorem [2]. This number theoretic result is central when studying the existence of non-trivial Collatz cycles [18, 19] and has a clear relationship with numerous Diophantine equations linked to the *abc* conjecture, e.g., Fermat last theorem [7, 13, 21]. Let us also point out that the set of Wieferich primes (namely, primes p such that $2^{p-1} - 1$ is a multiple of p^2) is another example of a topic that is somehow connected to both conjectures. On the one hand, the existence of infinitely many non-Wieferich primes, a question which is still open despite numerical evidence, is implied by the *abc* conjecture [17]. On the other hand, the set of Wieferich numbers, a natural extension of the aforementioned set of primes, is at the heart of a partial solution to several variants of the Collatz conjecture [6]. Specifically, the existence of sequences not leading to 1 is established in the so-called “ $qx + 1$ ” variants for q a Wieferich number, whereas the existence of sequences that diverge to infinity remains elusive for every odd $q > 3$.

We start our study with a lower bound hypothesis (LBH) regarding the first term of Collatz sequences of finite length having a given number of odd and even terms. If proved true, such a statement is strong enough to settle the Collatz conjecture (Lemma 3.1 in [16]). Therefore, any progress regarding its validity might be of interest.

Hypothesis 1.2. (Lower Bound Hypothesis - LBH) *There is a real constant $C \geq 0$ such that for all positive integers j and n not both equal to 1, we have*

$$n \geq j^{-C} 2^{\left(1-H\left(\frac{q}{j}\right)\right)j} \quad (3)$$

where q is the number of odd integers in the vector $(n, T(n), \dots, T^{j-1}(n))$ of the first j iterates starting from n , and H is the binary entropy function $H(x) = -x \log_2(x) - (1-x) \log_2(1-x)$ for $0 < x < 1$, with $H(0) = H(1) = 0$.

In other words, it is stated that integers n whose parity vector

$$(n \bmod 2, \dots, T^{j-1}(n) \bmod 2)$$

has low binary entropy are unlikely to be found below a certain lower bound. This conjecture is known to hold true when $2q \leq j$ and when $q = j$.

¹In fact, Kaneda tried to apply the *n*-conjecture, which is a generalization of *abc* with n terms, first introduced in [3].

Like the *abc* conjecture, LBH is derived from heuristic considerations and supported by empirical data. In fact, both are linked to a particular representation of positive integers, their prime factorization for *abc* and their parity sequence for LBH [15]. Those representations may be viewed as a kind of information whose entropy is measurable and can be translated into probabilistic terms, then enabling accurate predictions.

First, we show in section 2 that the *abc* conjecture implies a lower bound reasonably close to (3) in the particular case of sequences where all terms but one are odd (that is $q = j - 1$). Then, in section 3, we take a closer look at the *abc* conjecture and study the set \mathcal{H}_μ of triples (a, b, c) of coprime positive integers such that $a + b = c$ and $\log c > \mu(abc)$ where

$$\mu(p_1^{n_1} \dots p_k^{n_k}) = \log(p_1 \dots p_k) + \log(n_1 \dots n_k)$$

for distinct primes p_1, \dots, p_k and positive integers n_1, \dots, n_k . It turns out that \mathcal{H}_μ is a very sparse subset of the *abc*-hits. In section 4, we prove that the lower bound (3) holds with $C = 1$ for the same Collatz sequences as in §2, unless maybe if we encounter an element of \mathcal{H}_μ . Finally, in section 5, we show that the question of whether \mathcal{H}_μ is infinite has a close connection with the number of Wieferich primes.

2 A particular type of Collatz sequences

For any integer $j \geq 2$, let us consider Collatz sequences of length j for which all terms but one are odd. Let $\mathcal{N}(j)$ denote the set of positive integers lower than 2^j initiating such sequences. The set $\mathcal{N}(j)$ contains exactly j elements. This is a simple consequence of the one-to-one correspondence, discovered by C. J. Everett [5] and by R. Terras [20], between the congruence classes modulo 2^j and the set of parity vectors of length j . E.g., one has

$$\mathcal{N}(10) = \{159, 239, 447, 511, 639, 681, 767, 795, 871, 1022\}.$$

We have shown in Lemma 6.1 of [16] that

$$n \geq 2^{j/(1+\rho)} - 2 \quad \text{for any } n \in \mathcal{N}(j) \tag{4}$$

with $\rho = \log_2 3 = 1.585 \dots$. Unfortunately, this result is far from the desired lower bound given by LBH in the particular case $q = j - 1$, which asserts that

$$n \geq j^{-C} 2^{(1-H(1-1/j))j} \quad \text{for any } n \in \mathcal{N}(j) \tag{5}$$

where the right-hand side grows asymptotically as $j^{-(C+1)} 2^j$ up to a multiplicative constant.

As long as LBH remains unsolved for this particular case, it only serves to specify the best lower bound we should expect to hold. However, it is so far unclear which value of

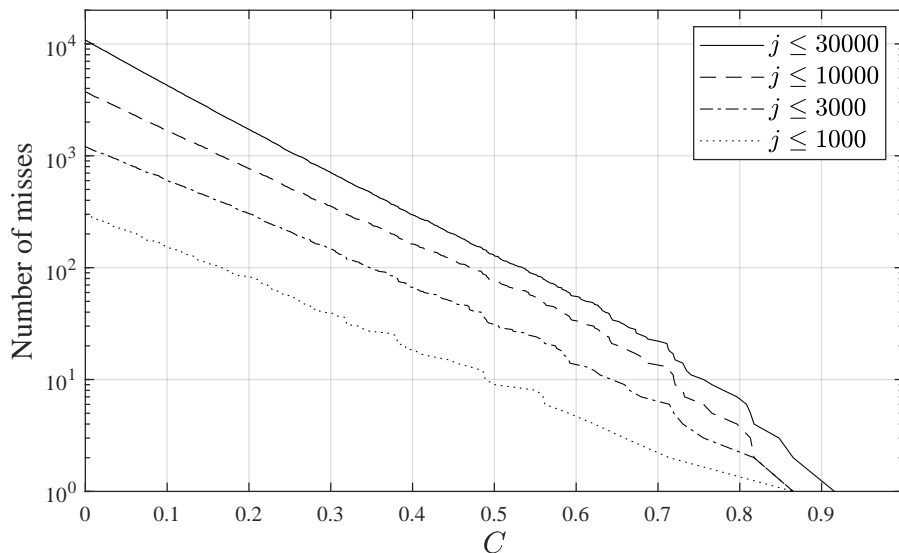


Figure 1: Log plot of the number of misses of Hypothesis 1.2 (LBH) with respect to the constant C arbitrarily fixed between 0 and 1, in the particular case of Collatz sequences of length j having exactly one even term ($q = j - 1$) for j up to 1000, 3000, 10000 and 30000.

C should be considered. Figure 1 shows that, for each $2 \leq j \leq 3 \cdot 10^4$, the inequality (5) is satisfied with $C = 1$ whereas it is increasingly falsified when C decreases below 0.9, suggesting that it is sharp with $C \approx 1$.

If we assume the *abc* conjecture true, we obtain a lower bound which is much better than (4) and comes close to (5).

Theorem 2.1. *Assume that the abc conjecture is true. Then, for every $\varepsilon > 0$, there exists a constant $K(\varepsilon) > 0$ such that*

$$n \geq K(\varepsilon) 2^{(1-\varepsilon)j} \quad \text{for all } j \geq 2 \text{ and } n \in \mathcal{N}(j). \quad (6)$$

Let us recall that the elements of $\mathcal{N}(j)$ are easy to compute by using, for each $0 \leq k \leq j - 1$, the congruence

$$n \equiv -1 - \left(\frac{2}{3}\right)^k \pmod{2^j} \quad (7)$$

given in [15, p4], where n is the element of $\mathcal{N}(j)$ such that $T^k(n)$ is even. Indeed, we have the congruences implied by (7)

- $T^i(n) \equiv -1 - \left(\frac{2}{3}\right)^{k-i} \pmod{2^{j-i}}$ for any $0 \leq i \leq k$,

- $T^i(n) \equiv -1 \pmod{2^{j-i}}$ as long as $k < i < j$,

so that all iterates $T^i(n)$, $i < j$, are odd except when $i = k$.

It is not difficult to see that Theorem 2.1 follows directly from the next lemma.

Lemma 2.2. *Assume that the abc conjecture is true. Then, for every $\varepsilon > 0$ and every integer j sufficiently large (i.e., for any $j \geq j_\varepsilon$ with j_ε depending only on ε), we have the lower bound*

$$n \geq \frac{1}{6} 2^{(1-\rho\varepsilon)j} - 1 \quad \text{for any } n \in \mathcal{N}(j) \quad (8)$$

with $\rho = \log_2 3$.

Proof. Let $j \geq 2$ and let $n \in \mathcal{N}(j)$ for which $T^k(n)$ is even with $0 \leq k \leq j-1$. Multiplying (7) by 3^k , we obtain

$$2^k + 3^k(n+1) = 2^j B$$

with B a positive integer. It turns out that $n+1$ is divisible by 2^k . Thus, we can write $n+1 = 2^k A$, which leads us to the equation

$$1 + 3^k A = 2^{j-k} B$$

with $3^k A$ and $2^{j-k} B$ relatively prime. Then, we apply the *abc* conjecture, assumed to be true, to the above triple for an arbitrary $\varepsilon > 0$. This gives the lower bound

$$2^{j-k} B \leq \text{rad} \left(2^{j-k} 3^k AB \right)^{1+\varepsilon},$$

except maybe for finitely many cases that can be ruled out if we assume j sufficiently large, that is $j \geq j_\varepsilon$ for a suitable j_ε . Since $\text{rad} \left(2^{j-k} 3^k AB \right) \leq \text{rad}(6AB) \leq 6AB$, we get

$$2^{j-k} \leq (6A)^{1+\varepsilon} B^\varepsilon$$

or, by taking the logarithm in base 2,

$$j - k \leq (1 + \varepsilon) \log_2(6A) + \varepsilon \log_2 B,$$

so that

$$\log_2 6(n+1) = k + \log_2(6A) \geq k + \frac{j - k - \varepsilon \log_2 B}{1 + \varepsilon}.$$

Note that $n \leq 2^j - 2$, as 2^j and $2^j - 1$ are not in $\mathcal{N}(j)$. Hence,

$$\begin{aligned} B &= \frac{3^k}{2^j}(n+1) + 2^{k-j} \leq \frac{3^k}{2^j}(2^j - 1) + 2^{k-j} \\ &= 3^k - \frac{3^k - 2^k}{2^j} \\ &\leq 3^k. \end{aligned}$$

Putting together all inequalities yields

$$\begin{aligned}
\log_2 6(n+1) &\geq k + \frac{j-k-\varepsilon\rho k}{1+\varepsilon} \\
&= \frac{j-\varepsilon(\rho-1)k}{1+\varepsilon} \\
&> \left(\frac{1-\varepsilon(\rho-1)}{1+\varepsilon}\right)j \\
&> (1-\varepsilon)(1-\varepsilon(\rho-1))j \\
&> (1-\varepsilon\rho)j.
\end{aligned}$$

□

Proof of Theorem 2.1. Let us fix $\varepsilon > 0$. We can further assume that $\varepsilon < 1$, otherwise (6) is trivially satisfied for $K(\varepsilon) = 1$. Lemma 2.2 applied to $\frac{\varepsilon}{\rho} > 0$ gives, for some j_0 taken sufficiently large,

$$n \geq \frac{1}{6} 2^{(1-\varepsilon)j} - 1 \quad \text{for any } j \geq j_0 \text{ and } n \in \mathcal{N}(j).$$

Now, we choose a positive integer j_1 such that

$$2^{-(1-\varepsilon)j_1} \leq \frac{1}{6} - K_1 \quad \text{for some } 0 < K_1 < \frac{1}{6}.$$

If we set $j_2 = \max(j_0, j_1)$, then we have

$$n \geq K_1 2^{(1-\varepsilon)j} \quad \text{for any } j \geq j_2 \text{ and } n \in \mathcal{N}(j).$$

Next, choose a constant $K_2 > 0$ such that

$$n \geq K_2 2^{(1-\varepsilon)j} \quad \text{for any } 1 < j < j_2 \text{ and } n \in \mathcal{N}(j).$$

This is always possible due to the finiteness of $\mathcal{N}(j)$. We obtain the lower bound (6) for the suitable constant $K(\varepsilon) = \min(K_1, K_2)$. □

One may object that the lower bound (6) derived from the *abc* conjecture is still weaker than the lower bound (5) derived from LBH. To help resolve this discrepancy, we propose in the next section to dive deeper into the *abc* conjecture and refine the notion of *abc*-hit.

3 A rare type of abc -hits

There exist infinitely many triples (a, b, c) of coprime positive integers such that $a + b = c$ and $c > \text{rad}(abc)$, which are called abc -hits or abc triples. To quantify how much $\text{rad}(abc)$ differs from c , one generally refers to the notion of *quality* defined by

$$q(a, b, c) = \frac{\log c}{\log \text{rad}(abc)},$$

which is a real number greater than 1 for any abc -hit. The abc conjecture asserts that, for every $\varepsilon > 0$, there exist only finitely many abc -hits with $q > 1 + \varepsilon$. For instance, the abc -hits of quality $q > 1.4$, assumed to be finite in number, are often called “good” abc triples. Their list is maintained on [1] and contains 241 good abc triples so far. Several algorithms have been proposed to search for triples of high quality, e.g., by using the continuous fraction expansion of algebraic numbers [3, 12].

Somehow, the quality q is not a metric well-suited to treat equations of the form

$$1 + 3^k A = 2^l B,$$

which are at the heart of the proof of Theorem 2.1. Indeed, the quality takes into account the size of the radicals, but the size of the exponents is poorly constrained.

We suggest replacing the radical by a different function where the size of the exponents also comes into play. Thus, let us introduce the function $\mu : \mathbb{Z}_{\geq 1} \rightarrow \mathbb{R}_{\geq 0}$ uniquely defined by two properties:

- (i) $\mu(p^n) = \log p + \log n$, for prime p and positive integer n ;
- (ii) $\mu(mn) = \mu(m) + \mu(n)$, for coprime positive integers m, n .

It follows from (ii) that $\mu(1) = 0$. The function μ is a crude measure of the number of digits (i.e., the amount of information) that are necessary to write the prime factorization of a positive integer, independently of the base of a given numeral system. One may apply a multiplicative constant to adjust this measure to a particular base. Recall that the amount of information of the leading digit should be weighted non-uniformly, according to Benford’s law. To this respect, the measure μ is consistent with Benford’s law. Furthermore, it satisfies the inequalities

$$\log \text{rad}(n) \leq \mu(n) \leq \log n, \tag{9}$$

and

$$\max(\mu(m), \mu(n)) \leq \mu(mn) \leq \mu(m) + \mu(n) \tag{10}$$

for all positive integers m, n . The properties (9) and (10) result from the well-known inequalities $a + b \leq ab \leq a^b$ for $a, b \geq 2$. If n is square-free, a double equality holds in

(9). Amazingly, one may interpret these properties in terms of the data compression that tends to occur when performing the prime factorization of integers. This compression is lossless and reversible, whereas taking the radical is not.

By analogy with the *abc*-hits, we define the set

$$\mathcal{H}_\mu = \{(a, b, c) \in \mathbb{Z}_{\geq 1}^3 : a + b = c, \gcd(a, b) = 1 \text{ and } \log c > \mu(abc)\}$$

and call μ -hits the elements of \mathcal{H}_μ . As a result of the first inequality in (9), the set \mathcal{H}_μ is trivially a subset of the *abc*-hits. It is also not empty as it contains the triple $(1, 239^2, 2 \cdot 13^4)$, which is the only μ -hit below one million.

To estimate the size of \mathcal{H}_μ , we analysed the exhaustive list of all the *abc*-hits below 10^{18} from the project ABC@home [1]. Among these 14 482 065 triples, we found only 464 μ -hits², of which 56 are also good *abc* triples. Additionally, we found 175 μ -hits among about 9 million *abc*-hits with c between 10^{18} and $2^{63} \approx 9.22 \cdot 10^{18}$ also identified by ABC@home.

In view of Figure 2, it appears that \mathcal{H}_μ contains an increasingly small proportion of the *abc*-hits as their size increases, and only a fraction of the good *abc* triples. Empirically, the number of μ -hits below x tends to follow a power law of the form $(x/x_0)^\alpha$ with $\alpha \approx \frac{2}{11}$ and $x_0 \approx 10^3$, unlike the good *abc* triples whose number is expected to be upper bounded according to the *abc* conjecture.

a	b	c	digits	q	g
1	239^2	$2 \cdot 13^4$	5	1.2540	0.139
2	$3^{10} \cdot 109$	23^5	7	1.6299	2.147
11	$7^3 \cdot 167^2$	$2 \cdot 3^{14}$	7	1.4283	0.389
1	$3 \cdot 5^5 \cdot 47^2$	$2^{18} \cdot 79$	8	1.4497	0.032
2^2	$3^{15} \cdot 5$	$17^4 \cdot 859$	8	1.3925	0.311

Table 1: List of the five smallest μ -hits (a, b, c) , number of digits of c , quality q and gain g .

Table 1 gives the list of the five smallest μ -hits, where the second one is the *abc*-hit with the highest known quality discovered by E. Reyssat [1, 3]. Much like the quality for *abc*-hits, it may be relevant to define a specific metric for μ -hits. We propose to use

$$g(a, b, c) = \log c - \mu(abc)$$

which measures the gain of, say, “digital information” when expressing the prime factorization of *abc* relatively to the standard writing of c . It is interesting to observe that about

²The list of μ -hits below 10^{18} is available at https://www.ipgp.fr/~rozier/abc/mu_hits_1018.txt.

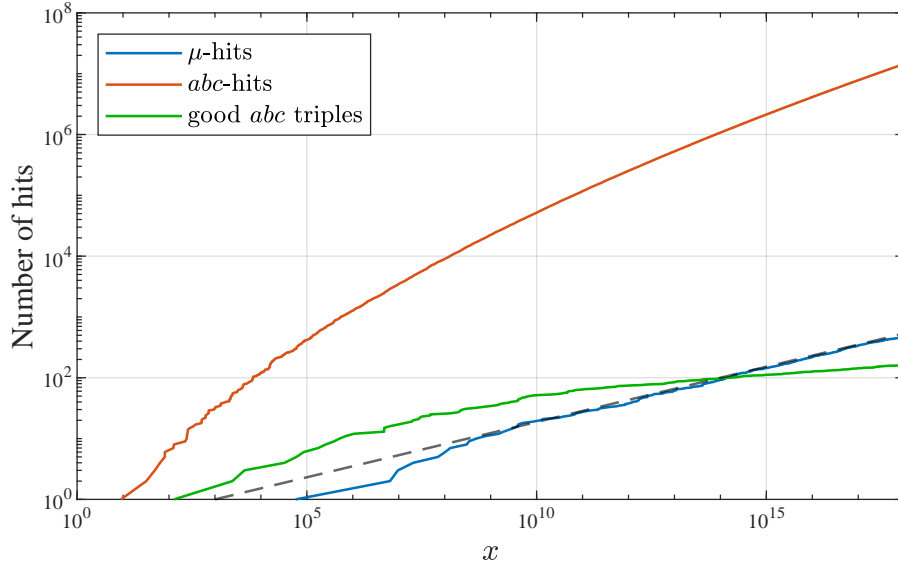


Figure 2: Log-log plot of the number N of μ -hits (blue), abc -hits (red) and good abc triples (green) below x . The dashed line corresponds to the power law $N = (x/x_0)^\alpha$ with $\alpha = 2/11$ and $x_0 = 1000$.

75% of the μ -hits below 10^{18} have a gain $g < 1$. The largest gain in this range is

$$g(19 \cdot 1307, 7 \cdot 29^2 \cdot 31^8, 2^8 \cdot 3^{22} \cdot 5^4) \approx 4.55$$

from a triple originally discovered by J. Browkin and J. Brzeziński [3], ranked third by quality in [1]. Moreover, when considering the list of 241 good abc triples from [1], we find only 128 positive gains of which the largest is

$$g(2^2 \cdot 11, 3^2 \cdot 13^{10} \cdot 17 \cdot 151 \cdot 4423, 5^9 \cdot 139^6) \approx 6.87$$

from a triple ranked 15 by quality and discovered by A. Nitaj [12].

First, one may ask if there exists a formula or an algorithm that gives infinitely many μ -hits, as is the case for abc -hits. Indeed, we know formulas like

$$(1, n^{2^k} - 1, n^{2^k}) \quad \text{where } n < 2^{k+1} \text{ is odd,} \quad (11)$$

given in [10] with $n = 3$ and for which every triple is a abc -hit. According to our computations, those formulas rarely generate μ -hits. In other words, the gain of the resulting triples is most often negative, but not by far in the case of (11) because $n^{2^k} - 1$ is divisible by 2^{k+2} . In fact, there seems to be no μ -hit of this form for $n < 239$. However, when

setting $n = 239$, we obtain a sequence of thirteen μ -hits for $k = 2, \dots, 14$ due to the divisibility of $n^{2^k} - 1$ by 2^{k+4} and by $n^2 + 1 = 2 \cdot 13^4$ which turns out to be the smallest μ -hit, as shown in Table 1. The gain g is hard to compute with certainty for $k \geq 7$, but is fairly easy to estimate, assuming that no power of a large prime appears when factorizing the second term of the triple. Hence, for $k \geq 2$, we predict that the gain is close to the lower bound

$$\log\left(\frac{2 \cdot 13^3}{239}\right) - \log(k + 4)$$

which is steadily decreasing and becomes negative for $k \geq 15$. It yields that there exist very large μ -hits with as much as 38968 digits when setting $k = 14$.

Next, we investigated this issue with other methods (e.g., the transfer method [11]) known to generate infinite sequences of abc -hits. But we found very little μ -hits by using these methods.

As a result of the above observations, one must keep in mind that μ -hits are far more difficult to generate than abc -hits.

4 From Collatz sequences to μ -hits

In this section, we show how the notion of μ -hit arises in the context of Collatz sequences. According to the following result, indeed, a lower bound stronger and more accurate than that of Theorem 2.1 should hold true unless maybe if a μ -hit of a given form is encountered.

Theorem 4.1. *Let T denote the Collatz function. For all $j \geq 2$, if n is a positive integer such that the sequence $n, T(n), \dots, T^{j-1}(n)$ contains exactly one even term, at least one of the following holds true:*

1. We have the lower bound

$$n \geq \frac{2^{j+1}}{3j^2} - 1. \tag{12}$$

2. There is a μ -hit of the form $(1, b, b + 1)$ with odd $b = T^k(n) + 1$ and $k < j$.

Proof. Let n be an element of $\mathcal{N}(j)$ as defined in §2, and let k be the smallest integer for which $T^k(n)$ is even.

The case $k = 0$ gives $n = 2^j - 2$ by using (7) and statement 1 is satisfied. Hereafter, we assume that $1 \leq k \leq j - 1$.

We follow the proof of Lemma 2.2 which leads to an equation of the form

$$1 + 3^k A = 2^{j-k} B$$

with $n = 2^k A - 1$ and A, B positive integers. The first k iterates are $T^i(n) = 3^i 2^{k-i} A - 1$ for $i \leq k$, so that

$$T^k(n) = 3^k A - 1.$$

Setting $b = T^k(n) + 1$, we ask whether the triple $(1, b, b + 1)$ is in \mathcal{H}_μ . If it is in \mathcal{H}_μ , we are led to statement 2. Otherwise, we have the inequality

$$\begin{aligned}
\log(b + 1) &\leq \mu(b(b + 1)) \\
&= \mu(b) + \mu(b + 1) \\
&= \mu(3^k A) + \mu(2^{j-k} B) \\
&\leq \mu(3^k) + \mu(2^{j-k}) + \mu(A) + \mu(B), && \text{by using (10),} \\
&= \log 3 + \log k + \log 2 + \log(j - k) + \mu(A) + \mu(B) \\
&\leq \log 6 + \log(k(j - k)) + \log A + \log B, && \text{by using (9).}
\end{aligned}$$

It yields that

$$\begin{aligned}
\log A &\geq \log(b + 1) - \log B - \log(k(j - k)) - \log 6 \\
&= \log(2^{j-k} B) - \log B - \log(k(j - k)) - \log 6 \\
&= (j - k) \log 2 - \log(k(j - k)) - \log 6,
\end{aligned}$$

so that

$$n = 2^k A - 1 \geq \frac{2^j}{6k(j - k)} - 1.$$

To show that statement 1 holds true, observe that $k(j - k) \leq (\frac{j}{2})^2$. □

Let us point out that Theorem 4.1 is of interest for two reasons. On the one hand, the effective bound (12) in statement 1 fully agrees with the expected bound (5) derived from Hypothesis 1.2 (LBH) when setting $C = 1$. One may further verify that (12) is slightly stronger than (5) for $C = 1$. On the other hand, Theorem 4.1 also provides a method based on the Collatz function for finding large μ -hits.

In Theorem 4.1, statement 2 seems unlikely to occur, except maybe on rare occasions in view of the low density expectations regarding \mathcal{H}_μ , so that statement 1 should hold true in most cases. This prediction is straightforward to verify for Collatz sequences of various lengths j by computing the j elements of $\mathcal{N}(j)$. In practice, we conducted a systematic search for all $j \leq 5000$. Due to the large size of the numbers considered, we used trial division to obtain a partial factorization with a size limit of 10^6 on prime factors. Therefore, we possibly missed μ -hits involving powers of large primes, although the probability is low. For all $5000 < j \leq 30000$, we restricted the search to sequences already associated to a “near miss” of LBH when setting $C = 0$ (see Figure 1). Such a restriction should not affect the verification of statement 1, but we probably missed many μ -hits for which equality holds in LBH with $C < 0$.

j	k	digits	pow(A)	pow(B)	q	g	C
19	16	9	-	53^3	1.474	1.313	-1.285
85	56	32	-	-	1.115	0.097	0.865
108	85	47	-	$7^4 \cdot 31^3$	1.137	2.861	-0.783
160	26	53	7^8	-	1.110	1.825	-1.151
294	127	111	37^2	7^6	1.052	0.176	-0.987
626	382	251	-	17^2	1.023	0.588	0.659
783	677	354	-	$13^3 \cdot 43^4$	1.022	3.566	-0.758
861	15	259	7^2	$5^2 \cdot 11^2$	1.022	1.098	0.105
874	45	271	7^9	-	1.024	1.987	-1.009
1056	921	479	-	$47^3 \cdot 109^2$	1.013	0.858	-0.600
1094	2	329	$7^2 \cdot 13^2 \cdot 1093^2$	$(B = 1)$	1.016	2.145	-0.829
1357	1174	614	7^5	5^4	1.011	0.194	-0.494
1367	296	463	$7^2 \cdot 67^4$	-	1.015	0.697	-0.768
1475	231	485	-	$11^3 \cdot 13^3 \cdot 97^3$	1.016	1.469	-1.130
2035	606	719	-	$5^4 \cdot 11^6$	1.010	0.015	-0.890
2186	2	658	$7^2 \cdot 13^2 \cdot 1093^2$	$(B = 1)$	1.008	1.452	-0.844
3279	3	987	$7^2 \cdot 13^2 \cdot 1093^2$	$(B = 1)$	1.006	1.739	-0.716
3514	4	1057	3511^2	$(B = 1)$	1.004	0.524	-0.584
4370	2	1315	$7^2 \cdot 13^2 \cdot 1093^2$	$(B = 1)$	1.004	0.758	-0.857
4393	436	1396	-	$13^4 \cdot 23^2$	1.005	0.519	-0.175
5461	488	1723	5^2	-	1.004	1.024	0.813
6962	396	2159	-	-	1.003	0.063	0.766
13056	11808	6002	-	17^2	1.001	1.284	0.735
27466	13732	10678	-	17^2	1.001	1.874	0.916
28107	27795	13351	7^4	5^3	1.001	0.721	0.066

Table 2: For μ -hits of the form $(1, 3^k A, 2^{j-k} B)$ related to Collatz, length j of the corresponding sequence and index k of its unique even term, number of digits of $2^{j-k} B$, prime powers in the factorization of A and B , quality q and gain g of the triple, value of C for which the equality holds in LBH on j iterations.

According to the computation results, inequality (12) is always satisfied in this range. Unexpectedly, we found a number of cases where statements 1 and 2 hold together. These cases are linked to a number of μ -hits detailed in Table 2. Remarkably, the smallest triple referenced in this table is $(1, 3^{16} \cdot 7, 2^3 \cdot 11 \cdot 23 \cdot 53^3)$, which is one of the abc -hit with the highest known quality (rank 22 by quality to date) discovered by A. Nitaj [1, 3, 12]. Moreover, bound (12) is very sharp (by less than 1%) for the second μ -hit in Table 2.

One may distinguish two kinds of μ -hits in Table 2, those for which (12) is rather sharp (that is, when the value C in the last column of Table 2 is close to 1), and those more numerous for which high powers of primes occur in the factorization of A and B .

It is amazing to observe that five μ -hits (for which $B = 1$) referenced in Table 2 are closely related to the known Wieferich primes, 1093 and 3511. They belong to a large family of μ -hits of the form $(1, b, b + 1)$ with

$$b = 2^{m(p-1)} - 1, \quad m \geq 1 \text{ and } p \text{ a Wieferich prime.}$$

The associated Collatz sequences, of length $j = m(p - 1) + k$ with k the 3-adic valuation of b , start from the integer $n = (2^j - 3^k - 2^k)/3^k$, which is generally well above the bound (12), then leading to $T^k(n) = 2^{j-k} - 2$ and to $T^j(n) = 3^{j-k-1} - 1$. In the next section, we briefly analyse the family of μ -hits linked to Wieferich primes.

5 Further insights from Wieferich primes

According to the computations of the previous section, all μ -hits of the form $(1, 2^n - 1, 2^n)$ seem to occur when n is divisible by $p - 1$ for some Wieferich prime p . Conversely, if p is a Wieferich prime, one may wonder how many μ -hits there are of the form $(1, b_m, b_m + 1)$ with $b_m = 2^{m(p-1)} - 1$ and m a positive integer. Assuming that $(1, b_1, b_1 + 1)$ is a μ -hit and taking into account the divisibility of b_m by b_1 is not sufficient to ensure that the triple $(1, b_m, b_m + 1)$ is also a μ -hit for any m . Therefore, it is so far unclear if we can generate infinitely many μ -hits of this form as m increases. By setting successively $p = 1093$ and $p = 3511$, we find that the respective proportions of such μ -hits with $m \leq 1000$ are at least 24% and 27%, by checking only the powers of small primes. Note that these proportions tend to decrease when considering larger values of m . The next lemma shows that μ -hits are far more likely to occur when combining several Wieferich primes.

Lemma 5.1. *Let $k \geq 1$ and let p_1, \dots, p_k be distinct Wieferich primes. If we put $L = \text{lcm}(p_1 - 1, \dots, p_k - 1)$, then we have*

$$g(1, 2^L - 1, 2^L) > \log \left(\frac{p_1 \cdots p_k}{2^{k+1} L} \right).$$

Proof. Assuming that p_1, \dots, p_k are k distinct Wieferich primes, then, for any $1 \leq i \leq k$, $2^L - 1$ is divisible by $2^{p_i-1} - 1$ and, consequently, by p_i^2 . We infer that

$$\begin{aligned} \mu(2^L - 1) &\leq \mu(p_1^2 \cdots p_k^2) + \mu\left(\frac{2^L - 1}{p_1^2 \cdots p_k^2}\right), \quad \text{by using (10),} \\ &< \mu(p_1^2 \cdots p_k^2) + \log\left(\frac{2^L}{p_1^2 \cdots p_k^2}\right), \quad \text{by using (9),} \\ &= (k + L) \log 2 - \log(p_1 \cdots p_k), \end{aligned}$$

so that

$$\begin{aligned} g(1, 2^L - 1, 2^L) &> \log(p_1 \cdots p_k) - k \log 2 - \mu(2^L) \\ &= \log\left(\frac{p_1 \cdots p_k}{2^{k+1} L}\right). \end{aligned}$$

□

As a consequence of Lemma 5.1, the gain of a triple of the form $(1, 2^{p-1} - 1, 2^{p-1})$ with p a Wieferich prime is bounded from below by $-\log 4$, which is not enough to conclude that such a triple is a μ -hit. However, if we apply this lemma to the known Wieferich primes $p_1 = 1093$, $p_2 = 3511$, and put $L = \text{lcm}(p_1 - 1, p_2 - 1) = 49140$, we obtain the lower bound

$$g(1, 2^L - 1, 2^L) > \log\left(\frac{p_1 p_2}{8L}\right) = 2.278 \dots$$

without taking into account the divisibility of $2^L - 1$ by the prime powers 3^4 , 5^2 , 7^2 , and 13^2 . By doing so, we get the much higher bound

$$g(1, 2^L - 1, 2^L) > 8.228 \dots$$

which is likely to be the actual gain of this triple whose second element has 14793 digits. From this particular example, it is possible to generate thousands of large μ -hits of the form $(1, 2^{mL} - 1, 2^{mL})$ by choosing adequate values of m . E.g., when setting $m = q^k$ with q an odd prime such that $q - 1$ divides L (that is, $q \in \{3, 5, 7, 11, 13, 19, 29, \dots, 24571\}$), we expect the gain of the resulting triples to decrease very slowly as k increases, due to the fact that $2^{q^k L} - 1$ is divisible by $3^4 \cdot 5^2 \cdot 7^2 \cdot 13^2 \cdot 1093^2 \cdot 3511^2 \cdot q^k$, where q^k can be replaced by q^{k+1} whenever $2^L - 1$ is not divisible by q^2 . The proof by induction on k is straightforward.

Regarding the question of whether there exist infinitely many μ -hits, we present a heuristic argument directly related to the Wieferich primes. The main point is that this set of primes is expected to be infinite although only two are known, according to simple statistical considerations [4]. Indeed, one expects the number of primes p below x having

presumably a given property with probability $1/p$ to grow like $\log \log x$, when applying the prime number theorem. Such a reasoning, whose prediction can hardly be neglected, was named the “log log Philosophy” by Jean-Pierre Serre (see [14, p413]).

Theorem 5.2. *If the set of Wieferich primes is infinite, then so is the set \mathcal{H}_μ of μ -hits.*

Proof. First, we set $p_1 = 1093$, $p_2 = 3511$, and $L_2 = \text{lcm}(p_1 - 1, p_2 - 1) = 49140$. If we assume the set of Wieferich primes to be infinite, then there exist distinct Wieferich primes p_3, \dots, p_k not equal to p_1, p_2 for any $k \geq 3$. Setting $L_k = \text{lcm}(p_1 - 1, \dots, p_k - 1)$, we have

$$\begin{aligned} L_k &= \text{lcm}(L_2, p_3 - 1, \dots, p_k - 1) \\ &\leq 2^{2-k} L_2 (p_3 - 1) (p_k - 1), \quad \text{since all terms are even,} \\ &< 2^{2-k} L_2 p_3 \cdots p_k. \end{aligned}$$

Applying Lemma 5.1, we get

$$\begin{aligned} g(1, 2^{L_k} - 1, 2^{L_k}) &> \log \left(\frac{p_1 \cdots p_k}{2^{k+1} L_k} \right) \\ &> \log \left(\frac{p_1 p_2}{8 L_2} \right) = 2.278 \dots \end{aligned}$$

Thus, we obtain an infinite sequence of μ -hits of the form $(1, 2^{L_k} - 1, 2^{L_k})$ for $k = 2, 3, \dots$. However, the exponents L_k are not necessarily distinct and it can happen that $L_{k+1} = L_k$ for some k . To show that we obtain infinitely many distinct triples, it suffices to observe that $L_k \geq p_k - 1$, so that the sequence of powers 2^{L_k} is unbounded. \square

Conjecture 5.3. *There are infinitely many μ -hits.*

Finally, we formulate the above conjecture that reflects the preceding arguments and findings. Although our results need not assume its validity, it is important in view of Theorem 4.1.

6 Conclusion

Somehow, the results presented so far show that the Collatz and *abc* conjectures are more connected than previously thought and suggest that they may well be of comparable hardness. Nevertheless, the validity of the *abc* conjecture is very unlikely to imply that of the Collatz conjecture. Indeed, the general formulas linked to T iterations involve an unbounded number of terms which are mostly sums of powers of 2 and 3. Therefore, a generalization of the *abc* conjecture might be required to treat expressions with much

more terms than just a, b, c . Various statements have been proposed (see [3, 11]), but they are not well suited in this context. Further studies are necessary to strengthen the link between these not-so-distant conjectures and investigate how far the notion of μ -hit can contribute.

Acknowledgements

We are grateful to Nik Lygeros for helpful comments.

References

- [1] ABC@home, data available at <https://www.math.leidenuniv.nl/~desmit/abc/>.
- [2] A. Baker, *Transcendental Number Theory*, Cambridge University Press (1975).
- [3] J. Browkin, J. Brzeziński, Some remarks on the abc -conjecture, *Math. Comp.* **62** (1994), 931–939.
- [4] R. Crandall, K. Dilcher, C. Pomerance, A search for Wieferich and Wilson primes, *Math. Comp.* **66** (1997), 433–449.
- [5] C. J. Everett, Iteration of the number theoretic function $f(2n) = n, f(2n+1) = 3n+2$, *Adv. Math.* **25** (1977), 42–45.
- [6] Z. Franco, C. Pomerance, On a conjecture of Crandall concerning the $qx+1$ problem, *Math. Comp.* **64** (1995), 1333–1336.
- [7] A. Granville, T. Tucker, It’s as easy as abc , *Notices Amer. Math. Soc.* **49** (2002), 1224–1231.
- [8] M. Kaneda, Two remarks on the Collatz cycle conjecture, preprint available at <https://arxiv.org/abs/1010.6206> (2010), 1–8.
- [9] J. C. Lagarias (editor), *The Ultimate Challenge: The $3x+1$ Problem*, Amer. Math. Soc., 2010.
- [10] S. Lang, Old and new conjectured Diophantine inequalities, *Bull. Amer. Math. Soc.* **23** (1990), 37–75.
- [11] G. Martin, W. Miao, abc triples, *Funct. Approx. Comment. Math.* **55** (2016), 145–176.
- [12] A. Nitaj, Algorithms for finding good examples for the abc and Szpiro conjectures, *Exp. Math.* **2** (1993), 223–230.

- [13] J. Oesterlé, Nouvelles approches du “théorème” de Fermat, *Astérisque* **161-162**, Sémin. Bourbaki (1988), 165–186.
- [14] P. Ribenboim, *The New Book of Prime Number Records*, Springer-Verlag New York (1996).
- [15] O. Rozier, Parity sequences of the $3x + 1$ map on the 2-adic integers and Euclidean embedding, *INTEGERS* **19** A8 (2019), 1–25.
- [16] O. Rozier, The $3x + 1$ problem: a lower bound hypothesis, *Funct. Approx. Comment. Math.* **56** (2017), 7–23.
- [17] J. H. Silverman, Wieferich’s criterion and the *abc*-conjecture, *J. Number Theory* **30** (1988), 226–237.
- [18] J. Simons, B. de Weger, Theoretical and computational bounds for m -cycles of the $3n + 1$ problem, *Acta Arith.*, **117** (2005), 51–70.
- [19] R. P. Steiner, A theorem on the Syracuse problem, *Proc. 7th Manitoba Conf. Num. Math.* (1977), 553–559.
- [20] R. Terras, A stopping time problem on the positive integers, *Acta Arith.* **30** (1976), 241–252.
- [21] M. Waldschmidt, Open Diophantine problems, *Mosc. Math. J.* **4** (2004), 245–305.

INSTITUT DE PHYSIQUE DU GLOBE DE PARIS, UNIVERSITÉ PARIS CITÉ, FRANCE
E-mail: rozier@ipgp.fr