Almost all Collatz orbits attain almost bounded values

Terence Tao

University of California, Los Angeles

Apr 15, 2020
Let \( \mathbb{N} = \{0, 1, 2, \ldots \} \) be the natural numbers.

Define the **Collatz map** \( \text{Col} : \mathbb{N} + 1 \to \mathbb{N} + 1 \) by setting
\[
\text{Col}(N) := \frac{N}{2} \quad \text{when } N \text{ is even and } \text{Col}(N) = 3N + 1 \quad \text{when } N \text{ is odd.}
\]

Each positive integer \( N \in \mathbb{N} + 1 \) then generates a **Collatz orbit**
\[
\text{Col}^N(N) := \{ N, \text{Col}(N), \text{Col}^2(N), \ldots \}
\]

Each Collatz orbit has a **minimal value**
\[
\text{Col}_{\text{min}}(N) := \inf \text{Col}^N(N).
\]
Examples:

- $\text{Col}^N(1) = \{1, 4, 2\}$.
- $\text{Col}^N(6) = \{6, 3, 10, 5, 16, 8, 4, 2, 1\}$.

Terence Tao
Almost all Collatz orbits attain almost bounded values
Collatz conjecture

We have $Col_{\text{min}}(N) = 1$ for all $N \in \mathbb{N} + 1$.

Also known as the “$3x + 1$ problem” or “Syracuse problem”. This conjecture is a prime example of what Frank Harary (1969) called a mathematical disease, as it obeys the following three criteria:

- Must be easy to state
- Must seem to be accessible, even to an amateur
- Must also have been repeatedly “solved” (often by the same person)

Almost all Collatz orbits attain almost bounded values
The Collatz conjecture states that if you pick a number, and if it’s even divide it by two and if it’s odd multiply it by three and add one, and you repeat this procedure long enough, eventually your friends will stop calling to see if you want to hang out.

XKCD, Randall Monroe, March 5, 2010
Epidemiology

- 1937: Lothar Collatz infected with $\text{Col}$ (Collatz, 1986)
- 1950: first reported person-to-person transmission (informal lecture, Collatz, ICM, Cambridge Mass.)
- 1952: independent introduction by Thwaites
- 1950s: multiple infections reported at Yale, U. Chicago, and Syracuse University
- 1963: first detection of $\text{Col}$ in a mathematical journal
- 1970: Coxeter states the Collatz conjecture and offers $50 for proof
- 1983: Erdős offers $500 for a proof
- 1996: Thwaites offers £1000 for a proof
- 1999: 197 published papers on the conjecture (Lagarias, 2011)
- 2009: 341 published papers on the conjecture (Lagarias, 2012)
- 2012: Lagarias abandons effort to maintain Collatz bibliography
“For about a month everyone at Yale worked on it, with no result. A similar phenomenon happened when I mentioned it at the University of Chicago. A joke was made that this problem was part of a conspiracy to slow down mathematical research in the U.S.” (Shizuo Kakutani, 1960)

“Mathematics is not yet ripe enough for such questions.”, (Erdős 1983)

“This is an extraordinarily difficult problem, completely out of reach of present day mathematics.” (Lagarias, 2010)
Selected partial results and obstructions

- $\text{Col}_{\text{min}}(N) = 1$ for all $1 \leq N \leq 10^{20}$ (yoyo@home project, 2017).

- Apart from the cycle $1, 4, 2$, all other Collatz cycles must have length at least $17,087,915$ (Eliahou, 1993). Note that the Collatz conjecture rules out such cycles completely.

- $\{N \leq x : \text{Col}_{\text{min}}(N) = 1\} \gg x^{0.84}$ for all large $x$ (Krasikov-Lagarias, 2003).

- There are variants $\tilde{\text{Col}}$ of $\text{Col}$ (piecewise affine on arithmetic progressions) and initial data $N_0$ for which the assertion $\tilde{\text{Col}}_{\text{min}}(N_0)$ is undecidable (Conway, 1987).

- The Collatz conjecture implies that there are finitely many solutions to $2^n - 3^m = k$ for each $k$ (T., 2011). All known unconditional proofs of the latter claim require some variant of Baker’s theorem in transcendence theory.
One can work on the easier **statistical** version of the Collatz conjecture, where one only seeks to prove statements for “most” choices of initial data $N$.

- One has $Col_{\text{min}}(N) < N$ for almost all $N$ (Terras, 1976)
- One has $Col_{\text{min}}(N) < N^{0.869}$ for almost all $N$ (Allouche, 1979)
- One has $Col_{\text{min}}(N) < N^{0.7925}$ for almost all $N$ (Korec, 1994)

Here “almost all” is in the sense of natural density (one has $\sum_{N \in E; N \leq x} 1 = o_{x \to \infty}(x)$ for the set $E$ of exceptions). As we shall see, these results are analogous to “almost sure local well-posedness” results in PDE.
Theorem (T., 2019)

If \( f : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) is such that \( \lim_{N \to \infty} f(N) = +\infty \), then one has \( \text{Col}_{\text{min}}(N) < f(N) \) for almost all \( n \).

Thus for instance \( \text{Col}_{\text{min}}(N) < \log \log \log \log \log N \) for almost all \( N \). Here “almost all” is in the sense of logarithmic density (one has \( \sum_{N \in E; N \leq x} \frac{1}{N} = o_{x \to \infty}(\log x) \) for the set \( E \) of exceptions). This result is analogous to an “almost sure almost global well-posedness” result in PDE.
As observed by Ben Green, an equivalent formulation is

**Theorem**

For every $\varepsilon > 0$ there exists $C > 0$ such that $\text{Col}_{\text{min}}(N) \leq C$ for $N$ in a set of logarithmic density at least $1 - \varepsilon$.

**Corollary**

There exists $C_0$ such that $\text{Col}_{\text{min}}(N) = C_0$ for all $N$ in a set of positive logarithmic density.

With more effort it may be possible to replace “logarithmic density” by “natural density”, and in principle it may be able to perform a finite computation to allow one to take $C_0 = 1$. 
The proof of the result proceeds in several stages:

1. Passing to the Syracuse formulation of the problem
2. Reduction to locating an “almost invariant” (or more precisely “almost self-similar”) measure
3. Reduction to establishing mixing of a “Syracuse probability measure” $\text{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$
4. Reduction to controlling Fourier coefficients of $\text{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$
5. Reformulation in terms of random walks hitting triangles
6. Exploiting a key separation property of triangles
**Syracuse formulation**

- The dynamics of the Collatz map $\text{Col}$ can be clarified by skipping the even numbers and looking only at the action on the odd numbers $2N + 1 = \{1, 3, 5, \ldots\}$.
- Define the **Syracuse map** $\text{Syr} : 2N + 1 \rightarrow 2N + 1$ by

$$
\text{Syr}(N) := \frac{3N + 1}{2^{\nu_2(3N+1)}},
$$

where $\nu_2(N)$ is the 2-valuation of $N$ (the largest $j$ such that $2^j$ divides $N$).

- It is easy to see that for any odd $N$ and $k \geq 0$, $\text{Syr}^N(N) \subset \text{Col}^N(2^k N)$ and $\text{Syr}_{\text{min}}(N) = \text{Col}_{\text{min}}(N)$. Hence the Collatz conjecture is equivalent to asserting that $\text{Syr}_{\text{min}}(N) = 1$ for all $N \in 2N + 1$, and the main theorem equivalent to asserting that $\text{Syr}_{\text{min}}(N) \leq f(N)$ for almost all $N \in 2N + 1$.  

Terence Tao

Almost all Collatz orbits attain almost bounded values
A heuristic proof of the (statistical) Collatz conjecture

- If $N$ is a “random” odd number, then $3N + 1$ is even, will be divisible by 4 “with probability $1/2$”, divisible by 8 “with probability $1/4$”, and so forth.

- This leads to the heuristic probabilistic distribution
  
  $\nu_2(3N + 1) \equiv \text{Geom}(2)$, where $\text{Geom}(2) \in \mathbb{N} + 1$ is a random variable with the geometric distribution of mean 2:

  $$\Pr(\text{Geom}(2) = k) = 2^{-k}$$

- We therefore morally have $\text{Syr}(N) \approx \frac{3}{2^{\text{Geom}(2)}} N$, or
  $$\log \text{Syr}(N) \approx \log N + \log 3 - \text{Geom}(2) \log 2.$$
  Thus we predict the dynamics $\log \text{Syr}^n(N)$, $n = 0, 1, 2, \ldots$ to behave like a random walk with drift

  $$\mathbb{E}(\log 3 - \text{Geom}(2) \log 2) = \log \frac{3}{4} < 0.$$ 

- The law of large numbers then predicts the orbit should almost surely drift down to 1.
A plot of $n \mapsto \log \text{Syr}^n(N)$, compared against the linear drift $n \mapsto \log N + n \log \frac{3}{4}$, with $N = 27382591$. 

Terence Tao
Almost all Collatz orbits attain almost bounded values
Can this heuristic be made rigorous?

- One can calculate

\[ \text{Syr}^n(N) = \frac{3^n}{2^{a_1+\cdots+a_n}} N + F_n(a_1, \ldots, a_n) \]

where \( a_1, \ldots, a_n \) are the valuations

\[ a_i := \nu_2(3\text{Syr}^i(N) + 1) \]

and \( F_n(a_1, \ldots, a_n) \) is the offset

\[ F_n(a_1, \ldots, a_n) := 3^{n-1} 2^{-a_1-\cdots-a_{n-1}} + 3^{n-2} 2^{-a_2-\cdots-a_{n-1}} + \cdots + 2^{-a_n} \]

- For short times \( n \leq c \log N \), \( c \) small, and \( N \) uniformly distributed from 1 to some large \( x \), it is not difficult to show that \( a_1, \ldots, a_n \) behave like independent copies of \textbf{Geom}(2). The law of large numbers gives \( a_1 + \cdots + a_n \approx 2n \) for almost all \( N \).
Thus we can basically show that

$$\text{Syr}^n(N) \approx \frac{3^n}{2^{2n}} N$$

for \( n \leq c \log N \) and almost all \( N \). This sort of “short time” control lets us show results of the form \( \text{Syr}_{\min}(N) \leq N^{1-c} \) for almost all \( N \) and some small \( c > 0 \), in the spirit of previous results of Allouche and Korec.

However, for long times \( n > c \log N \), we lose control on the distribution of the valuations \( a_n \).
What if we iterate?

- The above arguments show that for almost all $N$, we have can find an element $M = \text{Syr}^n(N)$ of the Syracuse orbit of $N$ with $M \leq N^{1-c}$. It is then tempting to iterate and claim that we can find an element $M' = \text{Syr}^{n'}(M)$ of the Syracuse orbit of $M$ with $M' \leq M^{1-c} \leq N^{(1-c)^2}$. This would imply that $\text{Syr}_{\text{min}}(N) \leq N^{(1-c)^2}$, and one could hope to use further iteration to improve the bound even more.

- But there is an obstruction: if $N$ is uniformly distributed on some interval $[1, x]$, the output $M$ is not going to be uniformly distributed (on, say, $[1, x^{1-c}]$). In particular it is conceivable that the Syracuse dynamics forces the output $M$ to lie in the small exceptional set of the short time almost sure result, thus preventing us from iterating.
The problem here is that the Syracuse map \( \text{Syr}(N) = \frac{3N+1}{2^\nu_2(3N+1)} \) is biased with respect to the uniform distribution: if \( N \) is a uniformly distributed random variable (on some interval), then \( \text{Syr}(N) \) will not be uniformly distributed.

For instance, it is obvious that \( \text{Syr}(N) \) is never a multiple of 3. Furthermore, for \( n \geq 1 \), \( \text{Syr}^n(N) \) will be biased to be twice as likely to equal 2 mod 3 than 1 mod 3.

For \( n \geq 2 \), \( \text{Syr}^n(N) \) acquires further biases modulo 3^2; for \( n \geq 3 \), there are biases modulo 3^3; and so forth.
Histogram of $N \mapsto \text{Syr}^n(N) \mod 3$ for odd $N \leq 10^4$ and $n = 0, 1$, together with the limiting Syracuse measure.
Histogram of $N \mapsto \text{Syr}^n(N) \mod 3^2$ for odd $N \leq 10^4$ and $n = 0, 1, 2$, together with the limiting Syracuse measure.
Histogram of $N \mapsto \text{Syr}^n(N) \mod 3^3$ for odd $N \leq 10^4$ and $n = 2, 3$, together with the limiting Syracuse measure.

Terence Tao

Almost all Collatz orbits attain almost bounded values
This data suggests that the Syracuse dynamics, when viewed modulo $3^n$, converges to an invariant measure in $\mathbb{Z}/3^n\mathbb{Z}$. This can in fact be proven, and the measure is the probability distribution of the Syracuse random variable

$$\text{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) := 3^{n-1}2^{-a_1}\cdots a_n + 3^{n-2}2^{-a_2}\cdots a_n + \cdots + 2^{-a_n} \mod 3^n$$

where $a_1, \ldots, a_n$ are iid copies of $\text{Geom}(2)$. 
One can use these measures to propose an “invariant measure” \( \mu_x \) on \([1, x]\). Roughly speaking, this is the logarithmic measure \( \sum_{n \leq x} \frac{1}{n} \delta_n \), weighted by the density of the Syracuse random variable for some \( n = \lceil c \log x \rceil \), and then normalised to be a probability measure.

If one can show that some (variable length) iteration of the Syracuse map pushes forward \( \mu_x \) to \( \mu_{x^{1-c}} \) with small error (e.g., \( O(\log^{-c} x) \) in total variation norm), then on iteration we can push forward \( \mu_x \) to be less than \( f(x) \) with probability \( 1 - o(1) \) for any \( f \) that goes to infinity as \( x \to \infty \).

This idea was motivated by arguments of Bourgain (1994) exploiting an invariant measure for the nonlinear Schrödinger equation to boost a almost sure local well-posedness result to an almost sure global well-posedness result.
After some routine calculations, we can check that this strategy works as long as the Syracuse random variables $\text{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ stabilise as $n \to \infty$. Specifically, it will suffice to show that

$$d_{TV}(\text{Syrac}(\mathbb{Z}/3^n\mathbb{Z}), \text{Ext}_{\mathbb{Z}/3^n\mathbb{Z}}(\text{Syrac}(\mathbb{Z}/3^m\mathbb{Z}))) \ll_A m^{-A}$$

for all $1 \leq m \leq n$ and $A > 0$, where $\text{Ext}_{\mathbb{Z}/3^n\mathbb{Z}}(\text{Syrac}(\mathbb{Z}/3^m\mathbb{Z}))$ is an element of $\mathbb{Z}/3^n\mathbb{Z}$ that reduces to $\text{Syrac}(\mathbb{Z}/3^m\mathbb{Z})$ modulo $3^m$, uniformly chosen amongst all such choices.
The Syracuse random variables obey a “skew-convolution” relation that roughly speaking takes the form

\[ \text{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = \text{Syrac}(\mathbb{Z}/3^m\mathbb{Z}) + 3^m 2^{-a_1 - \cdots - a_m} \text{Syrac}(\mathbb{Z}/3^{n-m}\mathbb{Z}) \]

where \( a_1, \ldots, a_m \) are the independent geometric variables involved in defining \( \text{Syrac}(\mathbb{Z}/3^m\mathbb{Z}) \).

Using (a rigorous version of) this relation and some Fourier analysis, one can obtain the desired stabilisation property if one can establish the Fourier uniformity property

\[ E \exp(2\pi i \xi \text{Syrac}(\mathbb{Z}/3^n\mathbb{Z})/3^n) \ll_A n^{-A} \]

whenever \( \xi \in \mathbb{Z}/3^n\mathbb{Z} \) is not divisible by 3.
The Syracuse random variable

\[ \text{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) := 3^{n-1}2^{-a_1-\cdots-a_n} + 3^{n-2}2^{-a_2-\cdots-a_n} + \cdots + 2^{-a_n} \mod 3^n \]

is not quite a sum of independent random variables, so the characteristic function

\[ \mathbb{E}\exp(2\pi i \xi \text{Syrac}(\mathbb{Z}/3^n\mathbb{Z})/3^n) \]

does not easily factor.
However, if we group the terms into pairs, and introduce the sums

\[ b_1 := a_1 + a_2, \quad b_3 := a_3 + a_4, \ldots \]

one can rewrite the above random variable as

\[ 3^{n-2} 2^{-b_2 + \cdots + b_{n/2}} \left( 3 \times 2^{-b_1} + 2^{-a_2} \right) \]
\[ + 3^{n-4} 2^{-b_3 + \cdots + b_{n/2}} \left( 3 \times 2^{-b_2} + 2^{-a_4} \right) \]
\[ + \cdots + \left( 3 \times 2^{-b_{n/2}} + 2^{-a_n} \right). \]

The key point is that once one conditions on the sums \( b_1, \ldots, b_{n/2} \), the summands are independent and now the characteristic function factorises!
\[3^{n-2} 2^{-b_2} + \cdots + b_{n/2}(3 \times 2^{-b_1} + 2^{-a_2})
+ 3^{n-4} 2^{-b_3} + \cdots + b_{n/2}(3 \times 2^{-b_2} + 2^{-a_4})
+ \cdots + (3 \times 2^{-b_{n/2}} + 2^{-a_n}).\]

Modulo a minor technicality (ignored here), the problem now boils down to this: uniformly for all \(\xi \in \mathbb{Z}/3^n\mathbb{Z}\) not divisible by 3, show that with high probability a large number of the quantities

\[3^{n-2} 2^{-b_2} + \cdots + b_{n/2} \xi / 3^n, 3^{n-4} 2^{-b_3} + \cdots + b_{n/2} \xi / 3^n, \ldots, \xi / 3^n\]

are not close to an integer.
This problem can be phrased geometrically as follows. Color a point \((j, k)\) in the lattice \(\mathbb{Z}^2\) “black” if \(3^{2j}2^{-k}\xi/3^n\) is close to an integer, and white otherwise. The task is then to show with high probability, the two-dimensional random walk

\[(0, 0), (1, b_1), (2, b_1 + b_2), \ldots\]

passes through a lot of white points. (Due to a technicality, we actually work with a subset of this random walk known as a two-dimensional renewal process, but we ignore this subtlety here.)
We are coloring points \((j, k)\) black when \(3^{2j}2^{-k}\frac{\xi}{3^n}\) is close to an integer.

Using elementary number theory, one can show that no matter what \(\xi\) is, the set of black points organise into disjoint right angled “triangles” (of slope \(\log 9/\log 2\)); the hypothesis that \(\xi\) is not a multiple of three places some upper limit on the size of the triangles. The question then boils down to showing that a certain random walk with drift will encounter a large number of non-black points.
Almost all Collatz orbits attain almost bounded values

\[ j = 1 \]

\[ j = \left\lfloor \frac{n}{2} - c \log \frac{1}{\epsilon} \right\rfloor \quad j = \left\lfloor \frac{n}{2} \right\rfloor \]
Small black triangles are not much of a problem, because they happen to be surrounded by a “moat” of white points. The enemy is the large black triangles, and in particular if the random walk hops from one large triangle to another without much of a gap of white points between the triangles.

A key geometric observation saves us: if a large triangle lies just beneath one or more other large triangles, the lower vertices of the latter triangles must be well separated from each other, so a random walk passing through the first large triangle is unlikely to immediately land on one of the latter triangles. Making this observation rigorous and quantitative concludes the argument.
Thanks for listening!