The Famous Collatz Conjecture

Dali Duraku

The Collatz conjecture is a fascinating problem in mathematics that has yet to be fully resolved. It has captured the attention of many mathematicians over the years and remains a topic of much interest and discussion in the field. The Collatz conjecture asks whether repeating two basic arithmetic operations will eventually transform any positive integer into 1. It is a famous unsolved problem in mathematics. Define a function f on the integers by

$$f(n)=\begin{cases} \frac{n}{2}, if\ n\ is\ even \\ 3n+1, if\ n\ is\ odd \end{cases}$$

Now setting our starting point to be $a_0=n$ for a natural number n we may define

$$a_n=f(a_{n-1})$$

to generate a sequence $a_0, a_1, a_2, \dots$. The Collatz conjecture is also known by other names such as the Syracuse problem, Kakutani's problem, Thwaites conjecture, and the Ulam conjecture, the 3x+1 conjecture, the 3n+1 conjecture. – postulates that this sequence will eventually reach 1. Some of the stories of these various names cast light on the difficulty of the problem at hand. For instance, S. Kakutani, hearing of the problem around 1960, circulated the problem around the mathematics department of Yale University. Due to the intractability of this problem a joke was made that this problem was part of a conspiracy to slow down mathematical research in the United States.

To demonstrate a Collatz sequence consider $a_0=7$ and observe that our sequence proceeds as

$$7 \rightarrow 22 \rightarrow 11 \rightarrow 34 \rightarrow 17 \rightarrow 52 \rightarrow 26 \rightarrow 13 \rightarrow 40 \rightarrow 20 \rightarrow 10 \rightarrow 5 \rightarrow 16 \rightarrow 8 \rightarrow 4 \rightarrow 2 \rightarrow 1.$$

If we take $a_0=n$, then the smallest i such that $a_i<a_0$ is called the stopping time of n. Similarly, the smallest k such that $a_k=1$ is called the total stopping time of n. From the definition of total stopping time we can see that another way of stating the Collatz conjecture is the statement that the total stopping time of every n is finite.

Although the Collatz conjecture has not been given a rigorous mathematical proof there is considerable empirical evidence for the conjecture. For instance, in a 2020 paper “Convergence verification of the Collatz problem” in *The Journal of Supercomputing*, Baring showed that the Collatz conjecture holds for all starting values up to 2^{68}. Furthermore, all Collatz sequences seem to end in the three-cycle (4; 2; 1). Of course, checking with a computer is not sufficient to establish the validity of the Collatz conjecture as there may be a counterexample when considering a number that is larger than all those we have previously checked.

Another way of viewing the conjecture is through the lens of graph theory. A directed graph is said to be weakly connected if it is connected when viewed as an undirected graph, i.e., for any two vertices there is path of edges jointing them, ignoring the directions on the edges. With this definition in hand we may give the Collatz conjecture as the statement that the Collatz graph of $T(n)$ on the positive integers is weakly connected.

We call the sequences of iterates $(n, T(n), T^2(n), \ldots)$ the trajectory of n. Thinking generally we may see that there are three possible behaviors of the trajectories when $n>0$: (i) the trajectory will be convergent with $T^k(n)=1$, (ii) the trajectory will be non-trivial and cyclic with the sequence $T^k(n)$ eventually becoming periodic with $T^k(n)$ not equal to 1 for any $k \geq 1$, or (iii) the trajectory will be divergent with $\lim_{k \to \infty} T^k(n)=\infty$.

The Collatz sequence for starting point $n=125$ is shown below.

125,376,188,94,47,142,71,214,107,322,161,484,242,121,364,182,91,274,137,412,206,103,310,
155,466,233,700,350,175,526,263,790,395,1186,593,1780,890,445,1336,668,334,167,502,251,
754,377,1132,566,283,850,425,1276,638,319,958,479,1438,719,2158,1079,3238,1619,4858,24
29,7288,3644,1822,911,2734,1367,4102,2051,6154,3077,9232,4616,2308,1154,577,1732,866,4
33,1300,650,325,976,488,244,122,61,184,92,46,23,70,35,106,53,160,80,40,20,10,5,16,8,4,2,1

The total stopping time for $n=125$ is seen to be 108. The seemingly haphazard trajectory is shown below in Figure 1.

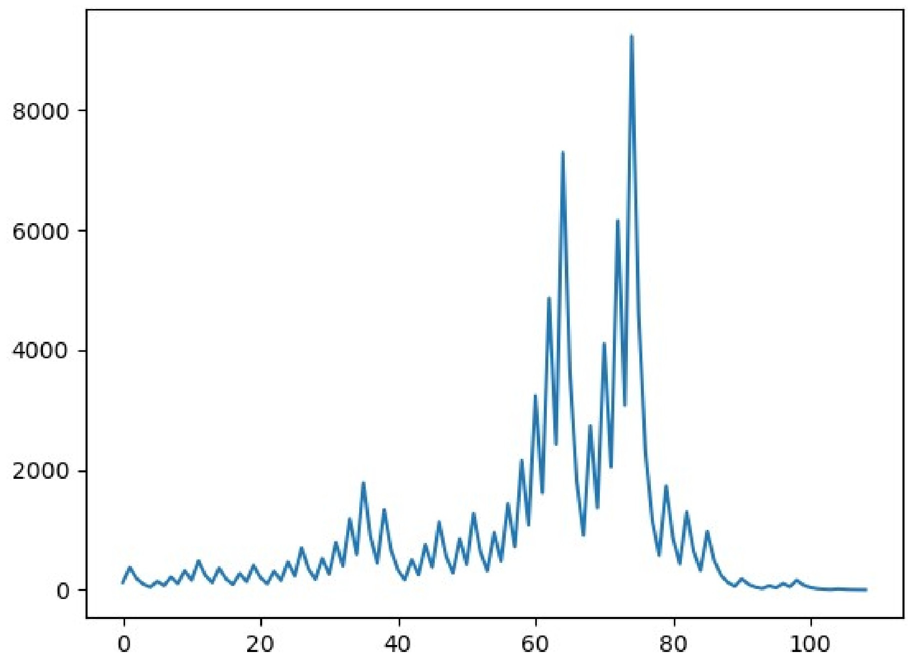

Figure 1: Visualization for trajectory with starting point n =125

For comparison examine the trajectory for starting points $n = 1223$ and $n = 2463$ both of which are shown in the figures below.

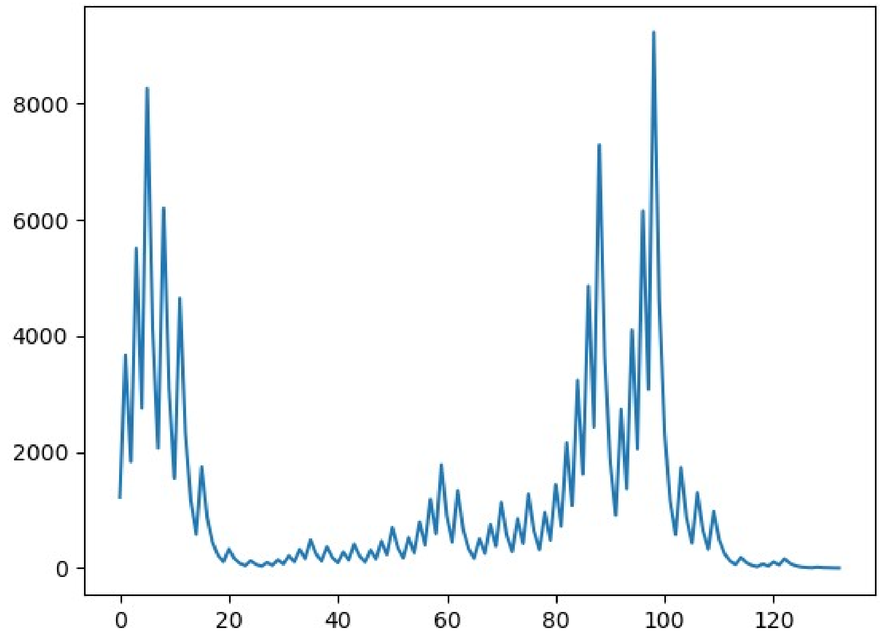

Figure 2: Visualization for trajectory with starting point n = 1223

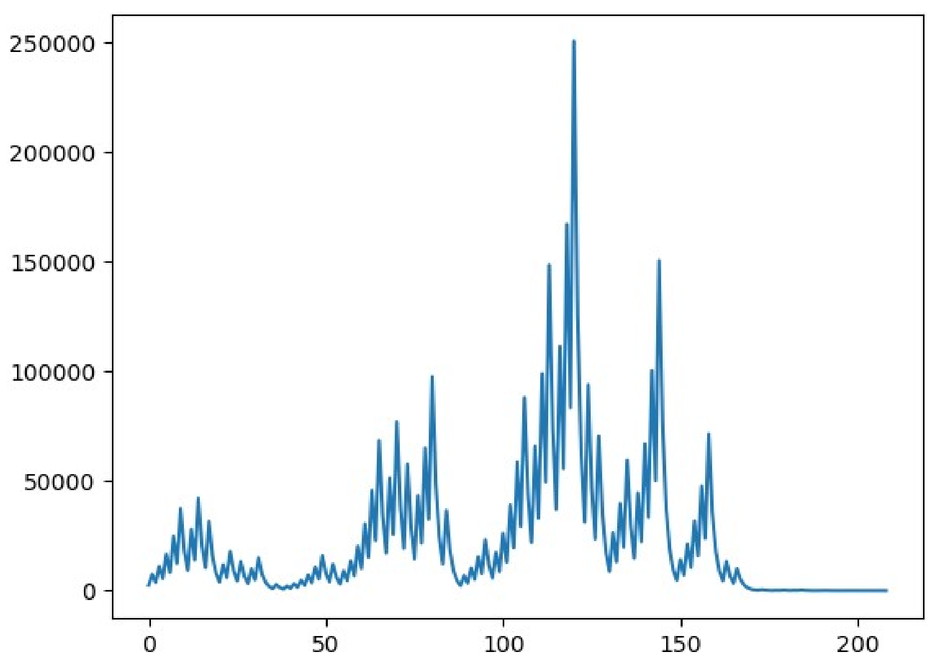

Figure 3: Visualization for trajectory with starting point n = 2463

An examination of the visualization of the trajectories above shows that although the trajectories are completely deterministic, in gazing at the trajectories one has the sense that one is gazing at random phenomena. As noted by Lagarias, the difficulty of resolving the Collatz conjecture is due to the fact that the sequences under consideration are seemingly random yet entirely deterministic. The task at hand is to find some hidden regularity within the seeming randomness in a general Collatz trajectory which will allow the conjecture to be resolved.

It is an observation of Terras that although the behavior of the total stopping time function is difficult to analyze, much can be said about the stopping time function. Terras proves the following result which is fundamental: The set of integers

$$S_k = \backslash\backslash\{ n : n \, has \, stopping \, time \leq k \}$$

has a limiting asymptotic density $F(k)$, i.e., the limit

$$F(k) = lim_{x \to \infty} \frac{\backslash\backslash\{ n : n \leq x \wedge \sigma(n) \leq k \}}{x}$$

exists. In addition, $F(k) \to 1$ as $k \to \infty$, so that almost all integers have a finite stopping time. This is a strong result about the stopping time function. Much less is known about the total stopping time function than about the stopping time function.

In searching for patterns about total stopping times one may observe the occurrence of many pairs and triples of integers which have the same total stopping time. For example the total stopping times for 20 and 21 is 6, the total stopping times for 12 and 13 is 7, and the total stopping times for 84 and 85 is 8. Indeed, for larger values of N, multiple consecutive values occur with the same total stopping time. For instance, there are 17 consecutive values of N with total stopping time equal to 40 for $7083 \leq N \leq 7099$. This observation further underscores that the Collatz trajectories are distinctly nonrandom though they might appear random. Another related phenomenon is that over short ranges of N the function for total stopping time tends to assume only a few values. For example, for $1000 \leq N \leq 1099$ we see that only 19 distinct total stopping times are observed.

These two phenomena have a plausible conceptual explanation. One can see that these phenomena are caused by distinct trajectories of different N's coalescing after a few steps. For example one may verify that the trajectories of $8k+4$ and $8k+5$ coalesce after 3 steps, for all $k \geq 0$.

A histogram of the total stopping times for the integers 1 through 100,000,000 is shown below in Figure 4

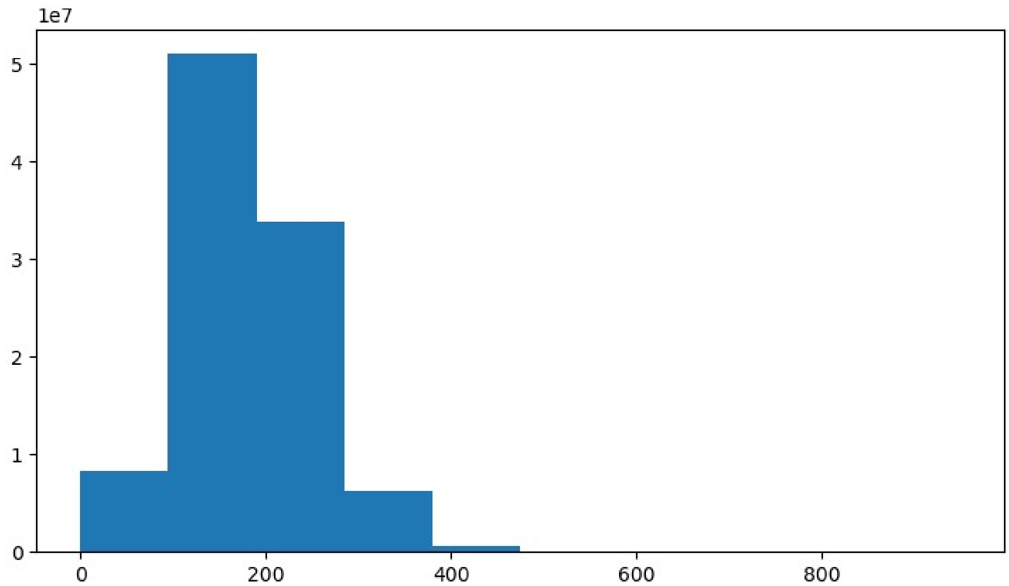

Figure 4: A histogram of total stopping times for N=1 through 100,000,000

One can see that the distribution takes a certain form which hints that perhaps this conjecture might be resolved with probabilistic methods.

The Collatz conjecture has remained remarkably intractable since it was first introduced and circulated by mathematicians around the leading research universities. One might ask, if the question is so intractable is it worthwhile to pursue such a question? Consider an aphorism related by Lagarias: “No problem is so intractable that something interesting cannot be said about it.” Indeed, often in attempting to resolve a problem, new fields of mathematics are developed and new insights into already existing fields of mathematics are developed. For example, consider the great advances in mathematics that occurred as a result of the attempts of various mathematicians to resolve Fermat’s Last Theorem. Hence, further study of the Collatz conjecture may should be encouraged as the pondering of any unresolved questions may lead to significant and interesting discoveries along the way.

Appendix: collatz module (Written by author)

The following python module is a collection of functions and methods we used to examine and visualize Collatz trajectories with various starting points.

```
# The collatz function returns the total stopping time of n

def collatz(n):
    count = 0
    while (n > 1):
        if n % 2 == 0:
            n = n/2
            count = count + 1
        else:
            n = 3 * n + 1
            count = count + 1
    return (count)

def collatz_seq(n):
    while (n > 1):
        print(int(n), end = ",")
        if n % 2 == 0:
            n = n/2
        else:
            n = 3 * n + 1
    print("1")

for i in range(1,100):
    print(collatz(i), end=",")
    if i == 99:
        print("...")

def collatz_list(n):
    list = []
    while (n > 1):
        list.append(int(n))
        if n % 2 == 0:
            n = n/2
        else:
            n = 3 * n + 1
    list.append(1)
    return(list)

# visualize the Collatz sequence and total stopping time for n =
125
# y = collatz_list(125)
# x = list( range(0, len(y) ) )
```

```
# plt.plot(x,y)

# visualize the Collatz sequence and total stopping time for n =
1223
# y = collatz_list(1223)
# x = list( range(0, len(y) ) )
# plt.plot(x,y)

# visuallize the Collatz sequence and total stopping time for n =
2463
# y = collatz_list(2463)
# x = list( range(0, len(y) ) )
# plt.plot(x,y)
```

References

Lagarias, Jeffrey (1985). "The 3x + 1 problem and its generalizations." *Amer. Math. Monthly* 92 (1985), no. 1, 3-23.

Terras, Riho (1976). "A stopping time problem on the positive integers". *Acta Arithmetica.* 30 (3): 241-252.