Density of Collatz Trajectories

J. Cooper Faile
The University of North Carolina at Chapel Hill

October 16, 2020

Abstract
In this paper I present a new method of studying the densities of the Collatz trajectories generated by a set \( S \subset \mathbb{N} \). This method is used to furnish an alternative proof that \( d(\{y \in \mathbb{N} : \exists k \text{ where } T^k(y) < cy\}) = 1 \) for all \( c > 0 \). Finally, I briefly discuss how the ideas presented in this paper could be used to improve the result that \( d(\{y \in \mathbb{N} : \exists k \text{ where } T^k(y) < y\}) = 1 \) for \( c > \log_4 3 \).

1 Introduction
Define \( T : \mathbb{N} \to \mathbb{N} \) by
\[
T(n) = \begin{cases} \frac{1}{2}n & \text{if } T^k(y) \text{ is odd} \\ \frac{3n+1}{2} & \text{otherwise} \end{cases}
\]
The Collatz conjecture states that for every \( n \in \mathbb{N} \) there exists \( k \in \mathbb{N} \) such that \( T^k(n) = 1 \). This conjecture is equivalent to the inductive statement that for each \( n \in \mathbb{N} + 1 \) there exists \( k \in \mathbb{N} \) such that \( T^k(n) < n \). This equivalent statement led to the study of the density of subsets \( S \subset \mathbb{N} \) which satisfy, for all \( n \in S \), \( T^k(n) < f(n) \) for functions \( f : \mathbb{N} \to \mathbb{R} \) (i.e. subsets of \( \mathbb{N} \) for which the sequence \( n, T(n), \ldots, T^k(n), \ldots \) contains a sufficiently “small” iterate.) Terras showed that the natural density of \( S \) was 1 for \( f(n) = n \), Venturini showed the same result for \( f(n) = c_1 n \), \( c_1 > 0 \), and Korek showed it for \( f(n) = n^{c_2} \), where \( c_2 > \log_4 3 \).

2 Notation and Prior Results
This paper requires the following notation from [1]. Given \( k, m, y \in \mathbb{N} \) and \( d \in \mathbb{R} \) we define
\[
X_k(y) = \begin{cases} 1 & \text{if } T^k(y) \text{ is odd} \\ 0 & \text{if } T^k(y) \text{ is even} \end{cases}
\]
\[
S_k(y) = X_0(y) + X_1(y) + \cdots + X_k(y)
\]
\[
U(m, d) = \{|y \in \mathbb{N} : 0 \leq y < 2^m \text{ and } S_m(y) \leq md|\}
\]

---

*I am a recent graduate who completed this work under the guidance of Professor Idris Assani.
We also need this lemma from [1]:

**Lemma 2.1.** We have that for any \( b \in \mathbb{N} \)

\[
U(m, d) = |\{ y : b \leq y < b + 2^m \text{ and } S_m(y) < md \}|
\]

and for any \( d > 1/2 \) we have

\[
limit_{m \to \infty} \frac{U(m, d)}{2^m} = 1.
\]

and another lemma, based on the proof in [3]:

**Lemma 2.2.** For any \( M \geq 3 \) there exists a \( d \in \mathbb{R} \) such that for all \( m > M \), we have that \( y \geq m2^m \) and \( S_m(y) < md \) implies \( T_m(y) < y/2 \).

**Proof.** Note that since \( y \geq m2^m \) we have that \( T_j(y) \geq y/2 \) for all \( 0 \leq j \leq m \). Letting \( k = S_m(y) \) we have the bound

\[
T_m(y) = y \cdot \frac{T(y)}{y} \cdot \cdots \cdot \frac{T_m(y)}{T_{m-1}(y)} < y \cdot \left( \frac{3m + 1}{2m} \right)^k \cdot \frac{1}{2^{m-k}}
\]

\[
= y \cdot \frac{3^k}{2^m} \cdot \left( 1 + \frac{1}{3m} \right)^k \leq y \cdot \frac{3^k}{2^m} \cdot \left( 1 + \frac{1}{3m} \right)^m < y \cdot \frac{3^k}{2^{m-1}}.
\]

The value \( y \cdot (3^k/2^{m-1}) < y/2 \) whenever \( 3^k < 2^{m-2} \), which is equivalent to

\[
S_m(y) = k < \frac{m - 2}{\log_2 3}.
\]

Since \( m > M \geq 3 \) we can select \( d \) such that \( dm < (m - 2)/\log_2 3 \) for \( m > M \), hence any \( d \) with \( 0 < d < (M - 2)/(M \log_2 3) \) satisfies this lemma.

---

### 3 Positive Density Preserving Functions

As a reminder,

**Definition 3.1.** Given \( A \subset \mathbb{N} \) the upper density of \( A \), denoted \( \overline{d}(A) \), is

\[
\lim_{n \to \infty} \frac{|A \cap [0, n]|}{n}.
\]

The lower density of \( A \), denoted \( \underline{d}(A) \), is

\[
\lim_{n \to \infty} \frac{|A \cap [0, n]|}{n}.
\]

If \( \overline{d}(A) = \underline{d}(A) \) we say \( A \) has **density** \( d(A) = \overline{d}(A) \).
To study the density of Collatz trajectories of \( S \subset \mathbb{N} \) we will consider functions which maintain positive density. This property guarantees that no set that is sufficiently large (i.e. \( \overline{d}(A) > 0 \)) has an image that is small (i.e. \( \overline{d}(f(A)) = 0. \)) This property can then be used to determine the density for certain sets.

**Definition 3.2.** A function \( H : 2^\mathbb{N} \to 2^\mathbb{N} \) is **positive upper density preserving** (which will be abbreviated \( \text{pdp} \)) if \( \overline{d}(A) > 0 \) implies \( \overline{d}(H(A)) > 0 \) for any \( A \subset \mathbb{N} \).

**Example 3.1.** \( T \) is \( \text{pdp} \). For any set \( A \subset \mathbb{N} \) with \( \overline{d}(A) > 0 \) let \( A = O \cup E \) where \( O, E \) are the odd and even elements of \( A \), respectively. We must have that one of \( \overline{d}(O) \) or \( \overline{d}(E) \) is greater than \( \overline{d}(A)/2 \). In these respective cases we have

\[
\begin{align*}
\overline{d}(T(A)) &\geq \overline{d}(T(E)) = 2\overline{d}(E) \geq \overline{d}(A) \\
\overline{d}(T(A)) &\geq \overline{d}(T(O)) = \frac{2}{3}\overline{d}(O) \geq \frac{1}{3}\overline{d}(A)
\end{align*}
\]

which shows that \( T \) is \( \text{pdp} \).

**Lemma 3.1.** If \( f \) and \( g \) is \( \text{pdp} \) then \( f \circ g \) and \( f^k \) is \( \text{pdp} \) for any \( k \in \mathbb{N} \).

**Proof.** If \( \overline{d}(A) > 0 \) then \( \overline{d}(g(A)) > 0, \overline{d}((f \circ g)(A)) > 0 \) showing \( f \circ g \) is \( \text{pdp} \). By induction the same reasoning shows that \( f^k \) is \( \text{pdp} \). \( \square \)

Now, given an increasing function \( f : \mathbb{N} \to \mathbb{R} \) we can define \( H_f : 2^\mathbb{N} \to 2^\mathbb{N} \) by

\[
H_f(A) = \{ T^k(n) : n \in A \text{ and } T^k(n) < f(n) \}.
\]

This map sends a subset \( A \subset \mathbb{N} \) to the set containing (partial) trajectories of every \( n \in A \). If this map is \( \text{pdp} \) for some \( f \) then it has some nice properties.

**Lemma 3.2.** If \( g(n) \leq f(n) \) for all \( n \in \mathbb{N} \) then \( H_g(A) \subset H_f(A) \) for all \( A \subset \mathbb{N} \). Additionally, if \( H_g \) is \( \text{pdp} \) then \( H_f \) is \( \text{pdp} \).

**Proof.** If \( m \in H_g(A) \) then there exist \( k \in \mathbb{N} \) and \( n \in A \) where \( m = T^k(n) < g(n) \). Since \( g(n) < f(n) \) we must have \( m \in H_f(A) \) and hence \( H_g(A) \subset H_f(A) \). If \( H_g \) is \( \text{pdp} \) then for all \( A \subset \mathbb{N} \) with positive upper density we have \( \overline{d}(H_f(A)) \geq \overline{d}(H_g(A)) > 0 \) showing that \( H_f \) is \( \text{pdp} \). \( \square \)

**Lemma 3.3.** For all \( A \subset \mathbb{N} \) and \( f, g : \mathbb{N} \to \mathbb{R} \) we have \( (H_f \circ H_g)(A) \subset H_{f \circ g}(A) \).

**Proof.** Any element of \( (H_f \circ H_g)(A) \) can be written as \( T^{k+k'}(n) \) for \( k, k', n \in \mathbb{N} \) satisfying \( T^k(n) < g(n) \) and \( T^{k+k'}(n) < f(T^k(n)) \) by the definitions of \( H_f \) and \( H_g \). Since \( f \) and \( g \) are increasing functions, we have \( T^{k+k'}(n) < f(T^k(n)) < f(g(n)) \) and hence \( T^{k+k'}(n) \in H_{f \circ g}(A) \). This establishes the inclusion. \( \square \)

**Corollary 3.1.** For all \( k \in \mathbb{N} \), \( A \subset \mathbb{N} \), and \( f : \mathbb{N} \to \mathbb{R} \) increasing we have \( (H_f)^k(A) \subset H_{f^k}(A) \). Furthermore, if \( H_f \) is \( \text{pdp} \) then so is \( H_{f^k} \).
Proof. \((H_f)^k(A) \subset H_{f^k}(A)\) follows directly from lemma 3.3. If \(H_f\) is \(\overline{dp}\) then by lemma 3.1 \((H_f)^k\) is \(\overline{dp}\) as well. Combining this with the inclusion \((H_f)^k(A) \subset H_{f^k}(A)\) we have \(\overline{d}((H_f)^k(A)) \leq \overline{d}(H_{f^k}(A))\) for all \(A \subset \mathbb{N}\). Hence \(H_{f^k}\) is \(\overline{dp}\).

**Lemma 3.4.** If \(H_f\) is \(\overline{dp}\) then the set

\[
M_f = \{y \in \mathbb{N} : \exists k \in \mathbb{N} \text{ where } T^k(y) < f(y)\}
\]

has density 1.

**Proof.** Let \(C = \mathbb{N} \setminus M_f\). We have \(T^k(n) \geq f(n)\) for every \(k \in \mathbb{N}\) and \(n \in C\). Hence, \(H_f(C) = \emptyset\) which has upper density zero, implying \(\overline{d}(C) = 0\) by contrapositive. Furthermore, \(\overline{d}(M_f) = 1 - \overline{d}(C) = 1\) establishing \(M_f\) has density 1.

The following theorem will be used with the previous lemmas to furnish an alternative proof of the result in [2].

**Theorem 1.** Let \(f : \mathbb{N} \rightarrow \mathbb{R}\) be \(f(y) = y/2\). Then \(H_f\) is \(\overline{dp}\).

**Proof.** Let \(A \subset \mathbb{N}\) be a set with \(\overline{d}(A) > 0\). From lemma 2.2 there exists \(d > 1/2\) such that for \(m\) sufficiently large all \(y \geq m2^m\) we have \(T^m(y) < y/2\) when \(S_m(y) < md\). Select \(m \in \mathbb{N}\) such that \(U(m, d)/2^m > 1 - \overline{d}(A)/4\) and it is large enough for the above implication. We can do this, as \(U(m, d)/2^m \rightarrow 1\) as \(m \rightarrow \infty\) for all \(d > 1/2\) due to lemma 2.1. Define the set

\[
L_{m,d} = \{y \in \mathbb{N} : S_m(y) < md\}.
\]

By definition of \(\overline{d}\) there exists a sequence \((a_i)_{i=1}^{\infty}\) where

\[
\frac{|A \cap [0, a_i]|}{a_i} \rightarrow \overline{d}(A).
\]

For any \(a_i > m2^m\) let \(b_i\) be the greatest integer for which \(b_i2^m \leq a_i\). Then the set \(A \cap [m2^m, b_i2^m]\) satisfies

\[
|A \cap [m2^m, b_i2^m]| = |A \cap [0, a_i]| - |A \cap [0, m2^m]| - |A \cap (b_i2^m, a_i]|
\leq |A \cap [0, a_i]| - |\mathbb{N} \cap [0, m2^m]| - |\mathbb{N} \cap [b_i2^m, a_i]|
\geq |A \cap [0, a_i]| - m2^m - 2^m.
\]

Dividing by \(a_i\) and taking the limsup we see

\[
\limsup_{i \rightarrow \infty} \frac{|A \cap [m2^m, b_i2^m]|}{a_i} \geq \overline{d}(A).
\]

Reindexing the sequence \((b_i)_{i=1}^{\infty}\), we can assume

\[
\frac{|A \cap [m2^m, b_i2^m]|}{a_i} \geq \frac{3\overline{d}(A)}{4}
\]
for all $i \in \mathbb{N}$. Additionally, note that due to our choice of $m$ and lemma [2.1] we have that
\[ L_{m,d}^c \cap [m2^m, b_i2^m] = (b_i - m) \cdot (2^m - U(m,d)) \leq \frac{\overline{d}(A)}{4}(b_i - m)2^m. \]

The set $A \cap [m2^m, b_i2^m] \cap L_{m,d}$ is a subset of $A$ for which $T^m(y) < y/2$ for of all it elements (due to the choice of $d$ and lemma [2.2]). Using the above bounds, the size of this set is at least
\[ |A \cap [m2^m, b_i2^m] \cap L_{m,d}| = |A \cap [m2^m, b_i2^m]| - |L_{m,d}^c \cap [m2^m, b_i2^m]| \geq \frac{3\overline{d}(A)}{4}a_i - \frac{\overline{d}(A)}{4}(b_i - m)2^m \geq \frac{\overline{d}(A)}{2}a_i. \]

Using this bound and that $|T^m(B)| \geq |B|/2^m$ for any $B \subset \mathbb{N}$ we have the following bound on the size of $H_f$ of this set:
\[ |H_f(A \cap [m2^m, b_i2^m])| \geq |\{T^m(y) : y \in A \cap [m2^m, b_i2^m] \text{ and } S_m(y) < md\}| \geq \frac{1}{2^m}|A \cap [m2^m, b_i2^m] \cap L_{m,d}| \geq \frac{a_i\overline{d}(A)}{2^{m+1}}. \]
Dividing this inequality by $a_i$ and noting that $H_f(A) \cap [0, a_i] \supset H_f(A \cap [0, a_i])$ gives the density estimate
\[ \frac{|H_f(A) \cap [0, a_i]|}{a_i} \geq \frac{|H_f(A \cap [m2^m, b_i2^m]|}{a_i} \geq \frac{a_i\overline{d}(A)}{2^{m+1}} = \frac{\overline{d}(A)}{2^{m+1}}. \]

This shows that $\overline{d}(H_f(A)) \geq \overline{d}(A)/2^{m+1} > 0$ and hence that $H_f$ is $\overline{pdp}$. □

**Corollary 3.2.** $H_f$ is $\overline{pdp}$ for all $f : \mathbb{N} \to \mathbb{R}$ where $f(y) = c_1y$, $c_1 > 0$.

**Proof.** From theorem 1 we know that $H_g$ is $\overline{pdp}$, where $g(y) = y/2$. For any $c_1$ there exists $n$ such that $1/2^n < c_1$. Then, by lemma [3.3] we have $H_{g^n}$ is $\overline{pdp}$. Since $g^n(y) = y/2^n < c_1y = f(y)$ we have $H_{g^n}(A) \subset H_f(A)$ for all $A \subset \mathbb{N}$ and hence by lemma [3.2] we have $H_f$ is $\overline{pdp}$. □

**Corollary 3.3.** The set $M_c = \{ y : \exists k \text{ such that } T^k(y) < cy \}$ has density 1 for all $c > 0$.

**Proof.** This follows directly from lemma [3.4] and corollary [3.2]. □

**Remark 3.1.** Suppose that $H_f$ is $\overline{pdp}$ where $f : \mathbb{N} \to \mathbb{R}$ is the map $f(y) = cy$ for a single $c < 1$. Applying lemma [3.6] (as we did in corollary [3.2]) will give us $H_g$ is $\overline{pdp}$ for $g(y) = cy$ for all $c' > 0$. This directly gives a way to improve the result in [3]. Proving that such a $H_f$ is $\overline{pdp}$ would require a bound better than $|T^m(B)| \geq |B|/2^m$ (as was used in the proof of theorem 1.)
4 Acknowledgements

I would like to thank Professor Idris Assani for bringing these problems to my attention and for his assistance in creating this paper.

References

[1] Riho Terras. “A stopping time problem on the positive integers”. eng. In: Acta Arithmetica 30.3 (1976), pp. 241–252. URL: http://eudml.org/doc/205476

[2] G Venturini. “On the 3x + 1 problem.” eng. In: Advances in Applied Mathematics 10 (3 1989), pp. 344–347.

[3] Ivan Korec. “A density estimate for the 3x+1 problem”. eng. In: Mathematica Slovaca 44.1 (1994), pp. 85–89. URL: http://eudml.org/doc/32414