On the Multifractality of Generalized Collatz Functions

January 27, 2025

Abstract

Collatz functions have received considerable attention for their connection to seemingly intractible problems within the fields of cellular automata (CA), number theory and computation. This paper will seek to prove certain properties of such functions by utilizing the theory of multifractals to derive a unified framework of past results. Finally, bounds on Hölder exponents for specific functions will be calculated, and in particular the information dimension.

1 Introduction

Generalized Collatz Functions, henceforth known as CFs, describe a series of affine transformations across modulo residue classes over the integers. Various connections have been found between such transformations and ergodic theory. Perhaps the most well known of these is the so called $3X+1$ function. The generalized problem can be given by Matthews [6, 8] as follows

$$T(x) = \frac{m_i x - r_i}{d} \quad \text{if } x \equiv i \pmod{d} \tag{1.1}$$

where $i = 0, ..., d-1$, and let $r_i \in \mathbb{Z}, m_i \in \mathbb{N}^*$ satisfy $r_i \equiv i m_i \pmod{d}$, then T is a function such that $T : \mathbb{Z} \to \mathbb{Z}$. This definition will utilize $gcd(m_i, r_i, d) = 1$ only, otherwise known as *relatively prime* CFs. Then, of particular interest is a full description of $T^K(x)$, and the motivation behind this paper is that iterating such Collatz-type functions defines a multifractal spectrum.

Multifractal systems were formalized by B.B. Mandelbrot in 1974 in his investigation of fully developed turbulence[7, 4, 10, 5]. Multifractal measures are generated by *multiplicative cascades*, and represent a measure around the boundary of a self-similar tree. They are described by multipliers on dyadic subintervals partitioned over $[0, 1]$.

2 Definitions

The description of the K-th iterate is given by Matthews as

$$T^K(x) = \frac{m_0(x)...m_{K-1}(x)}{d^K}\left(x - \sum_{i=0}^{K-1}\frac{r_i(x)d^i}{m_0(x)...m_i(x)}\right). \tag{2.1}$$

From now on we'll assume $X_0 \sim U[1, N]$ where U is the discrete uniform distribution on the interval from 1 to N. Observe that since $r_i \in \mathbb{Z}$ its domain is closed with a change of sign, thus we can instead consider $-r_i$ such that

$$-\sum_{i=0}^{K-1}\frac{-r_i(x)d^i}{m_0(x)...m_i(x)} = \sum_{i=0}^{K-1}\frac{r_i(x)d^i}{m_0(x)...m_i(x)}$$

which is the same as a reflection of the function about 0.

Then, with Matthew's definitions as a starting point, $m_K(x) = m_i$ for the K-th iterate, so we can group together similar m_i. This advances the expression of $T^K(X_0)$ to

$$T^K(X_0) = \left(\frac{m_0}{d}\right)^{k_0}\left(\frac{m_1}{d}\right)^{k_1}...\left(\frac{m_{d-1}}{d}\right)^{k_{d-1}}\left(X_0 + \sum_{i=0}^{K-1}\frac{r_i(x)d^i}{m_0(x)...m_i(x)}\right). \tag{2.2}$$

Where each k_i is just the frequency which the i-th residue appears up to and including K, and each will have the expression

$$k_i = \sum_{j=1}^{K}\mathbb{1}_{(i \pmod d)}(T^j(X_0))$$

$$\sum_{i=0}^{d-1} k_i = K.$$

Finally, we can then make the observation that $T^K(X_0)$ follows the *multifractal formalism*.

We denote β_K as the residue class for the K-th iterate. Lagarias originally proved this sequence is an adic number for the 3X+1 function, and this was later expanded for an arbitrary function and number d by Matthews, so $\beta_1\beta_2...\beta_K \in \mathbb{Z}_d$. The k_i count the number of elements in $[0, 1, ..., d-1]$ for an adic number's digit, and observe that Mandelbrot's construction defines a mass distribution off of d-adic intervals on $[0, 1]$.

To be specific, we assume there exists a positive, continuous Borel measure over $[0, 1]$. Then, we define the d-adic interval of the K-th generation for the digits $\beta_i \in [0, 1, ..., d-1]$ for an adic number such that

$$I_{0.\beta_1\beta_2...\beta_K} = \left[\sum_{i=1}^{K} \beta_i d^{-i}, d^{-K} + \sum_{i=1}^{K} \beta_i d^{-i}\right).$$

This defines a sequence of partitions of $[0, 1]$ we'll denote E_K with each having subintervals with measure d^{-K}. Then, a mass of 1 is assigned to E_0 which is then subdivided according to the ratio $w_0 : w_1 : ... : w_{d-1}$ for the subintervals $I_{K+1} \subset I_K$ for E_{K+1} with

$$w_0 = \frac{m_0}{r}, w_1 = \frac{m_1}{r}, ..., w_{d-1} = \frac{m_{d-1}}{r}, \quad \sum_{i=0}^{d-1} m_i = r$$

as we mandate the condition that

$$\sum_{i=0}^{d-1} w_i = 1.$$

This defines a Hausdorff measure over each interval which is $\mu(I_{\beta_1\beta_2...\beta_K})$, or for shorthand

$$\mu_{0.\beta_1\beta_2...\beta_K} = \prod_{j=1}^{K} w_{\beta_j} = w_0^{k_0} w_1^{k_1} ... w_{d-1}^{k_{d-1}}.$$

An interval $\epsilon \in [0, 1]$ has measure d^{-K} with an associated Hölder exponent α. Specifically the α is the degree of Hölder continuity for the measure. The coarse α is for an individual X_0's sequence is

$$\alpha(0.\beta_1\beta_2...\beta_K) = -\frac{k_0}{K}\log_d(w_0) - \frac{k_1}{K}\log_d(w_1) - ... - \frac{k_{d-1}}{K}\log_d(w_{d-1}).$$

Where k_i/K is the frequency of digits in X_0's associated adic number's address. Then a local exponent is described by

$$\alpha(0.\beta_1\beta_2...) = \lim_{K\to\infty}\frac{1}{K}\log_d(\mu(I_{0.\beta_1\beta_2...\beta_K})) = -\sum_{i=0}^{d-1}\varphi_i\log_d(w_i)$$

with $\varphi_i = k_i/K$ being the limiting frequency of digits in the measure's adic decimal. Asymptotically across the measure the number of intervals containing a specific exponent: $N_\epsilon(\alpha)$ scales according to

$$N_\epsilon(\alpha) \sim \epsilon^{-f(\alpha)}$$

where the $f(\alpha)$ function relates the measure's Hausdorff dimension to the scale exponent α.

The first exponent of interest for this paper is α_0 or $\alpha(0)$, which is the *global Hölder exponent*. The global exponent is given by Evertsz and Mandelbrot in MULTIFRACTAL MEASURES for the binomial/multinomial measure such that

$$\alpha_0 = -\frac{1}{d}\sum_{i=0}^{d-1}\log_d(w_i).$$

Which corresponds to $K = d, k_i = 1$ for the dimension of $f(\alpha_0)$, and this is the dimension of the μ's support, so

$$f(\alpha_0) = \dim_H(\mathrm{spt}\mu) = \max_\alpha f(\alpha).$$

In the case that $\lim_{\epsilon\to 0}$ it can be proven that the measure converges to a Cantor set with dimension μ, so the measure by extension is entirely contained within its supporting Cantor set. This is the second exponent of interest for this paper, and is denoted α_1 or $\alpha(1)$, which is the *information dimension*, and satisfies the identity

$$f(\alpha_1) = \alpha_1.$$

This is also the dimension of the measure μ, so

$$\alpha_1 = \dim_H(\mu) = \inf\{\dim_H(E) : E \text{ is a Borel set with } \mu(E) > 0\}.$$

Note that this is **not** the same as the dimension of μ's support, but rather the dimension of the measure μ itself. α_1 being significant because it "contains" the measure, meaning that as $K \to \infty$ the number of visits to this exponent limits towards 1.

Now we can get back to the original task at hand. Observe then that the value for the measure is different than the $\prod_{i=0}^{d-1}(m_i/d)^{k_i}$ term by an amount equal to $(r/d)^K$, so

$$T^K(X_0) = (\mu_{0.\beta_1\beta_2\ldots\beta_K})\left(\frac{r}{d}\right)^K \left(X_0 + \sum_{i=0}^{K-1} \frac{r_i(x)d^i}{m_0(x)\ldots m_i(x)}\right).$$

Note that it's only necessary to show that *a* multifractal measure exists which is equivalent to the $(m_i/d)^{k_i}$ terms for all K. On the real number line T can have any measure, even 0 measure, but its adic sequence generates a positive measure.

From here we need to look at the summand. Going back to Matthew's definitions the summation is equal to

$$\sum_{i=0}^{K-1} \frac{r_i(x)d^i}{m_0(x)\ldots m_i(x)} = \frac{\sum_{i=0}^{K-1} r_i(x)d^i m_{i+1}(x)\ldots m_{K-1}(x)}{m_0^{k_0}m_1^{k_1}\ldots m_{d-1}^{k_{d-1}}}.$$

It was proven $\sum_{i=0}^{K-1} r_i(x)d^i m_{i+1}(x)\ldots m_{K-1}(x) \in \mathbb{Z}_d$ by Möller [9], so we will define its value on the integers as a function of T's adic sequence such that

$$a_{0.\beta_1\beta_2\ldots\beta_K} = \sum_{i=1}^{K} r_{\beta_{i-1}} d^{i-1}\left(\prod_i^K m_{\beta_i}\right), \quad a_{0.\beta_1\beta_2\ldots\beta_K} \in \mathbb{Z}.$$

As a cosmetic choice the summation is shifted forwards by 1 to put it into the same index as μ, but it's trivial to tell a is equal to the summand term. Then

$$T^K(X_0) = (\mu_{0.\beta_1\beta_2\ldots\beta_K})\left(\frac{r}{d}\right)^K \left(X_0 + \frac{a_{0.\beta_1\beta_2\ldots\beta_K}}{m_0^{k_0}m_1^{k_1}\ldots m_{d-1}^{k_{d-1}}}\right)$$

and finally

$$T^K(X_0) = (\mu_{0.\beta_1\beta_2\ldots\beta_K})\left(\frac{r}{d}\right)^K (X_0) + \frac{a_{0.\beta_1\beta_2\ldots\beta_K}}{d^K}. \tag{2.3}$$

2.1 A Quick Remark

Scott Aaronson [1] considered a different type of function, where $\perp$ is taken as a halt symbol for some given modulo residue. Of note is the *Marxen-Buntrock* *machine*, which has some relationship with the Busy-Beaver function. This specific Collatz-type function has $d = 3$. If we consider the two modulo residues it contains which consist of affine functions alone, then the resultant multifractal measure will converge to the Lebesgue measure. However, a somewhat "natural" generalization is instead to take $w_0 = \frac{1}{2}, w_1 = \frac{1}{2}$ and $w_2 = 0$, where w_2 is a third weight which corresponds to $\perp$. Assuming this, it's fairly trivial to prove that at $K \to \infty$ the global exponent contains $\perp$, which should be a sufficient condition to prove convergence of the function.

There are 2 problems, however. The first is that this wouldn't exclude the existence of arbitrarily long orbits of the function, and second that the introduction of this extra weight causes the Hölder condition to be violated. As Mandelbrot stated in contrast however, for physical systems the condition that the Hölder exponent is finite actually can be violated. A full and complete proof of convergence of such a function is beyond the scope of this paper, however, as that would require more advanced notions in geometric measure theory.

3 Housekeeping

Theoretically, more rigmarole is required specifically because we took $a_{0.\beta_1\beta_2\ldots\beta_K}$ as arbitrary, then it is possible for it to attain an unbounded growth rate as $K \to \infty$.

Practically, however, it is only necessary to obtain the *maximal transformation* of a as the iteration is incremented, then we can produce an upper bound for T^K and determine the constant's value computationally. It is strongly suspected, though due to Conway's uncomputability proof [2], that it is impossible to find an exact expression for iteration.

As we will see later on, this is not so much a problem as it is just a mild annoyance which constrains the statements we can make, and indeed the main derivation is an upper bound. Some housekeeping is required first, however, before we can start evaluating specific CFs.

Theorem 1. *Assume T is not part of a cycle, then*

$$\frac{a_{0.\beta_1\beta_2\ldots\beta_K}}{d^K} \leq \frac{\bar{a}_{1+(K-1 \pmod d)}}{d^{1+(K-1 \pmod d)}} \left(\frac{m_0 m_1 \ldots m_{d-1}}{d^d}\right)^{\lfloor\frac{K-1}{d}\rfloor} + \bar{a}_d \frac{\left(\frac{m_0 m_1 \ldots m_{d-1}}{d^d}\right)^{\lfloor\frac{K-1}{d}\rfloor} - 1}{m_0 m_1 \ldots m_{d-1} - d^d}$$

for $K \in \mathbb{N}$, where $\bar{a}_{1+(K-1 \pmod d)}$ is a constant dependent on the modulo residue of K.

Proof. We can first define accelerated iteration, so we construct an iterated flow equation which will pass through the function's maximum, so

$$\underbrace{T^K \circ T^K \circ ... \circ T^K}_{\text{M times}}(X_0) = \left((\mu_{0.\beta_1\beta_2...\beta_K}) \left(\frac{r}{d}\right)^K \right)^M X_0 + (a_{0.\beta_1\beta_2...\beta_K}) \frac{\left((\mu_{0.\beta_1\beta_2...\beta_K}) \left(\frac{r}{d}\right)^K \right)^M - 1}{(\mu_{0.\beta_1\beta_2...\beta_K})(r)^K - d^K}. \tag{3.1}$$

This is simply the expression for $T^{KM}(X_0)$ for a fixed repeating sequence of modulo residues $\beta_1\beta_2...\beta_K$. However, observe that since this is an affine transformation we know the part independent of X_0 is equal to $a_{0.\beta_1\beta_2..\beta_{KM}}/d^{KM}$, so

$$a_{0.\beta_1\beta_2...\beta_K} \frac{\left((\mu_{0.\beta_1\beta_2...\beta_K}) \left(\frac{r}{d}\right)^K \right)^M - 1}{(\mu_{0.\beta_1\beta_2...\beta_K})(r)^K - d^K} = \frac{a_{0.\beta_1\beta_2...\beta_{KM}}}{d^{KM}}. \tag{3.2}$$

Choosing K and M corresponds to the addition of KM digits to the measure's adic decimal.

Note that while this may be sufficient for sequences of length KM, it's not so clear for sequences of length $i + KM$ for some initial choice of a sequence $\beta_1\beta_2...\beta_i$. For this we will offer two separate proofs, with the first one being trivial. First we can observe that $T^{KM} \circ T^i(X_0) = T^{i+KM}(X_0)$. One can then see that

$$T^{KM} \circ T^i(X_0) = \left((\mu_{0.\beta_1\beta_2...\beta_K}) \left(\frac{r}{d}\right)^K \right)^M \left((\mu_{0.\beta_1\beta_2...\beta_i}) X_0 + \frac{a_{0.\beta_1\beta_2...\beta_i}}{d^i} \right) + \frac{a_{0.\beta_1\beta_2...\beta_{KM}}}{d^{KM}}$$

$$= (\mu_{0.\beta_1\beta_2...\beta_{i+KM}}) X_0 + \frac{a_{0.\beta_1\beta_2...\beta_i}}{d^i} \left((\mu_{0.\beta_1\beta_2...\beta_K}) \left(\frac{r}{d}\right)^K \right)^M + \frac{a_{0.\beta_1\beta_2...\beta_{KM}}}{d^{KM}}.$$

Thus

$$\frac{a_{0.\beta_1\beta_2...\beta_{i+KM}}}{d^{i+KM}} = \frac{a_{0.\beta_1\beta_2...\beta_i}}{d^i} \left((\mu_{0.\beta_1\beta_2...\beta_K}) \left(\frac{r}{d}\right)^K \right)^M + a_{0.\beta_1\beta_2...\beta_K} \frac{\left((\mu_{0.\beta_1\beta_2...\beta_K}) \left(\frac{r}{d}\right)^K \right)^M - 1}{(\mu_{0.\beta_1\beta_2...\beta_K})(r)^K - d^K}$$

for a fixed repeating sequence $\beta_1\beta_2...\beta_K$ from β_{i+1} to β_{i+KM}. We can also prove it by utilizing the direct definition of a by going back to its introduction in section 1. So, assume that $a_{0.\beta_1\beta_2...\beta_{i+KM}}$ is fixed at a single exponent determined by $\beta_1\beta_2...\beta_K$ from residues β_{i+1} to β_{i+KM}, then

$$a_{0.\beta_1\beta_2...\beta_{i+KM}} = \sum_{j=1}^{i+KM} r_{\beta_{j-1}} d^{j-1} \left(\prod_j^{i+KM} m_{\beta_j} \right)$$

$$= \sum_{j=1}^{i} r_{\beta_{j-1}} d^{j-1} \left(\prod_{j}^{i+KM} m_{\beta_j} \right) + \sum_{j=i+1}^{i+KM} r_{\beta_{j-1}} d^{j-1} \left(\prod_{j}^{i+KM} m_{\beta_j} \right).$$

It is fairly trivial to show that the first summation is equal to

$$\sum_{j=1}^{i} r_{\beta_{j-1}} d^{j-1} \left(\prod_{j}^{i+KM} m_{\beta_j} \right) = (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^M \sum_{j=1}^{i} r_{\beta_{j-1}} d^{j-1} \left(\prod_{j}^{i} m_{\beta_j} \right)$$

$$= (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^M a_{0.\beta_1 \beta_2 ... \beta_i}.$$

Where the k_is are the frequencies of digits in the sequence $\beta_1 \beta_2 ... \beta_K$. For the remaining summation it's not as clear, however. Though one can observe that

$$\sum_{j=i+1}^{i+K} r_{\beta_{j-1}} d^{j-1} \left(\prod_{j}^{i+KM} m_{\beta_j} \right) = d^i (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^M \sum_{j=1}^{K} \frac{r_{\beta_{j-1}} d^{j-1}}{m_1(x) m_2(x) ... m_j(x)}$$

$$= d^i (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^M \sum_{j=0}^{K-1} \frac{r_{\beta_j} d^j}{m_0(x) m_1(x) ... m_j(x)}.$$

Going back to our definitions, observe, however, that the summation is simply

$$\sum_{j=0}^{K-1} \frac{r_{\beta_j} d^j}{m_0(x) m_1(x) ... m_j(x)} = \left(\frac{a_{0.\beta_1 \beta_2 ... \beta_K}}{m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}}} \right)$$

which then brings our original term equal to

$$d^i (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^{M-1} a_{0.\beta_1 \beta_2 ... \beta_K}.$$

In fact since the same sequence of modulo residues is repeated across each iteration, for the remaining intervals from $i + K + 1$ to $i + KM$ this value will be the same, so

$$\sum_{j=i+K+1}^{i+2K} r_{\beta_{j-1}} d^{j-1} \left(\prod_{j}^{i+KM} m_{\beta_j} \right) = d^{i+K} (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^{M-1} \sum_{j=1}^{K} \frac{r_{\beta_{j-1}} d^{j-1}}{m_1(x) m_2(x) ... m_j(x)}$$

$$= d^{i+K} (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^{M-2} a_{0.\beta_1 \beta_2 ... \beta_K}$$

$$\vdots$$

$$\sum_{j=i+KM-K+1}^{i+KM} r_{\beta_{j-1}} d^{j-1} \left(\prod_{j}^{i+KM} m_{\beta_j} \right) = d^{i+K(M-1)} (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}}) \sum_{j=1}^{K} \frac{r_{\beta_{j-1}} d^{j-1}}{m_1(x) m_2(x) ... m_j(x)}$$
$$= d^{i+K(M-1)} a_{0.\beta_1 \beta_2 ... \beta_K}.$$

So then finally we can observe that this entire sum is equal to

$$d^i a_{0.\beta_1 \beta_2 ... \beta_K} \left(\sum_{j=1}^{M} d^{K(j-1)} (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^{M-j} \right).$$

Rearranging this we get

$$d^i a_{0.\beta_1 \beta_2 ... \beta_K} \left(\sum_{j=0}^{M-1} d^{K(M-j-1)} (m_0^{k_0} m_1^{k_1} ... m_{d-1}^{k_{d-1}})^{j} \right)$$

$$= d^{i+KM} \left(\frac{a_{0.\beta_1 \beta_2 ... \beta_K}}{d^K} \right) \sum_{j=0}^{M-1} \left(\left(\frac{m_0}{d} \right)^{k_0} \left(\frac{m_1}{d} \right)^{k_1} ... \left(\frac{m_{d-1}}{d} \right)^{k_{d-1}} \right)^j$$

$$= d^{i+KM} \left(\frac{a_{0.\beta_1 \beta_2 ... \beta_K}}{d^K} \right) \sum_{j=0}^{M-1} \left(\mu_{0.\beta_1 \beta_2 ... \beta_K} \left(\frac{r}{d} \right)^K \right)^j$$

$$= d^{i+KM} a_{0.\beta_1 \beta_2 ... \beta_K} \frac{\left(\mu_{0.\beta_1 \beta_2 ... \beta_K} \left(\frac{r}{d} \right)^K \right)^M - 1}{\mu_{0.\beta_1 \beta_2 ... \beta_K} (r)^K - d^K}.$$

So then finally we can put it all together such that

$$\frac{a_{0.\beta_1 \beta_2 ... \beta_{i+KM}}}{d^{i+KM}} = \frac{a_{0.\beta_1 \beta_2 ... \beta_i}}{d^i} \left(\mu_{0.\beta_1 \beta_2 ... \beta_K} \left(\frac{r}{d} \right)^K \right)^M + a_{0.\beta_1 \beta_2 ... \beta_K} \frac{\left(\mu_{0.\beta_1 \beta_2 ... \beta_K} \left(\frac{r}{d} \right)^K \right)^M - 1}{\mu_{0.\beta_1 \beta_2 ... \beta_K} (r)^K - d^K}. \tag{3.3}$$

Otherwise there isn't much to say about this expression, but chiefly we can verify it's correct by observing that choosing $\beta_1 \beta_2 ... \beta_i \equiv \beta_1 \beta_2 ... \beta_K$ is exactly equivalent to taking $M + 1$.

It should be stated now that the i is used because we are interested in finding a mapping to the congruence class $K \to K \pmod{d}$. So since $T^m \circ T^n = T^{m+n}$ we're finding

$$T^K(X_0) \to \left\{ T^{1+dM}(X_0), T^{2+dM}(X_0), ..., T^{d+dM}(X_0) \right\}$$

which is an upper bound.

We are specifically interested in the maximal transformation, so the choice of K, k_i such that $a_{0.\beta_1\beta_2...\beta_K}/d^K$ is maximized. So, we can take the minimal covering of the measure's fractal set by taking $K = d, k_i = 1$ such that

$$\frac{a_{0.\beta_1\beta_2...\beta_{i+dM}}}{d^{i+dM}} \leq \frac{a_{0.\beta_1\beta_2...\beta_i}}{d^i}\mu(I_d^M)\left(\frac{r}{d}\right)^{dM} + a_{0.\beta_1\beta_2...\beta_d}\frac{\mu(I_d^M)\left(\frac{r}{d}\right)^{dM} - 1}{\mu(I_d^1)\left(r\right)^d - d^d}. \quad (3.4)$$

Note that this bound is valid even for an aperiodic sequence, since as Mandelbrot showed the coarse exponent will lie on the interval $(\alpha_{\min}, \alpha_{\max})$.

The I_d^M is used to denote the interval containing the global exponent, and indicates the measure is d-dimensional with M in the address for each dimension. Furthermore, from now on M will be utilized in iteration as opposed to K. This is a purely a cosmetic choice, but is done to emphasize that we have already chosen a value for K, so it's fixed.

The presence of the measure in 3.3. proves that a has nonzero Hausdorff dimension, and in fact it must have the same spectrum as μ, so an equivalent statement is simply that we are fixing the $a_{0.\beta_1\beta_2...\beta_K}/d^K$ at its Hausdorff/capacity dimension.

The problem then becomes about finding

$$S = \{\bar{a}_1, \bar{a}_2, ..., \bar{a}_d\}$$

which form a set of constants where $\bar{a}_i$ is the maximal constant among all the choices for $a_{0.\beta_1\beta_2...\beta_i}$ on $\mathbb{Z}$ in the path of the global exponent, so the sequence $\beta_1\beta_2...\beta_i$ is without repeated residues.

All of these indices can then be rescaled according to the floor function $\lfloor\frac{M-1}{d}\rfloor$ as the transformation for $i + dM \to i + 1 + dM$ is equivalent to swapping a constant such that $\bar{a}_{M \pmod d} \to \bar{a}_{M+1 \pmod d}$. The M is shifted backwards by 1 due to the initial iteration index being of length i. Then, we find

$$\frac{a_{0.\beta_1\beta_2...\beta_M}}{d^M} \leq \frac{\bar{a}_{1+(M-1 \pmod d)}}{d^{1+(M-1 \pmod d)}}\left(\frac{m_0m_1...m_{d-1}}{d^d}\right)^{\lfloor\frac{M-1}{d}\rfloor} + \bar{a}_d\frac{\left(\frac{m_0m_1...m_{d-1}}{d^d}\right)^{\lfloor\frac{M-1}{d}\rfloor} - 1}{m_0m_1...m_{d-1} - d^d}$$

for $M \in \mathbb{N}$ and $\bar{a}_{1+(M-1 \pmod d)} \in S$.

■

We can apply this back to our original definition in 2.3, so

$$T^M(X_0) \leq (\mu_{0.\beta_1\beta_2...\beta_M})\left(\frac{r}{d}\right)^M(X_0) + \frac{\bar{a}_{1+(M-1 \pmod d)}}{d^{1+(M-1 \pmod d)}}\left(\frac{m_0m_1...m_{d-1}}{d^d}\right)^{\lfloor\frac{M-1}{d}\rfloor}$$

$$+\bar{a}_d \frac{\left(\frac{m_0 m_1 ... m_{d-1}}{d^d}\right)^{\lfloor \frac{M-1}{d} \rfloor} - 1}{m_0 m_1 ... m_{d-1} - d^d}. \tag{3.5}$$

In practice, however, computing the $\bar{a}_i$ is easier said than done. We can replace the term with its upper bound as we know that $\max(S) = \bar{a}_d$. The proof of this is very simple, as since we already know that there's only one sequence of length $d+dM$ where the function achieves its absolute maximum it must be that $T^{i+dM}(X_0) < T^{d+dM}(X_0), i \in [1, d-1]$. This is where the function's coarse exponent is equal to the global exponent for all K.

Then let B be the set of all affine transformations $f_i = \frac{m_i x + r_i}{d}$ for a given T, then let $P(B)$ be the set of all permutations of B with each element composed together, then the value of $\bar{a}_d$ can be given by

$$\bar{a}_d = \max(S) = \max\{d^d v(x) - m_0 m_1 ... m_{d-1}(x) | v \in P(B)\}$$

which gives us the much simpler upper bound

$$\frac{a_{0.\beta_1\beta_2...\beta_M}}{d^M} \leq \bar{a}_d \frac{\left(\frac{m_0 m_1 ... m_{d-1}}{d^d}\right)^{1+\lfloor \frac{M-1}{d} \rfloor} - 1}{m_0 m_1 ... m_{d-1} - d^d}.$$

There is one more scenario, and that's if all the $\bar{a}_i$ are negative while $X_0 \in \mathbb{N}^+$ or all the $\bar{a}_i$ are positive while $X_0 \in \mathbb{N}^-$. in which case we can take $\bar{a}_i = 0$ for all i.

Conversely, suppose $X_0 \in \mathbb{N}^-$, then the minimal a_d, which we will denote $\underline{a}_d$, can instead be given by a simple change of sign for the $\bar{a}_d$. Otherwise, none of our assumptions are altered.

4 Multifractality

Matthews proposed a series of conjectures surrounding CFs, and it can be shown that these conjectures are resolved through multifractal analysis. This section is dedicated to that.

Theorem 2. *If $m_0 m_1 ... m_{d-1} > d^d$, then $\forall x_0$, $\lim_{M\to\infty} T^M(x_0) = \infty$ almost surely. If $m_0 m_1 ... m_{d-1} < d^d$, then $\forall x_0$, $\lim_{M\to\infty} T^M(x_0) \in \Omega$ where Ω is some finite cycle.*

Proof. Let Ω be a cycle of T, then evaluating the upper bound in 3.5 gives us

$$\lim_{M\to\infty} T^M(X_0) \to \Omega, \quad \frac{m_0 m_1 ... m_{d-1}}{d^d} < 1$$

or

$$\lim_{M\to\infty} T^M(X_0) \overset{as}{\to} \infty, \quad \frac{m_0 m_1 ... m_{d-1}}{d^d} > 1.$$

The proof of this is a fairly simple application of convergence of random variables. In particular, as Mandelbrot showed the measure can be interpreted as a random variable with the global exponent being it's expectation, and at the global exponent the multipliers will then grow according to $\left((m_0 m_1 ... m_{d-1})^{\frac{1}{d}}/d\right)^M$. It's then a straightforwards application of the law of large numbers which proves convergence. Note that $\mu > 0$ always, and this is for the same reason as given by Mandelbrot.

It is important to state that in the case $m_0 m_1 ... m_{d-1} > d^d$ there will be some exceptional set of x_0 for which T limits to Ω, which will have natural density 0. But as $M \to \infty$ the measure will either limit towards Ω or ∞ but can't exhibit other behavior as we are only looking at *relatively prime* CFs, meaning the measure is never Lebesgue. This being the same condition as $m_0 m_1 ... m_{d-1} \neq d^d$.

$\blacksquare$

While the upper bound tells us the conditions for when a function will always converge to a cycle, it does not tell us what the largest cycle is or how to find it. Though when $T^M(X_0) \in \Omega$ it will have an associated repeating d-adic sequence $\epsilon_1 \epsilon_2 ... \epsilon_{|\Omega|}$, and this will correspond to only a singular scaling exponent. As $M \to \infty$ then T's associated modulo residue frequencies will limit towards the value

$$\varphi_{|\Omega_i|} = \frac{|\Omega_i|}{|\Omega|}.$$

Where $\Omega_i \subseteq \Omega$ is the subset of elements in Ω such that $x \in \Omega_i, x \equiv i \pmod{d}$ with the convention $|\Omega| = \mathrm{card}(\Omega)$ for a set.

Assume $T^M(X_0)$ converges to Ω for some value of M, then T's coarse exponent will converge to the local exponent given by

$$\alpha_\Omega = \alpha(0.\epsilon_1\epsilon_2\epsilon_3...) = -\sum_{i=0}^{d-1} \varphi_{|\Omega_i|} \log_d(w_i).$$

This value will be indicated with α_Ω in the future. The principal problem being that we have no immediate way to derive Ω's invariant measure

$$\mu_{0.\epsilon_1\epsilon_2...\epsilon_{|\Omega|}} = \prod_{i=0}^{d-1} w_i^{|\Omega_i|}.$$

Which the $|\Omega_i|$ are clearly dependent on the individual choice of r_i and initial choice of x_0.

When T is in a cycle α_Ω ends up being the only non-vanishing exponent, or rather that T is infinitely more present on α_Ω than any other exponent. One might be led to believe that this is the information dimension α_1. However one can construct instances where $\alpha_\Omega = \alpha_0$, the most famous example being the $3X + 1$ function, which would then imply that $\alpha_0 = \alpha_1$, which clearly contradicts μ having a spectrum. The problem is that at ∞ the traditional definition of α_1 is not applicable as

$$\lim_{M\to\infty} (d^{-|\Omega|})^M = 0.$$

Which the "normal" α_1 is only really valid for when the measure is not part of a CF, in which case we would expect the sequence to remain Bernoulli at ∞. So at ∞ then μ is contained in a set of measure 0. We will instead need to look further to derive the information dimension.

Furthermore, T can have an arbitrary number of cycles each with its own exponent, each corresponding to the root of disjoint and infinite trees over $\mathbb{Z}$. So in reality μ describes a number of separate first order spectrums equal to the number of cycles present, and not just a single spectrum.

5 A Theory of Cycles

Shalom Eliahou in his 1991 paper [3] derived a very interesting inequality of the upper/lower bounds on the cycles of the $3X + 1$ function. Suppose we take T as the $3X + 1$ function and $\Omega_1 \subset \Omega$ as the set of odd elements of a cycle, then his derivation is that

$$\log_2(3 + M^{-1}) < \frac{\mathrm{card}(\Omega)}{\mathrm{card}(\Omega_1)} \leq \log_2(3 + m^{-1})$$

where $M = \max(\Omega)$ and $m = \min(\Omega)$.

This is a very interesting inequality, and perhaps surprising to some degree. However, the method's main problem is that it does not consider the multifractal and *scale invariant* properties of the function. Though, since this paper is dedicated to finding the cycles of CFs and not just the $3X + 1$ problem in particular, it's then interesting to consider taking Eliahou's method to its logical conclusion on how to derive upper/lower bounds on the cycles of *any* given function.

In the same manner as Eliahou's paper

$$\prod_{x\in\Omega} x = \prod_{x\in\Omega} T(x)$$

$$\prod_{x\in\Omega} \frac{T(x)}{x} = 1$$

$$\frac{T(x)}{x} = \frac{m_i - \frac{r_i}{x}}{d} \quad \text{if } x \equiv i \pmod d.$$

Again to make our lives easier we will only consider $-r_i$ from here on out as again, the r_i's domain $\mathbb{Z}$ is closed with respect to a change of sign, so

$$\prod_{x\in\Omega_0}\left(m_0 + \frac{r_0}{x}\right)\prod_{x\in\Omega_1}\left(m_1 + \frac{r_1}{x}\right)\dots\prod_{x\in\Omega_{d-1}}\left(m_{d-1} + \frac{r_{d-1}}{x}\right) = d^{|\Omega|}.$$

Then, let $P_i = \max(\Omega_i)$ and $\rho_i = \min(\Omega_i)$. The inequality will then follow

$$\prod_{i=0}^{d-1}\left(m_i + \frac{r_i}{P_i}\right)^{|\Omega_i|} \leq d^{|\Omega|} \leq \prod_{i=0}^{d-1}\left(m_i + \frac{r_i}{\rho_i}\right)^{|\Omega_i|}.$$

Then miraculously, and without even realizing it, we have now bounded Ω's Borel set simply by utilizing some basic assumptions.

Also observe that these two expressions will be equal if and only if $\max(\Omega_i) = \min(\Omega_i)$ for all i. Then finally we take the logarithm of this expression to extract $|\Omega|$, so

$$\sum_{i=0}^{d-1}|\Omega_i|\log_d\left(m_i + \frac{r_i}{P_i}\right) \leq |\Omega| \leq \sum_{i=0}^{d-1}|\Omega_i|\log_d\left(m_i + \frac{r_i}{\rho_i}\right). \tag{5.1}$$

Then we take the reciprocal

$$\frac{1}{\sum_{i=0}^{d-1}|\Omega_i|\log_d\left(m_i + \frac{r_i}{P_i}\right)} \geq \frac{1}{|\Omega|} \geq \frac{1}{\sum_{i=0}^{d-1}|\Omega_i|\log_d\left(m_i + \frac{r_i}{\rho_i}\right)},$$

and finally

$$-\frac{\sum_{i=0}^{d-1}|\Omega_i|\ln(w_i)}{\sum_{i=0}^{d-1}|\Omega_i|\ln\left(m_i + \frac{r_i}{P_i}\right)} \geq \alpha_\Omega \geq -\frac{\sum_{i=0}^{d-1}|\Omega_i|\ln(w_i)}{\sum_{i=0}^{d-1}|\Omega_i|\ln\left(m_i + \frac{r_i}{\rho_i}\right)}.$$

Note that the numerator is actually a positive number as $w_i < 1$ for all i, so the inequality is not reversed again. While this initially appears quite intimidating, recall that it is simply an interval on μ's large deviation spectrum. Some notation is in order, so we'll use

$$\alpha_T(\boldsymbol{P}) = -\frac{\sum_{i=0}^{d-1} |\Omega_i| \ln(w_i)}{\sum_{i=0}^{d-1} |\Omega_i| \ln\left(m_i + \frac{r_i}{P_i}\right)} \text{ and } \alpha_T(\boldsymbol{\rho}) = -\frac{\sum_{i=0}^{d-1} |\Omega_i| \ln(w_i)}{\sum_{i=0}^{d-1} |\Omega_i| \ln\left(m_i + \frac{r_i}{\rho_i}\right)}.$$

Now, to go futher we will need to go back to the notion of "measure". The α_Ω does not have measure, after all it's just a point! Conversely, the interval defined by the $\boldsymbol{P}$ and $\boldsymbol{\rho}$ does have positive measure when

$$\alpha_T(\boldsymbol{P}) - \alpha_T(\boldsymbol{\rho}) > 0.$$

Observe that this is fulfilled if and only if $\max(\Omega_i) \neq \min(\Omega_i)$ for at least one i, so if we assume this then $\alpha_T(\boldsymbol{P})$ and $\alpha_T(\boldsymbol{\rho})$ form an open interval on $(\alpha_{min}, \alpha_{max})$.

As we showed before, as $M \to \infty$ if μ's positive measure is concentrated on a single α, then it is not contained in a Borel set. However in the converse situation given a set $\alpha \subset (\alpha_{min}, \alpha_{max})$ there must exist some corresponding Borel set in $(0, 1)$ which contains μ. Recall that the dimension of the measure μ is given by

$$\dim_H(\mu) = \inf\{\dim_H(E) : E \text{ is a Borel set with } \mu(E) > 0\}.$$

Which then we must conclude

$$\dim_H(\mu) = \inf(f(\alpha_T(\boldsymbol{\rho})), f(\alpha_T(\boldsymbol{P}))) = f(\alpha_T(\boldsymbol{\rho})) = \alpha_T(\boldsymbol{\rho}).$$

Note that the $f(\alpha)$ corresponds to Rényi entropies. So, as the information dimension is the exponent for which $f(\alpha) = \alpha$, it must then satisfy $\alpha_T(\boldsymbol{\rho}) < 1$. Observe then it should trivially follow that

$$(0, 1) \cap (\alpha_T(\boldsymbol{\rho}), \alpha_T(\boldsymbol{P})) \neq \{\emptyset\}.$$

There is an exception, however, and that's for when $|\Omega_i| = 0$ for some i. The reason for this being somewhat subtle, as when this happens μ is no longer supported on its original Cantor set $\mathcal{C}$, and in this case

$$\text{spt}\mu \subset \mathcal{C}.$$

We cannot guarantee anything about μ as our assumptions are violated, so we will ignore cycles with this property in the future. Furthermore, we must also ignore cycles containing 0 as when this happens we cannot guarantee the expression in 5.1 is defined.

This serves as a good segue into a proof of the largest cycle's location, but first we need to establish some conditions.

Consider two cycles of the same function T which we will denote $\bar{\Omega}$ and $\underline{\Omega}$, where $\bar{\boldsymbol{\rho}}, \bar{\boldsymbol{P}} \in \bar{\Omega}$ and $\underline{\boldsymbol{\rho}}, \underline{\boldsymbol{P}} \in \underline{\Omega}$. If we view their respective Ω_i as vectors, then $\alpha_T : \Omega \to (\alpha_{min}, \alpha_{max})$ is a function space of the vector's L^∞-norm, thus the condition that $\alpha_T(\bar{\boldsymbol{\rho}}) > \alpha_T(\underline{\boldsymbol{P}})$ is equivalent to $\min(\bar{\Omega}) > \max(\underline{\Omega})$ on $\mathbb{N}^+$ and $\max(\bar{\Omega}) < \min(\underline{\Omega})$ on $\mathbb{N}^-$, which the latter follows from the norm being over the absolute value of Ω's elements. And this is possible because it is a function across finite sets, which is exactly why it's a function space and thus preserves the norm. Furthermore, this gives perhaps a more rigorous justification for the derivation of $\dim_H(\mu) = \alpha_T(\boldsymbol{\rho})$ from above, as under the L^∞-norm $\alpha_T(\boldsymbol{\rho})$ is uniformly convergent to the infimum.

We can then begin some analysis, so let

$$\hat{C} = \bigcup_{\alpha_T(\boldsymbol{P}) \geq 1} \Omega$$

be the union of all $\Omega \subset \mathbb{Z}$ for a given T such that $\alpha_T(\boldsymbol{P}) \geq 1$.

Theorem 3. *No cycles can exist on the interval* $(-\infty, \min(\hat{C})) \cup (\max(\hat{C}), \infty)$.

Proof. Suppose $\max(\hat{C}) \in \mathbb{N}^+$. We can denote a hypothetical cycle with a minimum larger than $\max(\hat{C})$ as $\bar{\Omega}$, meaning $\max(\hat{C}) < \min(\bar{\Omega})$. Conversely, suppose $\min(\hat{C}) \in \mathbb{N}^-$. We'll denote a hypothetical cycle with a maximum smaller than $\min(\hat{C})$ as $\underline{\Omega}$, meaning $\min(\hat{C}) > \max(\underline{\Omega})$. Yet, as we showed above these conditions lead to $\alpha_T(\bar{\boldsymbol{\rho}})$ and $\alpha_T(\underline{\boldsymbol{\rho}})$ being greater than or equal to 1, which furnishes the required contradiction. Clearly then $\bar{\Omega}$ and $\underline{\Omega}$ cannot exist.

There are a couple more situations, and that's if the function lacks a cycle with positive measure on either $\mathbb{N}^+$ or $\mathbb{N}^-$. However, in this case as again the norm is with respect to the absolute value of the cycle's elements, then we need only look for a cycle with $\alpha_T(\boldsymbol{P}) \geq 1$ on the opposing domain to bound the location of possible cycles.

The last situation is if there are no cycles with positive measure on either domain, in which case we cannot say anything. Recall these derivations are

only if $\max(\Omega_i) \neq \min(\Omega_i)$ for at least one i and $|\Omega_i| \neq 0$ for all i.
■

In all likelihood the latter restriction can be dropped with either a more careful derivation of the initial bounds on $|\Omega|$, or an altering of the information dimension's bound. A notable exception happens when $d = 2$, as if either $|\Omega_0|$ or $|\Omega_1|$ is 0 then the function must be at a fixed point and wont have positive measure. So far inequalities have been presented without evaluating any specific function. This section is dedicated to both providing examples and expanding upon the previous sections.

In theory these ρ_i, P_i may be able to be found analytically, but this is either incredibly difficult or impossible. However, theorem 3 entails it can be done empirically by examining a measurable cycle and calculating to determine if $\alpha_T(\boldsymbol{P}) > 1$. If not, then larger/smaller cycle's exponents must be computed. This algorithm is then continued until the above conditions are fulfilled.

References

[1] Scott Aaronson. The busy beaver frontier. *Electron. Colloquium Comput. Complex.*, TR20, 2020.

[2] J Conway. Unpredictable iterations. *The Ultimate Challenge: The 3x+1 Problem*, pages 49–52, 1972.

[3] Shalom Eliahou. The 3x+1 problem: new lower bounds on nontrivial cycle lengths. *Discrete Mathematics*, 118(1):45–56, 1993.

[4] Carl JG Evertsz and Benoit B Mandelbrot. Multifractal measures. *Chaos and fractals*, 1992:921–953, 1992.

[5] Kenneth Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2013.

[6] Jeffrey C. Lagarias. *The ultimate challenge: The 3X+1 problem*. American Mathematical Society, 2012.

[7] Benoit B Mandelbrot. *The fractal geometry of nature*. Macmillan, 1983.

[8] Keith R. Matthews. Generalized 3x+1 mappings: Markov chains and ergodic theory. 2010.

[9] Herbert Möller. Über hasses verallgemeinerung des syracuse-algorithmus (kakutanis problem). *Acta Arithmetica*, 34(3):219–226, 1978.

[10] Jacques Peyriére. Multifractal formalisms: Boxed versus centered intervals. *Analysis in Theory and Applications*, 19:332–341, 2003.