Self Referencing Sequences

Perry Kleinhenz, Fermi Ma, and Erik Waingarten

{pkleinhe,fermima,eaw}@mit.edu

1   Introduction

Written by Fermi Ma, edited by Perry Kleinhenz and Erik Waingarten

1.1   Problem Setup

Consider the sequence

$$1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, \ldots$$

We can break up the sequence into contiguous blocks, where each block is a stretch of repeated numbers:

$$[1], [2, 2], [1, 1], [2], [1], [2, 2], [1], [2, 2], [1, 1], [2], [1, 1], [2, 2], \ldots$$

The block lengths, read from left to right, reproduce the original sequence. Thus, this sequence is *self-referential*, and for the remainder of this paper we study properties of such sequences.

We say that this sequence is *generated* by the set of numbers used in the sequence, so the above sequence is generated by $\{1, 2\}$. We will refer to this set as the *generating set*. Observe that it is unambiguous which number to use next. After a block of 1's, the next block must be a block of 2's. However, in a sequence with generating set $\{1, 2, 3\}$, for example, a block of 1's could be followed by either a block of 2's or a block of 3's. Thus, we will assume that the next number to use is the following element (in cyclic order) in the generating set. (Note that because order matters, the *generating set* is technically not a set, but we do not call it a "sequence" to avoid possible confusion with its corresponding self-referencing sequence.)

With this rule, a self-referencing sequence generated by $\{1, 2, 3\}$ has blocks of 1's followed by blocks of 2's followed by blocks of 3's, followed by blocks of 1's, and so on. Thus, $\{1, 2, 3\}$ can generate

$$1, 2, 2, 3, 3, 1, 1, 1, 2, 2, 2, 3, 2, 1, 2, 2, 3, 3, 1, 1, \ldots$$

but not

$$1, 3, 3, 3, 2, 2, 2, 1, 1, 1, 3, 3, 2, 2, 1, 1, 3, 2, 1, 3, \ldots$$

as the second sequence has the numbers in the wrong order.

1.2   Overview

In this paper, we address a variety of questions regarding these sequences. In Section 2 we show that if we are given a generating set and a starting number for the sequence, the sequence is uniquely determined. After that, we consider how the starting number for the sequence affects what the sequence can be. In particular, in Section 3, we show that there exist limits on how much two sequences with the same generating set can resemble each other if they start with different numbers. In the following Section 4, we consider the limiting behavior of such sequences, and we conjecture that the limiting behavior of a sequence is determined only by the generating set, and not the starting number.

In Section 5, we take a different approach to analyzing these sequences by showing that there is an alternate way of generating them with an iterative process. In Section 6, we show how a large number of these sequences can actually be generated with a simple set of substitution rules. We then introduce the density problem in Section 7, and use ideas developed in Section 5 and Section 6 to answer the question for certain types of generating sets. Unfortunately, we find that the open problem of determining the density for the sequence of 1's and 2's is hard to solve using our methods.

2   Determinism

Written by Fermi Ma, edited by Perry Kleinhenz and Erik Waingarten

We claim that a self-referencing sequence is uniquely determined by the generating set and the starting number. To see why this is true, we can simply give an algorithm to build the sequence. Assume the generating set is $\{a_1, a_2, \ldots, a_m\}$ and that we start with some a_i. If we ever write down some a_k where $k > m$, we set this equal to $a_{k(\bmod\ m)}$.

Algorithm:
Suppose the first number of the sequence is a_i. Then the first block of the sequence must be a block of a_i instances of a_i, so we extend the sequence to this block. If a_i is not 1, then there is a second element of that sequence. We read that second element, and then we add a block to the end of that sequence with length dictated by the value of the second element. If a_i is equal to 1, we write down a_{i+1} in the second spot, and use the value of a_{i+1} to dictate the length of the second block.

This procedure can continue deterministically so long as there is always an element to read. However, every time we write down a block, we read 1 element and write down at least 1 element, so the end of the sequence will always be after the spot we are reading. The only exception is when $a_i = 1$, but then since $a_{i+1} \neq 1$, this issue will never occur again.

Thus, a self-referencing sequence is uniquely determined by the generating set and the starting number. From this point on, we will denote by $s(A, a)$ the unique sequence generated by set A with starting number $a \in A$.

3   Prefixes

Written by Perry Kleinhenz, edited by Fermi Ma and Erik Waingarten

Definition 1. *We say that two sequences $\{a_i\}$ and $\{b_i\}$ differ by a prefix of length n if*

$$a_{n+i} = b_i,$$

for i any positive integer. if no such n exists then we say that the two sequences are independent.

Note that a pair of sequences can have prefixes of different length. For example if we have

$$\{a_i\} = \{1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, \ldots\}$$
$$\{b_i\} = \{7, 8, 9, 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, \ldots\},$$

then a_i and b_i differ by prefixes of length $3 + 6k$ for k a nonnegative integer. Because of this we make the following definition.

Definition 2. *We say that two sequences $\{a_i\}$ and $\{b_i\}$ differ by a minimal prefix of length n if they differ by a prefix of length n, but do not differ by a prefix of length m for all $0 < m < n$.*

We note that if a pair of sequences differs by a prefix they differ by a minimal prefix. In our above example a_i and b_i differ by a minimal prefix of length 3.

Theorem 1. *The self-referencing sequence beginning with a 1 generated over $\{1, c_1, c_2, \ldots, c_n\}$ differs from the self-referencing sequence beginning with c_1 generated over $\{1, c_1, c_2, \ldots, c_n\}$ by a minimal prefix of length 1.*

Proof. Let a_i refer to the ith term of the sequence beginning with a 1 and b_i refer to the ith term of the sequence beginning with c_1. We know that $a_1 = 1$, this means that the first block in a_i must be of length 1. In other words we must have $a_2 = c_1$, but since a_2 is now part of the second block of a_i the value of a_3 is not specified by a_1. Thus we could specify the rest of a_i only knowing the value of a_2, but $b_1 = c_1$, so this is exactly how the values of b_i is specified. Therefore

$$a_{1+i} = b_i,$$

so a_i and b_i differ by a prefix of length 1, but $a_1 \neq b_1$ so it is a minimal prefix. $\square$

Theorem 2. *If $1 < a < b$ then the self-referencing sequence beginning with a generated over $\{1, a, b\}$ and the self-referencing sequence beginning with b generated over $\{1, a, b\}$ are independent.*

Proof. We proceed by contradiction. If the two sequences are not independent then they must differ by some minimal prefix of length n. If we let a_i be the sequence beginning with a and b_i be the sequence beginning with b the

$$a_{n+k} = b_k.$$

We know that the first b terms of the sequence which begins with a b are b's. Because $b > a > 1$ no block can be longer then b. Thus the terms immediately preceding and following this block in the sequence a_i must not be b's. Because of this the first block of b_i is the $m + 1$th block of a_i, where m is some positive integer. Since the value of the sequence at position k is also the length of the kth block this means that $a_{m+1} = b_1$. In fact the kth block of b_i is the $(m + k)$th block of a_i and so

$$a_{m+k} = b_k$$

We now claim that the prefix has length strictly larger than m. Assume otherwise so $n \leq m$. This means the entry in the a_i which gives the length of the $m + 1$th block occurs at or after the beginning of the $m + 1$th block. By our proof that self-referential sequences are well defined we know that the second case cannot occur. Therefore the only other option would be for the $m + 1$th block to start at position $m + 1$, but this would require every block to have length exactly 1, but this is a contradiction because $a_1 = a > 1$, that is the first block has length greater than one. Therefore the prefix must have length strictly larger than m.

We note that based on a previous step the two sequences differ by a prefix starting at m. Since $m < n$ we have produced a prefix which starts before n which is a contradiction. Therefore the two sequences are independent. $\square$

4   Equivalence

Written by Fermi Ma, edited by Perry Kleinhenz and Erik Waingarten

In this section, we consider a slightly weaker definition of equivalence. Consider, for example, sequences generated by $\{1, 2, 3\}$. Theorem 2 shows that the sequence starting with 2

$$2, 2, 3, 3, 1, 1, 1, 2, 2, 2, 3, 1, 2, 3, 3, 1, 1, 2, 2, 3, 3, 3, \ldots$$

and the sequence starting with 3

$$3, 3, 3, 1, 1, 1, 2, 2, 2, 3, 1, 2, 3, 3, 1, 1, 2, 2, 3, 3, 3, \ldots$$

are independent in the sense that neither sequence is exactly contained in the other. However, if we delete the first two numbers of the sequence starting with 2, and the first number of the sequence starting with 3, we get

$$3, 3, 1, 1, 1, 2, 2, 2, 3, 1, 2, 3, 3, 1, 1, 2, 2, 3, 3, 3, \ldots$$

in both cases. (Need to further justify why these two are the same, probably with a reading distance argument)

Thus, since both sequences are essentially the same if certain prefixes are removed, we call these sequences *equivalent*. To formalize, this, we say that two sequences $\{a_i\}$ and $\{b_i\}$ are equivalent if there exists some integers $n, k \geq 0$ such that

$$a_{n+i} = b_{k+i}$$

We now give the following conjecture

Conjecture 1. Any sequence generated by a set S is equivalent to any other sequence generated by the same set S.

This conjecture is partly motivated by the fact that the property holds for $\{1, 2, 3\}$. For generating sets with larger numbers, we have been unable to verify or refute the conjecture. However, in Section **??**, we show that the starting number of a sequence does not affect the densities of each number, which provides some evidence for this conjecture.

5 An Iterative Process

Written by Fermi Ma, edited by Perry Kleinhenz and Erik Waingarten

We change focus in this section and consider how to generate self-referential sequences with an iterative process. We first look at the sequences generated by $\{1, 2\}$.

One possible iterative process is as follows. Let the sequence $\{s^0\}$ be 2, and let the sequence $\{s^{(i)}\}$ for $i \geq 1$ be the unique sequence that starts with 2 and is such that its block lengths, read from left to right, reproduce the sequence $\{s^{(i-1)}\}$. This gives:

$$\{s^{(0)}\} = 2$$
$$\{s^{(1)}\} = 2, 2$$
$$\{s^{(2)}\} = 2, 2, 1, 1$$
$$\{s^{(3)}\} = 2, 2, 1, 1, 2, 1$$
$$\{s^{(4)}\} = 2, 2, 1, 1, 2, 1, 2, 2, 1$$
$$\cdots$$

Let s^∞ be the limit of this process. We claim that s^∞ is the unique self-referential sequence $s(\{1, 2\}, 2)$ generated by $\{1, 2\}$ with starting number 2.

Proof. We prove this with induction. We show that if $s^{(i)}$ is a prefix of $s(\{1, 2\}, 2)$, then $s^{(i+1)}$ is a longer prefix of $s(\{1, 2\}, 2)$. The base case of the induction is satisfied, as $s^{(0)}$ is a prefix of $s(\{1, 2\}, 2)$.

For the inductive step, observe that a prefix of the sequence describes the block lengths of a *longer* prefix of the sequence. We know that it describes a longer prefix because the prefix it describes has length equal to the sum of the original prefix. In other words, the length of $s^{(i+1)}$ must be equal to the sum of of the numbers in $s^{(i)}$, which is strictly greater than the length of $s^{(i)}$ (since there are numbers that are greater than 1). This holds true because $s^{(i)}$ merely contains a collection of lengths, which are summed up to give the next prefix. □

It turns out that this same argument holds in general.

Theorem 3. *Let $\{s^0\} = a$ for some $a \in A$ where $a \neq 1$, and let $\{s^{(i)}\}$ for $i \geq 1$ be the unique sequence that starts with a and is such that its block lengths, read from left to right, reproduce the sequence $\{s^{(i-1)}\}$. Then the limit of this process, s^∞, is the sequence $s(A, a)$.*

Note that we must specify that $a \neq 1$, because that is the only case in which the length of $s^{(i+1)}$ is not strictly greater than the length of $s^{(i)}$. The proof of this theorem is omitted, as it is essentially the same proof as for the $A = \{1, 2\}, a = 2$ case.

It turns out that we can be less restrictive about what the starting sequence is. Going back to the case of $\{1, 2\}$, we can let $s^{(0)}$ be *any* sequence of 1's and 2's that starts with a 2. For example, suppose $s^{(0)} = 2, 1, 1, 1, 2$. The rules for generating large $s^{(i)}$ give:

$$\{s^{(0)}\} = \mathbf{2}, 1, 1, 1, 2$$
$$\{s^{(1)}\} = \mathbf{2}, \mathbf{2}, \mathbf{1}, 2, 1, 2, 2$$
$$\{s^{(2)}\} = \mathbf{2}, \mathbf{2}, \mathbf{1}, \mathbf{1}, \mathbf{2}, \mathbf{1}, 1, 2, 1, 1, 2, 2$$
$$\{s^{(3)}\} = \mathbf{2}, \mathbf{2}, \mathbf{1}, \mathbf{1}, \mathbf{2}, \mathbf{1}, \mathbf{2}, \mathbf{2}, \mathbf{1}, \mathbf{2}, 1, 1, 2, 1, 2, 2, 1, 1$$
$$\cdots$$

Here, we highlight in bold the numbers that form a prefix of $s(\{1, 2\}, 2)$. We note that $s^\infty = s(A, a)$ simply because $s^{(0)}$ contains a 2 at the beginning, whch then causes a 2,2 to appear at the start of $s^{(1)}$, which causes $2, 2, 1, 1$ to appear at the start of $s^{(2)}$, and so on. The rest of each $s^{(i)}$ that is not a prefix of $s(1, 2, 2)$ we call the *tail*.

We now consider two different forms of convergence. We know that $s^{(i)}$ in some sense approaches $s(A, a)$, as arbitrarily many elements at the beginning of the sequence will match up with the beginning of $s(A, a)$ as i increases. However, if the tail of the sequence never entirely disappears, then there does not exist an i where $s^{(i)}$ is exactly a prefix of $s(A, a)$. Thus, we can formalize two types of convergence: *weak convergence* and *strong convergence*. Weak convergence is satisfied if $s^\infty = s(A, a)$, and strong convergence is satisfied if $s^{(i)}$ is a prefix of $s(A, a)$ for all $i \geq k$ for some k.

6   Substitution Rules

7   Density

Written by Fermi Ma, edited by Perry Kleinhenz and Erik Waingarten
   The above analysis shows that, for example, the s