

# 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

$$1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, \dots$$

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

$$[1], [2, 2], [1, 1], [2], [1], [2, 2], [1], [2, 2], [1, 1], [2], [1, 1], [2, 2], \dots$$

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

$$1, 2, 2, 3, 3, 1, 1, 1, 2, 2, 2, 3, 2, 1, 2, 2, 3, 3, 1, 1, \dots$$

but not

$$1, 3, 3, 3, 2, 2, 2, 1, 1, 1, 3, 3, 2, 2, 1, 1, 3, 2, 1, 3, \dots$$

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

## 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

$$\begin{aligned} \{a_i\} &= \{1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, \dots\} \\ \{b_i\} &= \{7, 8, 9, 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, \dots\}, \end{aligned}$$

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

*Proof.* Let  $a_i$  refer to the  $i$ th term of the sequence beginning with a 1 and  $b_i$  refer to the  $i$ th 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  $a$  are  $b$ 's. Because  $b > a > 1$  no block can be longer than  $b$ . 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  $k$ th block this means that  $a_{m+1} = b_1$ . In fact the  $k$ th 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

## 5 An Iterative Process

## 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