SELF-SIMILAR SEQUENCES AND GENERALIZED WYTHOFF ARRAYS

DAVID GARTH AND JOSEPH PALMER

Abstract. Self-similar sequences are sequences that contain a proper subsequence identical to the original sequence. Kimberling [6] showed that the famous Wythoff array gives rise to an interesting self-similar sequence. In this paper we discuss the construction of an array similar to the Wythoff array for a more general class of numeration systems. We then give a construction of a self-similar sequence arising from this array. The self-similar sequences arising from this construction are also compared to some other well-known self-similar sequences.

1. Introduction

An infinite sequence of integers is *self-similar* if it contains a proper subsequence that is identical to the original sequence. Examples of self-similar sequences can be found among many well-known integer sequences ([10],[12]). As an example, consider the sequence $\{s_n\}_{n=0}^{\infty}$, whose n^{th} term is the number of ones in the binary expansion of n. Since multiplication of n by 2 concatenates a 0 onto the end of the binary expansion of n, it follows that $s_n = s_{2n}$. Thus, $\{s_n\}$ is self-similar under the action of removing the odd indexed terms. The first few terms of the sequence, with the odd indexed terms underlined, are

$$0 \quad \underline{1} \quad 1 \quad \underline{2} \quad 1 \quad \underline{2} \quad 2 \quad \underline{3} \quad 1 \quad \underline{2} \quad 2 \quad \underline{3} \quad 2 \quad \underline{3} \quad 3 \quad \underline{4} \quad 1 \ldots$$

Reducing the entries of the ones counting sequence modulo two maps the ones counting sequence onto the well-known Thue-Morse sequence

$$0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 1 \ldots$$

This sequence, denoted $\{t_n\}_{n=0}^{\infty}$, is automatically self-similar since it inherits the property $t_n = t_{2n}$ from the ones counting sequence.

As another example, recall that the Zeckendorf representation of a natural number n is the unique representation of n as a sum of nonconsecutive Fibonacci numbers. This gives rise to the Fibonacci numeration system in which the natural numbers are represented by strings over $\{0, 1\}$ which do not contain the string 11. More precisely, if

$$n = c_0F_2 + c_1F_3 + \cdots + c_kF_{k+2}$$

is the Zeckendorf representation of n, where $c_k = 1$ and for $c_i \in \{0, 1\}$, $0 \leq i < k$, $c_i \in \{0, 1\}$, and $c_{i+1} = 0$ whenever $c_i = 1$, then the string $c_kc_{k-1} \cdots c_0$ is the Fibonacci representation of n. We say that a number has an odd Fibonacci representation if its Fibonacci representation ends in a 1. We can now define the *Fibonacci ones counting sequence* $\{s'_n\}_{n=0}^{\infty}$, where s'_n is the number of ones in the Fibonacci representation of n. It is not hard to see that the Fibonacci ones counting sequence is self-similar under the action of removing the terms indexed by numbers with an odd Fibonacci representation. The first few terms of this sequence, with removable terms underlined, are shown in the following table.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
s'_n	0	$\underline{1}$	1	1	$\underline{2}$	1	$\underline{2}$	2	1	$\underline{2}$	2	2	$\underline{3}$	1

The Fibonacci representations of the natural numbers can be used to construct the famous Zeckendorf array, also known as the Wythoff array [7], from which another self-similar sequence can be constructed. The Zeckendorf array is constructed as follows. The first column is the increasing sequence of natural numbers with odd Fibonacci representations. The successor of a number n in any row of the array is the number obtained by concatenating a zero to the end of the Fibonacci representation of n. The first five rows of the array are

$$\begin{array}{cccccccc} 1 & 2 & 3 & 5 & 8 & 13 & 21 & 34\dots \\ 4 & 7 & 11 & 18 & 29 & 47 & 76 & 123\dots \\ 6 & 10 & 16 & 26 & 42 & 68 & 110 & 178\dots \\ 9 & 15 & 24 & 39 & 63 & 102 & 165 & 267\dots \\ 12 & 20 & 32 & 52 & 84 & 136 & 220 & 356\dots \end{array}$$

Many properties of the Zeckendorf array are discussed in the literature ([3],[4],[5],[6],[7]). It follows from the construction that every positive integer appears exactly once in the array. The array is also a well-known example of an interspersion ([3],[5]), i.e., an array in which each row and column is an increasing sequence, and if $\{t_i\}$ and $\{u_i\}$ are two row sequences satisfying $t_m < u_n < t_{m+1}$ for some integers m and n, then $t_{m+1} < u_{n+1} < t_{m+2}$.

We define the sequence $\{r_n\}$, known as the *row sequence*, whose n^{th} term gives the row of the array containing n. The first few terms of this sequence are

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
r_n	1	1	1	2	1	3	2	1	4	3	2	5	1	6	4	3	7	2	8	5	1

Kimberling [6] showed that $\{r_n\}$ is self-similar under the action of removing the first occurrence of each integer in the sequence. Notice that this action, like the action of the ones counting sequence, also removes terms indexed by numbers with odd Fibonacci representations, despite the fact that the sequences are not the same.

In this paper we will generalize this result to show that the ones counting sequences and the row sequences corresponding to a large class of nonstandard numeration systems are self-similar in a manner similar to the examples just mentioned. We begin with definitions related to generalized numeration systems and in Section 2 we prove our main theorem.

2. Numeration Systems

Suppose that A be a nonempty finite set and denote by A^* the set of words over A. The length of a word x, denoted $|x|$, is defined to be the number of characters in the word. Also, if $x \in A$, then we write xA^* to denote the set of all words over A that begin with x. For the remainder of this paper we let $\Sigma_k = \{0, 1, 2, \dots, k-1\}$.

If A is a finite ordered set, then we can order the elements of A^* as follows. For $x, y \in A^*$ we say that $x < y$ if either $|x| < |y|$, or if $|x| = |y|$ and there exist $u, x', y' \in A^*$ such that $x = uax'$ and $y = uby'$ and $a, b \in A$ with $a < b$. This ordering is known as the *radix order* [9] and determines the ordering of numbers in typical numeration systems like the base 10 and binary numeration systems. The following definition is similar to definitions considered in [2], [8], and [11].

Definition 2.1. *Let $k \in \mathbb{N}$. A numeration system is an infinite subset $\mathcal{S}$ of the set Σ_k^* along with a bijective map $N : \mathcal{S} \to \mathbb{N}$ such that*

(i) *$\mathcal{S}$ does not contain the empty word.*
(ii) *No element of $\mathcal{S}$ begins with* 0.
(iii) *N maps the n^{th} word in $\mathcal{S}$ (under the radix order) to n.*

The map N is referred to as the evaluation map, while N^{-1} is the representation map. The set Σ_k is known as the alphabet of $(\mathcal{S}, N)$. If $N(w) = n$, then we say that w is the $\mathcal{S}$-representation of n.

In the literature it is common for a numeration system $\mathcal{S}$ to contain the empty word as the $\mathcal{S}$-representation of 0. For our purposes in this paper we do not need to consider representations of 0. Thus, requirement (i) in Definition 2.1 is a matter of convenience.

A familiar example of a numeration systems is the binary numeration system, in which $\mathcal{S} = 1\Sigma_2^*$. In the Fibonacci numeration system $\mathcal{S}$ is the set of all strings over Σ_2 that begin with 1 and do not contain two consecutive ones.

A desirable property for a numeration system $\mathcal{S}$ is that $x0 \in \mathcal{S}$ whenever $x \in \mathcal{S}$. We say such a numeration system is *right extendable* (see Chapter 7 of [9]). In base 10, concatenating a 0 to the end of a representation corresponds to multiplying by 10. Concatenating a 0 in the Fibonacci numeration system corresponds to a shift in the indices on the Fibonacci numbers in the Zeckendorf representation of n.

Let $\mathcal{S}$ be a right extendable numeration system over Σ_k. We associate an array $\mathcal{W}(\mathcal{S})$ to $\mathcal{S}$ as follows. The first row of the array is the sequence of natural numbers

$$N(v),\ N(v0),\ N(v00),\ N(v000), \ldots,$$

where $N(v) = 1$. The first element of any subsequent row is $N(w)$, where w is the minimal word in $\mathcal{S}$, under the radix order, such that $N(w)$ is not in any preceding row. The complete row is then

$$N(w),\ N(w0),\ N(w00),\ N(w000), \ldots$$

We refer to the array $\mathcal{W}(\mathcal{S})$ as the *generalized Wythoff array associated with* $\mathcal{S}$. Note that when $\mathcal{S}$ is right extendable, $\mathcal{W}(\mathcal{S})$ will contain infinitely many columns, but may contain only finitely many rows. For instance, if $\mathcal{S} = \{1, 10, 100, 1000, \ldots\}$, then $\mathcal{W}(\mathcal{S})$ contains only one row.

As an example in which $\mathcal{W}(\mathcal{S})$ contains infinitely many rows, in [2] the authors considered the numeration system $\mathcal{S}$ whose expansions consist of the greedy representations of the natural numbers with respect to the sequence $\{b_n\}$ defined by $b_0 = 1$, $b_1 = 2$, $b_2 = 3$, and $b_n = b_{n-1} + 2b_{n-2} - b_{n-3}$ for $n \geq 3$. (See [1] for a description of greedy numeration systems). The first few elements of $\mathcal{S}$ are listed in the following table.

n	1	2	3	4	5	6	7	8	9	10	11	12
$\mathcal{S}$-representation	1	10	100	101	110	1000	1001	1010	1100	10000	10001	10010

The first 5 rows of the array $\mathcal{W}(\mathcal{S})$ are as follows.

$$\begin{array}{cccccc}
1 & 2 & 3 & 6 & 10 & 19\ldots \\
4 & 8 & 13 & 25 & 43 & 80\ldots \\
5 & 9 & 16 & 29 & 52 & 94\ldots \\
7 & 12 & 22 & 39 & 71 & 127\ldots \\
11 & 21 & 36 & 67 & 118 & 216\ldots
\end{array}$$

We are now ready to state our main theorem.

Theorem 2.2. *Let $\mathcal{S}$ be a right extendable numeration system with evaluation map N. For $n \geq 1$ let r_n be the row of $\mathcal{W}(\mathcal{S})$ containing n. Then r_n is self-similar under the action of removing the first occurrence of each natural number.*

Proof. Let $\{r_{n_k}\}_{k=1}^{\infty}$ be the subsequence of $\{r_n\}$ obtained by removing the first occurrence of each natural number from $\{r_n\}$. Notice that $\{n_k\}_{k=1}^{\infty}$ is the sequence of natural numbers that are not in the first column of $\mathcal{W}(\mathcal{S})$, since the rows of $\mathcal{W}(\mathcal{S})$ are strictly increasing. Now define

$$\mathcal{S}' = \{N^{-1}(n_k) : k \geq 1\}.$$

Note that $\mathcal{S}'$ is a numeration system. Let $\mathcal{W}(\mathcal{S}')$ be the generalized Wythoff array associated with $\mathcal{S}'$. It follows that the row sequence of $\mathcal{W}(\mathcal{S}')$ is $\{r_{n_k}\}_{k=1}^{\infty}$. That is, r_{n_k} is the row of $\mathcal{W}(\mathcal{S}')$ containing k.

Notice that $\mathcal{S}' \subseteq \mathcal{S}$, and $\mathcal{S}'$ is precisely the set of $\mathcal{S}$-representations that are not the $\mathcal{S}$-representation of any number in the first column of $\mathcal{S}$. The first column of $\mathcal{W}(\mathcal{S})$ is made up of those numbers with $\mathcal{S}$ representations that either do not end in 0, or else are of the form $w0$, where $w \notin \mathcal{S}$. Since $\mathcal{S}$ is right extendable, it follows then that $\mathcal{S}'$ is the numeration system obtained by concatenating a 0 to the elements of $\mathcal{S}$. From this it follows that $\{r_{n_k}\} = \{r_n\}$. □

Notice that the proof of this theorem can be used to show that any sequence is self-similar where the action of removing terms depends only on digits which are not trailing zeros in the $\mathcal{S}$-representation. As an example, consider the ones counting sequence $\{s_n\}$ of the numeration system considered before the statement of Theorem 2.2. The first few terms of the ones counting sequence $\{s_n\}$ and the row sequence $\{r_n\}$ are shown in the following table, with the removable terms underlined.

$\{s_n\}$	$\underline{1}$	1	1	$\underline{2}$	$\underline{2}$	1	$\underline{2}$	2	2	1	$\underline{2}$	2	2	$\underline{3}$	$\underline{3}$	2	$\underline{3}$	$\underline{3}\ldots$
$\{r_n\}$	$\underline{1}$	1	1	$\underline{2}$	$\underline{3}$	1	$\underline{4}$	2	3	1	$\underline{5}$	4	2	$\underline{6}$	$\underline{7}$	3	$\underline{8}$	$\underline{9}\ldots$

Notice that these two sequences are not the same, but they have the same self-similar nature since the same terms must be removed to preserve the sequence. Initially, it may seem that the ones counting sequence has no relation to the row sequences discussed in Theorem 2.2, but here we see that when the sequence is viewed from the correct perspective, Theorem 2.2 does apply.

3. Discussion and Examples

Given a numeration system $(\mathcal{S}, N)$, our construction of the array $\mathcal{W}(\mathcal{S})$ and the resulting self-similarity properties are heavily dependent on the right extendability of the numeration system. Kimberling [6] gave a generalized construction of the Wythoff array that did not depend on right extendability but which applied only to based numeration systems which are greedy [1].

The numeration system $(\mathcal{S}, N)$ with alphabet Σ_k is said to be *based* if there exists a strictly increasing sequence $\{b_i\}_{i=0}^{\infty}$ of natural numbers, with $b_0 = 1$, and a function $\pi : \mathcal{S} \to \mathbb{N}$, such that

$$\pi(w_k w_{k-1} \cdots w_0) = \sum_{i=0}^{k} b_i w_i \tag{3.1}$$

for all $w_k w_{k-1} \cdots w_0 \in \mathcal{S}$ where $w_i \in \Sigma_k$ are the characters of the word. The sequence $\{b_i\}$ is the *base* for the numeration system. The set of Zeckendorf representations of the natural numbers mentioned in Section 1 and the numeration system mentioned prior to the statement of Theorem 2.2 are examples of based numeration systems.

If $(\mathcal{S}, N)$ is a based numeration system, then it follows from the construction that the first row of $\mathcal{W}(\mathcal{S})$ is its base sequence. Kimberling proved four necessary and sufficient conditions under which the base sequence gives rise to a generalized array $\mathcal{W}(\mathcal{S})$ in which the row sequence is self-similar. One of these conditions is that the numeration system is right extendable. It should be noted that Kimberling's result only applies to numeration systems that are based while Theorem 2.2 also applies to numeration systems for which there is no base sequence.

An example of such a numeration system is the numeration system defined by $\mathcal{S} = \{1^n0^m : n \geq 1, m \geq 0\}$. In this numeration system the first few values of the evaluation map N are

w	1	10	11	100	110	111	1000	1100	1110	1111	10000
$N(w)$	1	2	3	4	5	6	7	8	9	10	11

To see that this numeration system has no base sequence, notice that since $\mathcal{S}$ is right extendable, the base sequence must be the sequence $\{N(10^n)\}_{n=0}^{\infty}$. However, in this base the word 110 would evaluate to 6 instead of 5.

The first 5 rows of array $\mathcal{W}(\mathcal{S})$ are

$$\begin{array}{cccccccc} 1 & 2 & 4 & 7 & 11 & 16 & \dots \\ 3 & 5 & 8 & 12 & 17 & 23 & \dots \\ 6 & 9 & 13 & 18 & 24 & 31 & \dots \\ 10 & 14 & 19 & 25 & 32 & 40 & \dots \\ 15 & 20 & 26 & 33 & 41 & 50 & \dots \end{array}$$

The row sequence of $\mathcal{W}(\mathcal{S})$ is

$$1,\ 1,\ 2,\ 1,\ 2,\ 3,\ 1,\ 2,\ 3,\ 4,\ 1,\ 2,\ 3,\ 4,\ 5,\ \dots$$

which is easily seen to be self-similar under the action of removing the first occurrence of each natural number.

As another example of a non-based numeration system, consider the array constructed by subtracting 1 from all the elements of the original Wythoff array (disregarding the 0 term in the first row and shifting that row to the left). The first 5 rows of this array are

$$\begin{array}{cccccccc} 1 & 2 & 4 & 7 & 12 & 20 & 33 & 54\dots \\ 3 & 6 & 10 & 17 & 28 & 46 & 75 & 122\dots \\ 5 & 9 & 15 & 25 & 41 & 67 & 109 & 177\dots \\ 8 & 14 & 23 & 38 & 62 & 101 & 164 & 266\dots \\ 11 & 19 & 31 & 51 & 83 & 135 & 219 & 355\dots \end{array}$$

This array is the generalized array $\mathcal{W}(\mathcal{S})$ arising from the numeration system S of all binary words not containing the subword 11, minus the word 1. Since $1 \notin \mathcal{S}$, this numeration system cannot be based [2, Lemma 1]. The row sequence

$$1,\ 1,\ 2,\ 1,\ 3,\ 2,\ 1,\ 4,\ 3,\ 2,\ 5,\ 1,\ 6,\ 4,\ 3,\ \dots$$

is again self-similar, by Theorem 2.2.

It is also worth noting that Kimberling's result for based numeration systems only applies to base sequences in which the digit representation can be obtained by the standard greedy algorithm (see [1] for an explanation of greedy expansions with respect to an arbitrary base sequence). In general, we say that a based numeration system $(\mathcal{S}, N)$ with alphabet Σ_k is *greedy* if, whenever $w \in \mathcal{S}$ and $v \in \Sigma_k^*$, with $N(w) = \pi(w) = \pi(v)$, it follows that $w \geq v$ under the radix order. That is, for every $n \geq 1$ the $\mathcal{S}$-representation of n is the largest possible representation under the radix order (see [2], [8], and [11]).

In the event that a greedy based numeration system is right extendable, our results follow from those of Kimberling, but Theorem 2.2 also applies to based numeration systems in which the digits cannot be obtained via the greedy algorithm. For an example of such a numeration system, consider the numeration system $\mathcal{S}$ defined as follows. Let $\mathcal{S}$ be the minimal set such that:

(i) all binary words beginning with 1 that do not possess the subword 11 are in $\mathcal{S}$ and $1011 \in \mathcal{S}$;

(ii) if $w10000 \in \mathcal{S}$ and w is not the empty word then $w01011 \in \mathcal{S}$;
(iii) if $w \in \mathcal{S}$ then $w0 \in \mathcal{S}$.

It is straightforward to show that $\mathcal{S}$ is based using [2, Theorem 1]. Let $\{b_i\}$ be the base sequence. It can quickly be checked that $b_0 = 1$, $b_1 = 2$, $b_2 = 3$, and $b_3 = 5$. Furthermore, if $k \geq 4$ then 10110^{k-4} is the largest word of length k in $\mathcal{S}$, so the base sequence must satisfy

$$b_i = b_{i-1} + b_{i-3} + b_{i-4} + 1$$

for $i \geq 4$. It may be of interest to note that $\{b_i\}$ seems to be the sequence A097083 from [12], which satisfies no known recurrence relation. Moreover, $\mathcal{S}$ is not a greedy based numeration system, because if it were the representation of 8, for example, would be 1100 instead of 1011.

We note also that the set of representations of the natural numbers obtained from $\{b_i\}$ using the greedy algorithm is not right extendable. For example, the greedy representation of 14 under this base would be 11000. However, the word 110000 would be the representation of 24, which has a greedy representation of 1000000. Therefore, by Kimberling's result [6], the row sequence of this greedy numeration system is not self similar.

On the other hand, the numeration system $\mathcal{S}$ as defined above is not greedy but it is right extendable. The first few rows of the array $\mathcal{W}(\mathcal{S})$ are then

$$\begin{array}{ccccccc}
1 & 2 & 3 & 5 & 9 & 15 & 24\ldots \\
4 & 7 & 12 & 20 & 33 & 54 & 88\ldots \\
6 & 11 & 18 & 29 & 48 & 79 & 128\ldots \\
8 & 14 & 23 & 38 & 63 & 103 & 167\ldots \\
10 & 17 & 27 & 44 & 73 & 119 & 192\ldots \\
13 & 22 & 36 & 59 & 97 & 158 & 256\ldots
\end{array}$$

and the row sequence

$$1,\ 1,\ 1,\ 2,\ 1,\ 3,\ 2,\ 4,\ 1,\ 5,\ 3,\ 2,\ 6,\ 4,\ 1,\ 7\ldots$$

is self similar, by Theorem 2.2. These examples show that our result can be viewed as an extension of the result of Kimberling from [6].

References

[1] A. Fraenkel, *Systems of Numeration*, *Amer. Math. Monthly* **92** (1985), 105–114.
[2] Garth D., Ha T., Palmer J., *Self Generating Sets and Numeration Systems*, Combinatorial Number Theory, 41-56, Walter de Gruyter, Berlin, 2009.
[3] Kimberling, C., The First Column of an Interspersion, Fibonacci Quart. **32**, (1994), 301-314.
[4] Kimberling, C., Fractal Sequences and Interspersions, Ars Combinn. **45**, (1997), 157-168.
[5] Kimberling, C., *Interspersions and Dispersions*, Proc. Amer. Math. Soc. **117** (1993), 313-321.
[6] Kimberling, C., *Numeration Systems and Fractal Sequences*, Acta Arithmetica **73** (1995), 103-117.
[7] Kimberling, C., *The Zeckendorf Array Equals the Wythoff Array*, Fibonacci Quart. **33** (1995), no. 1, 3-8.
[8] P.B.A. Lecomte and M. Rigo, *Numeration systems on a regular language*, Theory of Comput. Syst. **34** (2001) 27–44.
[9] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, Cambridge, 2002.
[10] Pickover, C.A., *Fractal Horizons*, St. Martin's Press, New York, 1996.
[11] M. Rigo, *Numeration systems on a regular language: Arithmetic operations, recognizability, and formal power series*, Theoret. Comput. Sci. **269** (2001), no. 1-2, 469–498.
[12] Sloane, N.J.A., *Online Encyclopedia of Integer Sequences*, http://www.research.att.com/~njas/sequences/selfsimilar.html.

AMS Classification Numbers: 11B13, 11B75

Department of Mathematics and Computer Science, Truman State University, Kirksville, Missouri 63501

E-mail address: dgarth@truman.edu

Department of Mathematics, University of California, San Diego, La Jolla, California, 92093

E-mail address: j5palmer@ucsd.edu