Journal of Integer Sequences, Vol. 10 (2007),
Article 07.1.5

# Affinely Self-Generating Sets and Morphisms

David Garth and Adam Gouge
Division of Mathematics and Computer Science
Truman State University
Kirksville, MO 63501
USA
dgarth@truman.edu
acg938@truman.edu

**Abstract**

Kimberling defined a self-generating set $S$ of integers as follows. Assume 1 is a member of $S$, and if $x$ is in $S$, then $2x$ and $4x-1$ are also in $S$. We study similar self-generating sets of integers whose generating functions come from a class of affine functions for which the coefficients on $x$ are powers of a fixed base. We prove that for any positive integer $m$, the resulting sequence, reduced modulo $m$, is the image of an infinite word that is the fixed point of a morphism over a finite alphabet. We also prove that the resulting characteristic sequence of $S$ is the image of the fixed point of a morphism of constant length, and is therefore automatic. We then give several examples of self-generating sets whose expansions in a certain base are characterized by sequences of integers with missing blocks of digits. This expands upon earlier work by Allouche, Shallit, and Skordev. Finally, we give another possible generalization of the original set of Kimberling.

## 1 Introduction

Kimberling [4] defined a self-generating set $S$ as follows.

1. 1 belongs to $S$;
2. if $x$ belongs to $S$, then $2x$ and $4x-1$ belong to $S$; and
3. nothing else belongs to $S$.

Writing the integers in $S$ in increasing order gives the sequence

$$1,\ 2,\ 3,\ 4,\ 6,\ 7,\ 8,\ 11,\ 12,\ 14,\ 15,\ 16,\ 22,\ 23,\ 24,\ 27,\ 28,\ 30,\ 31,\ 32,\ldots,$$

which is Sloane's `A052499` [7]. Reducing the sequence modulo 2, we obtain

$$1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ldots$$

Kimberling found that this last sequence, starting at the second term, is the well-known Fibonacci word. Recall that the Fibonacci word can be obtained as the fixed point of the morphism $0 \to 01$, $1 \to 0$. The paper by Kimberling also noted that the sequence $R(n)$ of ranks past 1 of the even terms of $S$ is

$$R(n) = 1,\ 3,\ 4,\ 6,\ 8,\ 9,\ 11,\ \ldots,$$

which is sequence `A000201` [7].

Allouche, Shallit, and Skordev [2] made some additional interesting observations about $S$. In particular, they showed that the set $T := S - 1$, the set obtained from $S$ by subtracting 1 from all the elements of $S$, is also self-generating as follows:

1. 0 belongs to $T$
2. If $x$ belongs to $T$ then $2x + 1$ and $4x + 2$ belong to $T$
3. Nothing else belongs to $T$.

In addition, they showed that $T$ is the set of integers whose base 2 representation does not contain the block 00. These binary expansions correspond to the lazy Fibonacci expansions of the set of natural numbers.

Kimberling [5] generalized the notion of self-generating sets of integers in the following way. Let $F$ be a countable set of affine functions with integer coefficients. Let $S_F$ be the set of numbers satisfying the conditions that $1 \in S_F$, and if $x \in S_F$ and $f \in F$, then $f(x) \in S_F$.

Kimberling's paper contained several examples. The paper by Allouche, Shallit, and Skordev [2] cited some of these examples in connection with sequences of integers with missing blocks. For most of these examples they showed that the characteristic sequence of $S_F$ is 2-automatic. They then asked, "How far can the sequence proposed by Kimberling be generalized?" In particular they asked, "Under which conditions is the corresponding characteristic sequence automatic?" We address these questions for a general class of affine functions.

For the remainder of this paper, let $F$ be the family of functions defined as follows. Let $v$ and $a$ be positive integers, let $\ell_i$ be integers satisfying

$$1 = \ell_0 \le \ell_1 \le \ell_2 \le \cdots \le \ell_v$$

and let $b_i$ be integers satisfying

$$b_0 = 0, \quad \text{and} \quad 0 \le b_i \le a^{\ell_i} - 1 \quad \text{for} \quad 1 \le i \le v.$$

For $0 \le i \le v$ let $f_i(x) = a^{\ell_i} x - b_i$, and define $F = \{f_i : 0 \le i \le v\}$. Let $S$ denote the set $S_F$ described earlier. We prove the following two results.

**Theorem 1.** *Let $m$ be a positive integer. Then the sequence $S$ reduced modulo $m$ is the image, under a coding, of an infinite word that is a fixed point of a morphism over a finite alphabet.*

**Theorem 2.** *The characteristic sequence of $S$ is $a$-automatic.*

The family $F$ considered here is slightly more general than a similar family studied by Kimberling [5]. In particular, he added several restrictions to the functions in $F$ that ensure that the sets $f_j(S)$ are pairwise disjoint. In this case the numbers in $S$ are matched in one-to-one correspondence with the words over the alphabet $\{0, \ldots, v\}$. For such sets he also studied, for fixed $j$, the distribution of the numbers $\{f_j(x) : x \in S\}$ in $S$. As a result of Theorem 1 these frequencies, when they exist, are coordinates of a normalized eigenvector of the incidence matrix of the morphism corresponding to its Perron-Frobenius eigenvalue ([1], Theorem 8.4.6).

We prove Theorems 1 and 2 in Sections 3 and 4, respectively. The proofs give rise to simple algorithms for constructing the morphisms. The algorithms also reveal that Theorems 1 and 2 hold when the functions in $F$ are of the form $f_i(x) = a^{\ell_i}x + b_i$. In Section 5 we mention some examples of self-generating sets that have further connections with sequences of integers with missing blocks. In Section 6 we suggest another possible generalization of the Kimberling set in which these results apply. We begin, however, by proving some necessary lemmas.

## 2 Preliminary Lemmas for Theorem 1

In this section we show how to put a tree structure on $S$ that gives rise to the morphisms. Let $L = \max\{\ell_0, \ldots, \ell_v\}$, and let $B = \max\{b_0, \ldots, b_v\}$. Let $S = \{s_n\}_{n=1}^{\infty} = \{s_1 < s_2 < s_3 < \ldots\}$, and define $s_0 = 0$. For $n \geq 1$, define the set

$$R_n = \{x : as_{n-1} < x \leq as_n\}.$$

For $h \geq 0$, define a rooted tree $\tau_{n,h}$ of height $h$ with root $s_n$ as follows. Let $\tau_{n,0} = s_n$, and for $\tau_{n,1}$, the children of $s_n$ are the elements of $R_n$ in increasing order. Thus, if $R_n = \{s_{i_n}, \ldots, s_{i_n+p_n}\}$, then $\tau_{n,1}$ is as shown in Figure 1. (Notice that $s_{i_n+p_n} = as_n$).

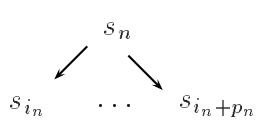



Figure 1: $\tau_{n,1}$

For $h > 1$, we continue this process recursively on the children of $s_n$ to obtain $\tau_{n,h}$.

**Example 1.** If $F = \{2x, 8x - 3, 16x - 4\}$, then $S = \{1, 2, 4, 5, 8, 10, 12, 13, 16, \ldots\}$ and $\tau_{1,4}$ is as presented in Figure 2. Notice that $\tau_{1,3}$, $\tau_{1,2}$, $\tau_{1,1}$, $\tau_{2,3}$, $\tau_{2,2}$, $\tau_{2,1}$, $\tau_{3,2}$, $\tau_{3,1}$, $\tau_{4,1}$, and $\tau_{5,1}$ are all contained in $\tau_{1,4}$.

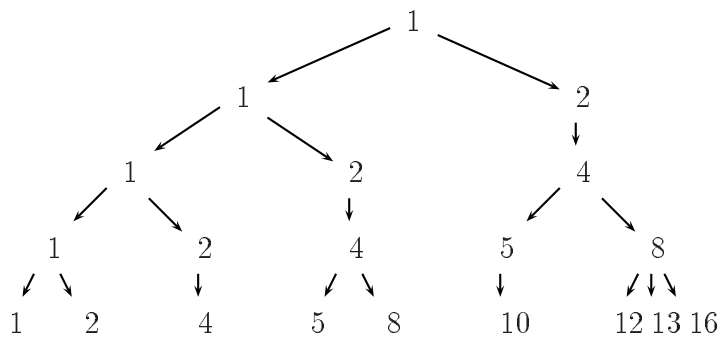



Figure 2: $\tau_{1,4}$ in Example 1

Notice that the vertices in level $h$ of $\tau_{1,h}$, in order, correspond to the first few elements of $S$. The following lemmas provide useful results about the structure of $\tau_{n,h}$. The first two are clear from the definitions. Note that it follows from the definition of $S$ and the fact that $b_0 = 0$ and $\ell_0 = 1$ that if $s \in S$, then $a^h s = f_0^h(s) \in S$ for all $h \geq 1$.

**Lemma 1.** *For $h \geq 1$, the smallest element of $S$ in level $h$ of $\tau_{n,h}$ is the successor in $S$ of $a^h s_{n-1}$, and the largest element in level $h$ is $a^h s_n$.*

**Lemma 2.** *Let $s \in S$. Then $a^h s_{n-1} < s \leq a^h s_n$ if and only if $s$ is a vertex in level $h$ of $\tau_{n,h}$.*

**Lemma 3.** *If $s_n \in S$ and $a^\ell x - b = f(x) \in F$, then $a^\ell s_n - b$ is a vertex in level $\ell$ of $\tau_{n,\ell}$.*

*Proof.* Note that $a^\ell s_n - b \in S$. By definition, $0 \leq b < a^\ell$, so

$$a^\ell s_{n-1} \leq a^\ell (s_n - 1) = a^\ell s_n - a^\ell < a^\ell s_n - b \leq a^\ell s_n$$

It follows from Lemma 2 that $a^\ell s_n - b$ is a leaf in the tree $\tau_{n,\ell}$. □

As a result of Lemmas 1 and 2, we can write an arbitrary vertex $s$ in level $h$ of $\tau_{n,h}$ in the form $a^h s_n - c$ for some constant $c$. The next Lemma establishes a bound for this constant.

**Lemma 4.** *Let $h \geq 1$, and let $a^h s_n - c$ be a vertex at level $h$ in the tree $\tau_{n,h}$. Then*

$$c \leq \left( \frac{a^h - 1}{a - 1} \right) B. \tag{1}$$

*Proof.* We use induction on $h$. Suppose $h = 1$. There exists an $s \in S$ and a function $f(x) = a^\ell x - b \in F$, with $0 \leq b < a^\ell$, such that $a^\ell s - b = f(s) = a s_n - c$. By definition of $S$, $a^{\ell-1} s \in S$, and it follows that $s_n \leq a^{\ell-1} s$. To see why, suppose $a^{\ell-1} s < s_n$. Then $a^{\ell-1} s \leq s_{n-1}$, and so $a^\ell s \leq a s_{n-1}$. Since $a s_n - c$ is a vertex in $\tau_{n,1}$, by Lemma 2 we have that $a s_n - c > a s_{n-1} \geq a^\ell s \geq a^\ell s - b = a s_n - c$, a contradiction. Thus, $a s_n \leq a^\ell s$, and since $a s_n - c = a^\ell s - b$, it follows that $c \leq b \leq B$.

Now suppose that (1) holds for some $h \geq 1$. Let $a^{h+1} s_n - c$ be a vertex in level $h+1$ of $\tau_{n,h+1}$. Then $a^{h+1} s_n - c$ is a child of some vertex $s_j$ in level $h$ of $\tau_{n,h}$. Write $s_j$ in the form $a^h s_n - c'$ for some constant $c'$. By hypothesis $c' \leq \left( \frac{a^h - 1}{a - 1} \right) B$. Also, $a^{h+1} s_n - c$ is a vertex in

$\tau_{j,1}$. We can then write $a^{h+1}s_n - c = as_j - c''$ for some $c''$. By the proof of the $h = 1$ case, we have that $c'' \leq B$. Also, since $s_j = a^h s_n - c'$, it follows that

$$a^{h+1}s_n - c = as_j - c'' = a^{h+1}s_n - ac' - c''.$$

It follows then that $c = ac' + c'' \leq a\left(\frac{a^h-1}{a-1}\right)B + B = \left(\frac{a^{h+1}-1}{a-1}\right)B$. □

The next Lemma shows that for $h \geq L-1$, the form of $\tau_{n,h}$ is completely determined by $\tau_{n,L-1}$, and in particular by the differences between successive vertices in each level of the tree $\tau_{n,L-1}$.

**Lemma 5.** *Assume $L \geq 1$, $k \neq 1$, and $n \neq 1$. Assume also that $\tau_{k,L-1}$ and $\tau_{n,L-1}$ have the property that $a^i s_k - c$ is a vertex in $\tau_{k,L-1}$ if and only if $a^i s_n - c$ is a vertex in $\tau_{n,L-1}$. Then $\tau_{k,L}$ and $\tau_{n,L}$ also possess this property.*

*Proof.* If $L = 1$ then for $k \neq 1$ and $n \neq 1$, $R_k = \{as_k - b_i : 0 \leq i \leq v\}$ and $R_n = \{as_n - b_i : 0 \leq i \leq v\}$, and the result is trivial. Assume then that $L > 1$. Let $s = a^L s_k - c$ be a vertex in level $L$ of $\tau_{k,L}$. We will show that $a^L s_n - c$ is a vertex in level $L$ of $\tau_{n,L}$. Since $s \in S$, there is an $s' \in S$ and a function $f(x) = a^\ell x - b \in F$ such that $f(s') = s$. By Lemma 3, $s$ is a vertex in level $\ell$ of the tree with root $s'$. Since $s$ is also in level $L$ of $\tau_{k,L}$, and since $\ell \leq L$, it follows that $s'$ is a leaf in the tree $\tau_{k,h}$, where $h := L - \ell$. Let $c'$ be such that $s' = a^h s_k - c'$. Since $f(s') = s$, we have that

$$a^L s_k - c = s = f(s') = f(a^h s_k - c') = a^{\ell+h}s_k - a^\ell c' - b = a^L s_k - a^\ell c' - b,$$

and so $c = a^\ell c' + b$. On the other hand, since $h \leq L-1$, by hypothesis we have that $a^h s_n - c'$ is a vertex in level $h$ of $\tau_{n,L-1}$. By Lemma 3 again, it follows that $f(a^h s_n - c')$ is a vertex in level $\ell$ of the tree with root $a^h s_n - c'$. Thus,

$$f(a^h s_n - c') = a^{\ell+h}s_n - a^\ell c' - b = a^L s_n - c$$

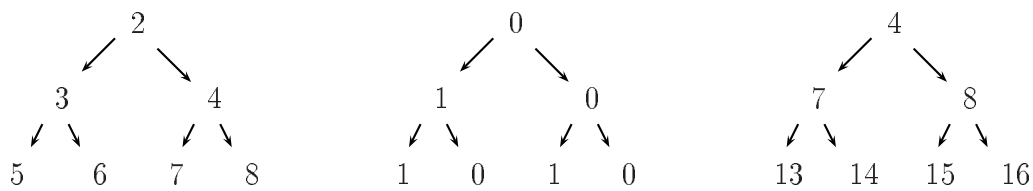
is a vertex in level $L$ of $\tau_{n,L}$. We have thus shown that if $a^L s_k - c$ is a vertex in $\tau_{k,L}$, then $a^L s_n - c$ is a vertex in $\tau_{n,L}$. Reversing the roles of $k$ and $n$, we can show that that if $a^L s_n - c$ is a vertex in $\tau_{n,L}$ then $a^L s_k - c$ is a vertex in $\tau_{k,L}$. This completes the proof. □

Note that it is a consequence of the definition of the trees that any two trees $\tau_{k,h}$ and $\tau_{n,h}$ satisfying the property of Lemma 5 have the same tree structure. Also notice that the requirement that $k \neq 1$ and $n \neq 1$ is only necessary when $L = 1$. This is because $n = 1$ is the only value of $n$ for which $s_n \in R_n$.

# 3 Proof of Theorem 1

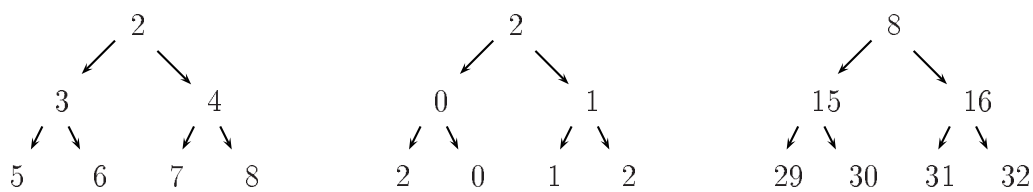
Define an equivalence relation $\sim$ on $S$ as follows. Let $\tau_{n,L-1} \bmod m$ be the tree obtained from $\tau_{n,L-1}$ by reducing the vertices of $\tau_{n,L-1}$ modulo $m$. We begin by defining $s_1 \sim s_1$. For $n \neq 1$ and $k \neq 1$, we say $s_n \sim s_k$ if and only if $\tau_{n,L-1}$ and $\tau_{k,L-1}$ satisfy the property of Lemma 5 and $\tau_{n,L-1} \bmod m = \tau_{k,L-1} \bmod m$. We pause from the proof for a brief example.

**Example 2.** Let $F = \{2x, 4x-1, 8x-3\}$. Then $2 \sim 4$ when reducing modulo 2, as the following tree structures show.




$\tau_{2,2}$ $\qquad$ $\tau_{2,2} \bmod 2 = \tau_{4,2} \bmod 2$ $\qquad$ $\tau_{4,2}$

When reducing modulo 3, note that $2 \nsim 4$, but $2 \sim 8$, as the following tree structures show.




$\tau_{2,2}$ $\qquad$ $\tau_{2,2} \bmod 3 = \tau_{8,2} \bmod 3$ $\qquad$ $\tau_{8,2}$

Now, by Lemma 4 there are only finitely many equivalence classes for this relation. We now define the morphism. Suppose $R_n = \{s_{i_n}, \ldots, s_{i_n+p_n}\}$. Let $\overline{s_n}$ denote the equivalence class of $s_n$, and let $A$ be a complete set of equivalence class representatives. Define $\sigma : A \to A^*$, where $A^*$ is the set of finite words over $A$, by

$$\sigma(\overline{s_n}) = \overline{s_{i_n}}\ \overline{s_{i_n+1}}\ \cdots\ \overline{s_{i_n+p_n}}.$$

To see that $\sigma$ is well-defined, suppose $\overline{s_n} = \overline{s_k}$ and $\sigma(\overline{s_k}) = \overline{s_{i_k}} \cdots \overline{s_{i_k+p_k}}$. If $L > 1$, then since $\overline{s_n} = \overline{s_k}$, it follows that $\tau_{n,L-1}$ and $\tau_{k,L-1}$ have the same tree structure, and therefore $p_n = p_k$. If $L = 1$, it was shown in the proof of Lemma 5 that $p_n = p_k$. In either case, it then follows from Lemma 5 that

$$\overline{s_{i_n}} = \overline{s_{i_k}}, \overline{s_{i_n+1}} = \overline{s_{i_k+1}}, \ldots, \overline{s_{i_n+p_n}} = \overline{s_{i_k+p_k}}.$$

Thus, $\sigma$ is well defined.

Extend $\sigma$ by concatenation to get a morphism from $A^* \to A^*$. It is clear that $\sigma(\overline{s_1}) = \overline{s_1} \cdots \overline{s_a}$. Thus, the infinite word given by

$$\lim_{i \to \infty} \sigma^i(\overline{s_1}) \tag{2}$$

is a fixed point of $\sigma$. Now define a map $\rho : A \to \{0, 1, \ldots, m-1\}$ by

$$\rho(\overline{s_n}) = s_n \bmod m.$$

Clearly $\rho$ is well defined. Extend $\rho$ by concatenation to obtain a morphism from $A^* \to \{0, 1, \ldots, m-1\}^*$. This morphism is then a coding [1]. It is clear that the sequence $S$ reduced modulo $m$ is the image, under $\rho$, of the fixed point (2). This completes the proof of Theorem 1.

**Example 3.** Let $F = \{3x, 3x-1\}$. For this $F$ the resulting sequence $S$ is `A032924`+1, where `A032924` is the sequence of natural numbers whose ternary representation contains no 0. We will produce the morphism for $S$ mod 3. We see that $L = 1$, so for $n \neq 1$ and $k \neq 1$, $s_n \sim s_k$ if and only if $s_n \bmod 3 = s_k \bmod 3$. The complete set of equivalence class representatives is then $\{\overline{1}, \overline{2}, \overline{3}\}$. The morphism is therefore

$$\begin{array}{lcl} \overline{1} & \to & \overline{1}\,\overline{2}\,\overline{3} \\ \overline{2} & \to & \overline{5}\,\overline{6} = \overline{2}\,\overline{3} \\ \overline{3} & \to & \overline{8}\,\overline{9} = \overline{2}\,\overline{3} \end{array}$$

Letting $a = \overline{1}$, $b = \overline{2}$, and $c = \overline{3}$, we get the morphism $a \to abc$, $b \to bc$, $c \to bc$. Iterating this morphism on $a$ produces a fixed point. Now define the morphism $\rho : \{a, b, c\}^* \to \{0, 1, 2\}^*$ by $\rho(a) = 1$, $\rho(b) = 2$, and $\rho(c) = 0$. The proof of Theorem 1 shows that the sequence $S$ mod 3 is the image of the fixed point of $\sigma$ under this coding.

**Example 4.** Let $F = \{2x, 4x-1, 8x-3\}$. We will find the morphism for $S$ mod 2. Using the method described above, we find

$$\begin{array}{lcl} \sigma(\overline{1}) & = & \overline{1}\,\overline{2} \\ \sigma(\overline{2}) & = & \overline{3}\,\overline{4} = \overline{3}\,\overline{2} \\ \sigma(\overline{3}) & = & \overline{5}\,\overline{6} = \overline{5}\,\overline{2} \\ \sigma(\overline{5}) & = & \overline{10} = \overline{2} \end{array}$$

Thus, the sequence $S$ mod 2 is the image by the coding $a \to 1$, $b \to 0$, $c \to 1$, $d \to 1$, of the fixed point of the morphism $a \to ab$, $b \to cb$, $c \to db$, $d \to b$.

**Example 5.** We use the same procedure to get the morphism for $S$ mod 3 when $F = \{2x, 4x-1, 8x-3\}$:

$$\begin{array}{ll} \sigma(\overline{1}) = \overline{1}\,\overline{2} & \sigma(\overline{6}) = \overline{11}\,\overline{6} \\ \sigma(\overline{2}) = \overline{3}\,\overline{4} & \sigma(\overline{7}) = \overline{13}\,\overline{2} \\ \sigma(\overline{3}) = \overline{5}\,\overline{6} & \sigma(\overline{11}) = \overline{21}\,\overline{4} \\ \sigma(\overline{4}) = \overline{7}\,\overline{2} & \sigma(\overline{13}) = \overline{2} \\ \sigma(\overline{5}) = \overline{4} & \sigma(\overline{21}) = \overline{6} \end{array} \qquad \text{or} \qquad \begin{array}{ll} a \to ab & f \to hf \\ b \to cd & g \to ib \\ c \to ef & h \to jd \\ d \to gb & i \to b \\ e \to d & j \to f \end{array}$$

with the coding $b \to 2$, $e \to 2$, $h \to 2$, $a \to 1$, $d \to 1$, $g \to 1$, $i \to 1$, $c \to 0$, $f \to 0$, $j \to 0$.

# 4 Proof of Theorem 2

Recall that $L = \max\{\ell_0, \ldots, \ell_v\}$ and $B = \max\{b_0, \ldots, b_v\}$. Let $\{u_n\}_{n=0}^{\infty}$ be the characteristic sequence of $S$. We will show that $\{u_n\}$ is the image, under a coding, of a fixed point of

a morphism of constant length. Specifically, we will show how to construct a morphism of length $a$ for which $\{u_n\}$ is the image, under a coding, of one of its fixed points. It will follow that $\{u_n\}$ is $a$-automatic ([1], Theorem 6.3.2).

Define an equivalence relation $\sim$ on $\{u_n\}$ as follows. Associate with each $u_n$ a rooted tree of height $L-1$, with $u_n$ as the root , in which each vertex $u_i$ has children

$$u_{ai-(a-1)}, \ldots, u_{ai-1}, u_{ai},$$

in order. It follows from this definition that the vertices of the tree of level $h$ are the terms

$$u_{a^h n-(a^h-1)}, \ldots, u_{a^h n-1}, u_{a^h n}.$$

Figure 3 shows a tree for $u_n$ when $L=3$ and $a=2$.

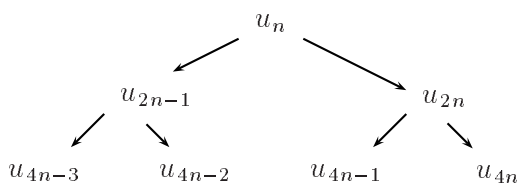



Figure 3: Tree associated with $u_n$ for $L=3$ and $a=2$

To define $\sim$, let $u_1 \sim u_1$. If $n \neq 1$ and $k \neq 1$, we say $u_n \sim u_k$ if and only if the corresponding vertices of the trees are equal. Clearly there are only finitely many equivalence classes for this relation. Let $\overline{u_n}$ be the equivalence class of $u_n$, and let $A$ be a complete set of equivalence class representatives. Define the mapping $\sigma : A \to A^*$ by

$$\sigma(\overline{u_n}) = \overline{u_{an-(a-1)}} \cdots \overline{u_{an-1}}\ \overline{u_{an}}.$$

To show that $\sigma$ is well-defined, assume that $\overline{u_n} = \overline{u_k}$. We need to show that $\overline{u_{an-i}} = \overline{u_{ak-i}}$ for $0 \leq i \leq a-1$. It suffices to consider the case that $n \neq 1$ and $k \neq 1$, since $u_1$ is the only element in the equivalence class for $u_1$. Extend the trees associated with $u_n$ and $u_k$ by one level to obtain rooted trees of height $L$. Figure 4 shows an extended tree for $L=3$ and $a=2$.

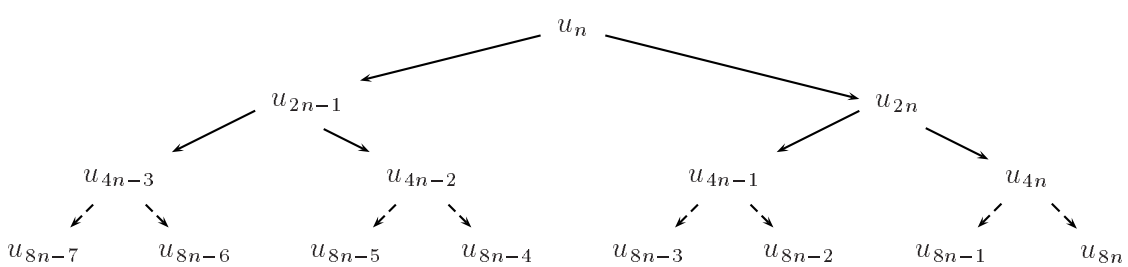



Figure 4: The extended tree for $u_n$ with $L=3$ and $a=2$.

Since $\overline{u_n} = \overline{u_k}$, the corresponding vertices in the extended trees associated with $u_n$ and $u_k$ are equal for levels 0 through $L-1$. We will show that the corresponding vertices in level $L$ of the extended trees are also equal.

Let $u_{a^L n-j}$ be a vertex at level $L$ of the extended tree for $u_n$, where $0 \leq j \leq a^L - 1$. Assume also that $u_{a^L n-j} = 1$. We will show that $u_{a^L k-j} = 1$. Since $u_{a^L n-j} = 1$, then $a^L n - j$ is an element of $S$. Since $n \neq 1$, then $a^L n - j \neq 1$, and so there is an $s \in S$ and an $f(x) = a^\ell x - b \in F$ such that

$$a^L n - j = f(s) = a^\ell s - b. \tag{3}$$

Recall that $0 \leq b < a^\ell$, and $\ell \leq L$.

If $\ell = L$, then $a^L s - b = a^L n - j$, and so $a^L(s - n) = b - j$. Since $0 \leq b, j < a^L$, this is only possible if $b = j$ and $s = n$. In this case we have that $n \in S$. Since $\overline{u_n} = \overline{u_k}$, it must be the case that $u_k = 1$, and so $k \in S$. It follows that $f(k) = a^L k - b = a^L k - j$ is also in $S$, and thus $u_{a^L k-j} = 1$.

Suppose then that $\ell < L$. It must be the case that $a^{L-\ell} n - (a^{L-\ell} - 1) \leq s \leq a^{L-\ell} n$. To see why, notice that if $s < a^{L-\ell} n - (a^{L-\ell} - 1)$, then $s \leq a^{L-\ell} n - (a^{L-\ell} - 1) - 1 = a^{L-\ell}(n-1)$, and so by (3) we have

$$\begin{aligned} a^L n - j = a^\ell s - b &\leq a^\ell(a^{L-\ell}(n-1)) - b \\ &= a^L(n-1) - b \\ &\leq a^L(n-1) \\ &< a^L(n-1) + 1 \\ &= a^L n - (a^L - 1) \\ &\leq a^L n - j. \end{aligned}$$

This is a contradiction, and so $a^{L-\ell} n - (a^{L-\ell} - 1) \leq s$. On the other hand, suppose that $s > a^{L-\ell} n$. Then $s \geq a^{L-\ell} n + 1$, and since $0 \leq b < a^\ell$, it follows from (3) again that

$$\begin{aligned} a^L n - j = a^\ell s - b &\geq a^\ell(a^{L-\ell} n + 1) - b \\ &= a^L n + a^\ell - b \\ &\geq a^L n + 1 \\ &> a^L n - j, \end{aligned}$$

a contradiction, and so $s \leq a^{L-\ell} n$. We have shown then that if $\ell < L$, then $a^{L-\ell} n - (a^{L-\ell} - 1) \leq s \leq a^{L-\ell} n$. It follows then that $u_s$ is a vertex in the tree of height $L - 1$ associated with $u_n$, and we can write

$$s = a^{L-\ell} n - m, \tag{4}$$

where $0 \leq m \leq a^{L-\ell} - 1$. Notice that we also have that $a^{L-\ell} k - m$ is a vertex in the tree of height $L-1$ associated with $u_k$. By (3) and (4) we have that $a^L n - j = f(s) = a^L n - a^\ell m - b$, and thus

$$j = a^\ell m + b. \tag{5}$$

Now, since $a^{L-\ell} n - m = s \in S$, $u_s = 1$, and since $\overline{u_n} = \overline{u_k}$, we have that $a^{L-\ell} k - m$ is in $S$. It also follows from (5) that $a^L k - j = a^L k - a^\ell m - b = f(a^{L-\ell} k - m)$. Thus, $a^L k - j$ is also in $S$, and therefore $u_{a^L k-j} = 1$.

We have shown that if $u_{a^L n-j} = 1$ then $u_{a^L k-j} = 1$. By reversing the roles of $n$ and $k$, the same proof shows that if $u_{a^L k-j} = 1$ then $u_{a^L n-j} = 1$. Thus, for $0 \leq j \leq a^L - 1$, $u_{a^L n-j} = 1$

if and only if $u_{a^L k-j} = 1$. This proves that $u_{a^L n-j} = u_{a^L k-j}$ for $0 \le j \le a^L - 1$. Thus, the extended trees for $u_n$ and $u_k$ are the same. Since, for $0 \le i \le a-1$, the trees associated with $u_{an-i}$ and $u_{ak-i}$ are subtrees of the extended trees for $u_n$ and $u_k$, respectively, it follows that $u_{an-i} = u_{ak-i}$ for $0 \le i \le a-1$. Thus, $\sigma$ is well-defined.

Now extend $\sigma$ to a mapping from $A^* \to A^*$ by concatenation. Since $\sigma(\overline{u}_1) = \overline{u_1} \cdots \overline{u_{an}}$, repeated iteration of $\sigma$ on $\overline{u}_1$ generates a fixed point. As in the end of the proof of Theorem 1, we can define a coding which associates $\overline{u_n}$ with $u_n$. The characteristic sequence $\{u_n\}$ is then the image of this fixed point. Since the morphism $\sigma$ is of length $a$, the characteristic sequence is $a$-automatic ([1], Theorem 6.3.2).

**Example 6.** We find the morphism generating the characteristic sequence of $S$ when $F = \{2x, 4x-1, 8x-3\}$.

$$\begin{aligned}
\sigma(\overline{u_1}) &= \overline{u_1}\,\overline{u_2} \\
\sigma(\overline{u_2}) &= \overline{u_3}\,\overline{u_4} = \overline{u_3}\,\overline{u_2} \\
\sigma(\overline{u_3}) &= \overline{u_5}\,\overline{u_6} = \overline{u_5}\,\overline{u_2} \\
\sigma(\overline{u_5}) &= \overline{u_9}\,\overline{u_{10}} = \overline{u_9}\,\overline{u_2} \\
\sigma(\overline{u_9}) &= \overline{u_{17}}\,\overline{u_{18}} = \overline{u}_9\,\overline{u_9}
\end{aligned}$$

This corresponds to the morphism $a \to ab$, $b \to cb$, $c \to db$, $d \to eb$, and $e \to ee$, whose fixed point produced by iteration on $a$ yields the characteristic sequence of $S$ via the coding $a \to 1$, $b \to 1$ $c \to 1$, $d \to 1$, $e \to 0$.

**Example 7.** Let $F = \{3x, 9x-1\}$. Then $L = 2$, and so each $u_n$ in the characteristic sequence $\{u_n\}$ of $S$ is associated with the tree of height 1 having children $u_{3n-2}$, $u_{3n-1}$, and $u_{3n}$. The morphism generating the characteristic sequence maps $\overline{u_n} \to \overline{u_{3n-2}}\,\overline{u_{3n-1}}\,\overline{u_{3n}}$. We find that

$$\begin{aligned}
\sigma(\overline{u_1}) &= \overline{u_1}\,\overline{u_2}\,\overline{u_3} \\
\sigma(\overline{u_2}) &= \overline{u_4}\,\overline{u_5}\,\overline{u_6} = \overline{u_2}\,\overline{u_2}\,\overline{u_2} \\
\sigma(\overline{u_3}) &= \overline{u_7}\,\overline{u_8}\,\overline{u_9} = \overline{u_2}\,\overline{u_8}\,\overline{u_3} \\
\sigma(\overline{u_8}) &= \overline{u_{22}}\,\overline{u_{23}}\,\overline{u_{24}} = \overline{u_2}\,\overline{u_2}\,\overline{u_3}
\end{aligned}$$

This corresponds to the morphism $a \to abc$, $b \to bbb$, $c \to bdc$, and $d \to bbc$. The fixed point of this morphism produced by iteration on $a$ gives the characteristic sequence via the coding $a \to 1$, $c \to 1$, $d \to 1$, $b \to 0$.

# 5 Examples

In this section we give some additional examples of self-generating sets which produce sequences of integers defined by missing blocks. It was mentioned already that the original sequence of Kimberling [4] is 1+`A003754`, where `A003754` is the sequence of integers whose binary expansion does not contain the block 00. The binary expansions of the integers in this sequence also correspond to the lazy Fibonacci expansions of the natural numbers. The following examples expand this idea. (For more on lazy expansions, see [3] and [6]).

**Example 8.** Let $S$ be the set generated by $F = \{2x, 4x-1, 8x-3\}$, starting with 1. By reasoning similar to that used for the original Kimberling sequence, $S$ is 1+`A003796`, where `A003796` is the sequence of integers whose binary expansion does not contain the block 000. The binary expansions of the integers in this sequence also correspond to the lazy expansions of the natural numbers with respect to the sequence of Tribonacci numbers, starting with $t_2 = 1$, $t_3 = 2$, $t_4 = 4$, and $t_n = t_{n-1} + t_{n-2} + t_{n-3}$ for $n \geq 5$. The sequence $S$ reduced modulo 2, starting with the second term, gives

$$0,\ 1,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 1,\ \ldots,$$

which is the image of the fixed point of the tribonacci morphism $a \to ab$, $b \to ac$, $c \to a$, under the coding $a \to 0$ and $b \to 1$, $c \to 1$.

**Example 9.** In general, if $F = \{2^k x - (2^{k-1} - 1) : 1 \leq k \leq n\}$, then $S - 1$ is the sequence of integers whose binary expansion does not contain blocks of $n$ zeros. These expansions correspond to the lazy expansions of the natural numbers with respect to the base given by the sequence of $n$-bonacci numbers. The sequence $S$ reduced modulo 2, starting at the second term, is the image of the fixed point of the morphism $a_1 \to a_1a_2$, $a_2 \to a_1a_3$, $\ldots$, $a_{n-1} \to a_1a_{n-1}$, and $a_n \to a_1$, under the coding $a_1 \to 0$ , $a_2 \to 1, \cdots, a_n \to 1$.

Familiarity with the method of the proof Theorem 1 reveals that the Theorem is also valid when the functions in $F$ are all of the form $f_i(x) = a^{\ell_i}x + b_i$. Moreover, if the minimal requirement is that $0 \in S$, then the requirement that $b_0 = 0$ can be removed.

**Example 10.** Let $F = \{3x+1, 3x+2, 9x+3, 9x+6\}$. If the minimal requirement for $S$ is that $0 \in S$, then it is not hard to see that $S$ is the sequence of integers whose ternary representation does not contain the block 00. These ternary representations correspond to the lazy representations of the natural numbers with respect to sequence `A028859`, which is given recursively by $r_1 = 1$, $r_2 = 3$, $\ldots$, $r_n = 2r_{n-1} + 2r_{n-2}$. It is also interesting to note that if the sequence $S + 1$ is reduced modulo 3 and then reduced again modulo 2, then the resulting sequence, starting from the second character, is the fixed point of the morphism $0 \to 001$, $1 \to 00$.

It is worth mentioning that the preceding examples can be easily generalized to generate the sequences of integers with missing blocks of zeros of any length in any base.

# 6 Further Generalizations

In this section we explore another possible generalization of Kimberling's set and its connections with morphisms and sequences of integers with missing blocks. Consider the following example.

1. Let $1, 2$ belong to $S$.

2. If $x$ belongs to $S$ then $3x$ belongs to $S$.

3. If $x$ belongs to $S$ and either $x = 3k$ or $x = 3k - 1$ for some $k \in \mathbb{Z}$, then $3x - 1$ and $3x - 2$ belong to $S$.

4. Nothing else belongs to $S$.

We say that $S$ is *conditionally self-generating*. Arranging the elements of $S$ in increasing order gives the sequence

$$S = 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 12,\ 13,\ 14,\ 15,\ 16,\ 17,\ 18,\ 21,\ 22,\ 23,\ \ldots$$

Letting $T := S - 1$, it is not hard to see that the integers in $T$ are those whose ternary representation does not contain the blocks 00 or 01. Moreover, these representations are precisely the Lazy representations of the natural numbers with respect to sequence `A001333`,

$$1,\ 3,\ 7,\ 17,\ 41,\ 99,\ 239, \ldots,$$

which is given by the recurrence $a_0 = 1$, $a_1 = 3$, $a_n = 2a_{n-1} + a_{n-2}$. Reducing $S$ modulo 3 gives the sequence

$$1,\ 2,\ 0,\ 1,\ 2,\ 0,\ 1,\ 2,\ 0,\ 0,\ 1,\ 2,\ 0,\ 1,\ 2,\ 0,\ 0,\ 1,\ 2,\ 0,\ 1,\ 2,\ 0,\ \ldots$$

It is not hard to see that the proof of Theorem 1 applies to this $S$. The method of Theorem 1 shows that the sequence $S$ reduced modulo 3 is the image by the map $a \to 1$, $b \to 2$, $c \to 0$, $d \to 1$, $e \to 2$ of the fixed point of the morphism $a \to abc$, $b \to dec$, $c \to dec$, $d \to c$, $e \to dbc$. Reducing this last sequence modulo 2 gives

$$1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ \ldots$$

Again, it is not hard to see that the proof of Theorem 1 also applies when reducing modulo 3 and then again modulo 2. This reduced sequence turns out to be the image of the fixed point of the morphism $a \to abb$, $b \to cbb$, and $c \to b$, under the coding $a \to 1$, $b \to 0$, $c \to 1$. It is interesting to note that omitting the first term, this sequence is actually the fixed point of the morphism:

$$0 \to 001 \quad \text{and} \quad 1 \to 0.$$

Finally, it is also not hard to see that the proof of Theorem 2 applies in this case, and the characteristic sequence of $S$ is 3-automatic, being the image of the fixed point of the morphism $a \to abb$, $b \to cbb$, $c \to ddb$, and $d \to ddd$, under the coding $a \to 1$, $b \to 1$, $c \to 1$, $d \to 0$.

# 7 Open Questions

We close with some remaining questions.

- Do Theorems 1 and 2 hold when the generating functions are characterized by "mixed base" rules? For example, do they hold when $F = \{2x, 3x + 1\}$? Allouche, Shallit, and Skordev [2] commented that in this case the corresponding characteristic sequence for $S$ is probably not $k$-automatic for any $k$. It seems likely that there is no analog of Lemmas 4 and 5 in the mixed base case.

- Do Theorems 1 and 2 hold for more general families of functions? The proofs of Theorems 1 and 2 rely heavily on the fact that if $f(x) = a^\ell x - b \in F$, then $0 \le b < a^\ell$. What if functions were allowed for which $b \ge a^\ell$?

## References


[1] Jean-Paul Allouche and Jeffrey Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.

[2] J.-P. Allouche, J. Shallit, G. Skordev, Self-generating sets, integers with missing blocks, and substitutions, *Discrete Math.* **292** (2005), 1–15.

[3] P. Erdős, I. Joó, V. Komornik, Characterization of the unique expansions $1 = \sum_{i=1}^{\infty} q^{-n_i}$ and related problems, *Bull. Soc. Math. France* **118** (1990), 377–390.

[4] C. Kimberling, *A self-generating set and the golden mean*, J. Int. Seq. **3** (2000) Article 00.2.8. http://www.cs.uwaterloo.ca/journals/JIS/VOL3/goldentext.html

[5] C. Kimberling, Affinely recursive sets and orderings of languages, Discrete Math. **274** (2004), 147–159.

[6] A. Petho and R.F. Tichy, On digit expansions with respect to linear recurrences, *J. Number Theory*, **33** (1989), 243-256.

[7] N.J.A. Sloane, On-line Encyclopedia of Integer Sequences, published electronically at http://www.research.att.com/~njas/sequences/


---


2000 *Mathematics Subject Classification*: Primary 11B99; Secondary 11B85.
*Keywords:* Kimberling sequence, self-generating set, morphism, characteristic sequence, automatic sequence, lazy expansion, integers with missing blocks


---

(Concerned with sequences A000201 A001333, A003754, A003796, A028859, A032924, and A052499.)

---


Received August 4 2006; revised versions received September 14 2006; October 19 2006. Published in *Journal of Integer Sequences*, December 30 2006.


---


Return to Journal of Integer Sequences home page.