Numeration Systems, Linear Recurrences, and Regular Sets

Jeffrey Shallit

Research Report CS-91-32
July 1991

Department of Computer Science
University of Waterloo
Waterloo, Ontario
Canada N2L 3G1

Faculty of Mathematics

University of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

Numeration Systems, Linear Recurrences, and Regular Sets

Jeffrey Shallit*
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
shallit@graceland.waterloo.edu

July 21, 1991

Abstract

A numeration system based on a strictly increasing sequence of positive integers $u_0 = 1, u_1, u_2, \ldots$ expresses a non-negative integer n as a sum $n = \sum_{j=0}^{i} a_j u_j$. In this case we say the string $a_i a_{i-1} \cdots a_1 a_0$ is a *representation* for n. If $\gcd(u_0, u_1, \ldots) = g$, then every sufficiently large multiple of g has some representation.

If the lexicographic ordering on the representations is the same as the usual ordering of the integers, we say the numeration system is *order-preserving*. In particular, if $u_0 = 1$, then the *greedy representation*, obtained via the greedy algorithm, is order-preserving. We prove that, subject to some technical assumptions, if the set of all representations in an order-preserving numeration system is regular, then the sequence $u = (u_j)_{j \geq 0}$ satisfies a linear recurrence. The converse, however, is not true.

The proof uses two lemmas about regular sets that may be of independent interest. The first shows that if L is regular, then the set of lexicographically greatest strings of every length in L is also regular. The second shows that the number of strings of length n in a regular language L is bounded by a constant (independent of n) iff L is the finite union of sets of the form wx^*y.

1 Introduction

Let Σ be a finite or infinite alphabet. A *numeration system* is a map $\mathbb{N} \rightarrow \Sigma^*$ that assigns a string, called the *representation* of n, to a nonnegative integer n. Common examples

*Supported in part by a grant from NSERC .

of numeration systems include radix-k representation (for k an integer ≥ 2), Fibonacci representation, factorial representation, etc. See [7, 8] for surveys on numeration systems.

One of the most common ways to construct a numeration system is to start with a strictly increasing sequence of positive integers $u_0, u_1, u_2, \ldots$ and try to express n as a non-negative integer linear combination of the u_j, say $n = \sum_{0 \leq j \leq i} a_j u_j$. If we can write n in this manner, we say n is *representable*, and one representation is the string $a_i a_{i-1} \cdots a_1 a_0$. We define

$$\operatorname{val}(a_i a_{i-1} \cdots a_1 a_0) = \sum_{0 \leq j \leq i} a_j u_j.$$

We say such a representation is *normal* if $a_i \neq 0$.

Note that if

$$\gcd(u_0, u_1, \ldots) = g,$$

then every sufficiently large multiple of g is representable.

For some choices of the sequence $u = (u_j)_{j \geq 0}$, the "digits" a_j may be required to be arbitrarily large. An example of this is the so-called *factorial representation*, where $u_j = (j+1)!$. See, for example, [7]. In this paper we only consider numeration systems with bounded digits. A necessary condition to ensure bounded digits is that the ratio u_j / u_{j-1} is bounded by a constant.

Since there may be many normal representations for a number n, it makes sense to try to identify one which is "canonical". This can be done in a variety of ways; for example, one could choose the *lexicographically least* or *lexicographically greatest* representation. (If $x = x_1 x_2 \cdots x_i$ and $y = y_1 y_2 \cdots y_j$ are strings, we say x is lexicographically greater than y, and write $x > y$, if $i > j$, or if $i = j$ and there exists an integer k, $1 \leq j \leq n$, such that $x_1 = y_1, x_2 = y_2, \ldots, x_{k-1} = y_{k-1}$, but $x_k > y_k$.)

A desirable property of a numeration system is that the mapping that sends an integer n to its (normal) canonical representation be *order-preserving*. More precisely, we require that for representable integers m, n, we have $m > n$ iff $\operatorname{rep}(m) > \operatorname{rep}(n)$. Here $\operatorname{rep}(m)$ denotes the canonical representation for m.

The *greedy representation* $\operatorname{grep}(n)$ of a positive integer n is defined as follows: let i be the largest index such that $u_i \leq n$. Then successively set $a_i \leftarrow \lfloor n/u_i \rfloor$, $n \leftarrow n - a_i u_i$, and $i \leftarrow i - 1$ until $i < 0$. If $n = \sum_{0 \leq j \leq i} a_j u_j$, then the greedy representation for n is the string $a_i a_{i-1} \cdots a_1 a_0$, and we say n is *greedily representable*. If not, then the representation for n is undefined. The greedy representation for 0 is defined to be ϵ, the empty string. The set of representations of all greedily representable integers is written $G(u)$.

Every non-negative integer is greedily representable iff $u_0 = 1$. If $u_0 \neq 1$, it is possible for a number to be representable, but not greedily representable. For example, consider expressing 4 in the numeration system $(u_0, u_1, u_2, \ldots) = (2, 3, 5, \ldots)$.

It is easy to see that the greedy representation is order-preserving; furthermore, it coincides with the lexicographically greatest representation if $u_0 = 1$.

We define $R(u)$ to be the set of canonical representations for all non-negative integers. More formally,

$$R(u) = \sum_{\substack{n \geq 0 \\ n \text{ representable}}} \text{rep}(n).$$

Example 1.

Let $u_j = k^j$, for k an integer ≥ 2. Then the set of greedy representations $G(u)$ is

$$\epsilon + (1 + 2 + \cdots + k - 1)(0 + 1 + \cdots + k - 1)^*.$$

Example 2.

Let $u_j = F_{j+2}$, where F_j is the jth Fibonacci number. Then it can be shown that the set of greedy representations $G(u)$ is

$$\epsilon + 1(0 + 01)^*.$$

(For more on Fibonacci representations, see [20, Ex. 1.2.8.34], [25], [4], [19], and [1].)

Example 3.

Let $u_j = 2^{j+1} - 1$. Then the set of greedy representations is

$$\epsilon + 1(0 + 1)^* + 1(0 + 1)^*20^* + 20^*.$$

See [3].

Example 4.

Let $u_j = 2^j$, and consider a numeration system using only the digits 1 and 2. Then it is easy to see that every non-negative integer can be written uniquely, and the set of representations is given by the regular set $(1 + 2)^*$. This numeration system, while not obtained via the greedy algorithm, is nonetheless order-preserving.

In this paper, we prove the following theorem: suppose the set of all representations $R(u)$ is regular. Then the sequence u satisfies a linear recurrence with integral constant coefficients.

The proof depends on two lemmas about regular sets, which may be of independent interest.

Remarks on the literature.

The point of view we will adopt in this paper is similar to that of Frougny, who has written extensively on this topic. See [9, 10, 11, 12, 14, 13].

In [24], I proved that the set of greedy representations is regular for the numeration systems with bounded digits considered by Fraenkel [7].

We note several other papers that have examined the relationship between ways of representing numbers and regular sets. See [15, 23, 6, 22, 16, 17]. However, these papers have adopted a very different point of view.

2 More Notation

Throughout this paper, Σ is a finite alphabet, and r, s, t denote regular expressions. The letters v, w, x, y, z denote strings. The lower-case letters $a, b, c, g, i, j, k, m, n$ and the upper-case letters A, B, C denote integers. We also use the letters a and b to represent elements of Σ. The capital letter L denotes a language and the capital letters W, X, Y denote finite languages.

3 Lexicographically Largest Strings

Suppose L is a regular language over a finite alphabet Σ. Suppose Σ has a total ordering; for example, suppose $\Sigma = \{0, 1, 2, \ldots, k-1\}$. If $x, y \in \Sigma^n$, we say $x > y$ if x is lexicographically greater than y. More precisely, we say $x > y$ if there exists an integer i, $0 \le i \le n-1$, such that $x_1 = y_1, x_2 = y_2, \ldots, x_i = y_i$, but $x_{i+1} > y_{i+1}$. Then let $B(L)$ be the union, over all $n \ge 0$, of the lexicographically largest string of length n in L. (By considering $L \cup 0^*$, we may assume without loss of generality that there is at least one string of every length in L.) More formally, define

$$B(L) = \bigcup_{n \ge 0} \{x \in L \cap \Sigma^n \ : \ (y \in L \cap \Sigma^n) \Rightarrow x \ge y\}.$$

Lemma 1 *If L is regular, then so is $B(L)$.*

Proof.

We show that $L - B(L)$, the relative complement of $B(L)$ in L, is a regular set. The result will then follow, since regular sets are closed under complement.

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a deterministic finite automaton accepting L. (See [18] for basic notions about automata and the notational conventions we use here.) The idea is to accept $L - B(L)$ with a nondeterministic finite automaton. We use two "fingers" to mimic the behavior of M on input w: the first "finger" imitates M precisely. The second "finger" nondeterministically simulates M on all possible inputs of length $|w|$, trying to find some string in L that is lexicographically greater than w. If we succeed, and w is accepted by M, then we have found a string in L that is not lexicographically greatest, and so we accept.

More formally, let $M' = (Q', \Sigma, \delta', q_0', F')$, a nondeterministic finite automaton, where $Q' = Q \times Q \times \{g, e, l\}$. Here g indicates that in the current state, we have already found a string lexicographically greater than the prefix of the input w seen so far. Similarly, e indicates equality, and l indicates less than. For each $a \in \Sigma$, define

$$\delta'([p, q, g], a) = \{[\delta(p, a), \delta(q, b), g] \ : \ b \in \Sigma^*\},$$

$$\delta'([p, q, e], a) = \{[\delta(p, a), \delta(q, b), x] \ : \ b \in \Sigma^*\},$$

where $x = g$ if $a > b$, $x = e$ if $a = b$, and $x = l$ if $a < b$, and

$$\delta'([p, q, l], a) = \{[\delta(p, a), \delta(q, b), l] \ : \ b \in \Sigma^*\}.$$

Finally, define $q_0' = [q_0, q_0, e]$ and

$$F' = \{[p, q, l] \ : \ p, q \in F\}.$$

We leave it to the reader to show that M' accepts w if and only if $w \in L - B(L)$. ■

Corollary 2 *If L is regular, then so is the set*

$$S(L) = \bigcup_{n \geq 0} \{x \in L \cap \Sigma^n \ : \ (y \in L \cap \Sigma^n) \Rightarrow x \leq y\}.$$

of lexicographically smallest strings of every length in L.

4 Bounded Regular Sets

In this section we prove that if L is a regular language such that the number of strings of length n in L is bounded by a constant (independent of n), then L is the finite union of sets of particularly simple form. More formally, we have

Lemma 3 *The following two statements are equivalent:*
(i) $L \subseteq \Sigma^*$ is regular and there exists a constant c such that $|L \cap \Sigma^n| \leq c$ for all $n \geq 0$
(ii) L is the finite union of sets of the form xy^*z, where $x, y, z \in \Sigma^*$.

Proof.
$(ii) \Rightarrow (i)$: Suppose

$$L = \sum_{i=1}^{c} x_i y_i^* z_i.$$

Then for each $n \geq 0$, $x_i y_i^* z_i$ contains at most one string of length n. Hence L contains at most c strings of length n.

$(i) \Rightarrow (ii)$: Let r be a non-trivial regular expression denoting L. The result is clearly true for $L = \emptyset$. Thus we may assume that the regular expression r does not contain $\emptyset$. We will show the following two "reduction" steps:

(a) If r contains a subexpression of the form t^*, then $L(t^*)$ can be rewritten in the form $X + Yz^*$, and

(b) If r contains a subexpression of the form $s_1^* t s_2^*$, where t contains no star, then $L(s_1^* t s_2^*)$ can be rewritten in the form Wz^*X.

The implication $(i) \Rightarrow (ii)$ will then follow.

(a) Suppose r contains a subexpression of the form t^*. Clearly if $|L(t)| \leq 1$, then t^* is already of the form xy^*z. Suppose $|L(t)| = 2$, say $L(t) = \{x, y\}$. Choose a positive integer m sufficiently large such that the linear Diophantine equation

$$a|x| + b|y| = m \tag{1}$$

has $\geq c+1$ solutions (a, b) in non-negative integers. (For example, it suffices to choose $m = c \operatorname{lcm}(|x|, |y|)$.) Then by the hypothesis, we must have

$$x^a y^b = x^{a'} y^{b'}$$

for some distinct pairs (a, b), (a', b') satisfying (1), for otherwise t and hence L would contain $\geq c+1$ strings of length n, for some n.

Without loss of generality we may assume $a \geq a'$, $b \leq b'$. Then

$$x^{a-a'} = y^{b'-b}.$$

By [21, Prop. 1.3.1] there exists a string z and integers i, j such that $x = z^i$, $y = z^j$. Hence

$$(x+y)^* = (z^i + z^j)^* = X + z^{\operatorname{lcm}(i,j)} z^*$$

for some finite set X. Thus we can replace the t^* in r by a set of the form $X + Yz^*$.

If $|L(t)| > 2$, we can repeat the argument above on pairs to obtain that *each* element of $L(t)$ is a power of some string z. Let us write

$$L(t) = \sum_{i=1}^{\infty} z^{a_i}.$$

Set $g = \gcd(a_1, a_2, \ldots)$. Then there is a finite subset of the a_i, say $b_1 \leq b_2 \leq \ldots b_k$, such that $g = \gcd_{1 \leq j \leq k} b_j$. Let

$$m = (b_1 - 1)(b_k - 1).$$

I claim that for some finite set X,

$$L(t) = X + z^m z^*.$$

For if $a_i \geq m$, we clearly have $z^{a_i} \in z^m z^*$, and there are only a finite number of distinct z^{a_i} that are not in $z^m z^*$. These we put in X. On the other hand, by a result of Brauer [2, Corollary to Thm. 1], we know that each $n \geq m$ is a nonnegative integer linear combination of the b_i. Hence every string z^j with $j \geq m$ must be in $(z^{b_1} + z^{b_2} + \cdots + z^{b_k})^*$, and hence in $L(t^*)$. This proves (a) and shows incidentally that L is of star-height 1. Hence by a theorem of Cohen [5, Lemma 3.1], we may assume that

$$L = X_1 + X_2 + \cdots + X_j,$$

where each X_i can be written in the form

$$w_1 x_1^* w_2 x_2^* \cdots w_k x_k^* w_{k+1}. \tag{2}$$

To prove (b), suppose r contains a subexpression of the form $s_1^* t s_2^*$. Then by the remark above, we may assume that $L(s_1) = \{x\}$, $L(s_2) = \{y\}$, and $L(t) = \{z\}$.

As above, choose m sufficiently large such that

$$a|x| + b|y| = m \tag{3}$$

has $\geq c+1$ solutions. Then by the hypothesis that L contains no more than c strings of length n for all n, we must have

$$x^a z y^b = x^{a'} z y^{b'}$$

for two distinct pairs (a, b), (a', b') satisfying (3). We may assume without loss of generality that $a > a'$ and $b < b'$. Then

$$x^{a-a'} z = z y^{b'-b}.$$

Using [21, Prop. 1.3.4], we see that there exist strings v, w and an integer e such that

$$x^{a-a'} = vw; \qquad y^{b'-b} = wv; \qquad z = v(wv)^e = (vw)^e v.$$

Hence

$$x^{a-a'} z y^{b'-b} = (vw)^{e+2} v.$$

Thus we see that

$$x^* z y^* = (\epsilon + x + x^2 + \cdots + x^{A-1})(vw)^*(vw)^e v(wv)^*(\epsilon + y + y^2 + \cdots + y^{B-1}),$$

where $A = a - a'$ and $B = b' - b$. Thus we have $x^* z y^* = X(vw)^* Y$ for finite sets X and Y.

To complete the proof of the lemma, we apply observation (b) repeatedly to terms of the form (2). At each stage, a term with $k > 1$ stars is reduced to a sum of terms with $k-1$ stars. The final result is a sum of terms with one star, and the result follows. ■

5 A Lemma on Linear Recurrences

Suppose we are given a sequence whose even and odd-numbered terms each satisfy a linear recurrence with integer coefficients, but not the same one, such as

$$\begin{aligned} A_{2n} &= 2A_{2n-1} + A_{2n-2} \\ A_{2n-1} &= 3A_{2n-2} + A_{2n-3} + A_{2n-4} - A_{2n-5}. \end{aligned} \tag{4}$$

Can we then conclude that the sequence A_n itself satisfies a linear recurrence with integer coefficients?

The answer is yes, as the following lemma shows:

Theorem 4 *Let k, d be positive integers, with $d \geq k$, and let $M = M_{ij}$ be a matrix of integers such that for all $n \geq 0$ we have*

$$\begin{aligned}
A_{kn} &= M_{11}A_{kn-1} + M_{12}A_{kn-2} + \cdots + M_{1d}A_{kn-d} \\
A_{kn-1} &= M_{21}A_{kn-2} + M_{22}A_{kn-3} + \cdots + M_{2d}A_{kn-d-1} \\
&\vdots \\
A_{kn-k+1} &= M_{k1}A_{kn-k-2} + M_{k2}A_{kn-k-3} + \cdots + M_{kd}A_{kn-k-d+1}.
\end{aligned}$$

Then the sequence A_n itself satisfies a linear recurrence with constant coefficients.

Proof.

Note that d is the maximum degree of the characteristic polynomials for the subsequences $A_{kn}, A_{kn+1}, \ldots, A_{kn+k-1}$.

By successively substituting the relations for $A_{kn-1}, A_{kn-2}, \cdots$ in the relation for A_{kn}, etc., we can find another matrix $P = P_{ij}$ such that for all n sufficiently large, we have

$$\begin{aligned}
A_{kn} &= P_{11}A_{kn-k} + P_{12}A_{kn-k-1} + \cdots + P_{1d}A_{kn-k-d+1} \\
A_{kn-1} &= P_{21}A_{kn-k} + P_{22}A_{kn-k-1} + \cdots + P_{2d}A_{kn-k-d+1} \\
&\vdots \\
A_{kn-d+1} &= P_{d1}A_{kn-k} + P_{d2}A_{kn-k-1} + \cdots + P_{dd}A_{kn-k-d+1}.
\end{aligned}$$

Note that P is a square matrix. Let $f(X)$ be the characteristic polynomial of P. Then each of the sequences $A_{kn}, A_{kn-1}, \ldots, A_{kn-k+1}$ satisfies the same linear recurrence, namely, the one whose characteristic polynomial is $f(X)$ (at least for n sufficiently large). Thus for n large enough, A_n satisfies the linear recurrence whose characteristic polynomial is $f(X^k)$. ∎

Example.

For example, for the recurrence specified by (4), we get

$$\begin{bmatrix} A_{2n} \\ A_{2n-1} \\ A_{2n-2} \\ A_{2n-3} \end{bmatrix} \begin{bmatrix} 7 & 2 & 2 & -2 \\ 3 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} A_{2n-2} \\ A_{2n-3} \\ A_{2n-4} \\ A_{2n-5} \end{bmatrix}.$$

The characteristic polynomial for the matrix is $X^4 - 8X^3 - X$, so A_n satisfies the recurrence $A_n = 8A_{n-2} + A_{n-6}$.

6 Proof of the Main Result

In this section, we prove the result mentioned in the introduction. The details of the proof are a little messy, so it may be helpful to first give the proof in the case of ordinary base-3 representation. In this case, the set of representations $R(u)$ is $\epsilon + (1+2)(0+1+2)^*$. It is easy to see that $B(R(u))$, the set of lexicographically greatest representations, is 2^*. Since

$$\mathrm{val}(1\overbrace{00\cdots0}^{k}) = 1 + \mathrm{val}(\overbrace{22\cdots2}^{k}),$$

it follows that

$$u_k = 1 + 2u_{k-1} + 2u_{k-2} + \cdots + 2u_0. \tag{5}$$

Similarly, we also have

$$u_{k+1} = 1 + 2u_k + 2u_{k-1} + \cdots + 2u_0. \tag{6}$$

Subtracting (5) from (6), we see

$$u_{k+1} - u_k = 2u_k,$$

and hence $u_{k+1} = 3u_k$.

We now state and prove the main result of the paper:

Theorem 5 *Let $u_0, u_1, \ldots$ be a strictly increasing sequence of non-negative integers such that*

(a) $\gcd(u_0, u_1, \ldots) = g$, and a canonical representation is provided for every sufficiently large multiple of g;

(b) for every sufficiently large n, there exists a representation in $R(u)$ of length n; and

(c) the numeration system based on u is order-preserving.

If $R(u)$ is regular, then the sequence $u = (u_n)_{n \geq 0}$ satisfies a linear recurrence with integral constant coefficients.

Before we begin the proof, let us explain the role of the technical hypotheses (a)-(b). For (a), if $\gcd(u_0, u_1, \ldots) = g$, then every sufficiently large multiple of g has some representation as a non-negative integer linear combination of the u_i. We wish to avoid the case where "most" representable integers simply do not have a canonical representation.

Hypothesis (b) is needed to exclude cases similar to the following: suppose our numeration system is

$$1, u_0, 10, u_1, 100, u_2, \ldots, \tag{7}$$

where u is any sequence that does not satisfy a linear recurrence, and $10^i < u_i \leq 10^{i+1}$. If we choose as our numeration system ordinary base-10 representation, and simply never use the u_i in any representation, we get a numeration system that is order-preserving, and a set of representations which is regular. However, the sequence (7) clearly does not satisfy a linear recurrence.

Note that hypothesis (b) is satisfied by both the greedy representation and the lexicographically greatest representation, for then the representation for u_n is $1\,0^n$. If $u_0 = 1$, then all three hypotheses are satisfied for the greedy algorithm.

Now let us begin the proof.

Proof (Sketch).

Let $g = \gcd(u_0, u_1, \ldots)$. Then there exists an integer C such that all $n \geq C$ are representable.

If $L = R(u)$ is regular, then by Lemma 1, the set $B(L)$ of lexicographically greatest strings of every length in L is also regular. By hypothesis (b) of the theorem, there exists C' such that $B(L)$ contains at most one string of length j for each $j \geq C'$. Thus Lemma 3 applies and so

$$B(L) = \sum_{i=1}^{k} x_i y_i^* z_i. \tag{8}$$

Similarly, by Corollary 2, the set $S(L)$ can also be written in the form

$$\sum_{i=1}^{k'} x'_i y'^*_i z'_i.$$

For simplicity in this proof sketch, we assume that $S(L) = 10^*$, although this assumption can easily be removed.

For a string $w = w_1 w_2 \cdots w_k$, define

$$(w, u) = \sum_{i=1}^{k} w_i u_{k-i}.$$

The main idea of the proof is as follows: let $w \in B(L)$ be sufficiently long such that $\text{val}(w) > C$. Let $v \in S(L)$ be such that $|v| = |w| + 1$. Then since the numeration system is order-preserving, we must have $\text{val}(v) = \text{val}(w) + g$. Hence

$$u_{|w|} = (w, u) + g. \tag{9}$$

Now let

$$g' = \text{lcm}_{\substack{1 \leq j \leq k \\ |y_j| \neq 0}} |y_j|.$$

By replacing y_i with $y_i^{g/|y_i|}$, adding extra terms to x_i and z_i, and renaming, we can rewrite (8) such that $|y_i| = g'$ for all i for which $|y_i| \neq 0$.

From (9) we have

$$u_{j|y|+|xz|} = (xy^j z, u) + g.$$

We also have

$$u_{(j+1)|y|+|xz|} = (xy^{j+1} z, u) + g.$$

Subtracting, we see

$$u_{(j+1)|y|+|xz|} - u_{j|y|+|xz|} = (xy^{j+1} z, u) - (xy^j z, u)$$

$$= (xy0^{j|y|+|z|}, u) - (x0^{j|y|+|z|}, u),$$

and the last expression on the right is a sum of terms of u such that the smallest non-zero index is $u_{j|y|+|z|}$. Hence for i sufficiently large, u_{gj+i} can be expressed as a linear combination of the g previous terms, and the particular linear combination depends only on the value of $i \pmod g$.

By Lemma 4, we can write u itself as a linear recurrence. This completes the proof. ■

7 Linear Recurrences and Non-Regular Sets

After seeing the main theorem, one immediately wonders if the converse is true. It is not, as the following theorem shows:

Theorem 6 *Suppose $u_j = (j+1)^2$ for $j \geq 0$. Then the set $G(u)$ of greedy representations is not a regular set.*

Proof.

Let $G(u)$ be the set of greedy representations, and assume it is regular. Then $G(u) \cap 10^*10^*$ would also be a regular set. However, it is easy to see that

$$\begin{aligned} G(u) \cap 10^*10^* &= \{10^a10^b \ : \ u_{b+a+2} > u_{b+a+1} + u_b\} \\ &= \{10^a10^b \ : \ b^2 < 2a+4\}, \end{aligned}$$

and this set is evidently not regular. ■

One would like a simple characterization of those sequence u for which $G(u)$ is regular. The next example shows that such a characterization based on the characteristic polynomial of the recurrence alone will not suffice.

Let $f_j = 2^j + 1$ for $j \geq 0$. In this numeration system, where the digits are bounded by 2, every integer except 1 has some representation. For our canonical representation, choose the lexicographically greatest representation. (Note that this system is *not* order-preserving.) Then

$$\begin{aligned} R(f) \cap 1^*0^* &= \{1^a0^b \ : \ f_{b+a} > f_{b+a-1} + f_{b+a-2} + \cdots + f_b\} \\ &= \{1^a0^b \ : \ a < 2^b + 1\}, \end{aligned}$$

and this set is clearly not regular. Hence $R(f)$ is not regular. However, f_j satisfies the same linear recurrence as the sequence $u_j = 2^j - 1$, discussed previously in Section 1, for which $G(u)$ is regular.

It is an open problem to give a sufficient condition for the regularity of $R(u)$.

8 Acknowledgments

An earlier version of Lemma 3 dealt with the case $c = 1$. I am grateful to E. Bach for having suggested the stronger formulation given in this paper. I would also like to thank A. Lubiw for suggesting an appropriate way to weaken the hypotheses in Theorem 5.

References

[1] J. Berstel. Fibonacci words – a survey. In G. Rozenberg and A. Salomaa, editors, *The Book of L*, pages 13–27. Springer-Verlag, 1986.

[2] A. Brauer. On a problem of Frobenius. *Amer. J. Math.*, 64:299–312, 1942.

[3] H. Cameron and D. Wood. Pm numbers, ambiguity, and multiway trees. Manuscript, 1991.

[4] L. Carlitz. Fibonacci representations. *Fibonacci Quart.*, 6(4):193–220, 1968.

[5] R. S. Cohen. Star height of certain families of regular events. *J. Comput. System Sci.*, 4:281–297, 1970.

[6] K. Culik II and A. Salomaa. Ambiguity and decision problems concerning number systems. *Inform. and Comput.*, 56:139–153, 1983.

[7] A. S. Fraenkel. Systems of numeration. *Amer. Math. Monthly*, 92:105–114, 1985.

[8] A. S. Fraenkel. The use and usefulness of numeration systems. *Inform. and Comput.*, 81:46–61, 1989.

[9] C. Frougny. Fibonacci numeration systems and rational functions. In *Math. Found. Comp. Sci.*, volume 233 of *Lecture Notes in Computer Science*, pages 350–359. Springer-Verlag, 1986.

[10] C. Frougny. Linear numeration systems of order two. *Inform. and Comput.*, 77:233–259, 1988.

[11] C. Frougny. Linear numeration systems, θ–developments and finite automata. In B. Monien and R. Cori, editors, *STACS 89*, volume 349 of *Lecture Notes in Computer Science*, pages 144–155. Springer-Verlag, 1989.

[12] C. Frougny. Systèmes de numération lineaires et automates finis (Thèse d'État). Technical Report 89-69, Laboratoire Informatique Théorique et Programmation, Université P. et M. Curie, Université Paris VII, September 1989.

[13] C. Frougny. Fibonacci representations and finite automata. *IEEE Trans. Inform. Theory*, 37:393–399, 1991.

[14] C. Frougny. Representations of numbers and finite automata. *Math. Systems Theory*, 1991. To appear.

[15] J. Honkala. Unique representation in number systems and L codes. *Disc. Appl. Math.*, 4:229–232, 1982.

[16] J. Honkala. Bases and ambiguity in number systems. *Theoret. Comput. Sci.*, 31:61–71, 1984.

[17] J. Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Informatique Théorique*, 20:395–403, 1986.

[18] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, 1979.

[19] W. H. Kautz. Fibonacci codes for synchronization control. *IEEE Trans. Inform. Theory*, IT-11:284–292, 1965.

[20] D. E. Knuth. *Fundamental Algorithms*, volume I of *The Art of Computer Programming.* Addison-Wesley, 1973.

[21] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and Its Applications.* Addison-Wesley, 1983.

[22] A. de Luca and A. Restivo. Representations of integers and language theory. In *Math. Found. Comp. Sci.*, volume 176 of *Lecture Notes in Computer Science*, pages 407–415. Springer-Verlag, 1984.

[23] H. A. Maurer, A. Salomaa, and D. Wood. L codes and number systems. *Theoret. Comput. Sci.*, 22:331–346, 1983.

[24] J. Shallit. A generalization of automatic sequences. *Theoret. Comput. Sci.*, 61:1–16, 1988.

[25] E. Zeckendorf. Répresentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas. *Bull. Soc. Royale des Sciences de Liège*, 3-4:179–182, 1972.