

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

December 10, 1991

Abstract

A numeration system based on a strictly increasing sequence of positive integers $u_0 = 1, u_1, u_2, \dots$ 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 $a_i a_{i-1} \dots a_1 a_0$ is a *representation* for n . If $\gcd(u_0, u_1, \dots) = g$, 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 xy^*z .

1 Introduction

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

*Supported in part by a grant from NSERC .

not every integer have a representation, in which case r is a partial function, rather than a function.

Common examples of numeration systems include radix- k representation (for k an integer ≥ 2), Fibonacci representation, factorial representation, etc. See [11, 12] for surveys on numeration systems.

In this paper we consider only *linear* numeration systems constructed as follows: given a strictly increasing sequence of positive integers u_0, u_1, u_2, \dots , we 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 *u-representable* (or *representable* if the sequence u is clear from the context), and one representation of n is the string $a_i a_{i-1} \cdots a_1 a_0$. We define the mapping val from strings back to integers as follows:

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

We say such a representation is *normal* if $a_i \neq 0$. In this paper, we consider only normal representations.

Note that if

$$\text{gcd}(u_0, u_1, \dots) = g,$$

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, [11]. In this paper we only consider numeration systems whose digits are bounded by some fixed constant. A necessary condition to ensure bounded digits is that the ratio u_j/u_{j-1} is bounded by a constant.

Since, in general, there may be many normal representations with bounded digits for a number n , we must identify *one* of these representations in order to specify the map r . This can be done in a variety of ways; for example, one could choose $r(n)$ to be the *lexicographically greatest* representation for n , among all nonnegative linear combinations of terms of the sequence u . The lexicographic order we use is defined as follows: 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 , $1 \leq k \leq i$, such that $x_1 = y_1, x_2 = y_2, \dots, x_{k-1} = y_{k-1}$, but $x_k > y_k$.¹

Another possible choice for $r(n)$ is the *greedy representation* $\text{gr}(n)$. We define $\text{gr}(n)$ for a positive integer n 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

¹Note that this ordering, called “radix order” by Berstel [1, p. 76] is a well-order (i.e. each nonempty set has a least element), and it differs from the lexicographic order defined by other authors [26, p. 64]; [1]. It is, however, the natural order for our problem.

to be ϵ , the empty string. The set of greedy representations for all greedily representable integers is written $G(u)$.

It is easy to see that 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 $(u_0, u_1, u_2, \dots) = (2, 3, 5, \dots)$. Note that if $u_0 = 1$, then the greedy representation is in fact the lexicographically greatest representation.

A desirable property of any numeration system is that the mapping r that sends an integer n to its representation be *order-preserving*. More precisely, we require that for integers m, n in the domain of r , we have $m > n$ iff $r(m) > r(n)$. It is easy to see that the greedy representation is order-preserving.

Given a numeration system r , based on the sequence u , we define $R(r, u) = R(u)$ to be the set of its representations for all non-negative integers. More formally,

$$R(u) = \sum_{\substack{n \geq 0 \\ n \in \text{Dom } r}} r(n). \quad (2)$$

Example 1.

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

$$\epsilon + (1 + 2 + \dots + k - 1)(0 + 1 + \dots + k - 1)^*,$$

which is nothing more than the ordinary radix- k representation.

If we let $u_j = j + 1$, then the set of greedy representations is simply $\epsilon + 10^*$. This numeration system is essentially “unary” notation.

Example 2.

Let $u_j = F_{j+2}$, where F_j is the j th 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 [25, Ex. 1.2.8.34], [31], [5], [24], and [2].)

Example 3.

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

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

See the recent paper of Cameron and Wood [4].

Suppose we are given an increasing sequence of positive integers $(u_j)_{j \geq 0}$, and suppose there is a constant C such that every non-negative integer n has a representation $n = \sum_{0 \leq j \leq i} a_j u_j$ with $0 \leq a_j \leq C$. Let $r(n)$ be defined as the *lexicographically least* representation among all such representations for n .

It is not difficult to show that this r is order-preserving. An example is as follows:

Example 4.

Let $u_j = 2^j$, and consider expressing integers n as $n = \sum_{0 \leq j \leq i} a_j u_j$ with $0 \leq a_j \leq 2$. Then Reznick has shown [29] that the number of such representations is $s(n+1)$, where $s(n)$ is the Stern function, defined by $s(0) = 0$; $s(1) = 1$; $s(2n) = s(n)$; and $s(2n+1) = s(n) + s(n+1)$ for $n \geq 1$.

Now define $r(n)$ to be the lexicographically least representation of this form for n ; for example, $r(13) = 221$. Then it is easy to see that $r(n)$ is a string of 1's and 2's; in fact, $R(u)$, the set of all such representations, is $(1+2)^*$. This corresponds to the well-known observation that every non-negative integer can be written uniquely in base 2, using only the digits 1 and 2. Note that this numeration system is *not* obtained via the greedy algorithm.

The main result of this paper is Theorem 6, which proves the following: suppose the numeration system r is such that the set $R(r, u)$ is regular. Then, subject to some technical conditions on r , the sequence u must satisfy a linear recurrence with integer 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 [13, 14, 15, 16, 17, 18, 19].

In [30], I proved that the set of greedy representations is regular for the numeration systems with bounded digits considered by Fraenkel [11].

We note several other papers that have examined the relationship between ways of representing numbers and regular sets. See [20, 27, 7, 8, 21, 22]. However, these papers have adopted a very different point of view.

2 More Notation

Throughout this paper, Σ is a finite alphabet, and q, s, t denote regular expressions. The letters v, w, x, y, z denote strings. The lower-case letters $a, b, c, d, e, g, i, j, k, m, n$ 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 letters L and Z denote languages 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 $\Sigma = \{0, 1, 2, \dots, k-1\}$. 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 , $1 \leq i \leq n$, such that $x_1 = y_1, x_2 = y_2, \dots, x_{i-1} = y_{i-1}$, but $x_i > y_i$.² Then we informally define $B(L)$ as follows: it is the union, over all $n \geq 0$, of the lexicographically largest string of length n in L .³ More formally, define

$$B(L) = \bigcup_{n \geq 0} \{x \in L \cap \Sigma^n : \forall (y \in L \cap \Sigma^n) \ x \geq y\}. \quad (3)$$

For example, $B(1(0+01)^*) = (10)^*(\epsilon+1)$.

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. From this the result follows: $B(L) \subseteq L$, and so

$$(\Sigma^* - (L - B(L))) \cap L = B(L).$$

But regular sets are closed under complement and intersection, and so $B(L)$ is regular.

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a deterministic finite automaton accepting L . (See [23] for the basic notions about automata and the notational conventions that 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 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 $w \in L - B(L)$. ■

²See footnote 1.

³Of course, it is possible that for some n , there is *no* string of length n in L . These n correspond to no string in $B(L)$. See Eq. (3).

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

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

of lexicographically smallest strings of every length in L .

It is perhaps worthwhile to note that we can use the technique of Lemma 1 to answer a question of D. Klarner (personal communication). He asked, suppose f is a *recognizable relation*. (For this concept, see Eilenberg [9] or Berstel [1].) Then is it true that the set of lexicographically smallest strings, one chosen from each nonempty equivalence class of f , must be regular? Using the technique above, it is easy to see that the answer is yes. For the case of f being a rational relation, I believe the question is still open.

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 q be a non-trivial regular expression denoting L . (When it is necessary to make the distinction, we will use the notation $L(q)$ to represent the language denoted by the regular expression q .)

The result is clearly true for $L = \emptyset$. Thus we may assume that the regular expression q does not contain \emptyset . We will show the following two “reduction” steps (recalling that the letters X, Y, W denote finite sets):

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

(b) If q 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 $X + Yz^*$. Suppose $|L(t)| = 2$, say $L(t) = \{x, y\}$. Choose a positive integer m sufficiently large such that the linear Diophantine equation

$$a|x| + b|y| = m \tag{4}$$

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 (4), 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 [26, Prop. 1.3.1] there exists a string z and integers i, j such that $x = z^i, y = z^j$. Hence

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

for some finite set X . Thus we can replace the t^* in q 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 . Thus, there is some (finite or infinite) set S such that

$$L(t) = \sum_{i \in S} z^{a_i}.$$

Set $g = \operatorname{gcd}(a_1, a_2, \dots)$. Then there is a finite subset of the a_i , say $b_1 \leq b_2 \leq \dots \leq b_k$, such that $g = \operatorname{gcd}_{1 \leq j \leq k} b_j$. Choose m sufficiently large such that eg is a non-negative integer linear combination of the b_i for all $e \geq m$; by a theorem of Brauer [3, Corollary to Thm. 1], we may in fact choose

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

Let

$$X = \{z^d \in L(t^*) : d < gm\};$$

note that X is a finite set. I claim that

$$L(t^*) = X + z^{gm}(z^g)^*.$$

For let $z^d \in L(t^*)$; if $d < gm$, then $z^d \in X$, by construction. Otherwise $d \geq gm$; clearly d must be a multiple of g and hence $z^d \in z^{gm}(z^g)^*$.

It remains to see that $z^{gm}(z^g)^* \subseteq L(t^*)$. We know that if $e \geq m$, then eg is a nonnegative integer linear combination of the b_i . Hence every string z^{eg} with $e \geq m$ must be in $(z^{b_1} + z^{b_2} + \dots + z^{b_k})^*$, and hence in $L(t^*)$. This completes the proof of (a) and shows incidentally that L is of star-height 1. Hence by a theorem of Cohen [6, Lemma 3.1], we may assume that

$$L = Z_1 + Z_2 + \dots + Z_j,$$

where each set Z_i can be written in the form

$$w_1 x_1^* w_2 x_2^* \dots w_k x_k^* w_{k+1}. \quad (5)$$

To prove (b), suppose q contains a subexpression of the form $s_1^* t s_2^*$. Then by Eq. (5) we may assume without loss of generality that $L(s_1) = \{x\}$, $L(s_2) = \{y\}$, and $L(t) = \{z\}$.

As above, choose m sufficiently large such that

$$a|x| + b|y| = m \quad (6)$$

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 (6). We may assume without loss of generality that $a > a'$ and $b < b'$. Then

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

Using [26, Prop. 1.3.4], we see that there exist strings v, w and an integer e such that

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

Hence

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

Thus we see that

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

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 (5). 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. ■

After the author proved Lemma 3, other proofs were shown to him by S. Yu and F. Fich.

5 Two Lemmas on Linear Recurrences

In this section we provide two useful lemmas about 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

$$\begin{aligned} A_{2n} &= 2A_{2n-1} + A_{2n-2} \\ A_{2n-1} &= 3A_{2n-2} + A_{2n-3} + A_{2n-4} - A_{2n-5}. \end{aligned} \tag{7}$$

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:

Lemma 4 *Let k, d be positive integers, with $d \geq k$, and let $M = [M_{ij}]$ be a $k \times d$ matrix of integers such that for all n sufficiently large we have*

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

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 $A_{kn}, A_{kn+1}, \dots, A_{kn+k-1}$.

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

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

Let $f(X)$ be the characteristic polynomial of P . Then each of the sequences

$$A_{kn}, A_{kn-1}, \dots, A_{kn-k+1}$$

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 the recurrence specified by (7), 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 $A_n = 8A_{n-2} + A_{n-6}$.

Our second lemma concerns integer linear recurrences with rational coefficients.

Lemma 5 *Suppose the infinite integer sequence u_0, u_1, u_2, \dots satisfies a linear recurrence relation*

$$u_n = a_1 u_{n-1} + a_2 u_{n-2} + \dots + a_d u_{n-d},$$

for all $n \geq 0$, where each a_i is a rational number. Then u_n satisfies a (possibly different) linear recurrence relation where all the a_i are actually integers.

Proof.

This follows immediately from the following lemma of Fatou [10] (also see [28, Part VIII, Problem 156]): if $v(z)$ is a rational function whose Taylor series has rational integer coefficients, then $v(z)$ can be written in the form $f(z)/g(z)$, where f and g are polynomials with integer coefficients and $g(0) = 1$. ■

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

$$\text{val}(1 \overbrace{00 \dots 0}^k) = 1 + \text{val}(\overbrace{22 \dots 2}^k),$$

it follows that

$$u_k = 1 + 2u_{k-1} + 2u_{k-2} + \dots + 2u_0. \tag{8}$$

Similarly, we also have

$$u_{k+1} = 1 + 2u_k + 2u_{k-1} + \dots + 2u_0. \tag{9}$$

Subtracting (8) from (9), 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 6 *Let u_0, u_1, \dots be a strictly increasing sequence of non-negative integers, and let r be a linear numeration system based on u . Let $R(u)$ be as in Eq. (2). Suppose that*

- (a) $\gcd(u_0, u_1, \dots) = g$, and $r(kg)$ is defined for all sufficiently large integers k ;
- (b) for all sufficiently large n , there exists a representation in $R(u)$ of length n ;

and

- (c) r is order-preserving.

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

Before we begin the proof, let us explain the role of the technical hypotheses (a)–(b). For (a), if $\gcd(u_0, u_1, \dots) = g$, 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 are not in the domain of the partial function r .

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

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

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 (10) clearly does *not* satisfy a linear recurrence.

Note that hypothesis (b) is satisfied by both the greedy representation and the lexicographically greatest representation, for in these cases we have the representation of u_n is 10^n . In fact, if $u_0 = 1$, then *all three* hypotheses are satisfied when r is the greedy representation.

Now let us begin the proof of Theorem 6.

Proof.

Let $g = \gcd(u_0, u_1, \dots)$. Then by hypothesis (a), there exists an integer C such that $r(n)$ is defined for all $n \geq C$ with $g \mid n$.

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 $B(L)$ exactly one string of length j for each $j \geq C'$. Thus Lemma 3 applies and so

$$B(L) = \sum_{i=1}^k x_i y_i^* z_i. \tag{11}$$

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

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

To avoid unnecessary complication, we first show how the proof goes in the case $S(L) = 10^*$. At the end of the proof we sketch what needs to be done when $S(L) \neq 10^*$. (Note that in fact $S(L) = 10^*$ for the most common case, when r is the greedy representation.)

The main idea of the proof is as follows: let $w \in B(L)$ be sufficiently long such that $\text{val}(w) > \max(C, C')$. Let $v \in S(L)$ be such that $|v| = |w| + 1$. Then by hypothesis (c) (i.e., r is order-preserving), we must have $\text{val}(v) = \text{val}(w) + g$. Since $S(L) = 10^*$, we have

$$u_{|w|} = \text{val}(w) + g. \quad (12)$$

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 (11) such that $|y_i| = g'$ for all i for which $|y_i| \neq 0$.

Consider a particular term of the sum in (11), say xy^*z . From (12) we have

$$u_{j|y|+|xz|} = \text{val}(xy^jz) + g.$$

We also have

$$u_{(j+1)|y|+|xz|} = \text{val}(xy^{j+1}z) + g.$$

Subtracting, we see

$$\begin{aligned} u_{(j+1)|y|+|xz|} - u_{j|y|+|xz|} &= \text{val}(xy^{j+1}z) - \text{val}(xy^jz) \\ &= \text{val}(xy0^{j|y|+|z|}) - \text{val}(x0^{j|y|+|z|}), \end{aligned}$$

and the last expression on the right is a linear combination of terms of u such that the smallest coefficient which is possibly non-zero corresponds to $u_{j|y|+|z|}$. Hence for j sufficiently large, the subsequence $u_{(j+1)g'+i}$ can be expressed as a non-trivial integer linear combination of the $g' + |x|$ previous terms of u , 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 in the case where $r(n) = 10^n$.

It remains to discuss what needs to be done to handle the more general case, when $S(L)$ is not necessarily equal to 10^* . The same techniques apply, except that since the leading coefficient of the recurrence we obtain is no longer necessarily 1, we see that u is a linear recurrence with *rational* coefficients. Then by Lemma 5, we see that the linear recurrence actually has *integer* coefficients. 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 7 *Suppose $u_j = (j + 1)^2$ for $j \geq 0$. Then the set $G(u)$ of greedy representations is not a regular set.*

Note that u_j , of course, satisfies the linear recurrence $u_{j+3} = 3u_{j+2} - 3u_{j+1} + u_j$ for all $j \geq 0$.

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. Let $r(n)$ denote the lexicographically greatest representation for n . (Note that $4 = \text{val}(2) > \text{val}(10) = 3$; hence r 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 $u_j = 2^{j+1} - 1$ discussed previously in Section 1 (namely $f_{j+2} = 3f_{j+1} - 2f_j$ for $j \geq 0$), 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 general discussion and for suggesting an appropriate way to weaken the hypotheses in Theorem 6. C. Stewart pointed out the lemma of Fatou.

References

- [1] J. Berstel. *Transductions and Context-Free Languages*. Teubner, Stuttgart, 1979.
- [2] J. Berstel. Fibonacci words – a survey. In G. Rozenberg and A. Salomaa, editors, *The Book of L*, pages 13–27. Springer-Verlag, 1986.
- [3] A. Brauer. On a problem of Frobenius. *Amer. J. Math.* **64** (1942), 299–312.
- [4] H. Cameron and D. Wood. Pm numbers, ambiguity, and regularity. Technical Report CS-91-47, University of Waterloo, Department of Computer Science, September 1991.
- [5] L. Carlitz. Fibonacci representations. *Fibonacci Quart.* **6**(4) (1968), 193–220.
- [6] R. S. Cohen. Star height of certain families of regular events. *J. Comput. System Sci.* **4** (1970), 281–297.
- [7] K. Culik II and A. Salomaa. Ambiguity and decision problems concerning number systems. *Inform. and Comput.* **56** (1983), 139–153.
- [8] A. de Luca and A. Restivo. Representations of integers and language theory. In *Math. Found. Comp. Sci.*, Vol. 176 of *Lecture Notes in Computer Science*, pages 407–415. Springer-Verlag, 1984.
- [9] S. Eilenberg. *Automata, Languages, and Machines (Vol. A)*. Academic Press, New York, 1974.
- [10] P. Fatou. Séries trigonométriques et séries de Taylor. *Acta Math.* **30** (1906), 335–400.
- [11] A. S. Fraenkel. Systems of numeration. *Amer. Math. Monthly* **92** (1985), 105–114.
- [12] A. S. Fraenkel. The use and usefulness of numeration systems. *Inform. and Comput.* **81** (1989), 46–61.
- [13] C. Frougny. Fibonacci numeration systems and rational functions. In *Math. Found. Comp. Sci.*, Vol. 233 of *Lecture Notes in Computer Science*, pages 350–359. Springer-Verlag, 1986.
- [14] C. Frougny. Linear numeration systems of order two. *Inform. and Comput.* **77** (1988), 233–259.
- [15] C. Frougny. Linear numeration systems, θ -developments and finite automata. In B. Monien and R. Cori, editors, *STACS 89*, Vol. 349 of *Lecture Notes in Computer Science*, pages 144–155. Springer-Verlag, 1989.
- [16] 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.
- [17] C. Frougny. Representations of numbers and finite automata. *Math. Systems Theory* (1991), to appear.

- [18] C. Frougny. Fibonacci representations and finite automata. *IEEE Trans. Inform. Theory* **37** (1991), 393–399.
- [19] C. Frougny. Representation of numbers in non-classical numeration systems. In P. Kornerup and D. W. Matula, editors, *Proc. 10th IEEE Symp. Computer Arithmetic*, pages 17–21. IEEE Computer Society Press, 1991.
- [20] J. Honkala. Unique representation in number systems and L codes. *Disc. Appl. Math.* **4** (1982), 229–232.
- [21] J. Honkala. Bases and ambiguity in number systems. *Theoret. Comput. Sci.* **31** (1984), 61–71.
- [22] J. Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Informatique Théorique* **20** (1986), 395–403.
- [23] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.
- [24] W. H. Kautz. Fibonacci codes for synchronization control. *IEEE Trans. Inform. Theory* **IT-11** (1965), 284–292.
- [25] D. E. Knuth. *Fundamental Algorithms*, Vol. I of *The Art of Computer Programming*. Addison-Wesley, 1973.
- [26] M. Lothaire. *Combinatorics on Words*, Vol. 17 of *Encyclopedia of Mathematics and Its Applications*. Addison-Wesley, 1983.
- [27] H. A. Maurer, A. Salomaa, and D. Wood. L codes and number systems. *Theoret. Comput. Sci.* **22** (1983), 331–346.
- [28] G. Pólya and G. Szegő. *Problems and Theorems in Analysis*, Vol. II. Springer-Verlag, Berlin, 1976.
- [29] B. Reznick. Some binary partition functions. In B. C. Berndt, H. G. Diamond, H. Halberstam, and A. Hildebrand, editors, *Analytic Number Theory (Proceedings of a Conference in Honor of Paul T. Bateman)*, pages 451–477. Birkhäuser, Boston, 1990.
- [30] J. Shallit. A generalization of automatic sequences. *Theoret. Comput. Sci.* **61** (1988), 1–16.
- [31] E. Zeckendorf. Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas. *Bull. Soc. Royale des Sciences de Liège* **3-4** (1972), 179–182.