Theory Comput. Systems **31**, 111–133 (1998)

Theory of Computing Systems

Greedy Numeration Systems and Regularity*

M. Hollander

2504 East Roy Street,
Seattle, WA 98112, USA
mijklh@accessone.com

Abstract. We find necessary and sufficient conditions for the regularity of greedy representations of the natural numbers. The proof relies on the periodicity properties of digit expansions with nonintegral bases.

1. Introduction and Definitions

1.1. *Numeration Systems*

Let $\mathcal{A} \subset \mathbf{N}$. Recall that $\mathcal{A}^*$ denotes the set of finite concatenations ("words") of members of $\mathcal{A}$. A *numeration system* is a map $G\colon \mathbf{N} \to \mathcal{A}^*$ that assigns a string $G(n)$, called the *representation* of n, to each nonnegative integer n.

Given a strictly increasing sequence of integers $A = \{A_1 = 1, A_2, A_3, \ldots\}$, the *greedy algorithm* can be used to construct a numeration system as follows: Let n be a nonnegative integer. If $n = 0$, let $G(n)$ be the empty word. Otherwise, let i be the largest index such that $n \geq A_i$. If $i = 1$, let $G(n) = n$. For $n \geq A_2$, let $d = \lfloor n/A_i \rfloor$ and write $G(n) = dG(n - dA_i)$. This numeration system is called the *greedy representation* based on A, and A is referred to as the *basis* of the numeration system. If conversely we define the *evaluation map* $E\colon \mathcal{A}^* \to \mathbf{Z}^+$ by

$$E(d_1 d_2 \cdots d_n) = \sum_{i=1}^{n} d_i A_{n-i+1}$$

we then have that $E(G(n)) = n$ for all integers n.

The set $L = G(\mathbf{Z}^+)$ of representations of all nonnegative integers is called the *language* of the numeration system. We place the following lexicographic ordering $\leq_{lex}$

* The work toward this paper was done while the author was a Ph.D. student at the University of Washington, and a part of the work appears in his dissertation [10].

on L: For $x, y \in L$ we take $x \leq_{lex} y$ if $|x| < |y|$ or if $|x| = |y|$ and $x_i < y_i$, where i is the lowest index where x and y differ. Under this ordering, the map G is a strictly order-preserving map from $\mathbf{N}$ to L.

A numeration system is called *regular* if the set of all representations L is regular; i.e., if it can be given as the output of a finite automaton (see Section 2 for details). In this paper we consider the following question: For which bases A are the associated numeration systems regular?

Shallit [16] has shown the following result:

Theorem 1.1 [16]. *If a greedy representation is regular, the basis must satisfy a linear recurrence relation*; *i.e., there exists a positive integer s and integers $u_1, \ldots, u_s$, such that $A_m = u_1 A_{m-1} + \cdots + u_s A_{m-s}$ for all $m > s$.*

1.2. *The Dominant Root Condition and the Beta Expansion*

If the basis A satisfies a linear recurrence, there is often a positive real β which dominates the growth of A (see Section 3 for details). We make the following definition:

Definition 1.1. Let A be a linearly recurrent basis. If $\lim_{m\to\infty}(A_m/A_{m-1}) = \beta$ for real $\beta > 1$, then A is said to satisfy the *dominant root condition* and β is called the dominant root of the recurrence.

In the case where a dominant root β exists, there is a strong relation between the maximum lexicographic sequence of length m in L and the expansion of 1 in base β. In this section we review expansions in real bases.

Assume β is a real number greater than 1. Let $x \in [0, 1]$. We construct the expansion of x base β, which we denote as $d(x, \beta)$, as a string of digits over the alphabet $\mathcal{A} = \{0, 1, \ldots, \lfloor\beta\rfloor\}$. The digits x_i, $i \geq 1$, of $d(x, \beta)$ are computed using the *greedy algorithm*. Begin by setting $x_1 = \lfloor x\beta \rfloor$. If we inductively assume, for any integer $i > 1$, that $x_1, \ldots, x_{i-1}$ have been assigned, we set

$$d = (x - x_1\beta^{-1} - \cdots - x_{i-1}\beta^{-i+1})\beta^i$$

and then let $x_i = \lfloor d \rfloor$. We note that if β is an integer, then $d(1, \beta)$ is just $.\beta$.

We call $d(x, \beta)$ *finite* if there exists l such that $x_m = 0$ for $m > l$. The minimal such l is called the *length* of the expansion. We call $d(x, \beta)$ *eventually periodic* if there exist two integers l and p such that, for all $k > l$, $x_{k+p} = x_k$. The minimal such l and p are called the *lead* and the *period* of the expansion.

In the case that $d(1, \beta)$ is finite or eventually periodic, there is a polynomial which naturally arises from the pattern of the digits. If $d(1, \beta)$ is eventually periodic and has lead l and period p, then

$$1 = \frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \cdots + \frac{d_l}{\beta^l} + \left(\frac{d_{l+1}}{\beta^{l+1}} + \cdots + \frac{d_{l+p}}{\beta^{l+p}}\right)\left(1 + \frac{1}{\beta^p} + \frac{1}{\beta^{2p}} + \cdots\right).$$

Therefore β satisfies the polynomial

$$b(x) = x^{l+p} - \sum_{i=1}^{l+p} d_i x^{l+p-i} - x^l + \sum_{i=1}^{l} d_i x^{l-i}.$$

Similarly, if $d(1,\beta)$ is finite, we let $d(1,\beta) = .x_1 \cdots x_l$ and write

$$b(x) = x^l - \sum_{i=1}^{l} d_i x^{l-i}.$$

The polynomial $b(x)$ has β as its maximum modulus root. We also note that it is dependent on the choice of l and p. We refer to any such polynomial as an *extended beta polynomial* for β. If we choose l and p to be minimal, the polynomial is referred to as the *canonical beta polynomial* for β.

Recall that an algebraic integer greater than 1 is called a *Pisot number* if all of its conjugates have modulus less than 1 and a *Salem number* if all of its conjugates have modulus less than or equal to 1 and at least one conjugate is of modulus 1. It is proved in [2] and [15] that if β is a Pisot number, then $d(x,\beta)$ is finite or eventually periodic for all $x \in \mathbf{Q}(\beta) \cap [0,1]$. Conversely, Schmidt proved in [15] that if all elements of $\mathbf{Q}[\beta] \cap [0,1]$ are finite or eventually periodic, then β is either a Pisot or a Salem number. It is also proved in [17] that if x is an algebraic integer greater than 1 which has a conjugate of modulus greater than the golden ratio $(1+\sqrt{5})/2$, then $d(1,\beta)$ is not finite or eventually periodic.

Properties of β expansions are strongly related to symbolic dynamics [4]. The closure of the set of infinite sequences appearing as β expansions is called the β shift. It is a symbolic dynamical system; that is, a closed invariant subset of $\mathcal{A}^{\mathbf{N}}$. A symbolic dynamical system is said to be *sofic* if the set of its finite factors is recognized by a finite automaton and is said to be a *shift of finite type* if it can be specified by forbidding a finite number of finite patterns (see [4]). A β-shift is sofic if and only if $d(1,\beta)$ is finite or eventually periodic and is a shift of finite type if and only if $d(1,\beta)$ is finite (see [4]). Also, Berend and Frougny [8], [1] have examined the question of whether an expansion in base β of an arbitrary real number x in $\mathbf{Z}[\beta^{-1}]$ into its (greedy) β-expansion can be realized by a finite automaton.

1.3. *Outline*

We shall prove the following:

Theorem 8.1. *Let A be a basis with dominant root $\beta > 1$. Then the numeration system L based on A can be regular only if the expansion of 1 base β, denoted $d(1,\beta)$, is finite or eventually periodic. If $d(1,\beta)$ is eventually periodic, L is regular if and only if A satisfies an extended beta polynomial for β. If $d(1,\beta)$ is finite, L is regular if A satisfies an extended beta polynomial for β; if L is regular, then A satisfies a polynomial of the form $(x^l - 1)B(x)$ where $B(x)$ is an extended beta polynomial for β and l is the length of $d(1,\beta)$.*

Example. Let the basis A be the Fibonacci sequence with initial values $A_1 = 1$ and arbitrary A_2. Then A satisfies the linear recurrence $A_n = A_{n-1} + A_{n-2}$ with minimal polynomial $p(x) = x^2 - x - 1$. The number β in this case is the golden mean $1.61\ldots$, and in base β we have $1 = .11$. Since the polynomial $p(x)$ is Pisot, the language L based on A is regular. Note that the language may depend on the initial value A_2. For example, if we take $A_2 = 2$, the language is $L_1 = \{a_1a_2\cdots a_n|a_ia_{i+1} \in \{00, 01, 10\}\}$. If we take $A_2 = 3$, the language is $L_2 = \{a_1a_2\cdots a_n|a_1a_2\cdots a_{n-1} \in L_1$ and $a_{n-1}a_n \in \{00, 01, 02, 10\}\}$.

In this paper we consider only cases where A_1 is taken to be 1. An example not of this form can be found in [16].

We begin in Section 2 by reviewing a few lemmas from automaton theory. In Section 3 we discuss the linear recurrences, minimal recurrence and the dominant root property. In Section 4 we show a general result regarding the convergence of the leading digits of the maximum lexicographic sequence to the leading digits of $d(1, \beta)$. In Section 5 we characterize regular languages for numeration systems which satisfy the dominant root property in terms of their maximum lexicographic sequences. In Section 6 we consider the case where $d(1, \beta)$ is not finite nor eventually periodic, and show that the language L cannot be regular. Finally, in Section 7, we consider the case where $d(1, \beta)$ is finite or eventually periodic. We exhibit necessary and sufficient conditions for the language L to be regular. We summarize in Section 8.

Apart from Shallit's result, some other cases have been considered. In [9] Frougny and Solomyak have considered the case where the dominant root β is Pisot and the minimal recurrence polynomial is irreducible. They have shown that the language is regular in this case. Bertrand-Mathis [3] and Loraud [11] have shown that the language L is regular in the dominant root case if the expansion of β is finite or eventually periodic and if certain conditions on the initial conditions of the recurrence hold.

2. Some Lemmas from Automaton Theory

We basically follow the exposition of [6] for the definition of finite automata over an alphabet. An *automaton over an alphabet* C, $\Lambda = (Q, C, E, I, T)$ is a directed graph labeled by elements of C; Q is the set of *states*, $I \subset Q$ is the set of *initial* states, $T \subset Q$ is the set of *terminal* states, and $E \subset Q \times C \times Q$ is the set of labeled *edges*. The automaton is said to be *finite* if Q is finite; this will always be the case in this paper. A subset L of C^* is said to be *regular* if there exists a finite automaton such that L is equal to the set of labels of paths starting in an initial state and ending in a terminal state.

A numeration system is called *regular* if the set of all representations L is regular.

Given a word w we define the set $w^* = \{w^i : i \geq 0\}$. The following lemmas will aid our study of the lexicographically maximum words.

Definition 2.1. Define $B(L)$, the set of lexicographically maximum words of every length, as follows:

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

Lemma 2.1 [13]. *If L is regular, then so is $B(L)$.*

Lemma 2.2 [16]. *The following two statements are equivalent*:

1. $L \subseteq \Sigma^*$ *is regular and there exists a constant* c *such that* $|L \cap \Sigma^n| \le c$ *for all* $n \ge 0$.
2. L *is the finite union of sets of the form* xy^*z, *where* $x, y, z \in \Sigma^*$.

Corollary 2.3. *Assume that* L *is regular and has exactly one word of each length* $m \ge 0$. *Then there exists an integer* n *such that*

$$L = L_0 \cup \bigcup_{i=1}^{n} x_i y_i^* z_i,$$

where $|y_i| = n$ *and* L_0 *is finite. Without loss of generality we can take* $|x_i| = n$. *Further, the union is disjoint, and each term in the union gives all sufficiently large words* w *in* L *whose lengths have a particular class modulo* n.

Proof. By Lemma 2.2, we can write $L = L_0 \cup \bigcup_{i=1}^{k} x_i y_i^* z_i$, where L_0 is finite and $|y_i| > 0$ for all i. If n is a common multiple of the $|y_i|$ we can split the ith term of the union into $n/|y_i|$ terms of the form $X_j Y_i^* z_i$ where $Y_i = y_i^{n/|y_i|}$ and $X_j = x_i (y_i)^{j-1}$, possibly adding terms to L_0. Thus without loss of generality we can assume $|y_i|$ to be n for all i. Clearly, we can assume that $|y_i| > |x_i|$. For each term i of the new union we can also increase the length of x_i by one without changing L by cycling y_i and adding a term to L_0. Thus without loss of generality we can also take $|x_i| = n$. $\square$

3. The Minimal Polynomial of the Recurrence

3.1. *Linear Recurrences*

In the statement of Theorem 1.1, the integer s is called the *degree* of the recurrence. If the basis satisfies a linear recurrence relation, it satisfies a unique linear recurrence relation with the smallest degree, which we call the *minimal recurrence* for the basis. This linear recurrence also has integral coefficients; for details, see pp. 4–6 of [14].

If A satisfies the linear recurrence with coefficients $u_1, \ldots, u_s$, we can construct an $s \times s$ matrix M, called the *recurrence matrix,* such that, for all positive m and k,

$$M^k \begin{bmatrix} A_m \\ A_{m+1} \\ \cdots \\ A_{m+s-1} \end{bmatrix} = \begin{bmatrix} A_{m+k} \\ A_{m+k+1} \\ \cdots \\ A_{m+k+s-1} \end{bmatrix}.$$

This matrix M is the companion matrix for the polynomial $p(x) = x^s - u_1 x^{s-1} - \cdots - u_s$. The basis A is said to *satisfy* the polynomial $p(x)$. Also, if u_i is the minimal recurrence for A, $p(x)$ is called the *minimal polynomial* for A.

Let β_i, $1 \le i \le r$, be the roots of $p(x)$, and let h_i be the multiplicity of root β_i. Then the entries of the root vectors for M are of the form $k^j \beta_i^{k-1}$, where $0 \le j < h_i$.

Therefore for any sufficiently large k we can write $[A_k, A_{k+1}, \ldots, A_{k+s-1}]^T$ as a linear combination:

$$\begin{bmatrix} A_k \\ A_{k+1} \\ \cdots \\ A_{k+s-1} \end{bmatrix} = M^{k-1} \begin{bmatrix} A_1 \\ A_2 \\ \cdots \\ A_{s-1} \end{bmatrix} = \sum_{i=1}^{r} \sum_{j=0}^{h_i-1} c_{ij} \begin{bmatrix} k^j \beta_i^{k-1} \\ (k+1)^j \beta_i^k \\ \cdots \\ (k+s-1)^j \beta_i^{k+s-2} \end{bmatrix}$$

for some $c_{ij} \in \mathbf{C}$. In particular we have the formula

$$A_k = \sum_{i=1}^{r} \sum_{j=0}^{h_i-1} c_{ij} k^j \beta_i^{k-1}.$$

The growth of a linearly recurrent basis A is seen from the above to be dominated by root sequences corresponding to roots of $p(x)$ having maximum modulus. If there is a single root of maximum modulus, we show below that the dominant root condition holds.

We also define the basis operator Δ_p associated with a polynomial p. For $p(x) = x^s - u_1 x^{s-1} - \cdots - u_s$, this is defined as

$$(\Delta_p(A))_m = A_{m+s} - u_1 A_{m+s-1} - \cdots - u_s A_m. \tag{1}$$

Then A satisfies the polynomial $p(x)$ if and only if $\Delta_p(A) = 0$. The following is immediate:

Lemma 3.1. *Let p and q be polynomials. Let A be a basis. Then*:

1. $\Delta_{pq} = \Delta_p \Delta_q$.
2. *If A satisfies $pq(x)$, then $\Delta_q(A)$ satsfies $p(x)$.*

3.2. *The Minimal Polynomial of the Recurrence*

Let

$$p(x) = \prod_{i=1}^{r} r_i^{h_i}(x)$$

be the minimal recurrence polynomial for A, where each r_i is irreducible and $i \neq j$ implies $r_i \neq r_j$. Let $g_i = \deg r_i$. Let β_{ij}, $1 \leq j \leq g_i$, be the roots of r_i. We define the *initial root vectors* for the recurrence as follows. For $\beta_{ij} \neq 0$, define

$$v_{ijk} = [1^k, 2^k \beta_{ij}, 3^k \beta_{ij}^2, \ldots, s^k \beta_{ij}^{s-1}]^T,$$

where $1 \leq i \leq r$, $1 \leq j \leq g_i$, $0 \leq k < h_i$, and s is the degree of $p(x)$. If $\beta_{ij} = 0$ for some i, j, define v_{ijk} to be the s-vector whose entries are all zero except for a one in the kth entry.

Lemma 3.2. *If* $1 \leq i \leq r$, $1 \leq j \leq g_i$, *and* $0 \leq k < h_i$, *define the sequence* a_{ijkn} *as follows*:

$$a_{ijkn} = n^k \beta_{ij}^{n-1} \qquad \textit{if} \quad \beta_{ij} \neq 0,$$

$$a_{ijkn} = \delta_{k,n} \qquad \textit{if} \quad \beta_{ij} = 0.$$

Then a_{ijkn} *satisfies the linear recurrence given by* u_i. *Furthermore, the vectors* v_{ijk}, *where* $1 \leq i \leq r$, $1 \leq j \leq g_i$, *and* $0 \leq k < h_i$, *form a basis for* $\mathbf{Q}^s$. *Also, let* z *be the initial vector, and write*

$$z = \sum_{i=1}^{r} \sum_{j=1}^{g_i} \sum_{k=0}^{h_i-1} c_{ijk} v_{ijk}.$$

Then

$$A_n = \sum_{i=1}^{r} \sum_{j=1}^{g_i} \sum_{k=0}^{h_i-1} c_{ijk} a_{ijkn}.$$

Proof. For $\beta_{ij} \neq 0$, the vector v_{ijk} is a root vector of the matrix of recurrence. Therefore the sequence a_{ijkn} satisfes the linear recurrence given by the u_i. If $\beta_{ij} = 0$ with multiplicity h_i, the linear recurrence does not depend on $A_1 \cdots A_{h_i}$, so a_{ijkn} satisfies the linear recurrence. Since the s vectors v_{ijk} are linearly independent, they are a basis for $\mathbf{Q}^s$. Also, since the sequence a_{ijkn} satisfies the linear recurrence and has the correct initial values, the last equality holds. □

We call the vector $[A_1, \ldots, A_s]^T$ the *initial vector* for the basis. We have the following lemma:

Lemma 3.3. *Let* $p(x) = \prod_{i=1}^{r} r_i^{h_i}(x)$ *be the minimal recurrence polynomial for the basis* A, *where* r_i *is irreducible for all* i *and* $r_i \neq r_j$ *for* $i \neq j$. *Let* $\{\beta_{ij}\}$ *be the* jth *root of* r_i, *and let* g_i *be the degree of* r_i. *Let* z *be the initial vector, and write*

$$z = \sum_{i=1}^{r} \sum_{j=1}^{g_i} \sum_{k=0}^{h_i-1} c_{ijk} v_{ijk}.$$

Then $c_{i,j,h_i-1} \neq 0$.

Proof. Fix i. Let $\mathbf{E} = \mathbf{Q}(\beta_{i1}, \ldots, \beta_{ig_i})$ be the Galois splitting field for r_i over $\mathbf{Q}$, and denote by G the Galois group for r_i. Note that G is transitive on the roots of r_i. We can extend $\mathbf{E}$ to a field $\mathbf{F}$ by adjoining all roots of p not already in $\mathbf{E}$; the members of G extend to $\mathbf{F}$, and since p has integral coefficients, the members of G map roots of p to roots of p.

Let $l = h_i - 1$, and assume $c_{ijl} = 0$ for some j, $1 \le j \le g_i$. Consider any j', $1 \le j' \le g_i$. Because G acts transitively on the roots of r_i, there exists $\sigma \in G$ such that $\sigma(\beta_{ij'}) = \beta_{ij}$. Because σ permutes the roots of p, we may also denote the permutation induced on the indices of β_{ij} by σ. Also, σ fixes $\mathbf{Z}$. Therefore we have

$$\begin{aligned} z = \sigma(z) &= \sum_{I=1}^{r} \sum_{J=1}^{g_i} \sum_{k=0}^{h_i-1} \sigma(c_{IJk} v_{IJk}) \\ &= \sum_{I=1}^{r} \sum_{J=1}^{g_i} \sum_{k=0}^{h_i-1} \sigma(c_{IJk}) v_{\sigma(IJ)k}. \end{aligned}$$

Because v_{IJk} is a basis for $\mathbf{Q}^s$, the coefficients c_{ijk} and $\sigma(c_{ij'k})$ must be equal. Therefore, $\sigma(c_{ij'l}) = 0$, so $c_{ij'l} = 0$. So by the transitivity of σ, $c_{ijl} = 0$ for all j. If this were the case, then $p(x)/r_i(x)$ would be a recurrence polynomial. This is a contradiction, since $p(x)$ is minimal. □

The preceding lemma motivates the definition of the dominant root condition (Definition 1.1).

Corollary 3.4. *Let A be a basis satisfying the linear recurrence given by minimal polynomial $p(x)$. Assume that $p(x)$ has a unique root β, possibly with multiplicity, of maximum modulus, and assume that β is real. Then A satisfies the dominant root condition with dominant root β.*

Proof. We freely use notation from above. Assume as in the statement that $p(x)$ has unique dominant root β of maximum modulus, and let i and j be such that $\beta_{ij} = \beta$. By Lemma 3.2, we have

$$A_n = \sum_{I=1}^{r} \sum_{J=1}^{g_I} \sum_{k=0}^{h_I-1} c_{IJk} a_{IJkn}.$$

By Lemma 3.3 and the expression for the a_{ijkn} we have

$$\lim_{n\to\infty} \frac{A_n}{a_{i,j,h_i-1,n}} = c_{i,j,h_i-1}.$$

Thus

$$\lim_{n\to\infty} \frac{A_n}{A_{n-1}} = \lim_{n\to\infty} \frac{a_{i,j,h_i-1,n}}{a_{i,j,h_i-1,n-1}} = \beta.$$

This is the dominant root condition with dominant root β. □

We assume for the remainder of the paper that the dominant root condition holds.

4. Convergence of the Maximum Lexicographic Sequence

We have the following lemmas relating the lexicographically maximum words of a given length to the expansion $d(1, \beta) = .d_1 d_2 \cdots$ of 1 in base β.

Lemma 4.1. *Assume A has dominant root β, and assume $d(1, \beta)$ has (possibly infinite) length l. For each N, $0 \le N < l$, there exists $M = M(N)$ and $C = C(N)$ such that, for $m \ge M$ and $1 \le k \le CA_m$, $G(A_m - k)_i = d_i$ for all i, $1 \le i \le N$.*

Proof. The lemma is trivially true for $N = 0$. Assume the lemma is true for $N - 1$. Let $m \ge M(N-1)$, and let $C' = C(N-1)$. By induction, if $1 \le k \le C'A_m$, $G(A_m - k)_i = d_i$ for $1 \le i < N$. To find the Nth digit, we first calculate the remainder after the first $N - 1$ digits. Let $s_m = d_1 A_{m-1} + \cdots + d_{N-1} A_{m-N+1}$. Then we have $G(A_m - k)_N = \lfloor D_{mk} \rfloor$, where $D_{mk} = (A_m - k - s_m)/A_{m-N}$.

Since

$$\lim_{m\to\infty} \frac{A_{m-i}}{A_{m-N}} = \beta^{N-i},$$

we have, for any c,

$$\begin{aligned}\lim_{m\to\infty} \frac{(A_m(1-c) - s_m)}{A_{m-N}} &= \beta^N(1-c) - (d_1\beta^{N-1} + \cdots + d_{N-1}\beta) \\ &= \beta^N(1 - (d_1\beta^{-1} + \cdots + d_{N-1}\beta^{-N+1})) - \beta^N c.\end{aligned}$$

Because $d(1, \beta)$ has length greater than N, the first term converges to a value strictly between d_N and $d_N + 1$. If c is close enough to zero, this is also true of the whole limit. Pick such a positive c, and let $C = \min(c, C')$.

If we then take $1 \le k \le CA_m$, we have

$$\frac{A_m - s_m - CA_m}{A_{m-N}} < D_{mk} < \frac{A_m - s_m}{A_{m-N}}.$$

So, for m sufficiently large, $G(A_m - k)_N = \lfloor D_{mk} \rfloor = d_N$. □

Lemma 4.2. *Assume A has dominant root β, and assume $d(1, \beta)$ has (possibly infinite) length l. Let K be a positive constant. For each N, $0 \le N < l$, there exists $M = M(N, K)$ such that, for $m \ge M$, $G(A_m - K)_i = d_i$ for i, $1 \le i \le N$.*

Proof. Lemma 4.1 gives C and M' such that, for $1 \le k \le CA_m$ and $m' > M'$, $G(A_{m'} - k)_i = d_i$ for i, $1 \le i \le N$. Take $M = M(N, K)$ large enough so that it exceeds M' and so that $K/A_M < C$. □

Lemma 4.3. *Assume A has dominant root β, and assume $d(1, \beta)$ has length $l < \infty$. Let $e_j = (d_1 d_2 \cdots d_{l-1}(d_l - 1))^{j-1}(d_1 d_2 \cdots d_l)0^\infty$. For each $N \ge 0$, there exists $M = M(N)$ such that for $m \ge M$, $G(A_m - 1)_i = (e_j)_i$ for some j and for all i, $1 \le i \le N$.*

Proof. As our inductive hypothesis, we claim that given $i \geq 1$, the proposition holds for $N = il - 1$. By Lemma 4.2, the inductive hypothesis is true for $i = 1$. We assume the inductive hypothesis holds for $i - 1$ and show that it holds for i.

Let R_{mjk} be the remainder after $k - 1$ digits from $A_m - 1$, assuming the expansion $G(A_m - 1)$ begins with the first $k - 1$ digits of e_j. Then we let $D_{mjk} = R_{mjk}/A_{m-k}$. That is,

$$D_{mjk} = \frac{(A_m - 1) - \sum_{n=1}^{k-1} (e_j)_n A_{m-n}}{A_{m-k}}.$$

Assume that the first $(i - 1)l - 1$ digits have been picked; fix $k = (i - 1)l$, and let j be as in the statement of the lemma for $N = k$.

If $j < i - 1$, then e_j has $jl < k$ nonzero digits, which are all used. Thus,

$$\begin{aligned}\lim_{m\to\infty} D_{mjk} &= \beta^k - (d_1\beta^{k-1} + \cdots + d_{l-1}\beta^{k-l+1} + (d_l - 1)\beta^{k-l}) \\ &\quad \cdot (1 + \beta^{-l} + \cdots + \beta^{-(j-2)l}) - (d_1\beta^{k-(j-1)l-1} + \cdots + d_l\beta^{k-jl}) \\ &= (\beta^k - d_1\beta^{k-1} - \cdots - d_l\beta^{k-l})(1 + \beta^{-l} + \cdots + \beta^{-(j-2)l}) \\ &\quad + \beta^{k-(j-1)l} - (d_1\beta^{k-(j-1)l-1} + \cdots + d_l\beta^{k-jl}) \\ &= 0.\end{aligned}$$

If $j \geq i - 1$, not all nonzero digits are used. We have

$$\begin{aligned}\lim_{m\to\infty} D_{mjk} &= \beta^k - (d_1\beta^{k-1} + \cdots + d_{l-1}\beta^{k-l+1} + (d_l - 1)\beta^{k-l}) \\ &\quad \cdot (1 + \beta^{-l} + \cdots + \beta^{-(j-2)l}) - (d_1\beta^{l-1} + \cdots + d_{l-1}\beta) \\ &= (\beta^k - d_1\beta^{k-1} - \ldots - d_l\beta^{k-l})(1 + \beta^{-l} + \cdots + \beta^{-(j-2)l}) \\ &\quad + (\beta^{l-1} - d_1\beta^{l-1} - \cdots - d_{l-1}\beta) \\ &= d_l.\end{aligned}$$

Let $B > \sup_n(A_n/A_{n-l+1})$. By induction, for m sufficiently large there exists j such that $G(A_m - 1)_i = (e_j)_i$ for all i, $1 \leq i < k$.

If $j < i - 1$, then $\lim_{m\to\infty} D_{mjk} = 0$, so, for m sufficiently large,

$$\frac{R_{mjk}}{A_{m-k}} = D_{mjk} < \frac{1}{B} \leq \frac{A_{m-il+1}}{A_{m-k}}.$$

Then, for such m, the remainder R_{mjk} is less than $A_{m-(k+l-1)}$, and so the kth through $(k + l - 1)$th digits of $G(A_m - 1)$ are zero.

Next consider the case that $j \geq i - 1$, and $D_{m,(i-1),k} \geq d_l$. Then the kth digit of $G(A_m - 1)$ is d_l. In this case we take without loss of generality $j = i - 1$, and consider the remainder after choosing k digits. Then, since $\lim_{n\to\infty} D_{mjk} = d_l$, for m sufficiently large,

$$\frac{R_{m,j,k+1}}{A_{m-k}} = \frac{R_{mjk} - d_l A_{m-k}}{A_{m-k}} < \frac{1}{B} < \frac{A_{m-(k+l-1)}}{A_{m-k}}.$$

Thus for such m the remainder is less than $A_{m-(k+l-1)}$, so the $(k+1)$th through $(k+l-1)$th digits are zero.

If $j \geq i-1$ and $D_{m,(i-1),k} < d_l$, the kth digit of $G(A_m - 1)$ is $d_l - 1$. In this case we take without loss of generality $j = i$. We note that for any $\delta > 0$ there exists M' such that, for $m > M'$,

$$|R_{mil} - d_l A_{m-l}| < \delta A_{m-l}.$$

Then

$$|R_{m,i,(l+1)} - A_{m-l}| < \delta A_{m-l}.$$

We have

$$\begin{aligned}
R_{m,i,k+1} &= A_m - 1 - \sum_{n=1}^{l} (e_i)_n A_{m-n} - \sum_{n=l+1}^{(i-1)l} (e_i)_n A_{m-n} \\
&= R_{m,i,l+1} - \sum_{n=l+1}^{(i-1)l} (e_i)_n A_{m-n} \\
&= R_{m,i,l+1} - A_{m-l} + \left(A_{m-l} - \sum_{n=l+1}^{(i-1)l} (e_i)_n A_{m-n} \right) \\
&= R_{m,i,l+1} - A_{m-l} + R_{m-l,i,k-l+1}.
\end{aligned}$$

Therefore, for sufficiently large m,

$$|R_{m,i,k+1} - R_{m-l,i,k-l+1}| < \delta A_{m-l}.$$

Continuing inductively, if m is sufficiently large,

$$\begin{aligned}
|R_{m,i,k+1} - A_{m-k} + 1| &= |R_{m,i,k+1} - R_{m-k,i,1}| \\
&< \delta(A_{m-l} + A_{m-2l} + \cdots + A_{m-k}) \\
&< \delta S A_{m-k},
\end{aligned}$$

where S is a constant independent of m. Thus,

$$(1 - \delta(S+1))A_{m-k} < R_{m,j,k+1} < A_{m-k}$$

for m sufficiently large.

By Lemma 4.1, for sufficiently small δ and for sufficiently large m the $(k+1)$th through $(k+l-1)$th digits of $G(A_m - 1)$ are $d_1 \cdots d_{l-1}$. Since we can make δ arbitrarily small, we are done. □

5. The Language Generated by the Lexicographically Maximal Sequence

We have the following equivalent condition to regularity:

Proposition 5.1. *Assume that L is the language of a numeration system based on basis A with dominant root β. Let $B(L)$ be the union of the lexicographically maximum words of each length. Then L is regular if and only if there exists p such that*

$$B(L) = L_0 \cup \bigcup_{i=1}^{p} x_i y_i^* z_i,$$

where $|y_i| = p$ and L_0 is finite.

Necessity is given by Corollary 2.3. We prove sufficiency below.

Let s_n denote the maximal lexicographic word of length n in L. Because $A_{n+1} - 1$ is the largest integer having representation of length less than $n+1$, we have that $s_n = G(A_{n+1} - 1)$. With $v = v_1 \cdots v_n$, define the set S_n by

$$S_n = \{v : \forall i, 1 \le i \le n-1, v_i v_{i+1} \cdots v_n \le_{lex} s_{n-i+1})\}.$$

We shall show that the *language of admissible words* $S = \bigcup_{n=0}^{\infty} S_n$ is equal to L, and, under the hypotheses of Proposition 5.1, we shall prove that S is regular.

Lemma 5.2. *Evaluation is strictly order-preserving from S to* $\mathbf{N}$.

Proof. Assume the lemma is not true. Then there is $w \in S$ such that $v \in S$ exists with $w <_{lex} v$ and $E(w) \ge E(v)$. Fix $w = w_1 w_2 \cdots w_n$ to be the lexicographically minimal such w, and fix $v = v_1 v_2 \cdots v_m$ to be the minimal v for w.

Clearly, $m \ge n$. First we consider the case $m > n$. We have $w \le_{lex} s_n$. If $w =_{lex} s_n$, then $w = s_n$, and so $E(w) = A_{n+1} - 1 \le A_m \le E(v)$. If $w <_{lex} s_n$, then $E(w) \ge E(v) \ge A_{n+1} > E(s_n)$, so v is not minimal for w. Therefore, $m = n$.

We claim that w has at least two nonzero digits. We have $E(w) \ge E(v) \ge G(v_1 A_n) \ge G(w_1 A_n)$. If w has only one nonzero digit, equality holds, and $w = w_1 A_n = v_1 A_n = v$.

Let i be the index of the second most significant nonzero digit in w. First, assume $w_1 = v_1$. Then v must have more than one nonzero digit. Let j be the index of the second most significant nonzero digit in v, and consider $w' = w_i w_{i+1} \cdots w_n$ and $v' = v_j v_{j+1} \cdots v_n$. Then $E(w') \ge E(v')$ and $w' <_{lex} v'$. This contradicts the lexicographic minimality of w.

Therefore we must have $w_1 < v_1$. Consider $w^- = w_i w_{i+1} \cdots w_n$ and $v^- = (v_1 - w_1)v_2 \cdots v_n$. Then $E(w^-) \geq E(v^-)$ and $w^- <_{lex} v^-$. However, w^- is lexicographically smaller than w, which contradicts the minimality of w. □

Lemma 5.3. *The language L for a greedy numeration system equals S.*

Proof. Let $L_n = L \cap \Sigma^n$. Because G is a greedy representation and $\max L_n = G(A_{n+1} - 1) = s_n$, it is clear that $L_n \subset S_n$. Therefore $L \subset S$. Choose $n \in \mathbf{N}$ and define $N = \{j \mid A_{n-1} \leq j \leq A_n - 1\}$. Let $i\colon L_n \to S_n$ be the inclusion map. Then $E \circ i \circ G$ is a one-to-one strictly order-preserving map from N to N. Thus i is onto, and $S_n \subset L_n$. □

Lemma 5.4. *Assume that*

$$B(L) = L_0 \cup \bigcup_{i=1}^{p} x_i y_i^* z_i,$$

where L_0 is finite, $|x_i| = |y_i| = p$. Then S is regular.

Proof. Let $M = \max|z_i|$. Define, for $1 \leq i < p$,

$$T_0 = \{w_1 \cdots w_n : w_{n-j+1} \cdots w_n \leq_{lex} s_j,\ j \leq M\},$$

$$T_i = \{w_1 \cdots w_n : w_{n-j+1} \cdots w_n \leq_{lex} s_j,\ j = M + i + kp, k \in \mathbf{N}\}.$$

Then $S = \bigcap_{i=0}^{p} T_i$. We claim that T_i is regular. Clearly T_0 is regular.

We construct an automaton similar to the β-shift automata of Parry [12]. Let the alphabet $\mathcal{A} = \{m : m \leq \sup A_{n+1}/A_n, n \in \mathbf{N}\}$. Fix i. Let $x_i = a = a_1 \cdots a_p$ and $y_i = b = b_1 \cdots b_p$. Let $Z = |z_i|$. We define the following subautomata:

1. Automaton φ accepts all words with alphabet $\mathcal{A}$ of length less than p.
2. Automaton ψ_1 accepts all words with alphabet $\mathcal{A}$ of length Z.
3. Automaton ψ_2 accepts all words of length Z which are lexicographically less than or equal to z_i.

The automata which accept T_i in the cases $a \geq_{lex} b$ and $a <_{lex} b$ are shown in Figures 1(a) and 1(b). We remark that in the second case, the language T_i can also be given as $T_i = \{w_1 \cdots w_n : w_{n-j+1} \cdots w_n \leq_{lex} a^{k+1},\ j = M + i + kp, k \in \mathbf{N}\}$. In these diagrams the following vertices have special significance:

1. The vertices α, β, and ω are encountered when there are no lexicographical constraints and there are $Z + kp$ symbols left in the word.
2. The vertices γ and δ are encountered after ab^* and a^*, respectively.

This automaton has the language T_i. Thus S is the finite intersection of regular languages, and is regular. □

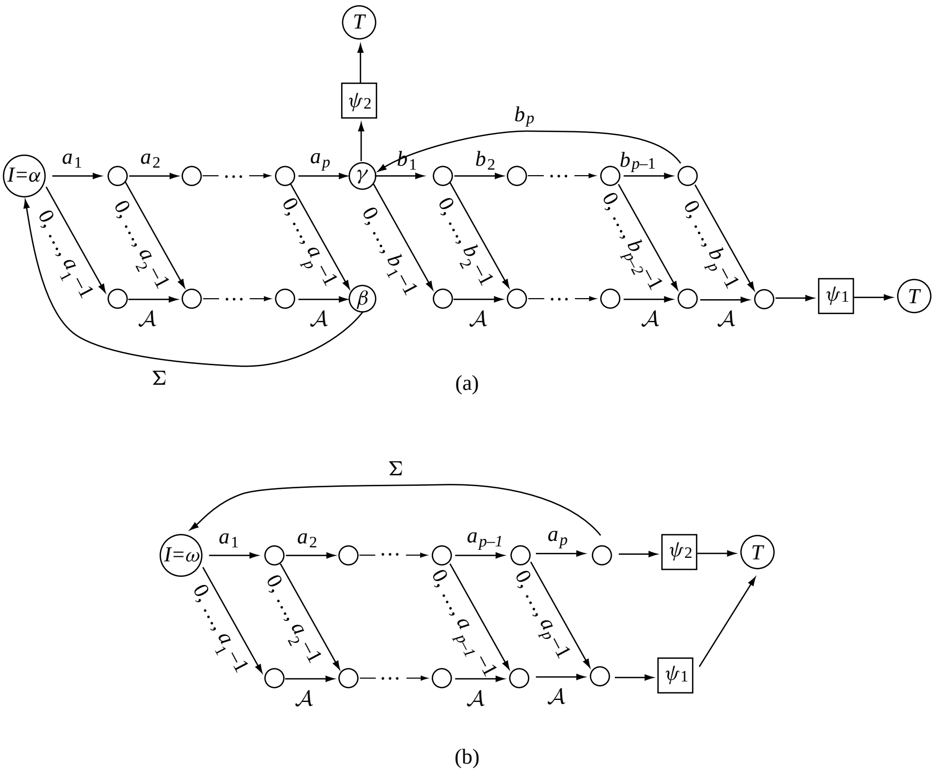

Fig. 1

6. Case Where the Expansion of Beta Is Not Eventually Periodic

In this case we show that L is not regular.

Proposition 6.1. *If $d(1, \beta)$ is not finite or eventually periodic, then L is not regular.*

Proof. Assume that $d(1, \beta) = .d_1 d_2 \cdots$ is not finite or eventually periodic, but L is regular. Then $B = B(L)$ is regular. By Lemma 2.2 we can write

$$B = \bigcup_{i=1}^{n} x_i y_i^* z_i .$$

Let $r = \sup_{i \le i \le n}\{|x_i|, |y_i|, |z_i|\}$. Let $w_m = d_1 d_2 \cdots d_m$. By Lemma 4.2, for each m, there exists n such that A_n begins with w_m. Thus there exists a word u_m such that $w_m u_m \in L$. Therefore, for $m > r$,

$$w_m u_m = x y^k z,$$

where $|x| \leq r$, $|y| \leq r$, and $|z| \leq r$. Thus $w_m = xy^{k'}Z$ where $k' \geq 0$ and either y or z begins with the word Z. In either case $|Z| \leq r$. Letting $m \to \infty$, we see that $d(1, \beta) = xy^{\infty}$, so it is finite or eventually periodic. So, by contradiction, L is not regular. □

7. Case Where the Expansion of Beta Is Eventually Periodic

7.1. *Beta Polynomials*

In this section we discuss the extended beta polynomials, which are polynomials derived from the relationship between the digits in $d(1, \beta)$.

First assume $d(1, \beta)$ is eventually periodic, and let l and p be the minimal lead and period for $d(1, \beta)$. Then we define the *canonical beta polynomial* $b(x)$:

$$b(x) = x^{l+p} - \sum_{i=1}^{l+p} d_i x^{l+p-i} - x^l + \sum_{i=1}^{l} d_i x^{l-i}.$$

In the case that $d(1, \beta)$ is finite, we let l be the length of $d(1, \beta)$ and define $b(x)$ as follows:

$$b(x) = x^l - \sum_{i=1}^{l} d_i x^{l-i}.$$

(Note that if $l = 1$, β is just d_1.)

We further define the set of *extended beta polynomials*. Let $b(x)$ be the canonical beta polynomial for β. Assume that $d(1, \beta)$ is eventually periodic (resp. finite) and let P be the minimal period (resp. the length) of the expansion. The set of extended beta polynomials for β is the set of polynomials of the form $b(x)q(x)$ where $q(x) = x^n(1 + x^P + \cdots + x^{(k-1)P})$ for $n \in \mathbf{Z}^+$, $k \in \mathbf{N}$.

The extended beta polynomials can be derived from the canonical beta polynomial by taking nonminimal values for l and p. (For example, if $d(1, \beta) = .4(31)^{\infty}$, the canonical beta polynomial is computed by taking $l = 1$ and $p = 2$. We could however write (for example) $1 = .43(1313)^{\infty}$ and take $l = 2$ and $p = 4$ to get an extended beta polynomial.) In the eventually periodic case we can take any lead $m \geq l$ and any period $p' = kp$, $k \in \mathbf{N}$. In the case that $d(1, \beta)$ is finite, we can consider the following noncanonical expansions of 1:

$$1 = .(d_1 \cdots d_{p-1}(d_l - 1))^{k-1}(d_1 \cdots d_l)$$

and

$$1 = .(d_1 \cdots d_{l-1}(d_l - 1))^{\infty}.$$

For example, let $b(x)$ be the canonical beta polynomial for an eventually periodic expansion having lead l, and let $b'(x)$ be the extended beta polynomial for the same expansion considered as having lead $l + k$. We have

$$\begin{aligned} b'(x) &= x^{l+k+p} - \sum_{i=1}^{l+k+p} d_i x^{l+k+p-i} - x^{l+k} + \sum_{i=1}^{l+k} d_i x^{l+k-i} \\ &= x^k b(x). \end{aligned}$$

Next, let $b(x)$ be as above, and let $b'(x)$ be the extended beta polynomial for the same expansion considered as having period kp. Then

$$\begin{aligned} b'(x) &= x^{l+kp} - \sum_{i=1}^{l+kp} d_i x^{l+kp-i} - x^{l} + \sum_{i=1}^{l} d_i x^{l-i} \\ &= (1 + x^p + \cdots + x^{(k-1)p}) b(x). \end{aligned}$$

Next we consider the noncanonical finite expansions. Using the first special finite expansion, the extended beta polynomial $b'(x)$ is

$$\begin{aligned} b'(x) &= \sum_{j=0}^{k-1} (x^{l(j+1)} - x^{lj} + x^{lj} - \sum_{i=1}^{l} d_i x^{l(j+1)-(k-1)+j(k-1)}) - 1 + 1 \\ &= (1 + x^l + \cdots + x^{(k-1)l}) b(x). \end{aligned}$$

In the second case the extended beta polynomial is the same as a beta polynomial in the eventually periodic case. If b' is the extended beta polynomial considering the representation of 1 as periodic with minimal period (i.e., lead 0 and period l) we have

$$b'(x) = x^l - \sum_{i=1}^{l} d_i x^{l-i} = b(x).$$

We can then vary the lead and period as in the eventually periodic case above.

We call an extended beta polynomial $B(x)$ the extended beta polynomial for lead $l + k$ and period kp (resp. length kp) if it can be derived from the (canonical) beta polynomial of length l and period p (resp. length l) by the above transformations.

We have the following:

Lemma 7.1. *Let $B(x)$ be an extended beta polynomial for $\beta > 1$, and assume that $B(x)$ is the minimal polynomial for basis A. Then A satisfies the dominant root condition for β. Further, β is a simple root of $B(x)$.*

Proof. By Corollary 3.4, it is sufficient to show that $B(x)$ has no root of modulus greater than β and that β is a simple root of $B(x)$. Let $b(x)$ be the canonical beta polynomial for β, and write $B(x) = b(x)q(x)$. Because all roots of $q(x)$ have modulus 0 or 1, it is sufficient to show the above for $b(x)$.

We set up the function $F(z)$, defined outside the unit disk:

$$F(z) = 1 - \sum_{i=0}^{\infty} \frac{d_{i+1}}{z^i}.$$

A straightforward calculation shows that if $|z| > 1$, $F(z) = 0$ if and only if z is a root of $b(x)$. Next we apply a result of Flatto [7, Theorem 4.2] which states that $F(z)$ has no zeros of modulus greater than or equal to β except for a single zero at $z = \beta$. □

7.2. *Case Where the Basis Satisfies an Extended Beta Polynomial*

In this section we make the additional restriction that the basis A satisfies the linear recurrence given by an extended beta polynomial $B(x) = b(x)q(x)$. It is sufficient that the minimal recurrence polynomial $p(x)$ is a divisor of $B(x)$. These restrictions are automatically satisfied if $p(x)$ is irreducible. We do the finite case first.

Lemma 7.2. *Assume $d(1, \beta)$ is finite, and let $d(1, \beta) = 0.d_1 d_2 \cdots d_l$. Let $b(x)$ be the canonical beta polynomial for β. Assume also that the basis A satisfies the recurrence given by the extended beta polynomial $B(x) = q(x)b(x)$ where $q(x) = x^n(1 + x^l + \cdots + x^{(a-1)l})$. Let*

1. *w_m^i denote the first m letters of $(d_1 d_2 \cdots d_{l-1}(d_l - 1))^{(i-1)} d_1 \cdots d_l 0^\infty$,*
2. *w_m^∞ denote the first m letters of $(d_1 d_2 \cdots d_{l-1}(d_l - 1))^\infty$.*

Then there exists M such that, for $m \geq M$,

$$s_m = w_{m-M}^{a_{m \bmod al}} W_{m \bmod al},$$

where W_i, $0 \leq i < al$, is a word of length M, $a_i \in \mathbf{Z}^+ \cup \{\infty\}$, and s_m is the lexicographically maximal word of length m in L. In particular, the language is regular.

Proof. First, we recall the definition of Δ_b from (1). For $m > l$, we let

$$e_m = \Delta_b(A)_{m-l} = A_m - d_1 A_{m-1} - \cdots - d_l A_{m-l}.$$

By Lemma 3.1, e_m satisfies the linear recurrence given by the polynomial $q(x)$. Since $q(x)$ divides $x^n(x^{al} - 1)$, for sufficiently large m we have $e_m = e_{m+al}$.

For simplicity of notation, we redefine e_m using its cyclic property (i.e., $e_m = e_{m+abl}$ for large enough b). Let $E = 1 + \sum_{j=0}^{al} |e_j|$.

The dominant root condition holds, and so by Lemma 4.2 for large enough M the first $l - 1$ letters of $G(A_M - E)$ equal w_l^1. We also want M to be large enough so that $A_{M+al} > E$, and so that e_m is cyclic for $m \geq M - al$.

Let $0 \leq i < al$, and choose $m > M$ such that $m \bmod al = i$.

First case: Assume that there is no $k < a$ such that $\sum_{j=0}^{k} e_{m-jl} > 0$. Then the lth digit of s_m must be $d_l - 1$ since the remainder after choosing $d_1 \cdots d_{l-1}(d_l - 1)$ is

$$A_m - 1 - \left(\sum_{k=1}^{l} d_l A_{m-k} \right) + A_{m-l} = A_{m-l} - 1 + e_m,$$

which is between 0 and A_{m-l} because $-A_{m-l} < -E < e_m - 1 < 0$. We apply Lemma 4.2 again to show that

$$s_m = G(A_m - 1) = d_1 d_2 \cdots d_{l-1}(d_l - 1) d_1 d_2 \cdots d_{l-1} \cdots .$$

The remainder after choosing $d_l - 1$ as the next digit is, by the same reasoning as above, $A_{m-2l} - 1 + e_m + e_{m-l}$. Since this is between 0 and A_{m-2l}, $d_l - 1$ is indeed the proper digit. Continuing by induction, we find that

$$s_m = (d_1 d_2 \cdots d_{l-1}(d_l - 1))^a G(r),$$

where

$$\begin{aligned} r &= A_m - 1 - \left(\sum_{i=1}^{al} (w_{al}^{\infty} A_{m-i}) \right) \\ &= A_{m-al} - 1 \end{aligned}$$

because the basis A satisfies the extended beta polynomial. We can then repeat the same argument, decreasing m by al, and so this case is done. We get w_m^{∞}.

Otherwise, let $k < a$ be the smallest integer such that $\sum_{j=0}^{k} e_{m-jl} = s > 0$. By the argument above, the first $(k-1)l$ letters of s_m are $d_1 d_2 \cdots d_{l-1}(d_l - 1))^{k-1}$ and

$$s_m = (d_1 d_2 \cdots d_{l-1}(d_l - 1))^{k-1} G(A_{m-(k-1)l} - 1 + s).$$

In this case the next l digits are $d_1 d_2 \cdots d_l$, and the remainder r is less than E and independent of i. Thus

$$s_m = (d_1 d_2 \cdots d_{l-1}(d_l - 1))^{k-1} (d_1 \cdots d_l) 0 \cdots 0 G(r).$$

Since everything in this depends only on the modulus class of m, the language is of the form given in the statement of the lemma. Finally, by Lemma 5.4, the language is regular. $\square$

Lemma 7.3. *Assume $d(1, \beta)$ is eventually periodic. Also assume that the basis A satisfies the recurrence given by the extended beta polynomial $B(x) = q(x)b(x)$. Write $d(1, \beta) = 0.d_1 d_2 \cdots d_l \ (d_{l+1} \cdots d_{l+p})^{\infty}$, where the lead l and period p are compatible with $B(x)$. Let $w_m = d_1 d_2 \cdots d_m$. Then there exists n and p-many length-N words $W_0, \ldots, W_{p-1}$ such that, for all $n \geq N$,*

$$s_n = w_{n-N} W_{n \bmod p},$$

where s_n is the lexicographically maximal word of length n. In particular the language is regular.

Proof. The dominant root condition holds from Lemma 7.1, and so Lemma 4.2 implies there exists $M = M(l + p, 1)$ such that, for $m \geq M$, the first $l + p$ letters of s_m equal w_{l+p}.

Let r_m be the remainder after the first l digits of $A_m - 1$. Because the basis satisfies the recurrence given by $B(x)$,

$$A_m - 1 = \sum_{i=1}^{l+p} d_i A_{m-i} + A_{m-p} - \sum_{i=1}^{l} d_i A_{m-p-i} - 1.$$

The first l digits of s_m are $d_1, d_2, \ldots, d_l$. Therefore,

$$\begin{aligned} r_m &= (A_m - 1) - \sum_{i=1}^{l} d_i A_{m-i} \\ &= \sum_{i=l+1}^{l+p} d_i A_{m-i} + A_{m-p} - \sum_{i=1}^{l} d_i A_{m-p-i} - 1. \end{aligned}$$

Since the next p digits are $d_{l+1}, \ldots, d_{l+p}$, for $m > M$, $G(r_m)$ has length $m - l - 1$ and begins with digits $d_{l+1} d_{l+2} \cdots d_{l+p}$.

We claim that if $m = M + ap + b, 0 \leq b < p$, $G(r_m) = (d_{l+1} \cdots d_{l+p})^{a+1} W_b$, where $W_b = G(r_{M+b})$. Fix b. The proposition is clearly true for $a = 0$. Let $a \geq 1$ and assume the proposition is true for $a - 1$.

We know that the first p digits of $G(r_m)$ are $d_{l+1} d_{l+2} \cdots d_{l+p}$. Then

$$r_m - \sum_{i=l+1}^{l+p} d_i A_{m-i} = A_{m-p} - \sum_{i=1}^{l} d_i A_{m-p-i} - 1 = r_{m-p}.$$

Therefore $G(r_m) = (d_{l+1} \cdots d_{l+p}) G(r_{m-p})$ and the claim is proved. Thus the lemma is proved for $N = M + p$.

Finally, by Lemma 5.4, the language is regular. $\square$

7.3. *Case Where the Recurrence Does Not Satisfy a Beta Polynomial*

We again split the results into finite and eventually periodic cases:

Lemma 7.4. *Let $d(1, \beta)$ be eventually periodic, and assume the language L is regular. Then the basis satisfies the linear recurrence given by an extended beta polynomial.*

Proof. We assume that L is regular. Write $d(1, \beta) = 0.d_1 d_2 \ldots d_l (d_{l+1} \ldots d_{l+p})^{\infty}$. Let $B = B(L)$. By Corollary 2.3, we may write

$$B = L_0 \cup \bigcup_{i=1}^{n} x_i y_i^* z_i,$$

where $n = kp$ is a positive integer and L_0 is finite.

By Lemma 4.2, we may choose $x_i = d_1 d_2 \cdots d_l$ and $y_i = (d_{l+1} \cdots d_{l+p})^k$. Note that x_i and y_i do not depend on i because the initial letters in the lexicographically maximal word of length m are fixed. Now fix i, and let $t = |z_i|$. We have, for m sufficiently large,

$$\begin{aligned} A_m - A_{m-kp} &= (A_m - 1) - (A_{m-kp} - 1) \\ &= (d_1 A_{m-1} + \cdots + d_{l+kp} A_{m-l-kp} + \cdots + z_{i1} A_t + \cdots + z_{it} A_1) \\ &\quad -(d_1 A_{m-kp-1} + \cdots + d_l A_{m-kp-l} + \cdots + z_{i1} A_t + \cdots + z_{it} A_1). \end{aligned}$$

Because the last t digits of $G(A_m - 1)$ and $G(A_{m-kp} - 1)$ are the same, we have

$$A_m - A_{m-kp} = (d_1 A_{m-1} + \cdots + d_{l+kp} A_{m-l-kp}) - (d_1 A_{m-kp-1} + \cdots + d_l A_{m-kp-l}).$$

The above gives a recurrence relation for the A_m with m sufficiently large of the form $r(x) = (1 + x^p + \cdots + x^{(k-1)p}) b(x)$. Therefore the basis A satisfies a recurrence relation of the form $x^N r(x)$, which defines an extended beta polynomial for β. □

Lemma 7.5. *Let $d(1, \beta) = 0.d_1 \cdots d_p$ be finite, and assume the language L is regular. Then the basis satisfies a linear recurrence given by $B'(x) = r(x) B(x)$ where $B(x)$ is an extended beta polynomial for β and $r(x)$ divides $x^p - 1$.*

Proof. Define, for m sufficiently large, (see proof of Lemma 7.2),

$$e_m = \Delta_b(A)_{m-l} = A_m - d_1 A_{m-1} - \ldots - d_l A_{m-l}.$$

We will show that, for sufficiently large m, e_m is periodic. This will show, by Lemma 3.1, that A_m satisfies a recurrence relationship of the form $(x^{al} - 1) b(x) = (x^l - 1)(1 + x^l + \cdots + x^{(a-1)l}) b(x)$, which is sufficient.

We proceed as in the proof of Lemma 7.4: Assume that L is regular. Let $B = B(L)$. We may write

$$B = L_0 \cup \bigcup_{i=1}^{n} x_i y_i^* z_i,$$

where n is a positive integer, $|y_i| = n$, L_0 is finite, and x_i, y_i, and z_i are in Σ^*.

By Lemma 4.3, we can take $n = kl$, and for each $i \le n$ we have either

1. $x_i = y_i = (d_1 \cdots d_{l-1}(d_l - 1))^k$, or
2. $x_i = (d_1 \cdots d_{l-1}(d_l - 1))^{k_i - 1}(d_1 \cdots d_l)$, $y_i = 0^n$.

We take $d' = (d_1 d_2 \cdots d_{l-1}(d_l - 1))^\infty$. Fix i, and let m be large such that $m \bmod n = i$.

First we assume that i is in the first case. Then the recurrence is analogous to the eventually periodic case (Lemma 7.4) with no lead digits, so, for m sufficiently large such that $m \bmod n = i$,

$$A_m - A_{m-kl} = d'_1 A_{m-1} + \cdots + d'_n A_{m-kl}.$$

Thus we define the polynomial $c(x) = q(x)b(x)$ where $q(x) = (1 + x^l + \cdots + x^{(k-1)l})$, and define $e'_j = (\Delta_c(A))_{j-l}$. Then, for m sufficiently large such that $m \bmod n = i$, e'_j is constant. Since $\Delta_c(A) = \Delta_q \Delta_b(A) = \Delta_q(e)$, for such m, e_m is periodic.

Next we consider i in the second case. For any i in the second case we have the following:

$$A_m = \sum_{j=1}^{k_i l} d'_j A_{m-j} + A_{m-k_i l} + E(z_i) + 1,$$

where E denotes evaluation.

We claim that if i is in the second case, for $t < k_i$, $i - tl$ is in the second case and $k_{i-tl} \leq k_i - t$. This is because the remainder from $A_m - 1$ after tl digits is less than or equal to $A_{m-tl} - 1$, and thus $G(A_{m-tl} - 1) \geq_{lex} (d_1 \cdots d_{l-1}(d_l - 1))^{k_i-1-t}(d_1 \cdots d_l)$.

Therefore there exists a set of integers $\{t_1, \ldots, t_r\}$ such that, for $j < r$, $i - t_j l$ is in the second case and $t_1 = 1$, $t_2 = t_1 + k_{i-t_1 l}$, $t_3 = t_2 + k_{i-t_2 l}, \ldots, t_r = t_{r-1} + k_{i-t_{(r-1)} l} = k_i$. So, for m sufficiently large such that $m \bmod n = i$,

$$\begin{aligned}
e_m &= A_m - d_1 A_{m-1} - \cdots - d_l A_{m-l} \\
&= \sum_{j=1}^{k_i l} d'_j A_{m-j} + A_{m-k_i l} + E(z_i) + 1 - d_1 A_{m-1} - \cdots - d_l A_{m-l} \\
&= -A_{m-l} + \sum_{j=l+1}^{k_i l} d'_j A_{m-j} + A_{m-k_i l} + E(z_i) + 1 \\
&= -\left(A_{m-t_1 l} - \sum_{j=1+t_1 l}^{t_2 l} d'_j A_{m-j} - A_{m-t_2 l}\right) \\
&\quad -\left(A_{m-t_2 l} - \sum_{j=1+t_2 l}^{t_3 l} d'_j A_{m-j} - A_{m-t_3 l}\right) - \cdots \\
&\quad -\left(A_{m-t_{r-1} l} - \sum_{j=1+t_{r-1} l}^{t_r l} d'_j A_{m-j} - A_{m-t_r l}\right) - A_{m-k_{t_r} l} + A_{m-k_i l} + E(z_i) + 1 \\
&= E(z_i) + E(z_{i-t_1 l}) + \cdots + E(z_{i-t_{r-1} l}) + t.
\end{aligned}$$

Since this depends only on i, we have that e_m is periodic. Thus we are done. □

In the finite case, regularity can indeed depend on initial conditions. We give an example. Let the minimal polynomial for the recurrence be $B(x) = (x - 1)(x - 3)$. First we consider initial conditions $A_1 = 1$, $A_2 = 4$. Then $A_m = 3A_{m-1} + 1$, and so $G(A_m - 1) = 30 \cdots 0$, and the language is regular. Next we consider initial conditions $A_1 = 1$, $A_2 = 2$. We claim that this is not regular. First note that $A_m = 3A_{m-1} - 1$, we use only the digits 0, 1, and 2. Using root vectors, we also have that $A_m = (3^m + 1)/2$. Let n be the largest integer such that $G(A_m - 1)$ begins with n twos. So n is the largest integer such that

$$\frac{3^m+1}{2} > 2\left(\frac{3^{m-1}+1}{2}+\cdots+\frac{3^{m-n}+1}{2}\right) = \frac{3^m-3^{m-n}}{2}+n.$$

We rewrite this as

$$3^{m-n}+1 > 2n$$

and

$$3^{m-n}+1+2(m-n) > 2m.$$

Thus as $m \to \infty$, both $n \to \infty$ and $m-n \to \infty$. Therefore this basis is not regular.

8. Remarks

8.1. *Conclusion*

We have proven the following:

Theorem 8.1. *Let A be a linearly recurrent sequence. Assume that A has dominant root $\beta > 1$. Then the numeration system L based on A can be regular only if the expansion of 1 base β, denoted $d(1,\beta)$, is finite or eventually periodic. If $d(1,\beta)$ is eventually periodic, L is regular if and only if A satisfies an extended beta polynomial for β. If $d(1,\beta)$ is finite, L is regular if A satisfies an extended beta polynomial for β; if L is regular, then A satisfies a polynomial of the form $(x^p-1)B(x)$ where $B(x)$ is an extended beta polynomial for β and p is the length of $d(1,\beta)$.*

Proof. If $d(1,\beta)$ is neither finite nor eventually periodic, Proposition 6.1 implies that the numeration system L is not regular. If $d(1,\beta)$ is eventually periodic, Lemmas 7.3 and 7.4 show that L is regular if and only if the basis satisfies a linear recurrence given by an extended beta polynomial for β. If $d(1,\beta)$ is finite and the minimal polynomial is of the form above, Lemma 7.2 shows that L is regular. Finally, Lemma 7.5 gives the above necessary condition for regularity. □

8.2. *Things Not Considered or Not Proven*

In the case that $\beta = |\beta_2|$ the limit $\lim_{i\to\infty}(A_i/A_{i-1})$ need not exist (e.g., $p(x) = x^3-k$). However, we conjecture that L can be regular only if there exists n such that $\lim_{i\to\infty}(A_{ni+k}/A_{(n(i-1)+k)})$ exists and is independent of k, and that $p(x)=q(x^n)$ where $q(x)$ is the minimal polynomial for a recurrence which gives a regular language. The only irreducible $p(x)$ for which $\beta=|\beta_2|$ are of the form $p(x)=q(x^n)$ for a polynomial q [5].

Acknowledgments

We would like to thank those people who helped inspire this paper. Among those most helpful were Boris Solomyak, who introduced us to the problem and thoroughly read the paper, Christiane Frougny, Doug Lind, and David Boyd. We would also like to thank the referees of the paper, who pointed out several errors in previous drafts of the paper.

References

[1] D. Berend and Ch. Frougny. Computability by finite automata and pisot bases. *Math. Systems Theory*, 27:275–282, 1994.

[2] A. Bertrand-Mathis. Développement en base de pisot et répartition modulo 1. *C. R. Acad. Sci. Paris*, 285:419–421, 1977.

[3] A. Bertrand-Mathis. Comment écrire les nombres entiers dans une base qui n'est pas entière. *Acta Math. Acad. Sci. Hungar.*, 54:237–241, 1989.

[4] F. Blanchard. β-expansions and symbolic dynamics. *Theoret. Comput. Sci*, 65:131–141, 1989.

[5] D. Boyd. Irreducible polynomials with many roots of maximal modulus. *Acta Arith.*, 68(1):85–88, 1994.

[6] S. Eilenberg. *Automata, Languages and Machines*, volume A. Academic Press, New York, 1974.

[7] L. Flatto. z-numbers and β-transformations. *Contemp. Math.*, 135:181–201, 1992.

[8] Ch. Frougny. Representations of numbers and finite automata. *Math. Systems Theory*, 25:713–723, 1992.

[9] Ch. Frougny and B. Solomyak. On representation of integers in linear numeration systems. In *Ergodic Theory of* $\mathbf{Z}^d$*-Actions*, pages 345–368. LMS Lecture Notes 228. Cambridge University Press, Cambridge, 1996.

[10] M. Hollander. Linear Numeration Systems, Finite Beta Expansions, and Discrete Spectrum of Substitution Dynamical Systems. Ph.D. thesis, University of Washington, 1996.

[11] N. Loraud. β-shift, systemes de numerations, et automates. *J. Théor. Nombres Bordeaux*, 7(2):473–498, 1995.

[12] W. Parry. On the β-expansions of real numbers. *Acta Math. Acad. Sci. Hungar.*, 8:477–493, 1957.

[13] J. Sakarovitch. Deux remarques sur un théorème de s. eilenberg. *RAIRO Inform. Théor.*, 17:23–48, 1983.

[14] R. Salem. *Algebraic Numbers and Fourier Analysis*. Heath, Boston, MA, 1963.

[15] K. Schmidt. On periodic expansions of pisot numbers and salem numbers. *Bull. London Math. Soc.*, 12:269–278, 1980.

[16] J. Shallit. Numeration systems, linear recurrences, and regular sets. *Inform. and Comput.*, 113:331–347, 1994.

[17] B. Solomyak. Conjugates of beta-numbers and the zero-free domain for a class of analytic functions. *Proc. London Math. Soc.* (3), 68:477–498, 1994.

Received July 1995, *and in final form August* 1997.