

H. A. CAMERON

D. WOOD

Pm numbers, ambiguity, and regularity

Informatique théorique et applications, tome 27, n° 3 (1993),
p. 261-275

http://www.numdam.org/item?id=ITA_1993__27_3_261_0

© AFCET, 1993, tous droits réservés.

L'accès aux archives de la revue « Informatique théorique et applications » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

P_m NUMBERS, AMBIGUITY, AND REGULARITY (*) (1)

by H. A. CAMERON (2) and D. WOOD (2)

Communicated by C. CHOFRUT

Abstract. — We introduce the pseudo- m -ray (P_m) number system, in which syms of the form $\sum_{i \geq 0} a_i(m^{i+1}-1)/(m-1)$ are the representations of numbers. We characterize the P_m representations that are produced by the greedy algorithm and show that they form a regular set. In addition, we show that the set of P_m representations that are the sole representations for their corresponding numbers is also a regular set.

Résumé. — Nous introduisons le système de numérotation pseudo- m -aire (P_m), dans lequel les sommes de la forme $\sum_{i \geq 0} a_i(m^{i+1}-1)/(m-1)$ sont les représentations des nombres. Nous caractérisons les représentations de P_m qui sont obtenues par l'algorithme vorace et nous montrons qu'elles forment un langage rationnel. De plus, nous montrons que l'ensemble des représentations de P_m des nombres qui ont une unique représentation, est un langage rationnel.

1. INTRODUCTION

Many number systems can be viewed as ways of representing integers based on finite or infinite integer sequences $1 = u_0 < u_1 < u_2 < \dots$. A common method of finding a representation of an integer in any such number system is the greedy algorithm; see Fraenkel [Fra85]. To find the greedy representation of an integer N , we find the largest u_i that is no larger than N and then repeatedly we set $a_i \leftarrow \lfloor N/u_i \rfloor$, $N \leftarrow N - a_i u_i$, and $i \leftarrow i - 1$, until $i = 0$. In some number systems, some integers may have representations other than the one obtained via the greedy algorithm. (A number that has more than one

(*) This work was supported under a Natural Sciences and Engineering Research Council of Canada Grant and under a grant from the Information Technology Research Centre.

(1) Received September 1991, accepted September 1992.

(2) Department of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

representation in the given number system is said to be *ambiguous*; otherwise, it is *unambiguous*.)

There appears to be a close relationship between the properties of number systems and the properties of formal languages; see Shallit [Sha91], for example. Two intriguing problems about this relationship are:

PROBLEM 1.1: *For which number systems are the sets of greedy representations regular?*

PROBLEM 1.2: *For which number systems are the sets of unambiguous numbers regular?*

We introduce the *pseudo- m -ary (P_m) number system* and show that the set of greedy representations and the the set of representations of unambiguous numbers in the P_m number system are regular sets. For any fixed integer $m > 1$, the P_m number system is based on the sequence

$$\frac{m^1 - 1}{m - 1}, \frac{m^2 - 1}{m - 1}, \frac{m^3 - 1}{m - 1}, \dots$$

As we will see, all integers are representable in the P_m number system.

The P_2 number system (the P_m number system when $m = 2$) has been studied previously. Allouche, Betrema, and Shallit [ABS89] characterized the set of integers that can be represented by P_2 representations using only the digits 0 and 1. Their interest in the P_2 number system arose from a study of the sequence of parentheses occurring in the recursive definition of the integers.

We have used the characterization of the greedy representations in the P_2 number system in Cameron [Cam91] and Cameron and Wood [CW93] to establish an upper bound result for a class of binary trees. Every binary tree can be viewed as a perfect binary tree (a binary tree whose leaves all appear on one level; see *fig. 1*) with some perfect binary subtrees removed. Each node of a perfect binary tree has two perfect binary subtrees, so each remaining node has 0, 1, or 2 perfect binary children removed by the pruning; see *figure 2*. A perfect binary subtree contains $2^h - 1$ nodes, where h is the height of the tree (the distance of the leaves from the root of the tree). Thus, we became interested in numbers of the form $\sum_{i \geq 0} a_i (2^{i+1} - 1)$, where $a_i = 0, 1,$ or 2, because they give the total size of the subtrees we have removed by pruning. These numbers are exactly the P_2 representations.

Similarly, each node of a perfect m -ary tree has m perfect m -ary subtrees. Pruning such a tree removes 0, 1, 2, ..., or m perfect m -ary subtrees from

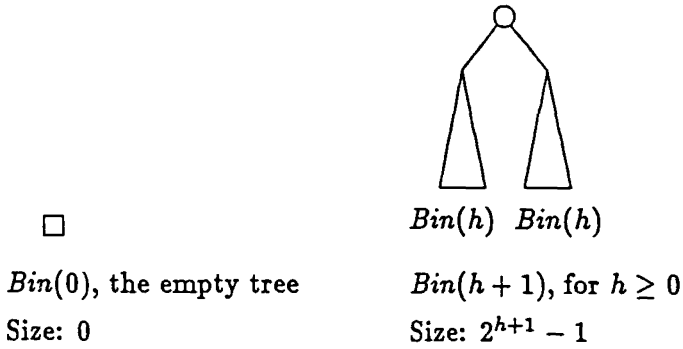


Figure 1. – A recursive definition of the perfect binary tree of height h (*Bin*(h)).

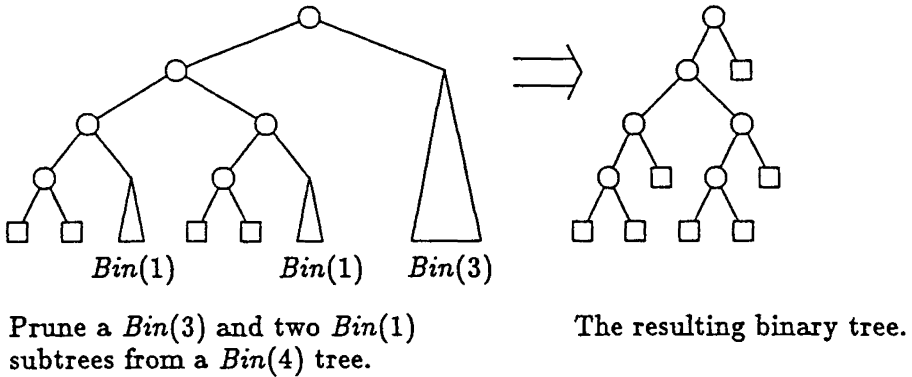


Figure 2. – Pruning a complete binary tree.

each remaining node. Again, because a perfect m -ary tree of height h contains $(m^h - 1)/(m - 1)$ nodes, we have a relationship between the number of nodes pruned from a perfect m -ary tree and sums of the form $\sum_{i \geq 0} a_i (m^{i+1} - 1)/(m - 1)$, where $a_i = 0, 1, 2, \dots$, or m ; that is, between m -ary trees and Pm representations.

In the following sections, all numbers discussed are assumed to be nonnegative integers, and we assume that m is some fixed integer greater than 1.

2. THE Pm NUMBER SYSTEM AND THE GREEDY ALGORITHM

In this section, we define the Pm number system and introduce the Pm representations obtained via the greedy algorithm.

The base m number system, for some integer $m > 1$, is based on the integer sequence $1 = m^0 < m^1 < m^2 < \dots$. If we wish to represent an integer in base m , then we use the digits $0, \dots, m-1$, and the i -th digit of a base m representation is the coefficient of m^i . (The least significant digit corresponds to index 0, and we count up from there.) We consider the pseudo- m -ary (Pm) number system, which is based on the integer sequence $1 = (m^1 - 1)/(m - 1) < (m^2 - 1)/(m - 1) < (m^3 - 1)/(m - 1) < \dots$. It uses the digits $0, \dots, m$, and the i -th digit of a Pm representation is the coefficient of $(m^{i+1} - 1)/(m - 1)$.

Thus, a Pm representation is either ε or a sequence of integers of the form $a_n \dots a_0$, where $n \geq 0$, $1 \leq a_n \leq m$, and $0 \leq a_i \leq m$, for all i , $0 \leq i < n$. The value of the Pm representation ε is 0. The value of any other Pm representation $a_n \dots a_0$ is denoted by $\text{value}(a_n \dots a_0)$ and is defined to be $\sum_{i=0}^n a_i (m^{i+1} - 1)/(m - 1)$. If we consider all non-zero Pm representations with exactly $n + 1$ digits, for some $n \geq 0$, the Pm representation consisting of a 1 digit followed by n zero digits (that is, the Pm representation 10^n , using the formal language notation 0^n to mean a string of n zeros) has the smallest value among all non-zero Pm representations with exactly $n + 1$ digits. Similarly, the Pm representation consisting of $n + 1$ digits equal to m (the Pm representation m^{n+1}) has the largest value among all non-zero Pm representations with exactly $n + 1$ digits. Thus, the value of the non-zero Pm representation $a_n \dots a_0$ is bounded by

$$\frac{m^{n+1} - 1}{m - 1} \leq \text{value}(a_n \dots a_0) \leq \frac{m}{m - 1} \left(\frac{m^{n+2} - 1}{m - 1} - n - 2 \right).$$

Now, we show that each non-negative integer is representable in the Pm number system. We will use the greedy algorithm and a result of Fraenkel [Fra85].

The greedy algorithm produces a representation $a_n \dots a_0$ (if one is possible) in a number system $1 \leq u_0 < u_1 < u_2 < \dots$ for a positive integer N as follows:

Find the largest index n such that $u_n \leq N$.

$i \leftarrow n$

Repeat

$a_i \leftarrow \lfloor N/u_i \rfloor$

$N \leftarrow N - a_i u_i$

$i \leftarrow i - 1$

Until $i = 0$.

Note that

$$\sum_{i=0}^k a_i u_i < u_{k+1}, \quad \text{for all } k, 0 \leq k \leq n,$$

because n is the largest index such that $u_n \leq N$ and because we remove as many multiples of u_i as possible from what remains of N before considering lower-order digits in the greedy representation, for all $i, 0 \leq i \leq n$. The following result of Fraenkel [Fra85] implies that, in certain number systems, the greedy representation is the only representation to satisfy

$$\sum_{i=0}^k a_i u_i < u_{k+1}, \quad \text{for all } k, 0 \leq k \leq n,$$

and that every non-negative integer has such a representation in these number systems.

PROPOSITION 2.1 (Fraenkel): *Let $1 = u_0 < u_1 < u_2 < \dots$ be any finite or infinite sequence of integers. Any non-negative integer N has precisely one representation in the system $S = \{u_0, u_1, u_2, \dots\}$ of the form $N = \sum_{i=0}^n a_i u_i$, where the a_i are non-negative integers that satisfy*

$$a_k u_k + a_{k-1} u_{k-1} + \dots + a_0 u_0 < u_{k+1} \quad (k \geq 0).$$

By Proposition 2.1, we see that, in the Pm number system, the greedy representation $a_n \dots a_0$ for a non-negative integer p is the only representation for p that satisfies

$$a_k \geq 0$$

and

$$\sum_{i=0}^k a_i \frac{m^{i+1} - 1}{m - 1} < \frac{m^{k+2} - 1}{m - 1},$$

for all $k, 0 \leq k \leq n$, and that there is such a representation for every non-negative integer.

3. THE REGULARITY OF GREEDY REPRESENTATIONS

We will show that the following regular language captures exactly the Pm representations that are produced by the greedy algorithm.

DEFINITION 3.1: Let L_G be the regular language

$$L_G = \{1, \dots, m-1\} \{0, \dots, m-1\}^* \\ + \{1, \dots, m-1\} \{0, \dots, m-1\}^* m 0^* + m 0^* + \varepsilon.$$

The regular language $\{1, \dots, m-1\} \{0, \dots, m-1\}^* + \varepsilon$ is the set of Pm representations that do not have any digit equal to m . The regular language $\{1, \dots, m-1\} \{0, \dots, m-1\}^* m 0^* + m 0^*$ is the set of Pm representations that have exactly one digit equal to m and all lower-order digits are zero. Note that if $a_n \dots a_0$ is in L_G , then $a_k \dots a_0$, where $a_k > 0$, for some $0 \leq k < n$, is also in L_G . Also, if we consider the Pm representations with exactly $n+1$ digits in L_G , then we see that the Pm representation that consists of the digit m followed by n zero digits has the largest value among them all; that is, the value of the Pm representation $a_n \dots a_0$ in L_G is bounded from above by

$$\text{value}(a_n \dots a_0) \leq \frac{m(m^{n+1} - 1)}{m - 1}.$$

Now, we will show that L_G is the set of all Pm representations produced by the greedy algorithm.

THEOREM 3.1: *The regular language L_G consists of exactly the Pm representations produced by the greedy algorithm for non-negative integers. Hence, the set of Pm representations produced by the greedy algorithm for non-negative integers is regular.*

Proof: We first show that any Pm representation not in L_G is not a greedy representation. Let $a_n \dots a_0 \in \bar{L}_G$. Then, some digit, a_k say, is m and some lower order digit, a_j say, is non-zero. That is, $a_n \dots a_0$ has the form $a_n \dots a_{k+1} m a_{k-1} \dots a_j \dots a_0$, where $a_j > 0$. By Proposition 2.1, $a_n \dots a_0$ must satisfy

$$\text{value}(a_n \dots a_0) = \sum_{i=0}^r a_i \frac{m^{i+1} - 1}{m - 1} < \frac{m^{r+2} - 1}{m - 1},$$

for all r , $0 \leq r \leq n$, if $a_n \dots a_0$ is the representation produced by the greedy algorithm for value $(a_n \dots a_0)$. We now show that $a_n \dots a_0$ violates that

condition at $r=k$. Consider

$$\begin{aligned} \text{value}(a_k \dots a_0) &\geq \text{value}(m0^{k-j-1}a_j0^j) \\ &\geq \text{value}(m0^{k-1}1) \\ &= \frac{m^{k+2}-1}{m-1}. \end{aligned}$$

Therefore, $a_n \dots a_0$ is not the representation produced by the greedy algorithm for value $(a_n \dots a_0)$.

Now we show that any Pm representation in L_G is produced by the greedy algorithm for the corresponding number. We show that every Pm representation $a_n \dots a_0$ in L_G satisfies $a_i \geq 0$ and

$$\sum_{j=0}^i a_j \frac{m^{j+1}-1}{m-1} < \frac{m^{i+2}-1}{m-1},$$

for all $i \geq 0$, thus proving, by Proposition 2.1, that every element of L_G is a greedy representation. Let $a_n \dots a_0$ be in L_G . Clearly, $a_k \geq 0$, for all k , $0 \leq k \leq n$. Also, for any k , $0 \leq k \leq n$, the Pm representation $a_k \dots a_0$ (ignoring leading zeros) is in L_G . Now, the value of the Pm representation $a_k \dots a_0$ in L_G is bounded from above by value $(a_k \dots a_0) \leq m(m^{k+1}-1)/(m-1)$. But

$$m(m^{k+1}-1)/(m-1) < (m^{k+2}-1)/(m-1),$$

so value $(a_k \dots a_0) < (m^{k+2}-1)/(m-1)$, as required. \square

4. THE Pm REPRESENTATIONS OF THE UNAMBIGUOUS NUMBERS

There are many Pm representations that are not in L_G ; namely, all those that have a digit equal to m and some other lower-order non-zero digit. Since the value of a Pm representation that is not in L_G is also the value of some Pm representation that is in L_G , the numbers corresponding to Pm representations that are not in L_G are ambiguous. For example, the Pm representation $m0^{n-2}1$, which is not in L_G , has value $m(m^n-1)/(m-1)+1 = (m^{n+1}-1)/(m-1)$ and so does the Pm representation 10^n , which is in L_G . Thus, the number $(m^{n+1}-1)/(m-1)$ is ambiguous in the Pm number system. We prove that the set of numbers that are unambiguous in the Pm number system is a regular set.

DEFINITION 4.1: Let L_U be the regular language

$$L_U = \{1, \dots, m-1\}^+ [0m + \{1, \dots, m-1\} \{0, \dots, m\} + m0] + \{\varepsilon, 1, \dots, m0\}.$$

Thus, L_U contains all Pm representations of length at most two that fall, in lexicographic order, between (and including) ε and $m0$, and L_U contains all Pm representations $a_n \dots a_0$, for $n \geq 2$, such that the last two digits $a_1 a_0$ fall, in lexicographic order, between (and including) $0m$ and $m0$, and $0 < a_i < m$, for all i , $2 \leq i \leq n$.

Clearly, L_U is a subset of L_G ; that is, a Pm representation in L_U has at most one digit equal to m , and, if it has a digit equal to m , then all lower-order digits are zero. Furthermore, if $a_n \dots a_0$ is in L_U , for some $n \geq 2$, then $a_{n-1} \dots a_0$ is in L_U , too.

We will show that the Pm representations in L_U are exactly the Pm representations of the unambiguous numbers. To do this, we first show that no two Pm representations in $\{0, 1, \dots, mm\}$ have the same value and then we bound the values of the Pm representations in L_U .

LEMMA 4.1: Let S_2 be the set of Pm representations with one or two digits; that is, let $S_2 = \{\varepsilon, 1, \dots, mm\}$. If x and y are in S_2 and $x \neq y$, then $\text{value}(x) \neq \text{value}(y)$.

Proof: For convenience, we treat all Pm representations in S_2 as if they have two digits, by adding leading zeros if necessary. Let $a_1 a_0$ and $b_1 b_0$ be in S_2 and let $a_1 a_0 \neq b_1 b_0$. There are two cases to consider: either $a_1 \neq b_1$, or $a_1 = b_1$ and $a_0 \neq b_0$.

If $a_1 \neq b_1$, then assume, without loss of generality, that $a_1 < b_1$. Consider $\text{value}(a_1 a_0) = a_1(m^2 - 1)/(m - 1) + a_0$. Since $a_1 b_1$ and $a_0 \leq m$, we have $\text{value}(a_1 a_0) \leq (b_1 - 1)(m^2 - 1)/(m - 1) + m = b_1(m^2 - 1)/(m - 1) - 1$. Since $b_0 \geq 0$, we have $\text{value}(a_1 a_0) < b_1(m^2 - 1)/(m - 1) + b_0 = \text{value}(b_1 b_0)$.

If $a_1 = b_1$ and $a_0 \neq b_0$, assume, without loss of generality, that $a_0 < b_0$. Then, $\text{value}(a_1 a_0) \leq b_1(m^2 - 1)/(m - 1) + (b_0 - 1) < b_1(m^2 - 1)/(m - 1) + b_0 = \text{value}(b_1 b_0)$.

In both cases, $\text{value}(a_1 a_0) \neq \text{value}(b_1, b_0)$. \square

Note that this result does not establish the unambiguity of the numbers with representations in S_2 because it does not consider Pm representations with more than two digits. Indeed, the numbers corresponding to some

two-digit Pm representations are ambiguous. For example, the number $m(m^2 - 1)/(m - 1) + 1$ is represented in the Pm number system by $m1$ and 100 .

LEMMA 4.2: *Let $a_n \dots a_0$ be in L_U . If $n < 2$, then*

$$0 \leq \text{value}(a_n \dots a_0) \leq m(m^2 - 1)/(m - 1).$$

Otherwise,

$$\frac{1}{m - 1} \left(\frac{m^{n+2} - 1}{m - 1} - 2m - n \right) \leq \text{value}(a_n \dots a_0) \leq \frac{m^{n+2} - 1}{m - 1} - n.$$

Proof: The set of Pm representations in L_U with zero, one, or two digits is $L_U(2) = \{\varepsilon, 1, \dots, m0\}$ and this set consists of all Pm representations of length at most two that fall, in lexicographic order, between (and including) ε and $m0$. By Lemma 4.1, no two of these representations have the same value. If we list the elements of $L_U(2)$ in lexicographic order, their values are strictly increasing. To see this, consider the Pm representation that comes after $a_1 a_0$ (we add leading zeros as necessary to obtain two digits). If $a_0 < m$, then the next representation is $a_1(a_0 + 1)$ and

$$\begin{aligned} \text{value}(a_1 a_0) &= a_1(m^2 - 1)/(m - 1) + a_0 \\ &< a_1(m^2 - 1)/(m - 1) + (a_0 + 1) \\ &= \text{value}(a_1(a_0 + 1)). \end{aligned}$$

If $a_0 = m$, then the next representation is $(a_1 + 1)0$ and

$$\begin{aligned} \text{value}(a_1 a_0) &= a_1(m^2 - 1)/(m - 1) + m \\ &< a_1(m^2 - 1)/(m - 1) + (m + 1) \\ &= (a_1 + 1)(m^2 - 1)/(m - 1) \\ &= \text{value}((a_1 + 1)0). \end{aligned}$$

Therefore, if $a_1 a_0 \in L_U$, then

$$\text{value}(\varepsilon) = 0 \leq \text{value}(a_1 a_0) \leq m(m^2 - 1)/(m - 1) = \text{value}(m0).$$

If $a_n \dots a_0 \in L_U$ and $n \geq 2$, then, by similar arguments about the last two digits of this representation,

$$\text{value}(a_n \dots a_2 0m) \leq \text{value}(a_n \dots a_0) \leq \text{value}(a_n \dots a_2 m0).$$

If some $a_i > 1$, where $2 \leq i \leq n$, then we can subtract 1 from a_i to create a Pm representation in L_U with smaller value than $\text{value}(a_n \dots a_0)$. Thus, value

$(1^{n-1}0m) \leq \text{value}(a_n \dots a_0)$, where 1^{n-1} represents a string of $n-1$ ones. Similarly, if some $a_i < m-1$, where $2 \leq i \leq n$, then we can add 1 to a_i to create another Pm representation in L_U with greater value than $\text{value}(a_n \dots a_0)$. Thus, $\text{value}(a_n \dots a_0) \leq \text{value}((m-1)^{n-1}m0)$, where $(m-1)^{n-1}$ represents a string of $m-1$'s of length $n-1$. \square

THEOREM 4.3: *A number is unambiguous in the Pm number system if and only if it has a representation in L_U . Hence, the set of Pm representations of unambiguous numbers is regular.*

Proof: We split the proof into two parts.

CLAIM 1: Each Pm representation in L_U is the only Pm representation with its value.

Clearly, ε is the only Pm representation for θ . Consider the Pm representations $a_n \dots a_0$ in L_U with positive values. The proof is by induction on n .

BASIS: The set $L_U(2) = \{1, \dots, m0\}$ contains the only Pm representations in L_U , for $n=0$ and $n=1$. Any Pm representation with three or more digits has value at least $\text{value}(100) = (m^3-1)/(m-1)$. The Pm representations $m1, m2, \dots, mm$ (the only Pm representations with at most two digits that are not in $L_U(2)$) have values at least $\text{value}(m1) = (m^3-1)/(m-1)$. By Lemma 4.2, a Pm representation in $L_U(2)$ has value at most $m(m^2-1)/(m-1) < (m^3-1)/(m-1)$, for all $m > 1$. By Lemma 4.1, no two of the Pm representations in $L_U(2)$ have the same value. Therefore, each representation in $L_U(2)$ is unambiguous.

INDUCTION HYPOTHESIS: Assume that each Pm representation $a_k \dots a_0$ in L_U is the only Pm representation for $\text{value}(a_k \dots a_0)$, for all $k < n$, for some $n > 1$.

INDUCTION STEP: Let $a_n \dots a_0$ be a Pm representation in L_U . Assume that there exists some other Pm representation $b_k \dots b_0$ (not necessarily in L_U or L_G) with the same value. There are three possibilities: either $k > n$, $k < n$, or $k = n$.

$k > n$. Since $L_U \subseteq I_G$, the representation $a_n \dots a_0$ is the greedy representation for $\text{value}(a_n \dots a_0)$. The greedy representation of a number has the longest length of any representation of that number, so this case cannot occur.

$k < n$. We show that the difference $\text{value}(b_k \dots b_0) - \text{value}(a_{n-1} \dots a_0)$ is not the same as the difference $\text{value}(a_n \dots a_0) - \text{value}(a_{n-1} \dots a_0)$.

$$\text{value}(a_n \dots a_0) - \text{value}(a_{n-1} \dots a_0) = a_n(m^{n+1} - 1)/(m - 1).$$

Thus, the Pm representations $a_n \dots a_0$ and $b_k \dots b_0$ cannot have the same value, a contradiction.

Since $b_k \dots b_0$ is a Pm representation of length $k + 1$,

$$\begin{aligned} \text{value}(b_k \dots b_0) &\leq \frac{m}{m-1} \left(\frac{m^{k+2} - 1}{m-1} - k - 2 \right) \\ &\leq \frac{m}{m-1} \left(\frac{m^{n+1} - 1}{m-1} - n - 1 \right). \end{aligned}$$

Furthermore, since $a_{n-1} \dots a_0 \in L_U$, by Lemma 4.2,

$$\frac{1}{m-1} \left(\frac{m^{n+1} - 1}{m-1} - 2m - n + 1 \right) \leq \text{value}(a_{n-1} \dots a_0).$$

Therefore,

$$\begin{aligned} &\text{value}(b_k \dots b_0) - \text{value}(a_{n-1} \dots a_0) \\ &\leq \frac{m}{m-1} \left(\frac{m^{n+1} - 1}{m-1} - n - 1 \right) - \frac{1}{m-1} \left(\frac{m^{n+1} - 1}{m-1} - 2m - n + 1 \right) \\ &= \frac{m^{n+1} - 1}{m-1} - (n-1) \\ &< \frac{m^{n+1} - 1}{m-1}, \end{aligned}$$

since $m > 1$ and $n \geq 2$. Thus,

$$\text{value}(b_k \dots b_0) - \text{value}(a_{n-1} \dots a_0) \neq a_n(m^{n+1} - 1)/(m-1),$$

a contradiction.

$k = n$. We know that $a_n \dots a_0$ is in $L_U \subseteq L_G$; that is, $a_n \dots a_0$ is produced by the greedy algorithm when it is given value $(a_n \dots a_0)$. Therefore, $a_n = \lfloor \text{value}(a_n \dots a_0)(m-1)/(m^{n+1} - 1) \rfloor$. But this implies that b_n cannot be larger than a_n ; that is, $b_n \leq a_n$.

Suppose $b_n = a_n$. Then, $b_{n-1} \dots b_0$ is not equal to $a_{n-1} \dots a_0$ and $\text{value}(b_{n-1} \dots b_0) = \text{value}(a_{n-1} \dots a_0)$. Now, $a_{n-1} \dots a_0$ is in L_U and, by the induction hypothesis, it is the only Pm representation for value $(a_{n-1} \dots a_0)$. Therefore, we must have $b_{n-1} \dots b_0 = a_{n-1} \dots a_0$, a contradiction.

Now, if $b_n < a_n$, then the Pm representations $b_{n-1} \dots b_0$ and $(a_n - b_n)a_{n-1} \dots a_0$ are two different Pm representations with the same value.

Since $a_n - b_n > 0$ and $a_n \dots a_0$ is in L_U , the Pm representation $(a_n - b_n)a_{n-1} \dots a_0$ is also in L_U . We have already shown above that we cannot have some Pm representation $(a_n - b_n)a_n \dots a_0$ in L_U and some other Pm representation $b_{n-1} \dots b_0$ such that $\text{value}(a_n \dots a_0) = \text{value}(b_k \dots b_0)$. Thus, this case is not possible either.

Each possibility leads to a contradiction; therefore, our assumption that there exists some other Pm representation $b_k \dots b_0$ that has the same value as $a_n \dots a_0 \in L_U$ must be false. Thus, each Pm representation in L_U is the only Pm representation with the corresponding value.

CLAIM 2: Each number that does not have a representation in L_U is ambiguous.

Suppose the Pm representation $a_n \dots a_0$ is not in L_U . We construct another Pm representation for value $(a_n \dots a_0)$ to show that $\text{value}(a_n \dots a_0)$ is ambiguous. There are two cases to consider: either $a_n \dots a_0$ is in L_G or $a_n \dots a_0$ is not in L_G .

If $a_n \dots a_0$ is not in L_G , then, by Theorem 3.1, there exists some Pm representation $b_k \dots b_0$ in L_G such that $\text{value}(b_k \dots b_0) = \text{value}(a_n \dots a_0)$. Thus, $\text{value}(a_n \dots a_0)$ is ambiguous.

Otherwise, $a_n \dots a_0$ is in L_G and we use the equality

$$\frac{m^{k+1} - 1}{m - 1} = m \frac{m^k - 1}{m - 1} + 1$$

to build another Pm representation with the same value as $a_n \dots a_0$. There are two subcases to consider: either there are digits $a_{j-1} = 0$ and $a_j > 0$, for some j , $2 < j \leq n$, or there are not.

Two such digits, a_{j-1} and a_j , exist. If $a_0 = m$, then, by the definition of L_G , since a_0 is non-zero, a_1 cannot be m . Consider the Pm representation $b_n \dots b_0$, where

$$\begin{aligned} b_j &= a_j - 1, \\ b_{j-1} &= a_{j-1} + m = m, \\ b_1 &= a_1 + 1, \\ b_0 &= 0, \text{ and} \\ b_i &= a_i, \text{ otherwise.} \end{aligned}$$

Since $j > 2$, we have not defined digit b_1 twice, so,

$$\begin{aligned} \text{value}(b_n \dots b_0) &= \text{value}(a_n \dots a_0) - \frac{m^{j+1} - 1}{m - 1} \\ &\quad + m \frac{m^j - 1}{m - 1} + \frac{m^2 - 1}{m - 1} - m \\ &= \text{value}(a_n \dots a_0). \end{aligned}$$

If $a_0 < m$, consider the Pm representation $b_n \dots b_0$, where

$$\begin{aligned} b_j &= a_j - 1, \\ b_{j-1} &= a_{j-1} + m = m, \\ b_0 &= a_0 + 1, \text{ and} \\ b_i &= a_i, \text{ otherwise.} \end{aligned}$$

We have

$$\begin{aligned} \text{value}(b_n \dots b_0) &= \text{value}(a_n \dots a_0) - \frac{m^{j+1} - 1}{m - 1} \\ &\quad + m \frac{m^j - 1}{m - 1} + 1 \\ &= \text{value}(a_n \dots a_0). \end{aligned}$$

Thus, $\text{value}(a_n \dots a_0)$ is ambiguous.

Two such digits, a_{j-1} and a_j , do not exist. Then, either $n \leq 2$, or $n > 2$ and $a_j > 0$, for all j , $2 \leq j \leq n$.

Let us first consider $n \leq 2$. (We add leading zeros as required to make all representations under consideration exactly three digits long.) Since $a_2 a_1 a_0$ is in L_G , if any digit is m , then all lower-order digits are zero. Since $a_2 a_1 a_0$ is not in L_U , either $a_2 = m$ or $0 < a_2 < m$ and $a_1 a_0$ is in $\{00, 01, \dots, 0(m-1), m1, m2, \dots, mm\}$ or $a_2 = 0$ and $a_1 a_0$ is in $\{m1, m2, \dots, mm\}$. Combining these two restrictions, we see that if $n \leq 2$, then $a_2 a_1 a_0$ is in $\{m00\} + \{1, 2, \dots, m-1\} \{00, 01, \dots, 0(m-1)\}$. Since $a_2 > 0$, $a_1 = 0$, and $a_0 < m$ in each case, the representation $(a_2 - 1)m(a_0 + 1)$ is

a valid Pm representation and

$$\begin{aligned} \text{value}((a_2 - 1)m(a_0 + 1)) &= \text{value}(a_2 a_1 a_0) - \frac{m^3 - 1}{m - 1} \\ &\quad + m \frac{m^2 - 1}{m - 1} + 1 \\ &= \text{value}(a_2 a_1 a_0). \end{aligned}$$

Now let us consider $n > 2$ and $a_j > 0$, for all j , $2 \leq j \leq n$. Since $a_n \dots a_0$ is in L_G , if any digit is in m , then all lower-order digits are zero. Thus, since $a_j > 0$, for all $2 \leq j \leq n$, we have $a_j \neq m$, for all j , $2 < j \leq n$. Since $a_n \dots a_0$ is not in L_U , either $a_2 = m$ (in which case $a_1 a_0 = 00$, since $a_n \dots a_0$ is in L_G), or $0 < a_2 < m$ and $a_1 a_0 \notin [0m + \{1, \dots, m-1\}\{0, \dots, m\} + m0]$ (in which case $a_1 a_0 \in \{00, 01, \dots, 0(m-1)\}$, since $a_n \dots a_0$ is in L_G). Since $a_2 > 0$, $a_1 = 0$, and $a_0 < m$ in each case, the representation $a_n \dots a_3 (a_2 - 1)m(a_1 + 1)$ is a valid Pm representation and

$$\begin{aligned} \text{value}(a_n \dots a_3 (a_2 - 1)m(a_1 + 1)) &= \text{value}(a_n \dots a_0) - \frac{m^{3+1} - 1}{m - 1} \\ &\quad + m \frac{m^2 - 1}{m - 1} + 1 \\ &= \text{value}(a_n \dots a_0). \end{aligned}$$

Thus, once again $\text{value}(a_n \dots a_0)$ is ambiguous.

Therefore, each number that does not have a Pm representation in L_U is ambiguous. \square

5. CONCLUSION

We have characterized the set of Pm representations that are constructed by the greedy algorithm and the set of numbers that are unambiguous in the Pm number system and shown that these are regular sets.

One question that we have not answered is whether we need all the digits $0, 1, \dots, m$. For instance, if we are not allowed to use the digit m , would some integer have no Pm representation? We see that L_U uses all the digits from $\{0, 1, \dots, m\}$ and each number with a representation in L_U has only one Pm representation. Thus, we need all the digits $0, 1, \dots, m$, if all non-negative integers are to be represented. This observation leaves an open

problem: Characterize the integers that have Pm representations if the digit set is restricted to some subset of $\{0, 1, \dots, m\}$.

As noted in the introduction, another more general problem that remains is: Characterize the number systems for which the set of greedy representations and the set of representations of unambiguous numbers are regular.

REFERENCES

- [ABS89] J.-P. ALLOUCHE, J. BETREMA and J. O. SHALLIT, Sur des points fixes de morphismes d'un monoïde libre, *Informatique Théorique et Applications/Theoretical Informatics and Applications*, 23, (3), 1989, pp. 235-249.
- [Cam91] H. A. CAMERON, *Extremal Cost Binary Trees*, Ph. D. thesis, University of Waterloo, 1991.
- [CW93] H. A. CAMERON and D. WOOD, The maximal path length of binary trees, *Discrete Applied Mathematics*, to appear, 1993.
- [Fra85] A. S. FRAENKEL, Systems of numeration, *The American Mathematical Monthly*, 92, (2), 1985, pp. 105-114.
- [Sha91] J. SHALLIT, *Numeration systems, linear recurrences, and regular sets*, Technical Report CS-91-32, University of Waterloo, July 1991.