Annales Academiæ Scientiarum Fennicæ
Series A. I. Mathematica
Volumen 14, 1989, 149–156

# ON NUMBER SYSTEMS WITH NEGATIVE DIGITS

**Juha Honkala**

**Abstract.** We study arbitrary number systems which allow negative digits. We show that the set of nonnegative integers represented by a number system $N = (n, m_1, \ldots, m_v)$ is $n$-recognizable. Using the theory of $k$-recognizable sets, we prove that the equivalence problem for number systems is decidable. We show that the degree of ambiguity of a given number system can be effectively computed.

## 1. Introduction

Recent work in the theory of codes and $L$ codes (see Maurer et al. (1983)) has led to problems relating to the representation of positive integers in arbitrary number systems. Here "arbitrary" means that the digits may be larger than the base and that completeness is not required, i.e., every integer need not have a representation in the system.

Because we do not require completeness, our work differs from the classical theory of number systems (see Knuth (1981) and Matula (1982)). Thus, classical results do not seem to be applicable. However, also this research area is well motivated for several reasons. The main motivation comes from $L$ codes and related topics in language and automata theory. Undoubtedly, the study of general number systems also increases our understanding of number systems in general. For further motivation, the reader is referred to Maurer et al. (1983) and Culik–Salomaa (1983).

Some basic facts about number systems were established in Culik–Salomaa (1983). A natural generalization is to consider number systems having negative digits, too. This generalization was briefly discussed in Culik–Salomaa (1983). This paper continues the work. We study ambiguity, bases and equivalence of number systems.

The reader is assumed to know the basic facts concerning gsm-mappings (see Salomaa (1973)) and $k$-recognizable sets (see Eilenberg (1974)).

## 2. Definitions and examples

By a *number system* we mean a $(v+1)$-tuple $N = (n, m_1, \ldots, m_v)$ of integers such that $v \geq 1$, $n \geq 2$ and $m_1 < m_2 < \cdots < m_v$. The number $n$ is referred to as the *base* and the numbers $m_i$ as the *digits* of the number system $N$. If $m_1 \geq 0$,

doi:10.5186/aasfm.1989.1411

then $N$ is called a *positive* number system. If 0 is not a digit of $N$, then $N$ is called 0-*free*. We denote $(n, -m_v, -m_{v-1}, \ldots, -m_1)$ by $-N$.

A nonempty word

$$m_{i_k} m_{i_{k-1}} \cdots m_{i_1} m_{i_0}, \qquad 1 \leq i_j \leq v, \tag{1}$$

over the alphabet $\{m_1, \ldots, m_v\}$ is said to *represent* the integer

$$[m_{i_k} \cdots m_{i_0}] = m_{i_k} \cdot n^k + m_{i_{k-1}} \cdot n^{k-1} + \cdots + m_{i_1} \cdot n + m_{i_0}. \tag{2}$$

The word (1) is said to be a *representation* of the integer (2). The set of all integers represented by $N$ is denoted by $S(N)$. We denote by $\operatorname{Pos} S(N)$ the set

$$S(N) \cap \{0, 1, 2, \ldots\}$$

and by $\operatorname{Neg} S(N)$ the set

$$S(N) \cap \{0, -1, -2, \ldots\}.$$

Whenever in $S(N)$, zero is included in $\operatorname{Pos} S(N)$. (Otherwise, zero should be treated separately in each case.) The same applies to $\operatorname{Neg} S(N)$. It is convenient to include zero in both sets, e.g., in the proof of Corollary 3.2.

A set $A$ of integers is said to be *representable by a number system*, RNS for short, if there exists a number system $N$ such that $A = S(N)$. An integer $n$ is called a *base of an* RNS *set* $A$ if there is a number system with base $n$ representing $A$.

The degree of ambiguity of a number system is defined as follows: A number system $N$ is *ambiguous of degree* $m \geq 1$ if at least one integer has $m$ distinct representations and no integer has more than $m$ representations. If no such $m$ exists, $N$ is ambiguous of degree $\infty$. A number system is called *unambiguous* (or *ambiguous*) if its degree of ambiguity is 1 (or $> 1$).

An RNS set is termed *unambiguous* if it is represented by an unambiguous number system. Otherwise it is termed *inherently ambiguous*.

**Example 2.1.** The only finite RNS set is $\{0\}$.

**Example 2.2.** We show that $\mathbf{Z} \setminus \{0\}$ is not representable by any number system.

Assume, on the contrary, that $\mathbf{Z} \setminus \{0\} = S(N)$ for a number system $N = (n, m_1, \ldots, m_v)$. Then there exists an index $i$ such that $m_i \equiv 0$ modulo $n$. Thus $m_i/n \in \mathbf{Z} \setminus \{0\}$ and $-m_i/n = a_0 + a_1 \cdot n + \cdots + a_k \cdot n^k$, where $k \geq 0$ and $a_0, a_1, \ldots, a_k$ are digits of $N$. Hence, $0 = m_i + a_0 \cdot n + a_1 \cdot n^2 + \cdots + a_k \cdot n^{k+1}$, which is impossible.

This example shows that there are cofinite subsets of $\mathbf{Z}$ which are not representable. On the contrary, it is known that every cofinite set of positive integers is representable by a positive 0-free number system (see Culik–Salomaa (1983)).

**Example 2.3.** If $A$ is an RNS set and 0 is an element of $A$, then $A$ is inherently ambiguous. Thus, for example, the sets $\mathbf{Z} = S(2,-1,0,1)$ and $2\mathbf{Z} = S(2,-2,0,2)$ are inherently ambiguous. (We omit double parenthesis in the notation.)

It is an open question whether or not the inherent ambiguity of RNS sets is decidable.

## 3. Pos $S(N)$ is $n$-recognizable

In this section we prove a generalization of the translation lemma of Culik–Salomaa (1983).

Let $w = d_0 d_1 \cdots d_m$ be a word over the alphabet $\{0,1,\ldots,k-1\}$, where it is assumed that each $d_i$ is a letter and that $k \geq 2$. Denote $\sum_{i=0}^{m} d_i k^{m-i}$ by $\nu_k(w)$. The notation $\nu_k$ is extended to languages in the obvious way. A subset $A$ of the set of nonnegative integers is *$k$-recognizable* if there exists a regular language $L$ over the alphabet $\{0,1,\ldots,k-1\}$ such that $A = \nu_k(L)$. A subset $A$ of the set of nonnegative integers is *recognizable* if $A$ is a finite union of arithmetic progressions.

**Theorem 3.1.** *The set* Pos $S(N)$ *is $n$-recognizable for every number system* $N = (n, m_1, \ldots, m_v)$.

*Proof.* Let $t$ denote $\max\{|m_1|, m_v\}$. When $x$ is an integer and $x = a + bn$, where $0 \leq a < n$, denote $a$ by $f(x)$ and $b$ by $g(x)$.

We construct a generalized sequential machine $M$ as follows. The state set of $M$ is $\{q_{-t}, q_{-t+1}, \ldots, q_0, q_1, q_2, \ldots, q_t, q_F\}$, where $q_0$ is the initial state and $q_F$ is the only final state. The input alphabet is $\{m_1, \ldots, m_v, \#\}$ and the output alphabet $\{0,1,\ldots,n-1\}$. The transitions are (we use the formalism of Salomaa (1973))

$$\begin{aligned} q_j m_i &\to f(j+m_i) q_{g(j+m_i)} && (i = 1,\ldots,v; j = -t, -t+1, \ldots, t) \\ q_j \# &\to w_j q_F && (j = 0,\ldots,t), \end{aligned}$$

where $w_j$ is the reversed $n$-ary representation of $j$. (If $j = 0$, then $w_j = \lambda$.) The transitions are well defined because

$$g(j+m_i) = \frac{j + m_i - f(j+m_i)}{n} \tag{1}$$

and because the right-hand side of (1) is at most $(t + t - 0)/n \leq t$, since $n \geq 2$, and at least $\big(-t - t - (n-1)\big)/n \geq -t - (n-1)/n > -t - 1$.

Assume that

$$m_{i_0} m_{i_1} \cdots m_{i_k} \# \tag{2}$$

is a word over the input alphabet of $M$ and that

$$q_0 m_{i_0} m_{i_1} \cdots m_{i_k} \# \Rightarrow^* a_0 a_1 \cdots a_k q_i \# \tag{3}$$

is a computation of $M$, where $a_0, a_1, \ldots, a_k \in \{0, 1, \ldots, n-1\}$. $M$ accepts (2) exactly when $i \geq 0$ in (3). By the definition of $M$, $[m_{i_k} m_{i_{k-1}} \cdots m_{i_0}]$ equals $a_0 + a_1 \cdot n + \cdots + a_k \cdot n^k + i \cdot n^{k+1}$. This is easily seen by an inductive argument. Furthermore, $i \geq 0$ if and only if $a_0 + a_1 \cdot n + \cdots + a_k \cdot n^k + i \cdot n^{k+1} \geq 0$, since $a_0 + a_1 \cdot n + \cdots + a_k \cdot n^k \leq (n-1)(1 + n + \cdots + n^k) = n^{k+1} - 1$. Hence $M$ accepts (2) exactly when $[m_{i_k} m_{i_{k-1}} \cdots m_{i_0}] \geq 0$.

Thus $\nu_n(miM(\{m_1, \ldots, m_v\}^+ \#)) = \operatorname{Pos} S(N)$. □

If $A$ is a subset of $\mathbf{Z}$, let $-A$ denote $\{-x | x \in A\}$.

**Corollary 3.2.** *The set* $-\operatorname{Neg} S(N)$ *is* $n$*-recognizable for every number system* $N = (n, m_1, \ldots, m_v)$.

*Proof.* The assertion follows from Theorem 3.1, because $-\operatorname{Neg} S(N) = \operatorname{Pos} S(-N)$. □

**Remark 3.1.** Theorem 3.1 and the results established later on can be generalized to the case of a negative base.

## 4. Corollaries

We say that integers $k > 0$ and $l > 0$ are *multiplicatively dependent* if $k^p = l^q$ for some integers $p > 0$, $q > 0$. Otherwise $k$ and $l$ are said to be *multiplicatively independent*.

For a proof of the following theorem, see Cobham (1969).

**Theorem 4.1.** *If* $k$ *and* $l$ *are multiplicatively independent, every set which is both* $k$*- and* $l$*-recognizable is recognizable.*

Theorem 3.1, Corollary 3.2 and Theorem 4.1 imply the following result.

**Corollary 4.2.** *Let* $N = (n, m_1, \ldots, m_v)$ *be a number system. Let* $m$ *be the least positive integer such that* $n = m^t$ *for some positive integer* $t$. *If either* $\operatorname{Pos} S(N)$ *or* $-\operatorname{Neg} S(N)$ *is not recognizable, every base of* $S(N)$ *is of the form* $m^k$, $k > 0$.

It is decidable whether or not $\operatorname{Pos} S(N)$ is recognizable (see Honkala (1986)).

For a positive 0-free number system $N = (n, m_1, \ldots, m_v)$, the set $S(N)$ has only the base $n$ if $S(N)$ has arbitrarily long gaps (Honkala (1984)). This result does not hold for arbitrary positive number systems because, by Lemma 4.3 below, the set $S(3, 0, 1)$ has arbitrarily long gaps and has bases $3^k$, $k > 0$.

**Lemma 4.3.** *Let* $N = (n, m_1, \ldots, m_v)$ *be a number system having the digit* 0. *Then the set* $S(N)$ *possesses at least the bases* $n^k$, $k > 0$.

*Proof.* Let $N_k$ denote $(n^k, a_1, \ldots, a_u)$, where $\{a_1, \ldots, a_u\} = \{b_0 + b_1 \cdot n + \cdots + b_{k-1} \cdot n^{k-1} \mid b_0, b_1, \ldots, b_{k-1}$ are digits of $N\}$. Clearly $S(N) = S(N_k)$. □

**Example 4.1.** In this example we show that the positive part of an RNS set is not always representable by a positive number system. (A trivial counterexample is obtained with $\operatorname{Pos} S(N) = \emptyset$.)

Let $N = (3, -1, 2)$. Then $\operatorname{Pos} S(N) = \{2, 5, 8, 14, 17, 23, 26, 41, 44, 50, 53, \ldots\}$. Because $\operatorname{Pos} S(N)$ has arbitrarily long gaps, the set $\operatorname{Pos} S(N)$ is not recognizable.

Suppose $\operatorname{Pos} S(N) = S(N_1)$ for a positive number system $N_1 = (n, m_1, \ldots, m_v)$. Then $n = 3^k$ for a positive integer $k$. Assume first that $k > 1$. Then $N_1$ must have the digits 2, 5 and 8. This is impossible because it is easily seen that for every $p > 2$ the closed interval $[p, p + 6]$ contains at most two elements of $S(N)$. Thus $k = 1$. Then $N_1$ must have the digits 2 and 5. But this is also impossible because $5 + 2 \cdot 3 = 11$ is not an element of $S(N)$.

The following lemma is a straightforward generalization of the same result for positive 0-free number systems (see Maurer et al. (1983), Honkala (1984)).

**Lemma 4.4.** *Let $N = (n, m_1, \ldots, m_v)$ be a number system. $N$ is unambiguous if the digits $m_i$ lie in different residue classes modulo $n$ and $0 \notin S(N)$. If $v > n$, $N$ is ambiguous of degree $\infty$.*

**Theorem 4.5.** *The degree of ambiguity of a given number system $N = (n, m_1, \ldots, m_v)$ can be effectively computed.*

*Proof.* Let $M_1$ be the generalized sequential machine constructed in the proof of Theorem 3.1 and $M_2$ be the generalized sequential machine constructed in the proof of Corollary 3.2. Let $M_i$ have $u_i$ states, $i = 1, 2$.

**Claim 1.** *It is decidable whether or not there is an integer having at least two representations of the same length.*

*Proof.* We decide first whether or not there are two words of equal length mapped into the same word by $M_1$. We claim that such words exist if and only if there exist such words of length at most $u_1^2 + 1$. This condition is clearly decidable.

To establish our claim, we let $w_1\#$ and $w_2\#$ be words of equal and minimal length mapped into the same word by $M_1$. Let $p_{1i}, p_{2i}, \ldots$ be the states $M_1$ is brought to when reading the word $w_i$ letter by letter. If $|w_1| = |w_2| > u_1^2$, there are $j_1$ and $j_2$ such that $1 < j_1 < j_2$ and $(p_{j_1,1}, p_{j_1,2}) = (p_{j_2,1}, p_{j_2,2})$. This means that we can remove from each $w_i$ every letter between and including the $(j_1+1)$st letter and $j_2$th letter, and the resulting words $w_i'\#$ are still mapped into the same word by $M_1$. Because the first letters of $w_1$ and $w_2$ are different, the words $w_1'$ and $w_2'$ are different. This contradiction establishes our claim.

In the same way we can decide whether or not there are two words of equal length mapped into the same word by $M_2$. □

If there is an integer having two representations of the same length, the degree of $N$ is $\infty$. In what follows we suppose that no integer has two representations of the same length. (This assumption is needed in the proof of Claim 2.)

**Claim 2.** *Given $k \in \mathbf{N}$, it is decidable whether or not there exists an integer having at least $k$ representations.*

*Proof.* Proceeding as in the proof of Claim 1 (see Honkala (1984), p. 67), we see there to be $k$ words $w_1\#, \ldots, w_k\#$ such that the translations of $w_1\#, \ldots, w_k\#$ are the same when zeros in the ends of the translations are disregarded if and only if there are $k$ words with the mentioned property such that the length of one of them is at most $u_1^k + 1$. This condition is shown to be decidable after the proofs of two more claims. (At this stage we could show the decidability using inverse gsm mappings. Below, however, we get a faster decision method with no extra effort.)

The same holds when $M_1$ is replaced by $M_2$. □

Denote $K = \max |m_i|$. Define $A_n = \{x \in S(N) \mid x$ has a representation of length $n\}$ and $B_n = A_n \cap \{-K, -K+1, \ldots, 0, \ldots, K\}$, $n = 1, 2, \ldots$.

**Claim 3.** *It is decidable whether or not there is an integer $x$ such that $|x| \leq K$ and $x$ has infinitely many representations.*

*Proof.* We first show that $B_i = B_j$, $i < j$, implies $B_{i+1} = B_{j+1}$.

Suppose $B_i = B_j$, $i < j$. Let $z = z_1 + nz_2$ be an element of $B_{i+1}$, where $z_1$ is a digit and $z_2 \in A_i$. Hence $|z| \leq K$. If $|z_2|$ were greater than $K$, we would have $z_1 + nz_2 > m_1 + K \cdot n \geq K$ or $z_1 + nz_2 < m_v - K \cdot n \leq -K$. Hence $|z_2| \leq K$, which implies that $z_2$ belongs to $B_i$. Consequently, $z_2$ belongs to $B_j$, which implies that $z_2 \in A_j$. Hence $z$ belongs to $B_{j+1}$. This shows, by symmetry, that $B_i = B_j$ implies $B_{i+1} = B_{j+1}$.

Now the claim immediately follows: Form sets $B_i$ until $i$ and $j$ are found such that $B_i = B_j$ and $i < j$. An integer $x$ with the properties mentioned in Claim 3 exists if and only if $B_i \neq \emptyset$. □

In what follows we assume there to exist an integer $q$ such that $B_q = \emptyset$.

**Claim 4.** *There exists (effectively) an integer $R(N)$ such that no positive integer has more than $R(N)$ representations according to $N$.*

*Proof.* Denote $R_0 = \min(A_q \cap \mathbf{N})$ and $R_1 = \max A_q$. Suppose that $z > 0$ belongs to $A_s$ and $A_t$, where $s > t \geq q$. Hence

(1)
$$m_1 + m_1 \cdot n + \cdots + m_1 \cdot n^{s-q-1} + y \cdot n^{s-q} \leq z \leq m_v + m_v \cdot n + \cdots + m_v \cdot n^{t-q-1} + R_1 \cdot n^{t-q},$$

where $y \in A_q$. Because $z > 0$, we obtain $y \geq R_0$. Hence (1) implies

$$\left(R_0 + \frac{m_1}{n-1}\right) \cdot n^{s-q} \leq \left(R_1 + \frac{m_v}{n-1}\right) \cdot n^{t-q} + \frac{m_1 - m_v}{n-1}.$$

Consequently,

$$s - t \leq R',$$

where $R' = \log_n \Big( \big(R_1 + m_v/(n-1)\big)\big(R_0 + m_1/(n-1)\big)^{-1} \Big)$. Hence we can choose $R(N) = R' + q$. □

Denote $R = \max\{R(N), R(-N)\}$. By Claim 4, no integer has more than $R$ representations. Furthermore, if an integer has a representation of length $t$, it does not have representations of length greater than $t + R$. Hence the condition in the proof of Claim 2 is decidable. Theorem 4.5 now follows from Claim 2. □

## 5. Decidability of equivalence

We need the following two lemmas from Eilenberg (1974).

**Lemma 5.1.** *If $k > 1$ and $p > 1$ are integers and $L$ is a regular language over the alphabet $\{0, 1, \ldots, k-1\}$, one can effectively construct a regular language $L'$ over the alphabet $\{0, 1, \ldots, k^p - 1\}$ such that $\nu_k(L) = \nu_{k^p}(L')$.*

**Lemma 5.2.** *If $m > 1$ is an integer and $A$ is a recognizable set of nonnegative integers, one can effectively construct a regular language $L$ over the alphabet $\{0, 1, \ldots, m-1\}$ such that $A = \nu_m(L)$.*

**Lemma 5.3.** *Let $k > 1$ and $m > 1$ be integers. If $L_1$ is a regular language over the alphabet $\{0, 1, \ldots, k-1\}$ and $L_2$ is a regular language over the alphabet $\{0, 1, \ldots, m-1\}$, it is decidable whether or not $\nu_k(L_1) = \nu_m(L_2)$.*

*Proof.* If $R_1$ and $R_2$ are languages, denote $\{w \mid R_1 w \cap R_2 \neq \emptyset\}$ by $R_1^{-1} R_2$.

If $k = m$, construct the languages $L_1' = 0^*\big((0^*)^{-1} L_1\big)$ and $L_2' = 0^*\big((0^*)^{-1} L_2\big)$. Clearly, $\nu_k(L_1) = \nu_m(L_2)$ if and only if $L_1' = L_2'$.

If there are integers $p$, $q$, $r$ such that $k = r^p$ and $m = r^q$, construct regular languages $L_3$ and $L_4$ such that $\nu_k(L_1) = \nu_{r^{pq}}(L_3)$ and $\nu_m(L_2) = \nu_{r^{pq}}(L_4)$. Then continue as in the case of $k = m$.

Assume finally that $k$ and $m$ are multiplicatively independent. By Cobham's theorem, $\nu_k(L_1) \neq \nu_m(L_2)$ or both $\nu_k(L_1)$ and $\nu_m(L_2)$ are recognizable. To decide which is the case run two algorithms concurrently. First, try to find an element from the symmetric difference of $\nu_k(L_1)$ and $\nu_m(L_2)$. Second, try to find a recognizable set $A$ such that $A = \nu_k(L_1)$. To check whether $A = \nu_k(L_1)$, form a regular language $L$ over the alphabet $\{0, 1, \ldots, k-1\}$ such that $A = \nu_k(L)$, Lemma 5.2, and check whether or not $\nu_k(L) = \nu_k(L_1)$. If a recognizable set $A$ such that $A = \nu_k(L_1)$ is found, form a regular language $L'$ over the alphabet $\{0, 1, \ldots, m-1\}$ such that $A = \nu_m(L')$, Lemma 5.2, and check whether or not $\nu_m(L') = \nu_m(L_2)$. □

**Theorem 5.4.** *It is decidable whether or not two given number systems $N_1$ and $N_2$ are equivalent.*

*Proof.* By Theorem 3.1, Corollary 3.2 and Lemma 5.3 it can be decided whether or not

$$\operatorname{Pos} S(N_1) = \operatorname{Pos} S(N_2)$$

and

$$\operatorname{Neg} S(N_1) = \operatorname{Neg} S(N_2). \square$$

**Acknowledgement.** I would like to express my deep gratitude to Professor Arto Salomaa for many rewarding discussions concerning this paper.

**References**

COBHAM, A.: On the base-dependence of sets of numbers recognizable by finite automata. - Math. Systems Theory 3, 1969, 186–192.

CULIK, K. II, and A. SALOMAA: Ambiguity and decision problems concerning number systems. - Information and Control 56, 1983, 139–153.

EILENBERG, S.: Automata, languages and machines, Volume A. - Academic Press, New York, 1974.

HONKALA, J.: Bases and ambiguity of number systems. - Theoret. Comput. Sci. 31, 1984, 61–71.

HONKALA, J.: A decision method for the recognizability of sets defined by number systems. - RAIRO 20, 1986, 395–403.

KNUTH, D.E.: The art of computer programming, Volume 2: Seminumerical algorithms, 2nd edition. - Addison–Wesley, Reading, Mass., 1981.

MATULA, D.W.: Basic digit sets for radix representation. - JACM 29, 1982, 1131–1143.

MAURER, H., A. SALOMAA, and D. WOOD: $L$ codes and number systems. - Theoret. Comput. Sci. 22, 1983, 331–346.

SALOMAA, A.: Formal languages. - Academic Press, New York, 1973.

University of Turku
Department of Mathematics
SF–20500 Turku
Finland

Received 22 April 1988