ELSEVIER
Theoretical Computer Science 290 (2003) 1433–1444
Theoretical Computer Science
www.elsevier.com/locate/tcs

The definable criterion for definability in Presburger arithmetic and its applications

An.A. Muchnik*
Institute of New Technologies Education, Nizhnaya Radishevskaya 10, Moscow 109004, Russia

Received 23 September 2000; received in revised form 17 September 2001; accepted 24 September 2001
Communicated by B. Durand

Abstract

In Section 1 of present paper we construct a formula $\varphi_n(A)$ of Presburger arithmetic (integers with addition and order) with additional n-ary predicate variable A. This formula is true if and only if predicate A is definable in Presburger arithmetic (Theorem 2).

This formula is used to prove the following facts: (1) given a finite synchronous automaton recognizing a set of n-tuples of integers written in positional notation one can effectively decide whether this set is definable in Presburger arithmetic; (2) every predicate (set of n-tuples of integers) recognizable in two essentially different positional systems is definable in Presburger arithmetic. The last result was proved by Cobham (Math. Systems Theory, 3(2) (1969) 186) for the case $n=1$. In general case both (1) and (2) were proved by Semenov (Ph.D. Thesis, Moscow State University; Siberian Math. J. 18(2) (1977) 403) (Semenov's proofs are very difficult).

Keywords: Definability; Presburger arithmetic; Cobham–Semenov theorem

0. Background[1]

The present work appeared while author was reading Semenov's thesis [9]. Using induction over n, Semenov constructs an algorithm that for a given recognizable set of n-tuples of integers decides whether this set is definable by a formula of Presburger arithmetic. The case $n=1$ is very simple: One-dimensional set is definable if and

* Fax: +7-0959156963.
E-mail address: amuchnik@int.glas.apc.org (A.A. Muchnik).

[1]This paper is a translation of preprint [8] published 10 years ago. Since then several articles that apply logical approach to automata theory appeared, see [2,4,6,7].

0304-3975/02/$ - see front matter
PII: S0304-3975(02)00047-6

only if it is ultimately periodic. This property can be expressed directly in Presburger arithmetic. Then we use the effective closeness of the family of recognizable sets under Presburger operations. A similar argument works for every n, but we need to use Theorem 2 below.

At first sight Theorem 2 is not plausible: Even two-dimensional Presburger-definable sets may have subsets with different periods that cannot be reduced to one period (as in one-dimensional case).

The property of self-definability ("to have a definable criterion for definability") is rather interesting. If we add to a self-definable structure new predicates such that the theory of the new structure is decidable then given a formula of the new structure one can effectively decide whether this formula is equivalent to some formula of the old structure. Unfortunately, we do not know any other examples of nice self-definable structures.

Structures with unsolvable elementary theory are usually mutually interpretable with the arithmetic of addition and multiplication of integers, the non-self-definability of which is proved in [1,11] (using category and measure arguments, respectively).

We believe that the structure formed by algebraic real numbers (with addition and multiplication) is not self-definable; however, a formal proof is missing (and seems to be rather complicated).

(Note that it is easy to prove that the structure formed by all real numbers with addition and multiplication is not self-definable. Indeed, let us assume that $\varphi_1(A)$ is true if and only if A is definable. Now we replace $A(x)$ by $x=y$. The new formula $\varphi'(y)$ is true if and only if y is algebraic. But we can eliminate quantifiers in $\varphi'(y)$ and get a finite union of segments. So we come to a contradiction.)

1. The definable criterion of definability

Let $A\subseteq\mathbb{Z}^n$ be some set of n-tuples of integers. We say that a vector $v\in\mathbb{Z}^n$ is A's *period* if $x\in A\Leftrightarrow x+v\in A$ for every $x\in\mathbb{Z}^n$. Let $W\subseteq\mathbb{Z}^n$. We say that v is a period of set A in W, if the equivalence

$$x\in A \quad\Leftrightarrow\quad x+v\in A$$

holds when $x\in W$ and $x+v\in W$ (two points in W that differ by v either both belong to A or both do not belong to A). Note that if v is a period of A in W, then $-v$ is also a period of A in W.

Presburger arithmetic is the elementary theory of integers with addition and order. The length $|x|$ of vector x is the sum of modules of its components.

Let $A\subseteq\mathbb{Z}^n$. We will call a *section of A* any set of the form

$$A_{i,l}=\{\langle x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n\rangle \mid \langle x_1,\ldots,x_{i-1},l,x_{i+1},\ldots,x_n\rangle\in A\},$$

where $i\in\{1,\ldots,n\},\ l\in\mathbb{Z}$. The set $A_{i,l}$ is called the (i,l)-*section* of A.

Theorem 1. *A set $A \subseteq \mathbb{Z}^n$ is definable in Presburger arithmetic if and only if all its sections are definable and there exists a finite set $V \subset \mathbb{Z}^n$ of non-zero vectors* (*called "possible periods"*) *with the following property*:

$$(*) \quad \begin{array}{l} \textit{for every } k \textit{ there exists } l \textit{ such that for every point } x \in \mathbb{Z}^n \textit{ with} \\ |x| > l \textit{ the set } A \textit{ is periodic in the } k\textit{-neighborhood of point } x \textit{ with} \\ \textit{some period that belongs to } V. \end{array}$$

Proof. I. Let us prove that every definable set has property $(*)$ (the definability of sections of a definable set is obvious). Using quantifiers' elimination, we may assume that formula $\psi(x_1 \ldots x_n)$ which defines A is a Boolean combination of expressions of types

$$t_i(x_1 \ldots x_n) = c_i \bmod m$$

and

$$u_i(x_1 \ldots x_n) \geqslant d_i,$$

where $t_i(x_1 \ldots x_n)$, $u_i(x_1 \ldots x_n)$ are linear combinations of variables $x_1 \ldots x_n$ with integer coefficients and m, c_i, d_i are constants.

Hyperplanes $u_i(x_1 \ldots x_n) = d_i$ divide $\mathbb{Z}^n$ into several regions. Periodicity is obvious within each region: ψ is reduced to statements about divisibility only, and every variable can be increased or decreased by m.

The problem arises when point x is close to some hyperplane (k-neighborhood of the hyperplane contains x). Let us consider the set of all hyperplanes (of the form considered above) that intersect k-neighborhood of some point x. There are two possibilities. If this set of hyperplanes has the full rank (normals to hyperplanes generate $\mathbb{Q}^n$) then the intersection of k-neighborhoods of hyperplanes is bounded, so this possibility can be ignored for points x with large $|x|$. If not, then there exists a vector (with integer coordinates) that is parallel to all those hyperplanes. Multiplying this vector by m, we get one of the periods.

Now let us continue with a more formal account.

(1) Every inequality $u_i(x_1 \ldots x_n) \geqslant d_i$ has form $u_{i1}x_1 + \cdots + u_{in}x_n \geqslant d_i$. We denote the integer vector $\langle u_{i1}, \ldots, u_{in} \rangle$ by u_i and call u_i the *ith normal vector* (because u_i is orthogonal to hyperplane $(u_i, x) = d_i$). Here $i \in \{1, \ldots, s\}$ where s is the number of inequalities.
(2) Let E be an arbitrary subset of the set $\{1 \ldots s\}$ of indices. Let us say that E is a *set of the first type*, if the family of vectors $\{u_i \mid i \in E\}$ has full rank (generates $\mathbb{Q}^n$). Otherwise we call E a *set of the second type*.
(3) For every set E of the second type we choose a non-zero integer vector which is orthogonal to vectors u_i for all $i \in E$. Without loss of generality we may assume that all components of this vector are multiples of m. These vectors, chosen for different sets E of the second type, form a set V of possible periods.
(4) Let k be an arbitrary positive integer (the size of neighborhood). Consider an arbitrary set E of indices that belongs to the first type. As the family of normal

vectors has the full rank, the intersection of hyperplanes corresponding to E has at most one point. The k-neighborhood of a hyperplane is a finite union of parallel hyperplanes, therefore for every E the intersection of k-neighborhoods of all E-hyperplanes is finite. The union of such intersections for all sets E of first type is also finite, and we take as l the maximal length of its elements.

(5) Let $|x| > l$. Let us consider all hyperplanes that intersect the k-neighborhood of the point x; let E be the set of their indices. This set belongs to the second type (because of the choice of l). So there exists some vector $v \in V$ parallel to all these hyperplanes. As $v \in V$, all its components are multiples of m. And v is a period (all hyperplanes that do not belong to E do not intersect the neighborhood).

II. Let us prove now that if the property $(*)$ is fulfilled for some finite set V (in this case we say that set A is *V-periodic*) and all sections of A are definable, then A is definable. To prove this assertion we use induction on cardinality of the set V. (We identify v and $-v$ and count them as one element of V since the definition of period is symmetric.)

Induction base: Let $V = \{v\}$, $v = \langle v_1, \ldots, v_n \rangle$. As $v \neq 0$, we may assume that $v_1 > 0$. Let $l > 0$ be so large that for $|x| > l$, vector v is a period of A in $|v|$-neighborhood of x. Obviously, this implies that v is a period of A in the set $\{x \in \mathbb{Z}^n \mid |x| > l\}$. Consider the sections $A_{1,l}, A_{1,l+1}, \ldots, A_{1,l+v_1-1}$ (the first coordinate is fixed as $l, l+1, \ldots, l+v_1-1$, respectively). They define all the sections $A_{1,s}$ of the set A with $s > l + v_1 - 1$ because of periodicity. More precisely, if

$$A' = \{\langle x_1 \ldots x_n \rangle \in A \mid l \leqslant x_1 < l + v_1\},$$

then for a vector $x = \langle x_1 \ldots x_n \rangle$ with $x_1 > l$ the property $x \in A$ can be expressed in the equivalent form: "there exist $y = \langle y_1 \ldots y_n \rangle \in A'$ and $q \in \mathbb{Z}$ such that $x_1 \geqslant y_1$, $x_1 - y_1 = v_1 q$, $x_2 - y_2 = v_2 q, \ldots, x_n - y_n = v_n q$". As A' is definable (it consists of finite number of definable sections), the part A^* of set A which consists of vectors $x = \langle x_1 \ldots x_n \rangle$ with $x_1 > l$ is also definable. The part A^{**} of set A which consists of vectors $x = \langle x_1, \ldots, x_n \rangle$ with $x_1 < -l$ is also definable for similar reasons. The set $A \backslash (A^* \cup A^{**})$ consists of finite number of definable sections, therefore A is definable. The base of induction is proved.

Induction step: We start with the following definition: Let $A \subseteq \mathbb{Z}^n$ be a set and $v \in \mathbb{Z}$ be a vector. By the *boundary* $\mathrm{Bd}(A, v)$ of the set A in the direction v we mean the set $\{x \in A \mid x + v \notin A\}$. We will prove the following facts for any set $A \subseteq \mathbb{Z}^n$ and for any non-zero vector $v \in \mathbb{Z}^n$.

Lemma 1.0. *If all sections of a set A are definable, then all sections of* $\mathrm{Bd}(A, v)$ *are definable.*

Lemma 1.1. *If A is V-periodic, $v \in V$ and V consists of v and $-v$, then the set* $\mathrm{Bd}(A, v)$ *is $V \backslash \{v, -v\}$-periodic.*

Lemma 1.2. *A is definable in terms of* $\mathrm{Bd}(A, v)$, $\mathrm{Bd}(A, -v)$ *and a finite number of sections of set A.*

(This means that A can be defined by a formula that uses addition, order and unary predicates for sets $\mathrm{Bd}(A,v)$, $\mathrm{Bd}(A,-v)$ and for finite number of sections of A.)

Lemmas 1.0–1.2 enable us to make the step of induction. It remains to prove them.

Proof of Lemma 1.0. The (i,l)-section of $\mathrm{Bd}(A,v)$ is easily definable in terms of (i,l)-section and $(i,l+v_i)$-section of A.

Proof of Lemma 1.1. We have to prove that the set $\mathrm{Bd}(A,v)$ is $V\backslash\{v,-v\}$-periodic. Let k be an arbitrary size of a neighborhood.

Since A is V-periodic, one can find l such that in $k+|v|$-neighborhood of every point x with $|x|>l$ the set A has a period w which belongs to V. If $w=v$ (or $w=-v$), the set $\mathrm{Bd}(A,v)$ is empty in the k-neighborhood of point x (and is periodic with any period). If $w\neq\pm v$, then in k-neighborhood of x the set $\mathrm{Bd}(A,v)$ is also periodic with period w. Lemma 1.1 is proved. □

Proof of Lemma 1.2. Let x be an arbitrary point of A. We denote the point $x+iv$ by u_i (so that $u_0=x$, $u_1=x+v$, $u_{-1}=x-v$ and so on). There are four (mutually exclusive) possibilities:

(1) for some s and some t ($s\leqslant 0\leqslant t$) all points $u_s,\dots,u_t$ belong to A and $u_s\in\mathrm{Bd}(A,-v)$, $u_t\in\mathrm{Bd}(A,v)$;
(2) for some $s\leqslant 0$ all points $u_s,u_{s+1}\dots$ belong to A and $u_s\in\mathrm{Bd}(A,-v)$;
(3) for some $t\geqslant 0$ all points $\dots,u_{t-1},u_t$ belong to A and $u_t\in\mathrm{Bd}(A,v)$;
(4) the points u_i belong to A for every i.

Thus, the set A is divided into four disjoint parts A_1,A_2,A_3,A_4 and it is enough to prove that all of them are definable (in terms of $\mathrm{Bd}(A,v),\mathrm{Bd}(A,-v)$ and sections of the set A).

Parts A_1,A_2,A_3 are definable in terms of $\mathrm{Bd}(A,v)$ and $\mathrm{Bd}(A,-v)$. Indeed, the statement "$x\in A_1$" can be expressed as "there are points $y\in\mathrm{Bd}(A,-v)$ and $z\in\mathrm{Bd}(A,v)$, for which $z-x$ and $x-y$ are positive multiples of v (i.e. $z-x=pv$, $x-y=qv$ with $p,q\in\mathbb{N}$) and every point w between y and z such that $y-w$ (and $z-w$) are multiples of v, does not belong to $\mathrm{Bd}(A,v)$ and $\mathrm{Bd}(A,-v)$". Properties $x\in A_2$ and $x\in A_3$ can be expressed in a similar way.

However, the property "$x\in A_4$" is not equivalent to the property "all the points that differ from x by a multiple of v do not belong to $\mathrm{Bd}(A,v)$ and $\mathrm{Bd}(A,-v)$", because the latter property is true also for points x such that all the points $u_i=x+iv$ do not belong to A. To distinguish between these two possibilities, we have to use sections of A (as we did while proving induction base). Lemma 1.2 is proved. □

Theorem 1 is proved. □

Theorem 2. *There exists a formula $\varphi_n(A)$ of Presburger arithmetic with additional n-ary predicate symbol A which is true if and only if predicate A is definable in Presburger arithmetic* (*without additional predicates*).

Proof. The phrase "there exists a finite set V" used in the criterion of definability (Theorem 1) cannot be expressed directly. But the property of V used in the criterion remains true if V is replaced by a bigger set.

Therefore, we may assume w.l.o.g that V is d-neighborhood of zero (the set of all vectors whose length does not exceed d) for some d. Then property $(*)$ becomes definable by a formula. An induction over n now completes the proof of Theorem 2 (we use that for sections of A the property "to be definable" is definable). □

Theorem 3. *Let M be a finite automaton whose input is n-tuple of natural numbers written in positional system (all numbers have the same base and are aligned; M reads n least significant digits at the first step). One can effectively decide whether the set recognized by M is definable in Presburger arithmetic.*

Proof. Let A be a (n-ary) predicate recognizable by M. Consider formula $\varphi_n(A)$ from Theorem 2. Ternary predicate $x+y=z$ and binary predicate $x\leqslant y$ (on $\mathbb{N}$) are recognizable by finite automata. The family of recognizable sets is closed under logical operations (complementation, unification, projection and cylindrification). Therefore, we can construct an automaton that corresponds to $\varphi_n(A)$; it recognizes empty set if $\varphi_n(A)$ is false and the set of all strings if $\varphi_n(A)$ is true. We can distinguish effectively between these two cases, so Theorem 3 is proved. □

Analyzing the proof of Theorem 1, we see that in fact a stronger result was proved: in the if-part of Theorem 1 the property $(*)$ can be replaced with the following one:

$(**)$ *the set A is periodic in every far enough neighborhood of radius k (where k is the sum of lengths of all vectors from V) with a period from V*

(i.e. it is sufficient to consider only one value of k, namely, the sum of lengths of vectors from V).

Indeed: (1) in the proof of the base of induction ($V=\{v\}$) it is enough to have periodicity in neighborhood of radius $|v|$, (2) if the set A is periodic in neighborhood of radius k with periods from V then the set $\mathrm{Bd}(A,v)$ is $V\setminus\{v\}$-periodic in neighborhood of radius $k-|v|$, and thus property $(**)$ remains true for $\mathrm{Bd}(A,v)$.

These arguments prove the following

Theorem 4. *If all sections of a set $A\subseteq\mathbb{Z}^n$ are definable and there exists a finite set $V\subset\mathbb{Z}^n$ of non-zero vectors such that property $(**)$ is fulfilled, then A is definable in Presburger arithmetic.*

This theorem will be used in the next section to prove the Cobham–Semenov's result as mentioned in the introduction.

2. Definability of sets recognizable in positional systems with two bases

In this section, we use the criterion of definability provided by Theorem 4 to prove that any predicate P recognizable by finite automata in two positional systems with bases p and q (where $\ln q/\ln p$ is irrational) is definable in Presburger arithmetic.

Let $P(x_1 \dots x_n)$ be a n-ary predicate on $\mathbb{N}$. Assume that $x_1, \dots, x_n$ are written in positional system with base p: $x_i = \sum_{j=0}^{m} x_{ij} p^j$, and $x_{im} \neq 0$ for some i. We form an input string for an automaton by combining digits $x_{1j} \dots x_{nj}$ into one input symbol. In this way, any tuple $\langle x_1 \dots x_n \rangle$ is represented by a string in an alphabet whose letters are columns of height n containing digits $0 \dots p-1$. (The most significant letter differs from the zero column). Let us consider for every n-ary predicate P and for every base p the set of strings that correspond to all n-tuples satisfying P. If this set is recognizable, we say that P a is *recognizable predicate in p-based system.*

It is well known that every predicate definable in Presburger arithmetic is recognizable in p-based system for any p. On the other hand, the set of powers of p, which is recognizable in p-system, is not definable. It is easy to show that if p and q are powers of the same number (this condition is equivalent to $\ln q / \ln p \in \mathbb{Q}$), recognizable predicates are the same for bases p and q [3].

Theorem 5. *If a predicate $P(x_1 \dots x_n)$ is recognizable in both p-based and q-based systems and $\ln q / \ln p$ is irrational then predicate P is definable in Presburger arithmetic.*

Proof. Consider "mixed" (p, q)-based system that uses digits $0 \dots p-1$ with subscript p as well as digits $0 \dots q-1$ with subscript q. If digits of only one type are used, we get p-based or q-based system. In general case, the value of the string that includes both types of digits is defined as follows: if string s represents number x, the string si_p represents $px + i$ and string sj_q represents $qx + j$. As an example,

$$\overline{1_8 2_8 7_{10}} = \overline{1_8 2_8} \times 10 + 7 = (\overline{1_8} \times 8 + 2) \times 10 + 7 = 107 = \overline{1_{10} 3_{10} 3_8}.$$

The (p, q)-based system can be used to represent n-tuples of natural numbers in the same way as it was done for one base. An additional requirement is needed: every column must contain digits of the same type (only p-digits or only q-digits). We come to the following:

Definition. A predicate P is (p, q)-recognizable if the set of all strings that represent (in (p, q)-based system) n-tuples for which P is true, is recognizable by a finite automaton.

It is clear that every (p, q)-recognizable predicate is recognizable in both systems (with bases p and q). It turns out that the converse assertion is also true.

Lemma 5.1. *If predicate P is recognizable in p-based and q-based systems then P is (p, q)-recognizable.*

Proof. We start with some observations. Consider a string that contains k digits with base p and l digits with base q. This string may represent any number between 0 and $p^k q^l - 1$. If we fix the places for p-digits and q-digits then this representation is unique.

Let S be a (p, q)-based representation of a number x, and T be a substring of S that (p, q)-represents some number y. Consider another string T' that represents the same number y and has the same number of p- and q-digits as T. It is easy to see that string S' that is obtained from S by replacement $T \to T'$, represents x. (It must be stressed that it is not enough for T' to have the same length as T. It is important that numbers of p- and q-digits in T and T' are the same: a string of k p-digits and l q-digits gives factor $p^k q^l$ for digits that precede it.)

Let us use the following criterion of recognizability. Let M be a set of strings in a fixed alphabet. Define the relation "strings S and T are k-equivalent with respect to M" as follows: $SR \in M \Leftrightarrow TR \in M$ for all strings R of length at most k. It is well known that a set M is recognizable if and only if the relations of k-equivalence with respect to M coincide for all sufficiently large values of k. (We may assume without loss of generality that S and T are non-empty.)

We know that predicate P is recognizable in p-based system. Consider the set of strings that corresponds to P is p-based system. This set is recognizable by a finite automaton. We use the criterion above and let k_p be the first number such that for all $k \geqslant k_p$ the relations of k-equivalence are the same. A similar number for q-based system is denoted by k_q.

Let us prove that for (p,q)-based system, relations of k-equivalence for all $k \geqslant k_p + k_q$ are the same.

Consider an alphabet whose letters are columns of p-digits and columns of q-digits (of height n). Strings over this alphabet (p,q)-represent n-tuples of natural numbers. We write $Z \in P$ for a string Z over this alphabet if Z represents a n-tuple of integers for which P is true. It is enough to show that if $SX \in P$, $TX \notin P$ for some strings S, T, X (where S and T are non-empty strings) and length of X exceeds $k_p + k_q$, then there exists a string Y shorter than X such that $SY \in P$ and $TY \notin P$.

If string T starts at the 0th column, then we can use an empty string Y. If not, let us count p- and q- columns in X. As the length of X exceeds $k_p + k_q$, either the number of p-columns exceeds k_p, or the number of q-columns exceeds k_q. Assume, for example, that the number of p-columns exceeds k_p. Consider a string X' that has the same length and the same value as X, but all the p-columns in X' are moved to the right (the number of p-columns remains the same). Then $X' = VW$ where V consists of q-columns and W consists of p-columns. Strings SX and SX' represent the same tuple of integers (as well as TX and TX'). Thus, $SX' \in P$, $TX' \notin P$. This means that $SVW \in P$, $TVW \notin P$. The string SV is (p,q)-representation of some tuple; let $\overline{SV}$ be a p-representation of this tuple. String TV is (p,q)-representation of another tuple; let $\overline{TV}$ be p-representation of this tuple. The string $\overline{SV}W$ is p-based representation of the same tuple as SVW, therefore, $\overline{SV}W \in P$. In a similar way $\overline{TV}W \notin P$. Since the length of W is greater than k_p, one can find a shorter string W' consisting of p-columns for which $\overline{SV}W' \in P$, $\overline{TV}W' \notin P$ (by definition of k_p). Then, $SVW' \in P, TVW' \notin P$. This means that there exists a string Y (namely, $Y = VW'$) which is shorter than X and for which $SY \in P$, $TY \notin P$. Lemma 5.1 is proved. $\square$

Let us now go on with the proof of Theorem 5. Lemma 5.1 shows that the predicate P is (p,q)-recognizable. We show that conditions of Theorem 4 are fulfilled and therefore P is definable.

We start with a proof sketch. For a sufficiently long vector $\langle x_1 \dots x_n \rangle$ we have to prove that the predicate P is periodic in the neighborhood of x (and the period is not very large).

At first we use only base p. For simplicity, we assume that P is an unary predicate (a set of natural numbers). Let us fix some number k and divide the set of natural numbers into segments of length p^k (the numbers in one segment have the same digits except for the last k digits). For each segment we consider a bit string of length p^k that is a restriction of P to this segment.

By our assumption P is recognizable (we assume here that input string is read starting from most significant digits), and the number of different bit strings of length p^k that are restrictions of P does not exceed the number of states. (Indeed, the state of automaton before it starts to read last k digits determines P's restriction on the corresponding segment.)

Now the question of periodicity is divided in two parts: (1) what happens if the neighborhood in question intersects the boundary between segments and (2) what happens if neighborhood is contained entirely inside one of the segment.

We reduce the first case to the second one using the following trick: consider the predicate $P(x)$ together with the predicate $P'(x)=P(x+c)$, where c is close to $p^k/2$. Predicate P' is also recognizable, and boundaries between P-segments correspond to midpoints of P'-segments.

For neighborhoods within segments we have to consider all possible restrictions of P (they correspond to different states of the automaton).

Recall that P is (p,q)-recognizable, so we can use p- and q-digits together. It is important for us that given number has many (p,q)-representations.

There are two methods that allow us to change a string x formed by p- and q-digits without changing the truth value of statement $x \in P$:

(1) we can replace x by another (p,q)-representation of the same integer;

(2) we can replace x by another string that puts automaton into the same state.

By our assumption $\ln q/\ln p$ is irrational. Therefore, we can find large natural numbers k and l such that $k \ln p - l \ln q$ is close to zero. Since $k \ln p - l \ln q = \ln(p^k/q^l)$, this means that p^k/q^l is close to 1 and the difference $p^k - q^l$ is small compared to p^k or q^l.

Before going on, let us consider an example. Let $p=10, q=2$; then $10^3=1000 \approx 1024=2^{10}$. Let $\overline{abcde}$ be a five-digit decimal number. Then $\overline{cde}$ is a natural number from 0 to 999; it can be written in binary as $v^9v^8 \cdots v^0$ (here $v^i \in \{0,1\}$). Consider now the $(10,2)$-representation $a_{10}b_{10}v_2^9v_2^8 \cdots v_2^0$. How much does the corresponding number differ from $\overline{abcde}$? It is easy to see that it exceeds $\overline{abcde}$ by $24\overline{ab}$. For number $a_{10}b_{10}v_2^9v_2^8 \dots v_2^0$ to be equal to $\overline{abcde}$ we need $v_2^9v_2^8 \dots v_2^0$ to be a binary representation of $\overline{cde} - 24\overline{ab}$ (this is possible only if $\overline{cde}$ is not too close to zero).

Now we return to our proof sketch. As we have said, the set of natural numbers is partitioned into segments of length p^k. We assume that $p^k \approx q^l$. we study the restriction of P onto one segment. This restriction is determined by the state of the automaton before reading last k digits. Let us assume that two different prefixes x and y put the automaton into the same state. Let us prove that corresponding restriction of P has period $(p^k - q^l)\ (\bar{x} - \bar{y})$ (where $\bar{x}$ and $\bar{y}$ are values of strings x and y in (p,q)-

based system). Indeed, let s be some k-digit string formed by p-digits. Let s_1 be the q-based representation of the number $\bar{s}+\bar{x}(p^k-q^l)$ that contains l digits. Then, xs_1 and xs represent the same number: $\bar{x}\cdot q^l+\bar{s}_1=\bar{x}\cdot p^k+\bar{s}$. Thus, $xs\in P\Leftrightarrow xs_1\in P$. And $xs_1\in P\Leftrightarrow ys_1\in P$ because strings x and y put the automaton in the same state.

Now consider the string s_2 which is the k-digit p-based representation of number $\bar{s}_1+\bar{y}(q^l-p^k)$. Then, ys_2 and ys_1 are the representations of the same number, therefore $ys_1\in P\Leftrightarrow ys_2\in P$. Because $ys_2\in P\Leftrightarrow xs_2\in P$ we finally get that $xs\in P\Leftrightarrow xs_2\in P$. As

$$\overline{xs}_2-\overline{xs}=\bar{s}_2-\bar{s}=\bar{y}(q^l-p^k)+\bar{x}(p^k-q^l)=(\bar{x}-\bar{y})(p^k-q^l),$$

the periodicity is proved. (This argument works if s is not too close to endpoints of the segment and x,y are not too big.)

This sketch of the proof should be filled with details. First of all we have to explain how to get two different prefixes x and y that put the automaton into the same state. Let us formulate corresponding statement for n-dimensional case. We assume that automaton recognizes P in (p,q)-based system. (The automaton's input is a string formed by columns of p-digits and q-digits of height n.) We call a string *reduced* if its first symbol is not a zero column.

Lemma 5.2. *Let j be the number of states of the automaton. Then for every reduced string x of length greater than j there exists another reduced string x' that puts automaton into the same state as x.*

Proof. Consider all the states that are passed while reading all non-empty prefixes of x. Some state appears twice. Throwing away the part between two instances of that state, we do not change the final state and the first symbol of x. Lemma 5.2 is proved. □

According to Lemma 5.2, all states are divided into two classes. The states of the first class appear only after reading reduced strings of length j or less. Those states can be neglected, as the criterion of definability deals only with points that are far enough from the origin. For every state of the second class there exists a pair of different reduced strings that put the automaton in this state. Those pairs of strings are used in the proof of the periodicity.

The proof of periodicity can be easily generalized to n-dimensional case. In that case $\mathbb{N}^n$ is partitioned into n-dimensional cubic cells with side p^k. Each cell consists of vectors which differ only in k last p-digits. These cells are divided into classes that correspond to different states of the automaton that recognizes P.

Lemma 5.3. *Let x and y be two strings that put the automaton into the same state. Consider the set of all points in the cell corresponding to this state that are at a distance at least $|\bar{x}+\bar{y}|\cdot|p^k-q^l|$ from cell boundaries. On this set predicate P is periodic with period $(\bar{x}-\bar{y})(p^k-q^l)$.*

Proof of Lemma 5.3. Repeats the above arguments.

For this lemma to be non-trivial, $|\bar{x}+\bar{y}|\cdot|p^k-q^l|$ must be small compared to p^k and q^l, i.e. $|p^k-q^l|/p^k$ must be much less than $1/|\bar{x}+\bar{y}|$. It can be done, as $|p^k-q^l|/p^k$ can be arbitrary small and $|\bar{x}+\bar{y}|$ is bounded by a constant that does not depend on k and is determined by the automaton. Thus, the inner part of the cell mentioned in Lemma 5.3 can be made large enough to contain 99% of the cell volume.

Also the sum of lengths of all vectors from the set of periods must be less than the size of periodic neighborhood (see the statement of Theorem 4). The number of periods is bounded by the number of states of the automaton and the length of any period is bounded by $|p^k-q^l|$ (up to a constant factor). Thus, the sum of lengths does not exceed $c|p^k-q^l|$ for some constant c. Hence, we can choose k and l in such a way that the sum of periods is $<1\%$ of the cell edge.

The only remaining problem is the behavior of P near the cell boundaries. As we have said, we solve this problem by shifting our predicate. It remains recognizable; however, the number of states can increase and all our constants will change (because the shift distance depends on k).

To avoid this difficulty we apply the following trick. Instead of predicate $P(x_1\ldots x_n)$, let us consider the predicate Q with doubled arity

$$Q(u_1,z_1,\ldots,u_n,z_n)=P(u_1+z_1,\ldots,u_n+z_n).$$

This predicate is also recognizable, and our arguments remain valid for Q. We choose k and l such that $p^k\approx q^l$ and let z_i be equal to 0 or $\lfloor p^k/2\rfloor$. In Lemma 5.3 we consider only automaton states that can be reached on strings in which all z_i are zeros. Then periods will have zero z-coordinates. Consequently, u-coordinates will form the periods of initial predicate. Thus, we have 2^n of partitions of $\mathbb{N}^n$ into cells (corresponding to different tuples of z_i: for every coordinate there exist two possibilities). For every point in $\mathbb{N}^n$ we can find a partition for which this point is "deep inside a cell" (i.e. belongs to the cell together with neighborhood of radius $p^k/5$). Now we can apply the criterion of definability. (The only thing that we have not checked is the definability of sections. As the sections of a recognizable predicate are recognizable, we can use induction over n.)

Theorem 5 is proved. □

Acknowledgements

Author is grateful to A. Shen for help in writing the draft version of the present paper.

References

[1] J.W. Addison Jr., The undefinability of the definable, Notices Amer. Math. Soc. (12) (1965) 347.

[2] A. Bés, Undecidable extensions of Büchi arithmetic and Cobham–Semenov theorem, J. Symbolic Logic 62 (4) (1997) 1280–1296.

[3] V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, Bull. de la Soc. Math. de Belgique 1 (1994) 191–238.

[4] J.R. Büchi, Weak second-order arithmetic and finite automata, Z. Math. Logik und Grundl. Math. 6 (1) (1960) 66–92.
[5] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, Math. Systems Theory 3 (2) (1969) 186–192.
[6] C. Michaux, R. Villemaire, Cobham's theorem seen through Büchi's theorem, Proc. ICALP, Lecture Notes in Computer Science, Vol. 700, 1993, pp. 325–334.
[7] C. Michaux, R. Villemaire, Presburger arithmetic and recognizability of sets of natural numbers by automata: new proofs of Cobham's and Semenov's theorems, Ann. Pure Appl. Logic 77 (1996) 251–277.
[8] An.A. Muchnik, The definable criterion for definability in Presburger arithmetic and its application, preprint, Institute of New Technologies, Moscow, 1991 (in Russian).
[9] A.L. Semenov, On definability in some solvable theories, Ph.D. Thesis, Moscow State University, 1975.
[10] A.L. Semenov, Predicates that are regular in two positional systems are definable in Presburger arithmetic, Siberian Math. J. 18 (2) (1977) 403–418.
[11] H. Tanaka, Some results in the effective descriptive set theory, Publ. Res. Inst. Math. Sci. Kyoto Univ. Ser. A (3) (1967) 11–52.