Theoretical Computer Science 8 (1979) 135–159.

ON THE REACHABILITY PROBLEM FOR 5-DIMENSIONAL VECTOR ADDITION SYSTEMS

John HOPCROFT

Computer Science Department, Cornell University, Ithaca, NY 14853, U.S.A.

Jean-Jacques PANSIOT

Centre de Calcul de l'Esplanade 7, rue René Descartes 67084 Strasbourg, France

Communicated by J. Ullman
Received July 1976
Revised January 1978

Abstract. The reachability sets for vector addition systems of dimension less than or equal to five are shown to be effectively computable semilinear sets. Thus reachability, equivalence and containment are decidable up to dimension 5. An example of a non-semilinear reachability set is given for dimension 6.

0. Introduction

Vector addition systems or equivalent formalisms like Petri Nets have been studied extensively as a model for parallelism and resource allocation in operating systems [6, 7]. Hack [4] and Rabin (see [1]) have shown that the equivalence and containment problems for arbitrary vector addition systems are undecidable. A number of other properties such as finiteness are known to be decidable [5, 7]. However, the reachability problem, i.e., given an initial configuration, can one reach a specified configuration, has been left open. Some partial results have been obtained, notably by Van Leeuwen [9] who proved that the reachability problem is decidable up to dimension 3 and several authors (see Cardoza [2]) who showed that for reversible or self-dual vector addition systems, the reachability problem (as well as the equivalence problem) is decidable. (This particular case corresponds to the word problem for commutative semigroups.) Both of these partial results depend, at least implicitly, on the fact that the reachability set is semilinear, which is not in general true. Sacerdote and Tenney [8] have claimed that the reachability problem is decidable but have not as yet provided a rigorous proof.

In this paper we show that the reachability set is an effectively computable semilinear set for dimensions less than or equal to 5. This proves that reachability, equivalence and containment are decidable up to dimension 5. An example of a non-semilinear reachability set is given for dimension 6. Thus results for higher dimension will need basically new approaches.

We introduce a variation of the vector addition system by adding a finite state control. The addition of states often reduces the dimension needed to model a given system. For a vector addition system with states the reachability set is semilinear up to dimension 2 but not, in general, semilinear for dimension 3 or higher. Since this model reduces the dimension at which non-semilinear sets arise, it is our hope that it will make it easier to prove further results for non-semilinear cases.

1. Preliminaries

We first give basic definitions and notation used throughout the paper. Let $\mathbf{N}$ denote the set of nonnegative integers $\{0, 1, \ldots\}$, $\mathbf{Z}$ denote the set of all integers $\{\ldots, -1, 0, 1, \ldots\}$ and $\mathbf{R}$ denote the rationals. Let $\mathbf{N}^n(\mathbf{Z}^n)$ denote the set of n-tuples of elements of $\mathbf{N}(\mathbf{Z})$. If t is an n-tuple, $\Pi_i(t)$ is the ith component of t. Unless otherwise specified, operations on tuples are componentwise extensions of the usual operations (e.g. for v and w in $\mathbf{N}^n$, $v+w$ is defined by $\Pi_i(v+w)=\Pi_i(v)+\Pi_i(w)$ for $1\leq i\leq n$). Similarly, when 0 is used as an n-tuple, it denotes the all zero tuple. An important exception is the relation $<$ between elements of $\mathbf{N}^n$. $v\leq w$ means $\Pi_i(v)\leq\Pi_i(w)$ $i=1,\ldots,n$ but $v<w$ means $\Pi_i(v)\leq\Pi_i(w)$ $i=1,\ldots,n$ and $\Pi_j(v)<\Pi_j(w)$ for some j, $1\leq j\leq n$. Also an obvious but important fact is that for $n>1$, $\leq$ is not a total order on $\mathbf{N}^n$. Any set of pairwise incomparable elements of $\mathbf{N}^n$ is finite, hence for any subset S of $\mathbf{N}^n$, $\min(S)$, the set of minimal elements of S, is finite.

An important concept which we use extensively is that of a semilinear set. For C and $P\subseteq\mathbf{N}^n$ let

$$\mathscr{L}(C,P)=\left\{x \mid \exists c \text{ in } C, \exists \alpha_1,\ldots,\alpha_k\in\mathbf{N} \text{ and } \exists p_1,\ldots,p_k\in P,\ x=c+\sum_{i=1}^{k}\alpha_i p_i\right\}.$$

For convenience we write $\mathscr{L}(c,P)$ for $\mathscr{L}(\{c\},P)$. If P is finite, then $\mathscr{L}(c,P)$ is said to be a *linear* set. P is the set of *periods*. A set is *semilinear* if it is a finite union of linear sets. The class of semilinear sets is closed under union, intersection and complementation [3]. For $L\subseteq\mathbf{N}^n$ and v in $\mathbf{Z}^n$ the *shift* of L with respect to v, denoted $L+v$, is the set $\{x+v \mid x \text{ in } L\}\cap\mathbf{N}^n$. The class of semilinear sets is closed under shift as seen in the following technical lemma.

Lemma 1.1. *If L is semilinear, then $L+v$ is semilinear.*

Proof. Since the class of semilinear sets is closed under union we need only show that $L+v$ is semilinear for L a linear set. Note that for $v \geq 0$, $\mathscr{L}(c, P)+v$ is just $\mathscr{L}(c+v, P)$. Let $P=\{p_1, \ldots, p_k\}$. Let A be the set of k-tuples $(\alpha, \ldots, \alpha_k) \in \mathbf{N}^n$ such that

$$x=c+\sum_{i=1}^{k} \alpha_i p_i+v \geq 0,$$

and let B be the set of points x corresponding to minimal k-tuples of A. B is finite and $\mathscr{L}(c, P)+v=\mathscr{L}(B, P)$, hence it is semilinear. □

Also it should be clear that $\mathscr{L}(c, P)+v$ is effectively computable.

The *cone* generated by a set of vectors $P=\{p_1, \ldots, p_k\}$ and a point b is the set

$$\mathscr{C}(b, P)=\left\{x \mid x \in \mathbf{N}^n, x=b+\sum_{i=1}^{k} \alpha_i p_i, \alpha_i \geq 0, \alpha_i \text{ in } \mathbf{R}\right\}.$$

We will make use of the fact that $\mathscr{L}(B, P)$ is semilinear even if B and P are infinite provided there exists a finite subset $P_f=\{p_1, \ldots, p_k\} \subseteq P$ such that $B \in \mathscr{C}(x_0, P_f)$ or some x_0, and $P \in \mathscr{C}(0, p_f)$. We first show that if P is finite, then $\mathscr{L}(B, P)$ is semilinear.

Lemma 1.2. *Let $B \subseteq \mathbf{N}^n$ be a possibly infinite set and $P=\{p_1, \ldots, p_k\}$ a finite subset of $\mathbf{N}^n$ such that B is contained in $\mathscr{C}(x_0, p)$ for some x_0. Then $\mathscr{L}(B, P)=\mathscr{L}(B', P)$ for some finite subset B' of B, and $\mathscr{L}(B, P)$ is semilinear.*

Proof. Let

$$A=\left\{x \in \mathbf{N}^n \mid x=\sum_{i=1}^{k} \gamma_i p_i, 0 \leq \gamma_i<1\right\}.$$

A is a finite subset of $\mathbf{N}^n$, and for all $b \in \mathscr{C}(x_0, P)$ there exists x, $x \in A$, such that

$$b=x_0+x+\sum_{i=1}^{k} n_i p_i, n_i \in \mathbf{N}.$$

Let

$$B=\left\{b \in B \mid b=x_0+x+\sum_{i=1}^{k} n_i p_i, n_i \in \mathbf{N}\right\}.$$

So $B=\bigcup_{x \in A} B_x$ and $B_x \subseteq \mathscr{L}(x_0+x, P)$. Let D_x be the set of k-tuples $\langle n_1, \ldots, n_k\rangle$ such that $x_0+x+\sum_{i=1}^{k} n_i p_i \in B_x$, and B_x' the finite subset of B_x corresponding to minimal k-tuples of D_x. Then

$$B_x \subseteq \mathscr{L}(B_x', P) \quad \text{and} \quad \mathscr{L}(B_x, P) \subseteq \mathscr{L}(\mathscr{L}(B_x' P), P)=\mathscr{L}(B_x', P),$$

hence $\mathscr{L}(B', P)=\mathscr{L}(B_x, P)$, and $\mathscr{L}(B, P)=\mathscr{L}(B', P)$ where B' is the finite set $\bigcup_{x \in A} B_x'$.

Lemma 1.3. *Let B and P be possibly infinite subsets of* $\mathbf{N}^n$ such that for some finite set $P_f \subseteq P$, $B \in \mathscr{C}(x_0, P_f)$ *for some* x_0, *and* $P \in \mathscr{C}(0, P_f)$. *Then* $\mathscr{L}(B, P)$ *is semilinear.*

Proof. $\mathscr{L}(B, P) = \mathscr{L}(\mathscr{L}(B, P), P_f)$ and hence is semilinear by the previous lemma since $\mathscr{L}(B, P) \subseteq \mathscr{C}(x_0, P_f)$.

Lemma 1.4. *Let $l = \mathscr{L}(c, \{a\})$ be a one dimensional subspace of* $\mathbf{N}^n$, *and assume that there is an infinite sequence of linear sets $\mathscr{L}(x_i P_i) \subseteq l$, such that $a \in P_i$ and $P_i \subseteq P_{i+1}$. Then*

$$\bigcup_{i \in \mathbf{N}} \mathscr{L}(x_i, P_i) = \bigcup_{i \in F} \mathscr{L}(x_i, P_i)$$

for some finite set F.

Proof. l is partitioned into a finite number of equivalence classes modulo a. Since a must be in each P_i, if x is in $\mathscr{L}(x_i, P_i)$ then all $y \geqslant x$ in the same equivalence class must also be in $\mathscr{L}(x_i, P_i)$. But there are only finitely many equivalence classes and for a given x there are only finitely many $y < x$ in the same class. Thus there are only a finite number of i such that $\mathscr{L}(x_i, P_i)$ contains an x not in any $\mathscr{L}(x_j, P_j)$, $j < i$. □

We use the term *boundary* to designate an hyperplane of the form $\{x \mid \Pi_i(\mathbf{x}) = 0, \mathbf{x} \in \mathbf{N}^n\}$ for some i. Boundaries separate $\mathbf{N}^n$ from the rest of $\mathbf{Z}^n$.

An n-dimensional *vector addition scheme* W is a finite subset of $\mathbf{Z}^n$. An n-dimensional *vector addition system* (VAS for short) is a pair (x, W) where x in $\mathbf{N}^n$ is called the *start* point and $W \subseteq \mathbf{Z}^n$. The *reachability set* of the VAS (x, W), denoted $R(x, W)$ is the set of all z, $z = x + v_1 + \cdots + v_j$, where each v_i is in W and for $1 \leqslant i \leqslant j$, $x + v_1 + \cdots + v_i \geqslant 0$. The sequence $v_1, \ldots, v_j$ is called a ***W-path*** or path when W is understood, *valid* at x; $v_1 + \cdots + v_j$ is the *displacement* of the path. A W-path is sometimes noted $p \in W^*$, using the notation of regular expressions. If $p \in W^*$, $w = \{w_1, \ldots, w_k\}$, the folding of p, $\chi(p)$ is a k-tuple whose ith component is the number of occurrences of w_i in p. Of course, a folding corresponds to many paths, and for a given start point some (or all) may be nonvalid.

The reachability problem is to determine for a VAS (x, W) and a point y whether y is in $R(x, W)$. It is an open problem whether there is a decision procedure for solving all instances of the problem. The problem is solvable up to dimension 3 [9] and in various special cases, for example when W is self-dual ($v \in W \Leftrightarrow -v \in W$) (see for example Cardoza [2]). Sacerdote and Tenney [8] have claimed that the reachability problem is decidable but have not as yet provided a rigorous proof.

2. Vector addition systems with states (VASS)

In this section we present a new model for vector addition systems that includes a finite state control. We first show that an n-dim VASS can be simulated by an $(n+3)$-dim VAS, hence the two formulations have the same power. Next we prove that a 2-dim VASS has an effective semilinear reachability set. Finally we give an example of a 3-dim VASS that generates exponentiation, hence its reachability set is not semilinear.

A *vector addition scheme with states* is a vector addition scheme W, together with a finite state control S. Transitions are in $S \to S \times W$. The transition $p \to (q, v)$ can be applied at the point x in state p and yields the point $x+v$ in state q, provided that $x+v \geqslant 0$.

A *vector addition system with states* (VASS for short) is a vector addition scheme with states $\langle W, S \rangle$ together with a starting point x_0 and a starting state $p_0 \in S$.

Lemma 2.1. *An* n-dim VASS *can be simulated by an* $(n+3)$-dim VAS.

Proof. We give the construction of the VAS. The last three coordinates encode the state while the first n coordinates are as in the VASS. Assume that the VASS has k states $q_1, \ldots, q_k$. Let $a_i = i$ and $b_i = (k+1)(k+1-i)$ for $i = 1$ to k. If the VASS is at v in state q_i then the VAS will be at $(v, a_i, b_i, 0)$. For each i the VAS has two dummy transitions t_i and t'_i defined so that t_i goes from $(v, a_i, b_i, 0)$ to $(v, 0, a_{k-i+1}, b_{k-i+1})$ and t'_i goes from $(v, 0, a_{k-i+1}, b_{k-i+1})$ to $(v, b_i, 0, a_i)$. Note that t_i and t'_i modify only the last three components. In addition there is a transition t''_i for each transition $i \to (j, w)$ of the VASS, defined by

$$t''_i = (w, a_j - b_i, b_j, -a_i).$$

Clearly any path of the VASS can be mimicked by the VAS. It remains to be shown that the VAS cannot do something unintended. We will only show that t''_i can only be applied if the last three components are b_i, 0 and a_i respectively. The other cases are similar. Observe that for each i and j, $a_i < a_{i+1}$, $b_i > b_{i+1}$, $a_i < b_j$ and $b_i - b_{i+1} = k+1 > a_j$. Let v''_i be the vector $(w, a_j - b_i, b_j, -a_i)$ which accomplishes the transition t''_i. Note that the $n+1$st and last components are negative. Hence t''_i cannot be applied when the last three coordinates are $(a_i, b_i, 0)$ or $(0, a_{k-i+1}, b_{k-i+1})$ since either the first or third components are 0. Let the last three coordinates be $(b_m, 0, a_m)$. Then if $m < i$, t''_i cannot be applied since $a_m - a_i < 0$. If $m > i$, then t''_i cannot be applied since $b_m + a_j - b_i \leqslant a_j - (k+1) < 0$. $\square$

Since an n-dim VASS can trivially simulate an n-dim VAS, the reachability problem for VAS is solvable if and only if the reachability problem for VASS is solvable.

We are now going to show that the reachability set for each 2-dim VASS is semilinear. The idea is the following. Start enumerating paths. On encountering a path containing a subpath starting and ending in the same state from some x to some y, $y \geqslant x$, we observe that the subpath can be repeated as often as we like, giving an infinite set of paths. Thus if z is any point reachable from y, we can reach the set of points $\{z + i(y - x) \mid i = 1, 2, \ldots\}$. We enumerate the reachability set by enumerating such linear sets continuing this process until a collection of linear sets is constructed which is closed under transitions of the VASS. Even though the above process does not in general terminate, in dimension 2 it does terminate implying that the reachability set of a 2-dim VASS is semilinear.

The intuitive reason why the enumeration terminates in the 2-dimensional case is as follows. If the process does not terminate, then there is an infinite path such that points along this path are not in previously computed linear sets. The set of periods for the linear sets corresponding to points on this path must eventually have arbitrarily large cardinality. By Lemma 1.3 this implies that the cones generated by the periods must be "widening" infinitely often. As we will see, in dimension 2 this implies that eventually periods parallel to the axis vectors can be added and hence the cones cannot widen further. This is not true in higher dimensions, for example, in the 3-dimensional VASS of Corollary 2.8, we can generate an infinite set of periods $\{(1, 0, 2^i)\}$ but we cannot get $(0, 0, 1)$.

In the following we make these ideas precise. A *short path* is a path with no repeating state except that the first and last state are the same. A ***short positive path*** is a short path with a positive displacement. Note that there are only finitely many short paths. An *axis* is a vector with one positive component and all other components zero.

We give an algorithm that constructs a tree labelled by 3-tuples $[x, p, A_x]$ where x is in $\mathbf{N}^2$, p is a state and $A_x \subseteq \mathbf{N}^2$. The label $[x, p, A_x]$ denotes the fact that every point in the linear set $\mathscr{L}(x, A_x)$ can be reached in state p from the start point x_0 and start state p_0. When a new vertex is added with label $[x, p, A_x]$ the displacement of any short positive path which is valid at x is added to the set of periods A_x. Also if there is a path valid at x whose displacement is an axis, the axis is added to A_x if a parallel axis vector is not already present. Each vertex inherits the periods of its father. If $\mathscr{L}(x, A_x)$ is contained in $\mathscr{L}(z, A_z)$ where the vertex labelled $[z, p, A_z]$ is an ancestor of $[x, p, A_x]$, the path is terminated at $[x, p, A_x]$ since any descendant of $[x, p, A_x]$ is equivalent to a descendant of $[z, p, A_x]$ which is closer to the root. In this case $[x, p, A_x]$ is marked.

Algorithm

Input: The set of transitions and the start point x_0 and start state p_0 forming a VASS.

Create root labelled $[x_0, p_0, \emptyset]$;

while there are unmarked leaves **do**

begin

Pick an unmarked leaf $[x, p, A_x]$;

Add to A_x all displacements of short positive paths from p to p valid at x;

if A_x is empty and there exists an ancestor $[z, p, A_z]$ with $z < x$, **then** add $x - z$ to A_x;

if there exists $c \in \mathbf{N}^2$, $c = (0, \gamma)$ or $(\gamma, 0)$ such that

(a) c is not colinear to any vector of A_x, and

(b) either (i) there exists an ancestor $[z, p, A_z]$ of $[x, p, A_x]$ such that $x - z = c$, or

(ii) for some short nonpositive path from p to p valid at x, with displacement a, and some $b \in A_x$, there exists α, $\beta \in \mathbf{N}$ such that $\alpha a + \beta b = c$

then add c to A_x;

if there exists an ancestor $[z, p, A_z]$ of $[x, p, A_x]$ such that $\mathscr{L}(z, A_z)$ contains x and $A_z = A_x$

then mark $[x, p, A_x]$

else for each transition $p \to (q, v)$ **do**

begin

Let $A_x = \{v, \ldots, v_k\}$

for each a, $a = \alpha_1 v_1 + \cdots + \alpha_k v_k$ where $(\alpha_1, \ldots, \alpha_k)$ is a minimal k-tuple such that $x + a + v \geqslant 0$, **do** construct a son $[y, q, A_y]$ where $y = x + a + v$ and $A_x = A_y$;

end;

if $[x, p, A_x]$ has no son **then** mark $[x, p, A_x]$;

end

Lemma 2.2 *There exists a constant b such that for each label $[x, p, A_x]$ of the tree, $|A_x| \leqslant b$. Moreover if $[x, p, A_x]$ is an ancestor of $[y, q, A_y]$ then $A_x \subseteq A_y$.*

Proof. A_x contains only displacements of short positive paths, at most 2 axis vectors, and possibly one more vector, the first one included in the first non-empty ancestor of A_x, hence there is a bound on $|A_x|$. If $[x, p, A_x]$ is an ancestor $[y, q, A_y]$ then $A_x \subseteq A_y$ since sons inherit the periods of the fathers.

Lemma 2.3. *The preceding algorithm always terminates, and the corresponding tree is finite and effectively computable.*

Proof. Assume the algorithm never terminates. All instructions inside the while loop are finite, so the only possibility is that the while loop itself never terminates. But each time the loop is executed a new vertex is visited, hence an infinite tree is constructed. Since the fan-out of the tree is finite, because $|A_x|$ is bounded, there must be an infinite path, by application of König's Lemma. We are now going to show that all paths must be finite, hence that the algorithm terminates.

Assume that there is an infinite path with vertices $[x_i, p_i, A_{x_i}]$ $i=0, 1, \ldots$. By Lemma 2.2 the A_{x_i}'s remain unchanged beyond some finite i_0. Thus there is an infinite path $[x_i, p_i, A]$, $i=i_0, i_0+1, \ldots$ for some A. We will show that there exists a cone $\mathscr{C}(y_0, A)$ such that all but a finite number of x_i's lie in the cone. But by Lemma 1.2 only a finite number of x_i's may lie in the cone and have distinct $\mathscr{L}(x_i, A)$, a contradiction. From this we conclude the path is finite. It remains to show the existence of the cone $\mathscr{C}(y_0, A)$. Note A cannot be empty, otherwise the sequence $x_1, \ldots, x_i, \ldots$ is non increasing.

We first show that only a finite number of the x_i's from the path may lie on the same horizontal or vertical line. Suppose $x_{i_1}, x_{i_2}, \ldots$ lie on the same horizontal or vertical line. The sequence $x_{i_1}, x_{i_2}, \ldots$ has a minimum say x_{i_m}. Also an axis vector colinear to the horizontal or vertical line must be in A since there must exist a pair of indices $i_j < i_{j'}$ for which $x_{i_j} < x_{i_{j'}}$. Hence for all k, $\mathscr{L}(x_{i_k}, A)$ lies in the cone $\mathscr{C}(x_{i_m}, A)$ and by Lemma 1.2

$$\bigcup_{i_k} \mathscr{L}(x_{i_k}, A) = \bigcup_{i_k \in F} \mathscr{L}(x_{i_k}, A)$$

for some finite F. Thus there exists $i_j < i_{j'}$ such that $x_{i_{j'}}$ is in $\mathscr{L}(x_{i_j}, A)$. Hence the last vertex should be marked and the sequence terminated, a contradiction. Thus only a finite number of x_i may lie on the same horizontal or vertical line.

Consider any fixed c_0 in $\mathbf{N}^2$ and let $D=\{x \mid X \geqslant c_0\}$. The region $\mathbf{N}^2 - D$ is composed of a finite number of vertical and horizontal lines and thus by the previous argument contains only a finite number of x_i from any infinite path in the tree. Choose c_0 sufficiently large so that all transitions and short paths are valid at c_0, hence at any point of D. Since the path in the tree has only a finite number of points outside D, there exists an index J_0 such that $x_{j_0}, x_{j_0+1}, \ldots$ are in D. In general $x_i = x_{i-1} + a_i + v_i$ where v_i is a transition and a_i is a minimal displacement such that $x_i \geqslant 0$. However at any point in the region D, v_i is valid, hence $a_i = 0$ and the path in the tree is also a path of the VASS.

We will show that for each state p, there can be only a finite number of vertices $[x_i, p, A]$ with x_i outside the cone $\mathscr{C}(x_{i_0}, A)$ where x_{i_0} is the first point on the path in state p, $i_0 \geqslant j_0$. Let $[x_i, p, A]$, $i > i_0$ be another point on the path in state p. If i does not exist, then our claim is trivially true. Clearly

$$x_i - x_{i_0} = \sum_{j=1}^{k} w_j$$

where the w_j's are displacements of short paths since there is a W-path from x_{i_0} to x_i and from state p to state p. We consider several cases.

Case 1. A does not contain any axis vector. Then either

(i) all w_j's are positive. Hence $w_j \in A$ and $x_i \in \mathscr{C}(x_{i_0}, A)$ (hence no x_i is outside the cone) or

(ii) all w_j's and all vectors of A are colinear, so $x_i \in \mathscr{C}(x_{i_0}, A)$ or $x_i < x_{i_0}$ (hence a finite number of x_i's are outside the cone).

Case 2. A contains both axes. Then $\mathbf{N}^2 - \mathscr{C}(x_{i_0}, A)$ is composed of finitely many vertical and horizontal lines, and we know that each of these lines can contain only finitely many points of the path (hence a finite number of x_i's are outside the cone).

Case 3. (See Fig. 1.) A contains one axis, say $c = (0, \alpha)$. Let v_1 be a vector of A with smallest slope (possibly infinite if $v_1 = c$, but not zero). Let $v_1 = (a, b)$ and

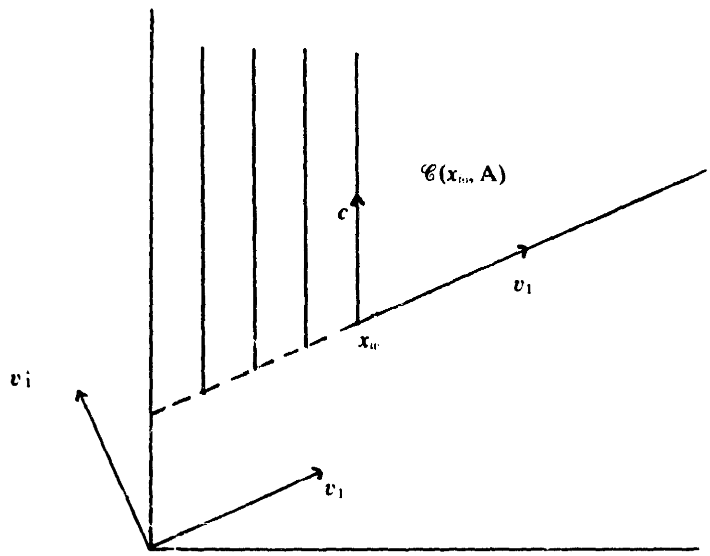

Fig. 1.

$v_1^\perp = (-b, a)$ be orthogonal to v_1, and assume that for some j, $w_j^\mathrm{T} \cdot v_1^\perp < 0$. Then either $w_j > 0$, w_j has a smaller slope than v_1, w_j is in A, a contradiction, or w_j is not positive and there exists integers c, d such that $cw_j + dv_i = (\gamma, 0)$ for some $\gamma > 0$. So a second axis is in A, a contradiction. Thus for all j's,

$$w_j^\mathrm{T} \cdot v_1^\perp \geqslant 0, \quad \text{and} \quad (x_i - x_{i_0})^\mathrm{T} \cdot v_1^\perp \geqslant 0,$$

hence all x_i's lie on the same side of the line generated by x_{i_0} and v_i. If $x_i \notin \mathscr{C}(x_{i_0}, A)$, x_i lies on one of a finite number $(\Pi_1(x_{i_0}))$ of vertical lines, and there are finitely many such x_i's.

In all three cases, there can be only a finite number of the x_i's reached in state p outside $\mathscr{C}(x_{i_0}, A)$. Let X be the set of x_i's such that x_i is reached in state p, $x_i \in \mathscr{C}(x_{i_0}, A)$. If X is finite then there is only a finite number of x_i's reached in state p. Suppose X is infinite. By Lemma 1.2 $\mathscr{L}(X, A) = \mathscr{L}(B, A)$ where B is a finite subset of X. Hence there exists $x_i \in B$, $x_j \in X$, $i < j$ such that $x_j \in \mathscr{L}(x_i, A)$. But then the path should terminate at x_j, a contradiction. So there can be only a finite number of the x_i's reached in state p. Since this is true for all states, the path is finite. Hence the tree is finite and the algorithm terminates. □

Let $T_p = \bigcup \mathcal{L}(x, A_x)$ where the union is over all vertices $[x, p, A_x]$ of the tree.

Lemma 2.4. *T_p is an effectively computable semilinear set.*

Proof. Clear since the tree is finite by the previous lemma.

The next lemma shows that indeed we compute the reachability set.

Lemma 2.5. *Let R_p be the set of points reachable in state p. Then $R_p = T_p$ for all states p.*

Proof. *Part* 1. $T_p \subseteq R_p$. We show by induction on the depth of x in the tree that for any node $[x, p, A_x]$, $\mathcal{L}(x, A_x) \subseteq R_p$.

Basis. Let $[x_0, p_0, A_{x_0}]$ be the label of the root. Let $w \in A_{x_0}$. Either w is the displacement of a short positive path from p_0 to p_0, valid at x_0, or w is an axis vector, $w = \alpha a + \beta b$ where a is the displacement of a short nonpositive path from p_0 to p_0, valid at x_0 and b is the displacement of a short positive path from p_0 to p_0, valid at x_0. In this case, we can apply first the β copies of b, followed by the α copies of a, so w is valid at x_0, and $w \geq 0$. In both cases, $x_0 - w$ is reachable in state p_0, and $x_0 + w \geq x_0$, whence $x_0 + a$, $a \in \mathcal{L}(0, A_{x_0})$ is reachable in state p_0, from x_0 and $\mathcal{L}(x_0, A_{x_0}) \subseteq R_{p_0}$.

Induction hypothesis. Assume that for each vertex $[x, p, A_x]$ of depth at most $n-1$, $\mathcal{L}(x, A_x) \subseteq R_p$. Let $[y, q, A_z]$ be a vertex of depth n, and let $[x, p, A_x]$ be its father. So $y = x + a_1 + v$, where $a_1 \in \mathcal{L}(0, A_x)$, and $p \to (q, v)$ is a transition. Let $z \in \mathcal{L}(y, A_y)$. We will show that

$$z \in R_q, \quad z = y + a, \quad a \in \mathcal{L}(0, A_y).$$

Since $A_x \subseteq A_y$, $a = a_2 + a_3$, $a_2 \in \mathcal{L}(0, A_x)$, $a_3 \in \mathcal{L}(0, A_y - A_x)$. Now $z = x + a_1 + a_2 + v + a_3$, and $x + a_1 + a_2 \in \mathcal{L}(x, A_x)$ is reachable in state p, by the induction hypothesis. $p \to (q, v)$ is valid at $x + a_1$, so it is also valid at $x + a_1 + a_2$, hence $x + a_1 + a_2 + v = y + a_2 \in R_q$. Now, by an argument similar to the one used in the basis part, all vectors of $A_y - A_x$ are positive displacements of paths from q to q valid at y, hence at $y + a_2$. So $z = y + a_2 + a_3 \in R_q$.

Part 2. (See Fig. 2.) $R_p \subseteq T_p$ for all p. We show by induction on the length of the W-path from x_0 to z (in state p) that $z \in T_p$.

Basis. If $z = x_0$ then $z \in \mathcal{L}(x_0, A_{x_0}) \subseteq T_{p_0}$.

Induction hypothesis. Assume that for each state p and any point z reachable in state p by a path of length $n-1$, $z \in \mathcal{L}(x, A_x)$ for some vertex $[x, p, A_x]$.

Let z be reachable in state p by a path of length n. Let $q \to (p, v)$ be the last transition of the path and $z = y + v$. By induction hypothesis, $y \in \mathcal{L}(x, A_x)$ for some

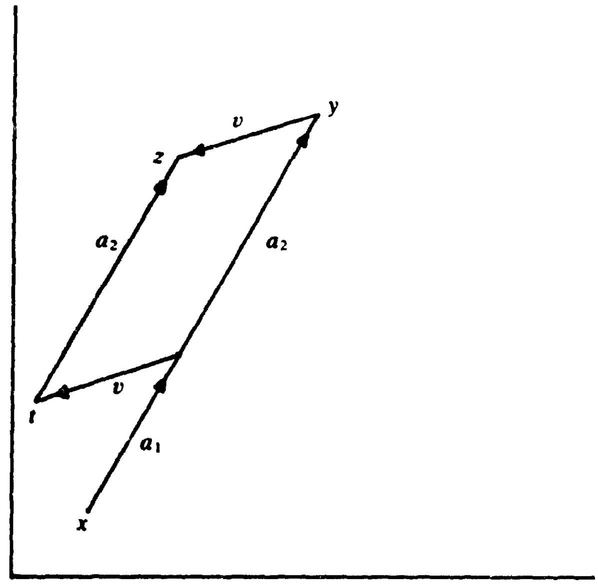

Fig. 2.

vertex $[x, q, A_x]$. So $y = x + c$, $c \in \mathscr{L}(0, A_x)$. We may assume that $[x, q, A_x]$ is not a leaf. Otherwise either

(i) no transition is applicable at x, a contradiction, or

(ii) $\mathscr{L}(x, A_x) \subseteq \mathscr{L}(x', A_x)$ for some interior vertex $[x', q, A_x]$, and we may replace x by x'.

Let $A_x = \{v_1, \ldots, v_k\}$. Then x has sons $x + a + v$ for each $a = \sum_{i=1}^{k} \alpha_i v_i$ corresponding to a minimum tuple $(\alpha_1, \ldots, \alpha_k)$ such that $x + a + v \geq 0$. Since $x + c + v \geq 0$ we can write $c = a_1 + a_2 \in \mathscr{L}(0, A_x)$ such that for $t = x + a_1 + v$ there is a son of $[x, q, A_x]$ labelled $[t, p, A_t]$, and $z = t + a_2$. But a_2 is in $\mathscr{L}(0, A_x) \subseteq \mathscr{L}(0, A_t)$. Hence z is in $\mathscr{L}(t, A_t)$.

By parts 1 and 2, $R_p = T_p$ for all p. □

Theorem 2.6. *In a* 2-dim VASS, *the set of points reachable in any given state is semilinear and effectively computable.*

Proof. Clear from Lemmas 2.4 and 2.5.

Corollary 2.7. *Equivalence and reachability are decidable for* 2-dim VASS.

We now give an example of a 3-dim VASS that generates exponentiation, hence in general, 3-dim VASS have non-semilinear reachability sets.

Lemma 2.8. *There exists a* 3-dim VASS *with a non-semilinear reachability set.*

Proof. Consider the following 3-dim VASS with two states, p and q. The start point and state are $x_0 = (0, 0, 1)$ and, $p_0 = p$. The transitions are:

$$t_1 : p \to (p, (0, 1, -1)), \qquad t_2 : p \to (q, (0, 0, 0)),$$

$$t_3 : q \to (q, (0, -1, 2)), \qquad t_4 : q \to (p, (1, 0, 0)).$$

Let condition (1) be $0 < x_2 + x_3 \leq 2^{x_1}$ and condition (2) be $0 < 2x_2 + x_3 \leq 2^{x_1+1}$.

Claim. $x = (x_1, x_2, x_3)$ is reachable in state p if and only if (1) holds, and $x = (x_1, x_2, x_3)$ is reachable in state q if and only if (2) holds.

We first show $\Rightarrow$. Note that initially (1) holds and

(i) if (1) holds and we apply t_1, (1) still holds,
(ii) if (1) holds and we apply t_2, (2) holds,
(iii) if (2) holds and we apply t_3, (2) holds,
(iv) if (2) holds and we apply t_4, (1) holds.

Hence any reachable point satisfies (1) or (2) depending upon the state. We now show $\Leftarrow$, i.e. if (1) or (2) holds, then we can reach x in the appropriate state. The proof is by induction on the first coordinate x_1.

Basis. If $x_1 = 0$ and (1) holds then either

(i) $x = (0, 1, 0)$ and x is reachable in state q by $t_1 t_2$
(ii) $x = (0, 0, 1)$ and x is reachable in state q by t_2
(iii) $x = (0, 0, 2)$ and x is reachable in state q by $t_1 t_2 t_3$.

So in all cases x is reachable in the right state.

Induction hypothesis. Assume that all points satisfying (1) or (2) and $x_1 \leq a_1 - 1$ can be reached in the appropriate state. Let $a = (a_1, a_2, a_3)$ and assume that (1) holds, i.e. $0 < a_2 + a_3 \leq 2^{a_1}$. We will show that a is reachable in state p.

Assume that $0 < a_2 + a_3 \leq 2^{a_1 - 1}$. Then by the induction hypothesis, $a' = (a_1 - 1, a_2, a_3)$ is reachable in state p, and by applying $t_2 t_4$, we reach a in state p. So now we assume that $2^{a_1 - 1} < a_2 + a_3 \leq 2^{a_1}$. Let $a_2 + a_3 = 2^{a_1 - 1} + b$ where $0 < b \leq 2^{a_1 - 1}$. By the induction hypothesis, there is a path to

$$a' = (a_1 - 1, b, 2^{a_1 - 1} - b)$$

since a' satisfies (1). But now, at a' we can apply $t_2 (t_3)^b t_4 (t_1)^{a_2}$ and we get $(a_1, a_2, 2^{a_1 - 1} + b - a_2)$ in state p. Since $2^{a_1 - 1} + b - a_2 = a_3$, we have reached a in state p.

A similar argument shows that if (2) holds we can reach a in state q. Hence, by induction, our claim is true. Clearly the reachability set is not semilinear, thus our lemma. □

Note that although the reachability set is not semilinear, we can specify it completely by recursive relations, hence reachability is decidable for this particular example. Indeed, a possible way to solve the reachability problem would be to find a mechanical way to compute such relations invariant by application of the transitions.

Remarks. The example we have presented is in some sense the simplest non-semilinear VASS. If we reduce by one the dimension, or the number of states, or even the number of transitions, we get a semilinear reachability set. Also, there is a very similar 3-dim VASS (2 states, 4 transitions) that generates squares.

Conclusion. In this section we have introduced VASS and showed that they have non-semilinear reachability sets for dimension as low as 3 (in the next section we will see that for VAS this happens only at dimension 6). So it might be easier to use VASS to prove results on non-semilinear systems. Also with VASS, it is possible to reduce the dimension of a system when one coordinate remains bounded, replacing each value of that coordinate by a state. In fact we use this property in the proof that 5-dim VAS have semilinear reachability sets.

Short of solving the general reachability problem, it would be interesting to solve the reachability problem for 3-dim VASS, since they have non-semilinear reachability sets.

3. 5-dim VAS have a semilinear reachability set.

In this section we show that the set of all z reachable from a given x is semilinear provided we restrict x and z to be sufficiently large. The first part, based on Van Leeuwen's results, is the case where x and z have $n-1$ large coordinates (larger than some computable constant b). In the second part we extend this to points having $n-2$ large coordinates (larger than some computable constant c). Finally, in part three we use these results to show that 5-dim VAS have semilinear reachability sets.

In this part we investigate some properties of paths and reachability sets when either endpoint of a path, or even an intermediate point, is sufficiently far from $n-1$ boundaries, in a sense to be defined later. This part is inspired by Van Leeuwen [9]. We first give some of his notations and results and then give some generalizations and improvements of these results. Many of these results use the fact that a nonvalid path can always be reordered so that at least one coordinate remains positive.

A *principal arcone* is a subset of $\mathbf{N}^n$ of the form $\{x \mid x \geq v\}$ for some $v \in \mathbf{N}^n$. It can be viewed as $\mathbf{N}^n$ shifted by a positive vector v.

Let $x \in \mathbf{N}^n$, $A \subseteq \mathbf{N}^n$. A *web* of x with respect to A is a set L of W-paths such that:

(i) each path in L is a valid path from x to some y in A.

(ii) if p_1 is a valid path from x to some y in A, then there exists p_2 in L, $\chi(p_2) \leq \chi(p_1)$.

(iii) if p_1, p_2 are in L, then $\chi(p_1)$ and $\chi(p_2)$ are incomparable.

Note that if F is the set of foldings of all valid W-paths from x to any point A, then $\min F$, the set of minimal elements of F is finite. A web L of x with respect to A is just a set of valid W paths such that for each α in $\min F$ there is exactly one p in L with $\chi(p) = \alpha$. Hence webs are always finite.

Lemma 3.1. (Van Leeuwen [9].) *For each $x \in \mathbf{N}^n$ and principal arcone A, the web of x with respect to A can be effectively determined.*

A *W-transformation area* S is a subset of $\mathbf{N}^n$ such that any (nonvalid) W-path between two points x and y of S can be rearranged into a valid path from x to y.

We now give a slightly generalized version of one of Van Leeuwen's theorems.

Lemma 3.2. *For each j, $1 \leq j \leq n$, there is an effectively computable v_j with $\Pi_j(v_j) = 0$ such that $A = \{x \mid x \geq v_j\}$ is a W-transformation area. Moreover, v_j can be chosen independently of the positive vectors of W. (That is, if we change the positive vectors of W, v_j remains unchanged).*

Proof. The first part of the lemma is Van Leeuwen's theorem. It remains to be shown that v_j can be chosen independently of the positive vectors of W. Note that this property is useful when dealing with linear starting sets $\mathscr{L}(x, P)$: We can then just add P to W, and consider x as the starting point. The transformation areas are unchanged.

Assume that we partition W into W_1 and W_2, W_1 containing all positive vectors of W and W_2 the rest. Using Van Leeuwen's construction, find a W_2-transformation area $A = \{x \mid x \geq v_j\}$. Then A is also a W-transformation area. Consider a (nonvalid) W-path from x to y, both x and y in A. Rearrange the path so that all positive vectors are at the beginning. We get a valid path from x to z, z in A, followed by a nonvalid W_2-path from z to y. Since both z and y are in a W_2-transformation area, this path can be rearranged into a valid path.

We can find transformation areas with an even more general form as the next lemma shows:

Lemma 3.3. (See Fig. 3.) *There is an effectively computable constant b such that the set of points having $n-1$ coordinates larger than b is a W-transformation area. Again, b is independent of the positive vectors of W.*

Proof. Note that this transformation area is of the form

$$S = \bigcup_{j=1}^{n} \{x \mid x \geq v_j'\}$$

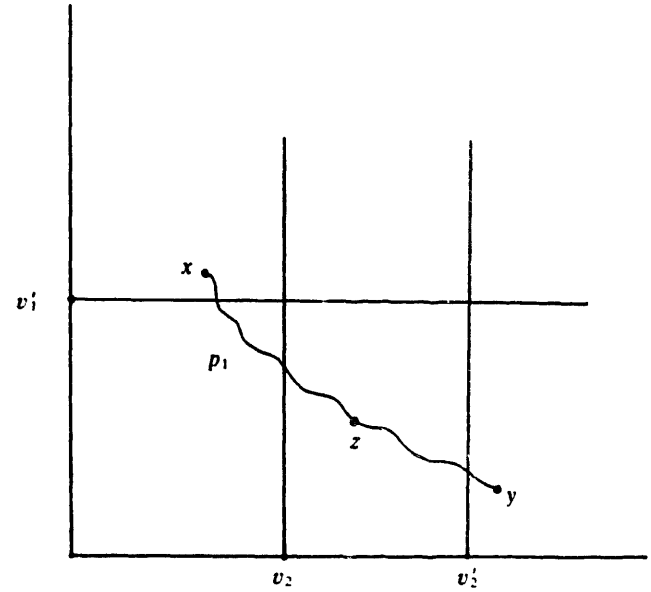

Fig. 3.

where $\Pi_i(v_j')=b$, $\Pi_j(v_j')=0$, $i \neq j$. Let v_j be as in the previous lemma. Let $c = \max_{i,j} \Pi_i(v_j)$ and

$$d = \max\{-\Pi_i(w) \mid \Pi_i(w) \leqslant 0,\ w \in W,\ 1 \leqslant i \leqslant n\}$$

and $b = c(d+1)$. We are going to show that this choice of b satisfies our lemma.

Assume there is a nonvalid path p from x to y, both in S. If x and y are larger than the same v_j, then by Lemma 3.2 we can rearrange p into a valid path. So without loss of generality, we can assume that $x \geqslant v_1'$, $\Pi_1(x) < \Pi_1(v_2) \leqslant c$, and $y \geqslant v_2'$, $\Pi_2(y) < \Pi_2(v_1) \leqslant c$. (This is because $v_1' > v_i 1 \leqslant i \leqslant n$).

Now since $\Pi_1(x) < \Pi_1(v_2) \leqslant c$ and $\Pi_1(y) \geqslant \Pi_1(v_2') = b > c$, p must contain some vectors with a positive first coordinate. In fact, a sequence of at most c vectors must increase the first coordinate from $\Pi_1(x)$ to at least c. Let p_1 be this sequence. If we put the sequence p_1 at the beginning, we get a valid path from x to some z followed by some path from z to y. But then

$$\Pi_1(z) \geqslant c \geqslant \Pi_1(v_2)$$

and

$$\Pi_1(z) \geqslant \Pi_i(x) - c \cdot d \geqslant b - c \cdot d \geqslant c \geqslant \Pi_i(v_2).$$

Hence both z and y are larger than v_2 so that path from z to y can be rearranged into a valid path. Hence our lemma. □

We now use these results to show that in some cases, the reachability set is semilinear. In the first case, the starting point has $n-1$ sufficiently large coordinates, and in the second case any intermediate point has $n-1$ large coordinates.

Given a vector addition scheme W, the set of points reachable from 0, by not necessarily valid paths is an effective linear set $L_0 = \mathscr{L}(0, P)$. Note that each period p of P can be generated by a nonvalid path from 0 to p. For any $x \in \mathbf{N}^n$, the set of points $y \geqslant x$ reachable from x by a (nonvalid) path is $L_0 + x = \mathscr{L}(x, P)$. If x is sufficiently large, the paths generating each period become valid when applied at x. Hence there exists a contant c_w^i such that for all $x_0 \geqslant c_w^i$, $R(x_0, W) \cap \{x \mid x \geqslant x_0\}$ is equal to $\mathscr{L}(x_0, P)$. Moreover c_w^i can be chosen so that $\Pi_j(c_w^i) = 0$, by reordering paths generating the periods of P, so that they are always valid in the jth dimension. We show that c_w^i has a stronger property.

Lemma 3.4. *There exists an effectively computable constant $c_w^i \in \mathbf{N}^n$, $\Pi_j(c_w^i) = 0$ such that for each $x_0 \geqslant c_w^i$, $R(x_0, W) = \mathscr{L}(B, P)$ for some finite, effectively computable B and P.*

Proof. Take c_w^i and P as defined before, and let $x_0 \geqslant c_w^i$. We already have

$$\mathscr{L}(x_0, P) = R(x_0, W) \cap \{x \mid x \geqslant x_0\}.$$

Also if $x \in R(x_0, W)$, then $\mathscr{L}(x, P) \subseteq R(x_0, W)$. To see this, note that if $p \in \mathscr{L}(0, P)$, $x + p = x_0 + p + (x - x_0)$. But $x_0 + p \in R(x_0, W)$ and $x - x_0$ is a valid path at $x_0 + p$ since it is valid at x_0, so $x + p \in R(x_0, W)$. To find B, we are going to close $\mathscr{L}(x_0, P)$ under shifts by vectors of W. (They give semilinear sets by Lemma 1.1.) To do that we construct a tree labelled by points x, where a son is a shift of its father. More precisely:

(i) The root is labelled x_0.

(ii) If x is an unmarked leaf, for all $w \in W$, shift $\mathscr{L}(x, P)$ by w. The shifted set is of the form $\mathscr{L}(D, P)$ for some finite D. If D is empty, mark x.

(iii) Create a son y for each $y \in D$. If $y \geqslant z$ for some ancestor z of y, mark y.

A path in this tree is labelled by a sequence of distinct non increasing points. Assume that along this path the first point is removed, and any point that is incomparable to all previously removed points is also removed. The set of removed points is made of pairwise incomparable points, hence is finite, as mentioned in Section 1. Moreover, all remaining points are smaller than a removed point, hence there are a finite number of them. From this we conclude that any path in the tree is finite, and since the fan out is finite, the tree is finite.

First $\mathscr{L}(B, P) \subseteq R(x_0, W)$ since all labels of the tree are clearly reachable from x_0. Also $\mathscr{L}(B, P)$ contains x_0, so it suffices to show that $\mathscr{L}(B, P)$ is closed under shift by any vector of W. Let x be a vertex of the tree. Either

(i) x is unmarked, then the shift of $\mathscr{L}(x, P)$ is included in $\mathscr{L}(B, P)$ since we create a son y of x for each constant of $\mathscr{L}(D, P)$, the shift of $\mathscr{L}(x, P)$.

(ii) x is marked and the shift of $\mathscr{L}(x, P)$ is empty, or

(iii) x is marked because it has some ancestor y, $y \leqslant x$. But then, since x is reachable from y, $x \geqslant y$, $x \in \mathscr{L}(y, P)$ and $\mathscr{L}(x, P) \subseteq \mathscr{L}(y, P)$ so the shift of $\mathscr{L}(x, P)$ is contained in the shift of $\mathscr{L}(y, P)$.

In each case, the shift of $\mathscr{L}(x, P)$ is included in $\mathscr{L}(B, P)$. $\square$

In the next lemma we give a stronger result, namely the set of points reachable from x_0 through a point having $n-1$ large coordinates is semilinear, no matter where x_0 is.

Lemma 3.5. *There exists an effectively computable constant K^i_w, $\Pi_j(K^i_w)=0$ such that for all x_0 in $\mathbf{N}^n$, the set of points reachable from x_0 through a point y, $y \geqslant K^i_w$, is an effective semilinear set.*

Proof. Let v_j, $\Pi_j(v_j)=0$ define a W-transformation area as in Lemma 3.2. Let c^i_w with $\Pi_j(c^i_w)=0$, be defined as in Lemma 3.4. Define K^i_w by

$$\Pi_j(K^i_w)=0 \quad \text{and} \quad K^i_w = \max(v_j, c^i_w).$$

Let $A=\{x \mid x \geqslant K^i_w\}$. We want to show that the set of points reachable through paths having at least one point in A is semilinear. Pick some starting point x_0, $x_0 \in \mathbf{N}^n$. Note that if $x_0 \in A$, we are done by Lemma 3.4. We can determine the web of x_0 with respect to A, by Lemma 3.1. Let $S=\{p_1, \ldots, p_m\}$ be this web. Let z_i, $i=1, \ldots, m$ be the points of A reached from x_0 by the paths p_i, $i=1, \ldots, m$. By definition of a web the p_i are valid paths at x_0. By Lemma 3.4, $R(z_i, W)$ is an effective semilinear set. We will show that any point reachable through A is also reachable through some z_i.

Let z be reachable from x_0, through some $y \in A$. Let p be a valid path from x_0 to y. By definition of a web, there exists $p_i \in S$ such that $\chi(p) \geqslant \chi(p_i)$, hence y is reachable from z_i by a (nonvalid) path. However A is also a W-transformation area, so this path can be rearranged into a valid path, hence y and z are in $R(z_i, w)$. So the set we are looking for is $\bigcup_{i=1}^{m} R(z_i, W)$, which is semilinear by Lemma 3.4. □

We can generalize this result even more.

Theorem 3.6. *There exists a constant K_w, such that for all $x_0 \in \mathbf{N}^n$, the set of points reachable from x_0, through a point having any $n-1$ coordinates larger than the corresponding coordinates of K_w is an effective semilinear set.*

Proof. Let

$$k = \max_{\substack{1 \leqslant j \leqslant n \\ 1 \leqslant i \leqslant n}} \Pi_i(K^j_w).$$

Then $K_w(k, k, \ldots, k)$ clearly satisfies the theorem since if x has $n-1$ coordinates larger than K_w's (all but the jth one) then $x \geqslant K^j_w$ and Lemma 3.5 applies. □

The interesting point of this theorem is that when we compute a reachability set we can restrict paths to have at least one small coordinate, that is paths running in some finite number of $n-1$ dimensional spaces. We will see next that for paths with at least two small coordinates we can only get a weaker result. Furthermore, these results do not hold at all for paths with 3 small coordinates, since we can simulate states.

We are now concerned with points far from $n-2$ boundaries. We are going to show (Theorem 3.11) that the set of points z reachable from some point x_0 is semilinear when z and x_0 are sufficiently far from the same $n-2$ boundaries. Without loss of generality, we now assume (up to Corollary 3.12) that the $n-2$ large coordinates are the last $n-2$. We first prove (Lemma 3.9) the existence and computability of a constant $C \in \mathbf{N}^n$, $\Pi_1(C)=\Pi_2(C)=0$ such that for all x and y greater than C there is a valid path from x to y if and only if there is a path valid in the first two dimensions. We then prove (Lemma 3.10) that the set of points reachable from some x_0 with a path valid in two dimensions is an effective semilinear set.

We first define C and then prove it has the required properties. Let $B=(b, b, \ldots, b) \in \mathbf{N}^n$ be a constant such that the set of points having $n-1$ coordinates larger than b is a W-transformation area, as in Lemma 3.3. Let $V=\{v_1, v_2, \ldots, v_e\}$ be the set of displacements v_i of paths of length at most b^2, such that $\Pi_1(v_i)=\Pi_2(v_i)=0$ (i.e. the projection of the path generating v_i along the first 2 coordinates is a loop). Find b' such that the set $\{x \mid \Pi_i(x) \geq b', i=3, \ldots, n\}$ is a V-transformation area as in Lemma 3.3. Note that V is equivalent to an $n-2$) dim system. Finally we choose c such that $c \geq 2|\Pi_i(p)|+\max(b, b')$ for all $i=3, \ldots, n$ and any path p of length at most b^2. Let $C=(0, 0, c, \ldots, c)$ and consider a path from x to z, valid in first two dimensions, x and y larger than C.

Lemma 3.7. *If the path from x to z contains a point y such that $\Pi_1(y) \geq b$ or $\Pi_2(y) \geq b$ then the path can be rearranged into a valid path.*

Proof. Let y_1 (possibly equal to x) be the first point along the path such that $\Pi_1(y_1) \geq b$ or $\Pi_2(y_1) \geq b$. Let y_2 (possibly equal to y_1 or z) be the last such point. Now consider the projection of the path along the first two dimensions. $\bar{x}$ means the projection of x.

Reorder the path from x to y, so that it starts with a loop-free path p_1 from $\bar{x}$ to $\bar{y}_1$. In $\mathbf{N}^n$, p_1 goes from x to some point y_1' where $\bar{y}_1' = \bar{y}_1$. Similarly reorder the path from $\bar{y}_2$ to $\bar{z}$ so that it ends with a loop-free path p_2 from $\bar{y}_2$ to $\bar{z}$. In $\mathbf{N}^n$, p_2 goes from some y_2' to z, where $\bar{y}_2' = y_2$. These two reorderings can always be made by shifting loops at one end. Now we have a path from x to y_1' to y_2' to z. p_1 and p_2 are such that their projections are loop-free and remain in the square $[0, b] \times [0, b]$ except for their first and/or last point, hence their length is less than b^2. p_1 and p_2 are valid in the first two dimensions because they are obtained from valid paths by removing loops, and they are valid in the other dimensions because $\Pi_i(x) \geq c$, $\Pi_i(z) \geq c$ and $c \pm \Pi_i(p) \geq 0$ for any path p of length less than b^2, $i=3, \ldots, n$.

Moreover $\Pi_i(y_1')\geq b$, $\Pi_i(y_2')\geq b$ for $i=3,\ldots,n$ from the definition of c and the fact that the length of p_1 and p_2 is at most b^2. Also by assumption $\Pi_1(y_1')\geq b$ or $\Pi_2(y_1')\geq b$ and $\Pi_1(y_2')\geq b$ or $\Pi_2(y_2')\geq b$. Hence y_1' and y_2' have $n-1$ coordinates larger than b, and by Lemma 3.3, the path from y_1' to y_2' can be reordered into a valid path. □

The square $[0,b]\times[0,b]$ is the set of pairs $(x_1,x_2): 0\leq x_1\leq b$, $0\leq x_2\leq b$.

Lemma 3.8. *Assume that the path from x to z is such that its projection along the first two dimensions lies in the square* $[0,b]\times[0,b]$ *and consists of a loop-free path from* $\bar{x}$ *to* $\bar{y}$ *followed by a number of simple loops from* $\bar{y}$ *to* $\bar{y}$, *followed by a loop-free path from* $\bar{y}$ *to* $\bar{z}$. *Then the path can be rearranged into a valid path.*

Proof. Paths from $\bar{x}$ to $\bar{y}$ and from $\bar{y}$ to $\bar{z}$ are of length less than b^2 since they are loop-free. Also simple loops from $\bar{y}$ to $\bar{y}$ are vectors of v. Let y_i be the ith point along the path such that $\bar{y}_i=\bar{y}$, $i=1,\ldots,m$ for some arbitrary m. In the n-dimensional space, we have a path from x to y_1 to $y_2\ldots$ to y_m to z. Also

$$\Pi_i(y_1)+\Pi_i(p)\geq\max(b,b')$$

and

$$\Pi_j(y_m)+\Pi_j(p)\geq\max(b,b')$$

for $j=3,\ldots,n$ and any p of length at most b^2.

Note that there is a V-path from y_1 to y_m and both points are in a V-transformation area, hence the V-path can be rearranged into a valid V-path. In fact we have a stronger property. Since $\Pi_i(y_1)+\Pi_i(p)\geq b'$ and $\Pi_i(y_m)+\Pi_i(p)\geq b'$, all points y' of the valid V-path are such that $\Pi_i(y')+\Pi_i(p)\geq 0$, for $i=3,\ldots,n$, and any p of length at most b^2. Consider the W-path induced by the valid V-path, i.e. consider each vector of V as a path in W. Any point on the W-path is of the form $y'+p$ where y' is on the V path and p is a portion of a simple loop, hence of length at most b^2. So the W-path is also valid in the ith dimension, $i=3,\ldots,n$. Moreover the W-path is still valid in the first two dimensions since we have just reordered loops around $\bar{y}$. Hence we have a valid path from x to z.

Lemma 3.9. *There exists an effective constant* $C\in\mathbb{N}^n$, $\Pi_1(C)=\Pi_2(C)=0$ *such that for all x, z greater than C, if there is a path from x to z valid in the first two dimensions, it can be rearranged into a valid path from x to z.*

Proof. We take C as before and consider a path from x to z valid in the first two dimensions, $x,z\geq C$. We are going to show that we can rearrange the path so that it satisfies either Lemma 3.7 or Lemma 3.8. In both cases, we can then rearrange the path into a valid one.

Assume that the projection of the path along the first two dimensions lies entirely in the square $[0, b]\times[0, b]$. If not, the conditions of Lemma 3.7 are satisfied.

Let y be a point along the path with largest first coordinate. From the path from $\bar{x}$ to $\bar{y}$, we can extract simple loops until we get a loop-free path from $\bar{x}$ to $\bar{y}$. Similarly we decompose the path from $\bar{y}$ to $\bar{z}$ into a number of simple loops and a loop-free path from $\bar{y}$ to $\bar{z}$. All we have to prove is that all the simple loops can be inserted at $\bar{y}$ and be valid (in the first two dimensions). For this we insert each loop at $\bar{y}$ starting with its point of smallest second coordinate. Now consider the position of the loop in the original path and in the reordered path: this is just a shift of vector $\bar{y}-\bar{y}'$ where $\bar{y}'$ is the point of smallest second coordinate in the original position of the loop. Since $\bar{y}$ has largest first coordinate in the original path, $\Pi_i(\bar{y}-\bar{y}')\geq 0$, hence the loop is still valid in the first dimension. Moreover all points of the loop have the second coordinate larger than $\Pi_2(\bar{y})$, hence the loop is still valid in the second dimension.

Once this is done, either one loop goes outside the square $[0, b]\times[0, b]$ and we can apply Lemma 3.7, or all loops remain in the square and we can apply Lemma 3.8. In both cases the path from x to z can be rearranged into a valid path.

Remark. The region $B=\{x\,|\,x\geq C\}$ has a property similar to, but weaker than a W-transformation area. A path from x to z, $x\in B$, $z\in B$, can be reordered into a valid path, but only if it is already valid in two dimensions.

Lemma 3.10. *The set of points reachable from some point x_0 with a path valid in the first two dimensions is an effective semilinear set.*

Proof. Let $W=\{v_1,\ldots,v_k\}$ be the vector addition scheme, $\bar{W}=\{\bar{v}_1,\ldots,\bar{v}_2\}$ and let A be the $n\times k$ matrix whose columns are the vectors of W. If p is a W-path define $y\in\mathbf{N}^k$ by

$$\Pi_i(y)=\Pi_i(\chi(p)).$$

Then a path from x_0 to x must satisfy $x_0+Ay=x$.

We now characterize the projection of such a path along the first two dimensions. Let $S=[0, b]\times[0, b]$ where b defines a $\bar{W}$-transformation area as in Lemma 3.3. Consider a valid path (in these 2 dimensions) from $\bar{x}_0$ to $\bar{x}$ and let $\bar{z}_1$ be the first point out of S ($\bar{z}_1$ possibly equal to $\bar{x}_0$) let $\bar{z}_2$ be the last point out of S ($\bar{z}_2$ possibly equal to $\bar{z}_1$ or $\bar{x}$). If there is no point out of S, let $\bar{z}_1=\bar{z}_2=\bar{x}$. The path from $\bar{x}_0$ to $\bar{z}_1$ is either null or inside S (except for its last point). Hence the set of foldings of such paths can be expressed as a semi-linear set $L_1(\bar{x}_0)$. L_1 is the union of b^2+1 sets, one for each point of S, and the empty set for x_0 outside S. There is a similar set $L_2(\bar{x})$ for paths from $\bar{z}_2$ to $\bar{x}$. But then the conditions:

$$\bar{x}_0+\bar{A}.y=\bar{x},\qquad y=y_1+y_2+y_3,\qquad y_1\in L_1(\bar{x}_0),\qquad y_2\in L_2(\bar{x})$$

are verified if and only if there exists a valid path $\bar{p}$ from $\bar{x}_0$ to $\bar{x}$ with $\chi(\bar{p})=y$. Together with the first condition we get:

$$x_0+A\cdot y=x, \qquad y=y_1+y_2+y_3,$$
$$y_1\in L_1(\bar{x}_0), \qquad y_2\in L_2(\bar{x}),$$

which is true if and only if there exists a path p from x_0 to x, such that $\chi(p)=y$, and the path is valid in the first two dimensions.

The conditions above have a semilinear set of solutions, hence our lemma.

Using Lemmas 3.9 and 3.10 we are able to prove the next theorem.

Theorem 3.11. *There exists a constant C, effectively computable, with $\Pi_1(C)=\Pi_2(C)=0$ such that the set of points z reachable from a given x_0, x_0 and z greater than C, is an effective semilinear set.*

Proof. Let x_0 be greater than C, where C is defined as before. By Lemma 3.9 the set of z's, z greater than C, reachable by a valid path, is the same as the set of z's reachable by a path valid in the first two dimensions. By Lemma 3.10, this set is the intersection of a semilinear set and $\{x \mid x\geqslant C\}$, hence it is semilinear. □

Corollary 3.12. *For the same constant C, and any set of periods P, the set of z's reachable from $\mathscr{L}(x_0, P)$, x_0 and z greater than C, is an effective semilinear set.*

Proof. $R(\mathscr{L}(x_0,P), W)=R(x_0, W\cup P)$. So the only question is whether the constant C is the same for schemes W and $W\cup P$. However if C satisfies Lemma 3.9 for W, it clearly satisfies Lemma 3.9 for $W\cup P$, hence our Corollary. □

We are now ready to prove that a 5-dim VAS has an effective semilinear reachability set. Informally, we can view $\mathbf{N}^5$ as a (nondisjoint) union of 5-dim subspaces, each one far from $n-2$ (i.e. 3) boundaries, together with a finite union of subspaces where three dimensions are bounded. The subspaces where three dimensions are bounded can be simulated by a 2-dim VASS. Since we can compute the reachability set in both types of subspaces (in a somewhat restricted sense), the main problem is with paths crossing from one subspace to another.

By extending Corollary 3.12 to all pairs of coordinates, we know that there exists an effectively computable c such that the set of points z reachable from $\mathscr{L}(x_0, P)$, x_0 and z having the same $n-2$ coordinates larger than c, is an effective semilinear set. Define C_{ij}, $i\neq j$ by $\Pi_k(C_{ij})=c$, $k\neq i$, $k\neq j$ and $\Pi_i(C_{ij})=\Pi_j(C_{ij})=0$. Let

$$R_{ij}=\{x \mid x\geqslant C_{ij}\} \quad \text{and} \quad R=\bigcup_{i,j} R_{ij}.$$

Let

$$\bar{R}_{ijk} = \{x \mid \Pi_i(x) \leq c, \Pi_j(x) \leq c, \Pi_k(x) \leq c\} \quad \text{for } i \neq j \neq k,$$

and let

$$\bar{R} = \bigcup_{i,j,k} \bar{R}_{ijk}.$$

Note that $R \cup \bar{R} = \mathbf{N}^5$, and each $\bar{R}_{ijk}$ is the union of c^3 "parallel" planes, hence $\bar{R}_{ijk}$'s can be simulated by 2-dim VASS, where the states record the ith, jth and kth coordinate.

The *extended intersection* E of the subspaces $\bar{R}_{ijk}$ is the set of points x such that $x \in \bar{R}_{ijk}$ and $x + v \in \bar{R}_{pqr}$ for some transition v, and index $p, q, rp \neq i$ or $q \neq j$ or $k \neq r$. Note that E contains the usual pairwise intersections of the $\bar{R}_{ijk}$'s, plus some other points. A W-path going directly from a region $\bar{R}_{ijk}$ to a region $\bar{R}_{pqr}$ must contain a point of E. Paths can also change region by going through R. E is a finite union of lines parallel to one of the axis (line stands for one dimensional subspace of $\mathbf{N}^5$).

A path *crosses through a line* $l \in E$ if the path has a point in $l \in \bar{R}_{ijk}$ and the next point along the path is in another region $\bar{R}_{pqr}$.

Lemma 3.13. *Let L_0 be a semilinear starting set. There is an effectively computable semilinear set of reachable points including all points reachable from L_0 by paths that never cross through E.*

Proof. We use a tree-based argument: We construct a tree labelled by semilinear sets. Edges correspond to shifts and closures. More precisely:

(1) The root is labelled by L_0. We assume without loss of generality that L_0 lies entirely in a region R_{ij} or $\bar{R}_{ijk}$.

(2) Let L be the label of an unmarked leaf. By induction L is included in a region R_{ij} or $\bar{R}_{ijk}$. Then *either*

(i) $L \subseteq \bar{R}_{ijk}$ for some i, j, k. Add to L the set of points, L', of $\bar{R}_{ijk}$ reachable from L by paths remaining in $\bar{R}_{ijk}$, using a 2-dim VASS to determine them. Then shift L by all vectors of W, getting L''. Create a new son of L, $L_{pq} = L'' \cap R_{pq}$ for each pair p, q. *Or*

(ii) $L \subseteq R_{ij}$ for some i, j. If L has an ancestor included in the same R_{ij}, mark L. Else add to L the set of points L' of R_{ij} reachable from L by any valid path, as guaranteed by Corollary 3.12. Then shift L by all vectors of W, getting L''. Create a new son of L, $L_{pq} = L'' \cap R_{pq}$ or $L_{pqr} = L'' \cap \bar{R}_{pqr}$ for each p, q, r.

Note that the tree is finite since along any path, closure under a region R_{pq}, for fixed p and q, can occur only once, and closures under region $\bar{R}_{ijk}$'s must lie between closures under regions of the form R_{pq}. Clearly the union of the labels of the tree is an effectively computable semilinear set of points reachable from L_0. We have to show that all points reachable without any crossing through E are in this semilinear set.

Consider a path from a point x_0 of L_0 to some z that never crosses through E. We will show that z is one of the sets labelling the tree. Let $x_0, x_1, \ldots, x_m = z$ be the points along the path. x_0 is in the label of the root. Assume that x_i is in a label L and let $x_{i+1} = x_i + v$ be the next point.

First we observe that if L is marked, L is included in some R_{pq}, and there is an ancestor L' of L included in the same R_{pq}, but then there is a point x_j in L' such that x_i is reachable from x_j. However L' is closed under any path from R_{pq} to $R_{pq'}$ hence $x_i \in L'$, L' unmarked. So we may assume that L is unmarked.

If x_{i+1} is in the same region as x_i, then x_{i+1} is in L, since L is closed under paths remaining in the same region.

If x_{i+1} is not in the same region than $x_{i+1} \in (L+v) \cap R_{pq}$ or $x_{i+1} \in (L+v) \cap \bar{R}_{pq2}$ which are labels of sons of L. Our lemma follows, by induction on the length of the path. □

Note that the semilinear set we have constructed may also contain some points reachable by crossing through E. This is because applications of Corollary 3.12 give points reachable by any kind of valid paths.

Let m be the number of lines of E. We are going to show, by induction on i, $0 \leq i \leq m$ that the set of points reachable from L_0 by paths crossing through i lines of E is an effective semilinear set. The previous lemma is the basis, and the result for $i = m$ gives our main Theorem:

Theorem 3.14. *Given a semilinear starting set L_0, the reachability, set of a 5-dim* VAS *is an effective semilinear set.*

Proof. By induction on the number of lines of E crossed.

Basis. This is the previous lemma, 3.13.

Inductive Step. Assume that for any set of k lines of E, $\{l_1, \ldots, l_k\}$, $0 \leq k \leq m-1$, and semilinear set L_0, we can compute a semilinear set L of points reachable from L_0, including all points reachable by paths crossing through $\{l_1, \ldots, l_k\}$.

Consider a set of $k+1$ lines of E, $l_0, l_1, \ldots, l_k$ and a starting set L_0. We are going to alternate closure under paths crossing through $l_1, \ldots, l_k$, and shifts $w \in W$ that create cro sings through l_0. Hence we keep getting new crossing points in l_0 until the whole process is closed under both operations. The problem is to do that in a finite number of steps.

For that we observe that a line $l_0 \in E$ is parallel to some axis, hence is a totally ordered subset of $\mathbf{N}^5$, and that if a path crosses infinitely often through distinct points of l_0, then there must be a pair x, y, $x \in l_0$, $y \in l_0$, $x < y$ and x occurs before y along the path. If so the linear set $\mathscr{L}(x, \{y-x\})$ is reachable. Moreover any successor of x gets the period $y - x$.

Again we are going to use a tree-based argument, since we have to keep track of the ancestors of a given set. Let L_0 be the starting set. By the induction hypothesis, we can compute a set L_1 of points reachable from L_0 including all those reachable by paths crossing only through $l_1, \ldots, l_k$. We intersect l_0 and L_1 and get a semi-linear set. We create a separate tree for each linear component $\mathscr{L}(x_0, P_0)$ of this set.

Consider one tree. The root is labelled $\mathscr{L}(x_0, P_0)$ (P_0 possibly empty). Let $\mathscr{L}(x_j, P_j)$ be an unmarked leaf. We shift $\mathscr{L}(x_j, P_j)$ by any vector of W that causes only crossing through l_0. Let L be the shifted set. Compute L', the set of points reachable from L by crossing through $l_1, \ldots, l_k$, and let L'' be $L' \cap l_0$. For each linear component $\mathscr{L}(x_{j+1}, P_{j+1})$ of L'', create a new son of $\mathscr{L}(x_j, P_j)$. If $\mathscr{L}(x_{j+1}, P_{j+1})$ is contained in the union of its ancestors, mark it. If not, add P_j to P_{j+1}. If P_{j+1} is still empty but $x_{j+1} > x_i$ for some ancestor x_i, add $x_{j+1} - x_i$ to P_{j+1}. Continue this process until all leaves are marked.

We are going to show that this tree is finite. Assume there is an infinite path labelled by $\mathscr{L}(x_j, P_j)$ $j = 0, \ldots, \infty$. If the P_j's remain empty then the x_j's must be strictly decreasing which is impossible. So P_j contains at least one vector v for sufficiently large j's. But by Lemma 1.4, there can be only a finite number of distinct $\mathscr{L}(x_j, P_j)$ in the one dimensional space l_0 hence one of the x_j's should be marked, a contradiction.

We conclude that the tree is finite, hence the process is effective. It should be clear that if L is the union of the labels of the tree, then L contains all points of l_0 reachable from L_0, by paths crossing through $l_0, l_1, \ldots, l_k$. By closing L under paths crossing through $l_1, \ldots, l_k$, we get all the points reachable from L_0, by paths crossing through $l_0, l_1, \ldots, l_k$. Hence the induction hypothesis is true for $k+1$, and by induction, for $k = m$. So our theorem holds.

Since the exponential example of Section 2 can be expressed with a 6-dim VAS (using Lemma 2.1), we have:

Theorem 3.15. *The reachability set of an n-dim VAS is an effect semilinear set for $n \leq 5$, and is not in general semilinear for $n \geq 6$.*

Corollary 3.16. *Reachability, Equivalence, Containment are all decidable for n-dim* VAS, $n \leq 5$.

From these results, we can draw some conclusions. First any new results on VAS must come from at least 6-dimensional systems. However, it is very hard to have any intuition on what can happen in $\mathbf{N}^6$ (but not in $\mathbf{N}^5$!). So the VASS model might be more interesting, since open problems arise as low as dimension 3.

Also, most results so far are based, at least implicitly, on the fact that the reachability set is semilinear. Clearly such approaches cannot be used directly for further results.

References

[1] H.G. Baker, Jr., Rabin's Proof of the Undecidability of the Reachability Set Inclusion Problem of Vector Addition Systems, C.S.G. Memo 79, Project MAC, M.I.T. (1973).

[2] E.W. Cardoza, Computational Complexity of the Word Problem for Commutative Semigroups, MAC Technical Memorandum 67, M.I.T. (October 1975).

[3] S. Ginsburg, *The Mathematical Theory of Context Free Languages* (McGraw-Hill, New York, 1966).

[4] M.H. Hack, The Equality Problem for Vector Addition Systems is Undecidable, C.S.G. Memo 121, Project MAC, M.I.T. (April 1975).

[5] M.H. Hack, Decision Problems for Petri Nets and Vector Addition Systems, MAC Technical Memorandum 53, Project MAC, M.I.T. (March 1975).

[6] A.H. Holt, Final Report of the Information System Theory Project, Technical Report RADC-TR-68-305, Rome Air Development Center, Griffiss Air Force Base, New York (1968).

[7] R.M. Karp and R.E. Miller, Parallel Program Schemata, *J. Comput. Syst. Sci.* **3** (1969) 147–195.

[8] G.S. Sacerdote and R.L. Tenney, The Decidability of the Reachability Problem for Vector Addition Systems, *Ninth Ann. ACM Symp. on Theory of Computing* (1977) 61–76.

[9] J. Van Leeuwen, A Partial Solution to the Reachability Problem for Vector Addition Systems, *Sixth Ann. ACM Symp. on the Theory of Computing* (1974) 303–309.