PACIFIC JOURNAL OF MATHEMATICS
Vol. 16, No. 2, 1966

SEMIGROUPS, PRESBURGER FORMULAS, AND LANGUAGES

SEYMOUR GINSBURG AND EDWIN H. SPANIER

An interesting class of subsets of lattice points in n-space arises naturally in the mathematical theory of (context free) languages. This is the class of semilinear subsets, a subset of lattice points being semilinear if it is the finite union of cosets of finitely generated sub-semigroups of the set of all lattice points with nonnegative coordinates.

The family of semilinear sets is here shown to be equivalent to the family of sets defined by modified Presburger formulas. A characterization of those semilinear sets which correspond to languages is then given. Finally, using the two preceding results and the known decidability of the truth of a modified Presburger sentence, a decision procedure is given for determining whether an arbitrary linear set corresponds to a language.

The class of semilinear sets, first considered in [3] was extensively studied in connection with the theory of bounded languages [1]. In [1] it was shown that the class of semilinear sets is closed with respect to Boolean operations. A consequence of these techniques (in particular, of the proof of Theorem 6.1 of [1]) is that the intersection of two finitely generated sub-semigroups of nonnegative lattice points in n-space is itself a finitely generated sub-semigroup.

The definition of a semilinear set as a finite union is an "internal" description of the set. More precisely, a semilinear set is defined by a finite set of nonnegative lattice points (called *constants*) to each of which is associated a finite set of nonnegative lattice points (called *periods*). The semilinear set is the set generated by adding to each constant an arbitrary finite sequence of its associated periods (allowing repetitions of the same period in the sequence).

Another class of subsets of nonnegative lattice points is defined by the modified Presburger formulas. This class is also closed with respect to Boolean operations [5]. The subsets in this class are defined by an "external" description. More precisely, each set in the class is defined as the extension of a modified Presburger formula with n free variables, i.e., the set of all n-tuples of nonnegative integers satisfying the given formula.

Section 1 contains a proof that the family of semilinear sets is

Received October 9, 1964. This work was supported in part by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract AF 19 (628)-485.

identical with the family of sets defined by modified Presburger formulas. (This was stated without proof in [1].) Thus each set in this family has both an internal and external description. Furthermore, each description can be effectively obtained from the other. The situation is somewhat analogous to the two ways of describing subspaces of a finite dimensional vector space, the internal description for vector spaces consisting of a finite subset of vectors which span the subspace and the external description consisting of a finite system of linear equations whose solution space is the subspace.

Our interest in semilinear sets stems from their relation to languages. Section 2 is devoted to this relation. It is shown that those semilinear sets which correspond to languages can be given semilinear descriptions of a particular form. For the special case of linear sets we then give a decision procedure for determining whether an arbitrary set corresponds to a language. The general case is still unresolved.

1. Semilinear sets and Presburger formulas. Let N denote the set of nonnegative integers and N^n the Cartesian product of N with itself n times. For $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$ in N^n define $x + y = (x_1 + y_1, \cdots, x_n + y_n)$ and, for t in N, define $tx = (tx_1, \cdots, tx_n)$. Then N^n is a semigroup and is partially ordered by the relation $x \leqq y$ if $x_i \leqq y_i$ for $1 \leqq i \leqq n$.

Given subsets C, P of N^n define $L(C; P)$ to be the set of all x in N^n which can be represented in the form

$$x = x_0 + x_1 + \cdots + x_m$$

with x_0 in C and $x_1, \cdots, x_m$ a (possibly empty) finite sequence of elements of P. C is called the set of *constants* and P the set of *periods* of $L(C; P)$. If C consists of a single element c and $P = \{p_1, \cdots, p_r\}$ we write $L(c; P)$ and $L(c; p_1, \cdots, p_r)$ for $L(\{c\}; P)$. A subset L of N^n is said to be *linear* if there exist an element c in N^n and a finite subset P of N^n such that $L = L(c; P)$. In this case P generates a finitely generated sub-semigroup S of N^n and L is the coset of S in N^n containing c. Thus L is linear if and only if it is a coset of a finitely generated sub-semigroup of N^n. A subset of N^n is said to be *semilinear* if it is a finite union of linear sets.

EXAMPLES. (1) In N^2 the set $A = \{(x, y) \mid x \geqq 1\}$ is a linear set, namely $A = L((1, 0); (1, 0), (0, 1))$. Clearly A is a sub-semigroup of N^2. Since no element of the form $(1, y)$ is a sum of other elements of A, A is not finitely generated.

(2) In N^2 the set $X = \{(x, y) \mid y \leqq x^2\}$ is a sub-semigroup which is not semilinear. To see this, note that every vertical line meets X in

a finite set. Thus each linear set contained in X can only have periods $\neq(0, 0)$ of the form (x, y) with $x > 0$. Let $L(c_1; P_1), \cdots, L(c_m; P_m)$ be a finite sequence of linear sets contained in X and let

$$M = \max \{y/x \mid (x, y) \text{ in } \cup P_j, (x, y) \neq (0, 0)\} .$$

Then the slope of the line joining any two points of $L(c_j; P_j)$ is $\leqq M$. If z_1, z_2 in N are such that $M \leqq z_1 < z_2$, then the slope of the line joining (z_1, z_1^2) and (z_2, z_2^2) is $z_1 + z_2 > 2M > M$. Therefore the set $L(c_j; P_j)$, $1 \leqq j \leqq m$, can contain at most one element of the set $\{(z, z^2) \mid z \geqq M\}$. Since $\{(z, z^2) \mid z \geqq M)$ is an infinite subset of X, it follows that $X \neq \cup_1^m L(c_j; P_j)$. Thus X is not semilinear.

From Theorem 6.1, Corollary 1 of Theorem 6.2, and Lemma 6.3 of [1] there follows

THEOREM 1.1. *The family of semilinear sets of N^n is closed with respect to union, intersection, and complementation. The projection of a semilinear set is semilinear.*

We now consider formulas (= statements) about nonnegative integers. If P is such a formula and has $n \geqq 1$ free variables $x_1, \cdots, x_n$, we also write it as $P(x_1, \cdots, x_n)$. *The set of Presburger formulas,*[1] denoted by $\mathscr{P}$, is the smallest class of formulas satisfying the following five conditions:

(a) For given nonnegative integers t_i, t_i', $0 \leqq i \leqq n$,

$$t_0 + \sum_1^n t_i x_i = t_0' + \sum_1^n t_i' x_i$$

is a formula in $\mathscr{P}$.

(b) If P_1, P_2 are in $\mathscr{P}$, so is their conjunction $P_1 \wedge P_2$.

(c) If P_1, P_2 are in $\mathscr{P}$, so is their disjunction $P_1 \vee P_2$.

(d) If P is in $\mathscr{P}$, so is its negation $\sim P$.

(e) If $P(x_1, \cdots, x_n)$ is in $\mathscr{P}$ and $1 \leqq i \leqq n$, then the formula $(\exists x_i)P(x_1, \cdots, x_n)$ is in $\mathscr{P}$.

A *Presburger sentence* is defined to be a Presburger formula with no free variables. One of the main results is the following [2].

THEOREM 1.2. *It is decidable whether an arbitrary Preburger sentence is true.*

REMARKS. (1) The formula

[1] To be precise, this is the set of *modified* Presburger formulas. The original Presburger formulas [4] were defined over all the integers.

$$t_0 + \sum_1^n t_i x_i \leqq t'_0 + \sum_1^n t'_i x_i$$

is regarded as a Presburger formula since it is equivalent to the Presburger formula

$$(\exists x_{n+1})\left(t_0 + \sum_1^n t_i x_i = t'_0 + \sum_1^n t'_i x_i + x_{n+1}\right).$$

(2) If $P(x_1, \cdots, x_n)$ is a Presburger formula, then the formula

$$(x_i)P(x_1, \cdots, x_n)$$

is regarded as a Presburger formula since it is equivalent to

$$\sim (\exists x_i)(\sim P(x_1, \cdots, x_n)).$$

Similarly if P and Q are Presburger formulas, $P \Rightarrow Q$ is regarded as a Presburger formula.

For $n > 0$, a set A is a *Presburger set in* N^n *if*

$$A = \{(x_1, \cdots, x_n) \mid P(x_1, \cdots, x_n) \text{ is true}\}$$

for some $P(x_1, \cdots, x_n)$ in $\mathscr{P}$. A *Presburger set* is a Presburger set in N^n for some n.

It follows from (b), (c), and (d) that the family of Presburger sets in N^n is closed under intersection, union, and complementation. From (e), the projection of a Presburger set is a Presburger set. Our interest in Presburger sets is due to the following result.

THEOREM 1.3. *The family of Presburger sets of N^n is identical with the family of semilinear sets of N^n. Furthermore, each description is effectively calculable from the other.*

Proof. It is obvious that every linear set, thus every semilinear set, in N^n is a Presburger set. To see the reverse, by Theorem 1.1 it suffices to show that the set of nonnegative solutions of

$$t_0 + \sum_1^n t_i x_i = t'_0 + \sum_1^n t'_i x_i \tag{1}$$

is semilinear. Let C be the set of minimal solutions in N^n of (1) and let P be the set of minimal solutions in $N^n - 0^n$ [2] of the associated homogeneous equation

$$\sum_1^n t_i x_i = \sum_1^n t'_i x_i . \tag{2}$$

Then C, P are finite (by the corollary of Lemma 6.1 of [1]) and effec-

[2] We use 0^n to denote $(0, \cdots, 0)$ in N^n.

tively calculable (by Lemma 6.5 of [1]). It follows from the method of proof of Theorem 6.1 of [1] that the Presburger set defined by the equation in (1) is $\cup_{c \text{ in } \sigma} L(c; P)$. This gives the result.

2. Semilinear sets and languages. Let us recall the basic ideas associated with context free languages. A *grammar* G is a 4-tuple (V, Σ, Q, σ) where V is a finite set, Σ is a subset of V, σ is an element of $V - \Sigma$, and Q is a finite set of ordered pairs of the form (ξ, w) with ξ in $V - \Sigma$ and w a string over V. (ξ, w) in Q is denoted by $\xi \rightarrow w$. For strings y, z over V, we write $y \Rightarrow z$ if $y = u\xi v$, $z = uwv$, and $\xi \rightarrow w$. We write $y \overset{*}{\Rightarrow} z$ if either $y = z$ or if there exists a sequence of strings $z_0, \cdots, z_r$, called a *derivation* of $y \overset{*}{\Rightarrow} z$, such that $y = z_0$, $z_r = y$, and $z_i \Rightarrow z_{i+1}$ for each i. The *language generated by* G, denoted by $L(G)$, is the set of strings over Σ, $\{w \mid \sigma \overset{*}{\Rightarrow} w\}$. A *context free language* (over Σ) is a language $L(G)$ generated by some grammar $G = (V, \Sigma, Q, \sigma)$. By a language we shall always mean a context free language.

Let $a_1, \cdots, a_n$ be distinct letters and let $a_1^* \cdots a_n^*$ denote the set of all words $w = a_1^{i_1} \cdots a_n^{i_n}$, $0 \leqq i_j$ for $j = 1, \cdots, n$. Let τ be the *Parikh mapping* of $a_1^* \cdots a_n^*$ into N^n defined by

$$\tau(a_1^{i_1} \cdots a_n^{i_n}) = (i_1, \cdots, i_n) .$$

Clearly τ is one-to-one. If $Z \subseteqq a_1^* \cdots a_n^*$ is a language, then $\tau(Z)$ is a semilinear subset of N^n [3]. We are interested in characterizing those semilinear sets $L \subseteqq N^n$ such that $\tau^{-1}(L)$ is a language.

A subset X of N^n is said to be *stratified* if the following two conditions are satisfied:

(a) Each element in X has at most two nonzero coordinates.

(b) There are no integers, i, j, k, m, with $1 \leqq i < j < k < m \leqq n$ and $x = (x_1, \cdots, x_n)$, $x' = (x', \cdots, x'_n)$ in X such that $x_i x'_j x_k x'_m \neq 0$. In other words, no two elements X have nonzero coordinates which "interlace."

LEMMA 2.1. *If $Z \subseteqq a_1^* \cdots a_n^*$ is a language, then $\tau(Z)$ can be represented as a finite union of linear sets each of which has a stratified set of periods.*

Proof. We prove the lemma by induction on n. For $n = 1$ or 2 every subset of N^n is stratified, and the result is valid. Let $n \geqq 3$ and assume the lemma holds for languages in $a_1^* \cdots a_{n-1}^*$. By Lemma 2.5 of [1] every language in $a_1^* \cdots a_n^*$ is a finite union of sets of the form

$$L(D, E, F) = \{a_1^i uva_n^j \mid a_n^j \text{ in } D, u \text{ in } E, v \text{ in } F\}$$

where D, E, F are languages in $a_1^* a_n^*$, $a_1^* \cdots a_q^*$, $a_q^* \cdots a_n^*$ respectively, and $1 < q < n$. Hence it suffices to prove the result for such a set $L(D, E, F)$.

Let τ': $a_1^* \cdots a_q^* \to N^q$ and τ'': $a_q^* \cdots a_n^* \to N^{n-q+1}$ be the appropriate Parikh mappings. Let μ be the mapping of $a_1^* a_n^*$ into N^n defined by $\mu(a_1^i a_n^j) = (i, 0, \cdots, 0, j)$. By the induction hypothesis, $\tau'(E)$ is a finite union of sets of the form $L(c'; P')$ where c' is in N^q and P' is a stratified subset of N^q. Also by the induction hypothesis, $\tau''(F)$ is a finite union of sets of the form $L(c''; P'')$ where c'' is in N^{n-q+1} and P'' is a stratified subset of N^{n-q+1}. Since $\mu(D)$ is semilinear, $\mu(D)$ is a finite union of sets of the form $L(c; P)$ where c is in N^n and every element of P has zero as the ith coordinate, $1 < i < n$. Consequently $\tau(L(D, E, F))$ is a finite union of sets of the form

$$L((c' \times 0^{n-q}) + (0^{q-1} \times c'') + c; (P' \times 0^{n-q}) \cup (0^{q-1} \times P'') \cup P) .$$

Since P', P'' are stratified, so are $P' \times 0^{n-q}$ and $0^{q-1} \times P''$. Now each element of $P' \times 0^{n-q}$ has its nonzero coordinates in the set $\{1, \cdots, q\}$, each element of $0^{q-1} \times P''$ has its nonzero coordinates in the set $\{q, \cdots, n\}$, and each element of P has its nonzero coordinates in the set $\{1, n\}$. Therefore $(P' \times 0^{n-q}) \cup (0^{q-1} \times P'') \cup P$ is a stratified set.

We now prove the converse.

LEMMA 2.2. *If the subset L of N^n is a finite union of linear sets each of which has a stratified set of periods, then $\tau^{-1}(L)$ is a language.*

Proof. It suffices to prove the result for a set $L = L(c; P)$ where P is a stratified set. We do this by induction on n. If $n = 1$ or 2, then each subset P of N or N^2 is stratified. From the corollary to Lemma 2.2 of [1], it follows that $\tau^{-1}(L(c; P))$ is a language if P is finite.

Assume $n > 2$ and that the lemma is true for $1 \leqq m < n$. We prove the result for $L(c; P)$ by induction on the number of periods in P. If P is empty, then $L(c; P)$ consists solely of c. Therefore $\tau^{-1}(L(c; P))$ is finite and hence a language. Assume P is nonempty and consider the following two cases.

Suppose P contains a period p whose first and nth coordinates p_1, p_n are both nonzero. Then $P' = P - \{p\}$ is stratified and has fewer elements than P. Therefore $\tau^{-1}(L(c; P'))$ is a language. Let $G' = (V', \Sigma, Q', \sigma')$ be a grammar generating $\tau^{-1}(L(c; P'))$. Let $G = (V, \Sigma, Q, \sigma)$ where σ is an element not in V', $V = V' \cup \{\sigma\}$, and $Q = Q' \cup \{\sigma \to \sigma', \sigma \to a_1^{p_1} \sigma a_n^{p_n}\}$.

Clearly $L(G) = \tau^{-1}(L(c; P))$, so that $\tau^{-1}(L(c; P))$ is a language.

Suppose that P contains no period having nonzero first and nth coordinates. If every period in P having a nonzero first coordinate has all of its other coordinates zero, let $q = 2$. If there exists a period in P having nonzero first and jth coordinate with $j > 1$, let q be the largest such j. Clearly $1 < q < n$. From the way q is chosen and the fact that P is stratified, it follows that every element in P either has zero ith coordinate for all $q < i \leqq n$ or zero ith coordinate for all $1 \leqq i < q$. Let P' be the set of those elements in P having zero ith coordinate for each $q < i \leqq n$ and $P'' = P - P'$. Let $c = (c_1, \cdots, c_n)$, $c' = (c_1, \cdots, c_q, 0, \cdots, 0)$, and $c'' = (0, \cdots, 0, c_{q+1}, \cdots, c_n)$. Clearly $c = c' + c''$ and

$$L(c; P) = L(c'; P') + L(c''; P'').^{3}$$

Let Π', Π'' be the projections of N^n onto N^q, N^{n-q+1} defined by projecting to the first q, last $n - q + 1$ coordinates respectively. Let $\tau'\colon a_1^* \cdots a_q^* \rightarrow N^q$ and $\tau''\colon a_q^* \cdots a_n^* \rightarrow N^{n-q+1}$ be the appropriate Parikh mappings. Then

$$L(c'; P') = L(\Pi'(c'); \Pi'(P')) \times 0^{n-q}$$

and

$$L(c''; P'') = 0^{q-1} \times L(\Pi''(c''); \Pi''(P'')) .$$

Therefore

$$\tau^{-1}(L(c; P)) = [\tau'^{-1}(L(\Pi'(c'); \Pi'(P')))][\tau''^{-1}(L(\Pi''(c''); \Pi''(P'')))] .$$

Since P is stratified, $\Pi'(P)'$ and $\Pi''(P'')$ are stratified subsets of N^q and N^{n-q+1} respectively. From the induction hypothesis,

$$\tau'^{-1}(L(\Pi'(c'); \Pi'(P'))), \quad \tau''^{-1}(L(\Pi''(c''); \Pi''(P'')))$$

are languages contained in $a_1^* \cdots a_q^*, a_q^* \cdots a_n^*$ respectively. Since the product of languages is a language, $\tau^{-1}(L(c; P))$ is a language. Hence the lemma.

On combining Lemmas 2.1 and 2.2 we obtain

THEOREM. 2.1. *Given a subset L of N^n, $\tau^{-1}(L)$ is a language if and only if L can be represented as a finite union of linear sets each having a stratified set of periods.*

We do not have a procedure for deciding whether a given semilinear subset L of N^n satisfies the condition of Theorem 2.1. The following results are directed toward a decision procedure for the special case when L is a linear subset of N^n.

[3] For $X, Y \subseteq N^n$, $X + Y = \{x + y \mid x \text{ in } X, y \text{ in } Y\}$.

LEMMA 2.3. *Let* $M = L(c_1; P_1) \cup \cdots \cup L(c_r; P_r)$ *be a semilinear subset of* N^n *and let* $L(c; p)$ *be a linear subset of* N^n, *with one period, which meets* M *infinitely often. Then there exists* $1 \leqq i \leqq r$ *and a positive integer* k *such that* kp *is a sum of positive multiples of some elements of* P_i.

Proof. Since $L(c; p)$ meets M infinitely often, there exists $1 \leqq i \leqq r$ such that $L(c; p)$ meets $L(c_i; P_i)$ infinitely often. Let $P - 0^n = \{q_1, \cdots, q_m\}$ and consider the set

$$X = \{(s, t_1, \cdots, t_m) \text{ in } N^{m+1} \mid c + sp = c_i + \sum_1^m t_j q_j\} .$$

By assumption, X is infinite. By Lemma 6.1 of [1] there exist distinct elements $(s, t_1, \cdots, t_m)$ and $(s', t'_1, \cdots, t'_m)$ in X such that $s \leqq s'$ and $t_j \leqq t'_j$ for $1 \leqq j \leqq m$. Then

$$(s' - s)p = \sum_1^m (t'_j - t_j)q_j .$$

Thus $s < s'$ or $t_j < t'_j$ for some j. In either case, $k = s' - s$ is positive and kp is a sum of positive multiples of some elements of P_i.

LEMMA 2.4. *Let* X *be a stratified subset of* N^n *and* Y *a subset of* N^n. *If for every* y *in* Y *there exists* x *in* X *and a positive integer* k *such that* $kx \geqq y$, *then* Y *is stratified.*

Proof. If $kx \geqq y$ then y can have nonzero coordinates only where x has nonzero coordinates. From this and the fact that X is stratified, it follows that Y is stratified.

LEMMA 2.5. *Let* L *be a linear subset of* N^n *with set of periods* P *and let* L' *be a linear subset of* L *with stratified periods. Then there exists a finite subset* F *of* L *and a stratified subset* Y *of* P *such that* $L' \subseteqq L(F; Y) \subseteqq L$.

Proof. Let $L' = L(c; X)$ where X is a finite set of stratified periods. Let Y be the set of all y in P having the property that there exists x in X and a positive integer k such that $kx \geqq y$. By Lemma 2.4, Y is stratified. For each x in X, since $L(c; x) \subseteqq L$ it follows from Lemma 2.3 that there is a positive integer k such that kx is a sum of positive multiples of some elements of P. Note that these elements of P are in Y. Let $X = \{x_1, \cdots, x_m\}$ and for $1 \leqq i \leqq m$ let k_i be a positive integer such that $k_i x_i$ is a sum of positive multiples of some elements of Y. Let F be the finite set

$$F = \left\{c + \sum_1^m t_i x_i \mid 0 \leqq t_i < k_i\right\}.$$

We complete the proof by showing that $L' \subseteqq L(F; Y) \subseteqq L$. Now each element of L' is of the form $c + \sum_1^m s_i x_i$. For each i there exists r_i and $0 \leqq t_i < k_i$ such that $s_i = r_i k_i + t_i$. Thus

$$c + \Sigma s_i x_i = c + \Sigma t_i x_i + \Sigma r_i k_i x_i$$

is in $L(F; Y)$. Hence the first inclusion holds. The second inclusion holds from the fact that $F \subseteqq L$ and $Y \subseteqq P$.

COROLLARY 1. *A linear set $L = L(c; P)$ is a finite union of linear sets with stratified periods if and only if it is a finite union of linear sets each of whose periods form a stratified subset of P.*

Proof. If $L = L_1 \cup \cdots \cup L_m$ where each L_i is a linear set with stratified periods, it follows from Lemma 2.5 that there exist finite subsets $F_1, \cdots, F_m$ of L and stratified subsets $Y_1, \cdots, Y_m$ of P such that $L = \bigcup_1^m L(F_i; Y_i)$. The corollary follows from this and the fact that, for $1 \leqq i \leqq m$, $L(F_i; Y_i)$ is a finite union of linear sets with set of periods Y_i.

In case the periods are linearly independent (as vectors over the rationals) we obtain the following result.

COROLLARY 2. *Let L be a linear subset of N^n with a linearly independent set of periods P. Then $\tau^{-1}(L)$ is a language if and only if P is stratified.*

Proof. If P is stratified, it follows from Theorem 2.1 that $\tau^{-1}(L)$ is a language. We prove the converse. If $\tau^{-1}(L)$ is a language, it follows from Theorem 2.1 and Corollary 1 above that $L = \bigcup_1^r L_i$ where each L_i is a linear set whose periods form a stratified subset of P. Let $L = L(c; P)$, with $P = \{p_1, \cdots, p_m\}$. Then $L(c; p_1 + \cdots + p_m) \subseteqq L$. By Lemma 2.3 there exist $1 \leqq i \leqq r$ and a positive integer k such that $k(p_1 + \cdots + p_m)$ is a sum of positive multiples of some periods of L_i. Thus $k(p_1 + \cdots + p_m) = t_1 p'_1 + \cdots + t_s p'_s$, where each p'_j is a period of L_i and $t_j > 0$. Since $\{p'_1, \cdots, p'_s) \subseteqq \{p_1, \cdots, p_m\} = P$ and P is linearly independent, $\{p'_1, \cdots, p'_s\} = P$ and $k = t_j$ for each j. Since $\{p'_1, \cdots, p'_s\}$ is stratified, so is P.

COROLLARY 3. *Let $L = L(c; P)$ be a linear subset of N^n. If $\tau^{-1}(L)$ is a language, then for every period p with more than two nonzero coordinates there is a positive multiple kp which is a sum of positive multiples of some stratified periods of L.*

Proof. By Corollary 1, $L(c; P) = \bigcup_1^r L(c_i; P_i)$, each P_i a stratified subset of P. By Lemma 2.3, there exists $1 \leqq i \leqq r$ and a positive integer k such that kp is a sum of positive multiples of some elements of P_i.

EXAMPLES. (1) We give a simple proof of Theorem 3.2 of [1]. That is, we show that if $L \subseteqq \{(i, j, k)\} \mid 0 \leqq i \leqq j,\ 0 \leqq k \leqq j\}$ and $L \cap L((1, 1, 1); (1, 1, 1))$ is infinite, then $\tau^{-1}(L)$ is not a language. Suppose the contrary, that is, suppose $\tau^{-1}(L)$ is a language. Then $L = \bigcup_{s=1}^r L_s$ where each L_s is a linear set with stratified periods. Since $L((1, 1, 1); (1, 1, 1))$ meets L infinitely often, by Lemma 2.3 there exists s such that some positive multiple of $(1, 1, 1)$ is a sum of positive multiples of some nonzero periods $p_1, \cdots, p_m$ of L_s. Let

$$p_1 = (p_{11}, p_{12}, p_{13}), \cdots, p_m = (p_{m1}, p_{m2}, p_{m3}) .$$

Then there exist positive integers $t, t_1, \cdots, t_m$ such that $t(1, 1, 1) = \sum_{h=1}^r t_h p_h$. Thus

$$\sum_1^m t_h p_{h1} = \sum_1^m t_h p_{h2} = \sum_1^m t_h p_{h3} .$$

Since $L \subseteqq \{(i, j, k) \mid 0 \leqq i \leqq j \text{ and } 0 \leqq k \leqq j\}$ it readily follows that

$$p_{h1} \leqq p_{h2} \text{ and } p_{h3} \leqq p_{h2}$$

for $1 \leqq h \leqq m$. Hence $p_{h1} = p_{h2} = p_{h3}$ for each h. Therefore $p_h = (p_{h1}, p_{h2}, p_{h3})$ has three nonzero coordinates, contradicting the condition that $\{p_1, \cdots, p_m\}$ is stratified.

(2) Let L be the linear set in N^3 with constant $(0, 0, 0)$ and periods $(1, 1, 1)$, $(1, 0, 0)$, $(0, 2, 3)$. Since these periods are linearly independent but not stratified, it follows from Corollary 2 above that $\tau^{-1}(L)$ is not a language.

(3) The set $X = \{a^i b^j c^i d^j \mid 0 \leqq i, j\}$ is not a language since $\tau(X) = L((0, 0, 0, 0,); (1, 0, 1, 0), (0, 1, 0, 1))$ whereas $(1, 0, 1, 0), (0, 1, 0, 1)$ are linearly independent and not stratified.

We now use Corollary 1 to obtain a decision procedure for determining of a linear set L whether $\tau^{-1}(L)$ is a language.

THEOREM 2.2. *It is decidable to determine of an arbitrary linear set L whether $\tau^{-1}(L)$ is a language.*

Proof. By Theorem 2.1 we are reduced to showing that it is decidable whether L is a finite union of linear sets each having stratified periods. Let $S_1, \cdots, S_m$ be all the stratified subsets of the periods of L. By Corollary 1, L is a finite union of linear sets with stratified periods if and only if there exist finite (possibly empty) subsets F_i of L for $1 \leqq i \leqq m$ such that $L = \bigcup_1^m L(F_i; S_i)$. Then $F = \bigcup^m F_i$ is finite. Since $F \subseteqq L$ and $S_i \subseteqq P$, $L = \bigcup^m L(F; S_i)$. Hence, by Theorem 1.2, we need only show there is a Presburger sentence whose truth is equivalent to the condition that there exists a finite subset F of L such that $L \subseteqq \bigcup^m L(F; S_i)$.

By Theorem 1.3 there is a Presburger formula $P(x)$ for L, where $x = (x_1, \cdots, x_n)$. If S_i consists of the elements

$$y_1^i = (y_{11}^i, \cdots, y_{1n}^i), \cdots, y_{r(i)} = (y_{r(i)1}^i, \cdots, y_{r(i)n}^i)$$

let $Q_i(z, x)$, where $z = (z_1, \cdots, z_n)$ be the Presburger formula

$$(\exists t_1) \cdots (\exists t_{r(i)}) \bigwedge_1^n \left(x_k = z_k + \sum_1^{r(i)} t_j y_{kj}^i\right).$$

Then $Q_i(z, x)$ is a Presburger formula with $2n$ free variables. The corresponding Presburger set is the set of all $2n$-tuples (z, x) such that x is in $L(z; S_i)$. It follows that the Presburger sentence

$$(\exists M_1) \cdots (\exists M_n)(x_1) \cdots (x_n)$$
$$\left[P(x) \Rightarrow (\exists z_1) \cdots (\exists z_n)\left[P(z) \wedge \bigwedge_1^n (z_i \leqq M_i) \wedge \bigvee_1^n Q_i(z, x)\right]\right]$$

is true if and only if $L \subseteqq \bigcup_1^m L(F; S_i)$ for the finite set

$$F = L \cap \{z \mid z \leqq (M_1, \cdots, M_n)\}.$$

Hence the result.

Our final result provides a condition for deciding for an arbitrary semilinear set L whether $\tau^{-1}(L)$ contains an infinite language.

THEOREM 2.3. *Given a semilinear subset L of N^n, $\tau^{-1}(L)$ contains an infinite language if and only if whenever L is represented as a finite union of linear sets one of them has a period with exactly one or two nonzero coordinates.*

Proof. If $L(c; P) \subseteqq L$ where P contains an element p having exactly one or two nonzero coordinates, then $L(c; p) \subseteqq L$. Therefore $\tau^{-1}(L(c; p)) \subseteqq \tau^{-1}(L)$. Clearly $\tau^{-1}(L(c; p))$ is infinite and, by Theorem 2.1, is a language. We prove the converse.

If $\tau^{-1}(L)$ contains an infinite language, it follows from Theorem 2.1 that L contains a set of the form $L(c; p)$ where p has exactly one or two nonzero coordinates. Given a representation of L as a union of linear sets $L_1, \cdots, L_m$, it follows from Lemma 2.3 that there exists $1 \leqq i \leqq m$ such that some positive multiple of p is a sum of positive multiples of periods of L_i. Then L_i has a period having exactly one or two nonzero coordinates.

The problem of finding a decision procedure for determining of an arbitrary semilinear set L whether $\tau^{-1}(L)$ is a language is open.

BIBLIOGRAPHY

1. S. Ginsburg and E. Spanier, *Bounded ALGOL-Like Languages*, Trans. Amer. Math. Soc. **113** (1964). 333-368.

2. D. Hilbert and P. Bernays, *Grundlagen der Mathematik*, Edward Brothers, Inc., Ann Arbor, Michigan, 1944.

3. R. J. Parikh, *Language-Generating Devices*, Quarterly Progress Report No. 60, Research Laboratory of Electronics, Massachusetts Institute of Technology, January 1961, pp. 199-212.

4. M. Presburger, *Über die Vollstandigkeit eines gewissen Systems der Arithmetic ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt*, Sprawozdanie z I Kongresu Matematykow Krajow Slowianskich, Warsaw 1930, pp. 92-101.

5. A. Robinson and E. Zakon, *Elementary Properties of Ordered Abelian Groups*, Trans. Amer. Math. Soc. **96** (1960), 222-236.

SYSTEM DEVELOPMENT CORPORATION, SANTA MONICA, CALIFORNIA
AND UNIVERSITY OF CALIFORNIA, BERKELEY