BOUNDED REGULAR SETS[1]

SEYMOUR GINSBURG AND EDWIN H. SPANIER

Introduction. Earlier [2] we introduced the bounded context free languages (also called "bounded ALGOL-like languages"). (The context free languages are good approximations to the syntactic classes which arise in currently used programming languages.) A most important class of context free languages is the class of regular sets.[2] Because of this it seems appropriate to consider the class of bounded regular sets. In this note we give two characterizations of these sets. One of these characterizations is related to a characterization of bounded context free languages given in [3]. We also relate certain bounded regular sets to their commutative closures.

1. Two characterizations. Let Σ be a finite nonempty set and Σ^* the free semigroup with identity ϵ generated by Σ. We shall deal with subsets of Σ^*. If A is a subset of Σ^*, we use A^* to denote the subsemigroup with identity (contained in Σ^*) generated by A. In case A consists of a single word w, we also write w^* for this subsemigroup.

A set X is said to be *bounded* if there exist a finite number of words $w_1, \cdots, w_k$ such that $X \subseteq w_1^* \cdots w_k^*$.[3]

We shall be concerned with bounded regular sets. In this section we present two characterizations of bounded regular sets.

A set X is said to be *commutative* if $xy = yx$ for all x, y in X.

LEMMA 1.1. *Let Z be the smallest family of sets which contain all finite sets, all sets w^* (w in Σ^*), and which is closed with respect to finite union and finite product. Then Z contains every commutative regular set.*

PROOF. Let R be a commutative regular set. By Lemma 5.2 of [2], there exists a word w such that $R \subseteq w^*$. Since R is regular, the set of integers $R' = \{n \mid w^n \text{ in } R\}$ is ultimately periodic.[4] Therefore

Received by the editors February 4, 1966.

[1] The research reported in this paper was sponsored in part by the Air Force Cambridge Reseach Laboratories, Office of Aerospace Research, under Contract AF 19(628)-5166, CRL-Algorithmic Languages Program.

[2] For the definitions of regular set and automaton, as well as their interrelations, see [9].

[3] If $X_1, \cdots, X_k$ are sets of words, then $X_1 \cdots X_k$, called the *product*, is the set $\{x_1 \cdots x_k \mid \text{each } x_i \text{ in } X_i\}$.

[4] In case w is a letter, this is proved in [6]. For arbitrary w, the result follows from the case for letters and the fact that the inverse of a homomorphism maps a regular set to a regular set [1].

there exist finite sets I, J of nonnegative integers and a positive integer k such that

$$R' = I \cup \{j + ik \mid j \text{ in } J,\ i \geqq 0\}.$$

Let $F = \{w^i \mid i \text{ in } I\}$. Then F is finite. For each j in J let $R_j = w^j(w^k)^*$. Since $R = F \cup \bigcup_{j \text{ in } J} R_j$, R is in Z.

COROLLARY. *Z is the smallest family of sets which contains all finite sets, all sets R^*, R a commutative regular set, and which is closed with respect to finite union and finite product.*

PROOF. Let Z' be the smallest family of sets which contain all finite sets, all sets R^*, R a commutative regular set, and which is closed with respect to finite union and finite product. Clearly $Z \subseteq Z'$. To prove the converse, it suffices to show that Z contains all sets R^*, R a commutative regular set. Let R be a commutative regular set. Then there exists a word w such that $R \subseteq w^*$. Since $R^* \subseteq (w^*)^* = w^*$, R^* is a commutative regular set. By Lemma 1.1, Z contains R^*.

We now derive our first characterization result.

THEOREM 1.1. *R is a bounded regular set if and only if R is in the family Z of Lemma* 1.1.

PROOF. Obviously every set in Z is regular. By Theorem 3.1 of [2], every set in Z is also bounded. Thus each set in Z is bounded regular.

To see the converse, let R be a bounded regular set. Then, since R is regular, there exists a finite sequence $X_1, \cdots, X_m$ of families of sets such that

(1) X_1 is a finite collection of finite sets.

(2) R is in X_m.

(3) For $2 \leqq i \leqq m$, X_i is obtained by adjoining to X_{i-1} a set A_i which is either the union or the product of two sets in X_{i-1}, or is E^* for some set E in X_{i-1}. For each i let Y_i be the collection of bounded sets in X_i. Then $Y_1 = X_1$. Since R is bounded, R is in Y_m.

Consider Y_i, $2 \leqq i \leqq m$. It is clear that either $Y_i = Y_{i-1}$ or $Y_i = Y_{i-1} \cup \{A_i\}$. If A_i is not bounded, then $Y_i = Y_{i-1}$. If A_i is bounded, then either $A_i = B_1 \cup B_2$ or $A_i = B_1B_2$ or $A_i = B_3^*$ for B_1, B_2 (or B_3) in X_{i-1}. In the first two cases, both B_1 and B_2 must be bounded, thus in Y_{i-1}. In the last case, since A_i is bounded, it follows from Lemma 5.3 of [2] that B_3 is commutative. In any case, either $Y_i = Y_{i-1}$ or Y_i is obtained from Y_{i-1} by adjoining a set A_i which is either the union or the product of two sets in Y_{i-1} or is C^* for some commutative regular set C in Y_{i-1}. Thus R is in Z by the corollary of Lemma 1.1.

COROLLARY. *R is bounded regular if and only if R is in the smallest family of sets containing all regular subsets of w^*, w in Σ^*, and closed with respect to finite union and finite product.*

PROOF. Let W be the smallest family of sets containing all regular subsets of w^*, w in Σ^*, and closed with respect to finite union and finite product. Then W contains Z of Lemma 1.1. Since each set in W is also bounded regular, $W=Z$.

We now derive our second characterization of bounded regular sets.

LEMMA 1.2. *Let $\Sigma=\{a_1, \cdots, a_n\}$. A subset X of $a_1^* \cdots a_n^*$ is regular if and only if it is a finite union of sets of the form*

(1) $A_1 \cdots A_n$,

where each A_i is a regular subset of a_i^*.

PROOF. If X is such a finite union, it is clearly regular. We prove the converse. Since multiplication is distributive with respect to union, it follows from the previous corollary that each bounded regular set is a finite union of sets of the form

(2) $B_1 \cdots B_m$,

where each B_i is a regular subset of w_i^* for some w_i in Σ^*. Thus any regular subset of $a_1^* \cdots a_n^*$ is a finite union of sets of the form (2). Obviously each corresponding word w_i is either a power of some a_j or is a product $a_{i_1} \cdots a_{i_k}$, with $i_1 \leqq i_2 \leqq \cdots \leqq i_k$. Thus each set of the form (2) can be written as

(3) $C_1 \cdots C_r$,

where each C_i is a regular subset of $a_{f(i)}^*$ for some $f(i)$, $1 \leqq f(i) \leqq n$. For each i, $1 \leqq i \leqq n$, let

(4) $A_i = C_{i1} \cdots C_{ig(i)}$,

the C_{ij} being just those C_k which are subsets of a_i^*. Let $A_i=\{\epsilon\}$ if there is no i such that $C_i \subseteq a_i^*$. Then

(5) $A_1 \cdots A_n = C_1 \cdots C_r$,

so that X has the prescribed form.

Using the language of [3], in particular, the notion of a linear set, we can reinterpret the lemma as follows.

COROLLARY. *Let $\Sigma=\{a_1, \cdots, a_n\}$. Then $R \subseteq a_1^* \cdots a_n^*$ is regular if and only if $\{(i(1), \cdots, i(n)) \mid a_1^{i(1)} \cdots a_n^{i(n)} \text{ in } R\}$ is a finite union of linear sets L_k, each period in each L_k having at most one nonzero coordinate.*

Using the lemma we now prove

THEOREM 1.2. *A subset X of $w_1^* \cdots w_r^*$ is regular if and only if it is a finite union of sets of the form $X_1 \cdots X_r$, where each X_i is a regular subset of w_i^*.*

PROOF. Let $b_1, \cdots, b_r$ be r distinct symbols not in Σ. Let h be the homomorphism which maps each b_i into w_i. Since the inverse of a homomorphism maps a regular set to a regular set [1], $h^{-1}(X)$ is regular. Thus $Y = h^{-1}(X) \cap b_1^* \cdots b_r^*$ is regular. Clearly

$$Y = \left\{ b_1^{i(1)} \cdots b_r^{i(r)} \,\middle|\, w_1^{i(1)} \cdots w_r^{i(r)} \text{ in } X \right\}.$$

By Lemma 1.2, Y is a finite union of sets of the form $Y_1 \cdots Y_r$, each Y_i a regular subset of b_i^*. Then $X = h(Y)$ is a finite union of sets of the form $h(Y_1) \cdots h(Y_r)$, each $h(Y_i)$ a regular subset of w_i^*.

From the corollary to Lemma 1.2 and the fact that

$$Y = \left\{ b_1^{i(1)} \cdots b_r^{i(r)} \,\middle|\, w_1^{i(1)} \cdots w_r^{i(r)} \text{ in } X \right\}$$

is regular we get the following which is related to Theorem 2.1 of [3].

THEOREM 1.3. *Let $X \subseteq w_1^* \cdots w_r^*$ each w_i in Σ^*. A necessary and sufficient condition that X be regular is that*

$$\left\{ (i(1), \cdots, i(r)) \,\middle|\, w_1^{i(1)} \cdots w_r^{i(r)} \textit{ in } X \right\}$$

be a finite union of linear sets L_k, each period in each L_k having at most one nonzero coordinate.

2. **Commutative closure.** We now give a condition for a subset X of $a_1^* \cdots a_n^*$, the a_i being distinct symbols, to be regular in terms of the "commutative closure" of X.

DEFINITION. The *commutative closure* $c(X)$ of $X \subseteq \Sigma^*$ is the set of all words $x_1 \cdots x_k$, each x_i in Σ, such that for some permutation τ of $\{1, \cdots, k\}$ the word $x_{\tau(1)} \cdots x_{\tau(k)}$ is in X. ($c(X)$ contains ϵ if and only if X contains ϵ.)

In general, X may be regular without $c(X)$ being regular. For example, let $\Sigma = \{a, b\}$. Then $X = (ab)^*$ is regular but $c(X)$ is not regular. For if $c(X)$ were regular, then $c(X) \cap a^*b^* = \{a^n b^n \mid n \geqq 0\}$ would be regular, a contradiction.

THEOREM 2.1. *Let $\Sigma = \{a_i \mid 1 \leqq i \leqq n\}$ and X a subset of $a_1^* \cdots a_n^*$. Then X is regular if and only if $c(X)$ is regular.*

PROOF. Suppose $c(X)$ is regular. Then $X = a_1^* \cdots a_n^* \cap c(X)$.

Since the intersection of regular sets is regular, X is regular.

Suppose X is regular. By Lemma 1.2, X is a finite union of sets of the form $X_1 \cdots X_n$, where each X_i is a regular subset of a_i^*. It thus suffices to prove that if X is of the form $X_1 \cdots X_n$, each X_i a regular subset of a_i^*, then $c(X)$ is regular. For each i, let $A_i = (K_i, \Sigma, \delta_i, s_{i0}, F_i)$ be an automaton such that $T(A_i) = X_i$. Let A be the automaton $(K, \Sigma, \delta, s_0, F)$, where $K = K_1 \times \cdots \times K_n$, $s_0 = (s_{10}, \cdots, s_{n0})$, $F = F_{10} \times \cdots \times F_{n0}$, and

$$\delta((s_1, \cdots, s_n), a_i) = (s_1, \cdots, s_{i-1}, \delta_i(s_i, a_i), s_{i+1}, \cdots, s_n).$$

It is readily seen that a word w in Σ^* is in $T(A)$ if and only if, for each i, $a_i^{h_i(w)}$ is in $T(A_i)$ where $h_i(w)$ is the number of occurrences of a_i in w. Therefore $T(A) = c(X)$, so that $c(X)$ is regular.

Note that Theorem 2.1 does not state that the commutative closure of a bounded regular set is regular. In fact, the example prior to Theorem 2.1 furnishes a bounded regular set whose commutative closure is not regular.

3. **Decidability.** We now consider the problem of deciding whether a given set is a bounded regular set. It follows from Theorem 5.2 of [2] that it is decidable whether or not a given regular set is bounded. We show that it is decidable of a given semilinear[5] subset L of N^n whether $\tau^{-1}(L)$ is regular (where τ is the function which maps a word $a_1^{i_1} \cdots a_n^{i_n}$ of $a_1^* \cdots a_n^*$ into the n-tuple $(i_1, \cdots, i_n)$ of N^n). In view of Theorem 2.1, this is equivalent to the condition that $c(\tau^{-1}(L))$ is regular.

We first obtain a necessary and sufficient condition on L for $c(\tau^{-1}(L))$ to be regular. For this we introduce the following relation on elements of N^n.

Notation. Let L be a subset of N^n. For x, x' in N^n, write $x \equiv^L x'$ if, for all u in N^n,

$$x + u \text{ is in } L \text{ if and only if } x' + u \text{ is in } L.$$

Clearly $\equiv^L$ is an equivalence relation.

Lemma 3.1. *Let L be a subset of N^n. Then $c(\tau^{-1}(L))$ is regular if and only if there are a finite number of equivalence classes generated by the equivalence relation $\equiv^L$.*

Proof. Let $X = c(\tau^{-1}(L))$. Now X is regular if and only if there are a finite number of equivalence classes of Σ^* generated by the following equivalence relation $\sim^X$: $w \sim^X w'$ if and only if, for all words

[5] The reader is referred to [3] for the notation and concepts used in this section.

w'' in Σ^*, ww'' is in X if and only if $w'w''$ is in X [**9**]. Since X is the commutative closure of $\tau^{-1}(L)$ it follows that if w is any word in Σ^* and w' is an arbitrary permutation of w, then $w \sim^X w'$. Therefore every equivalence class has a representative which is a word in $a_1^* \cdots a_n^*$. Thus X is regular if and only if there are a finite number of equivalence classes of $a_1^* \cdots a_n^*$ generated by $\sim^X$.

To complete the proof we verify that for w and w' in $a_1^* \cdots a_n^*$, $w \sim^X w'$ if and only if $\tau(w) \equiv^L \tau(w')$. Let w and w' be in $a_1^* \cdots a_n^*$. Suppose $w \sim^X w'$. Let u be an arbitrary element of N^n and y in $\tau^{-1}(u)$. Then $\tau(w)+u$ is in L if and only if some permutation of wy is in $\tau^{-1}(L)$, and $\tau(w')+u$ is in L if and only if some permutation of $w'y$ is in $\tau^{-1}(L)$. Therefore $\tau(w)+u$ (or $\tau(w')+u$) is in L if and only if wy (or $w'y$) is in X. Thus $\tau(w) \equiv^L \tau(w')$. Conversely, suppose $\tau(w) \equiv^L \tau(w')$. Then for each element y in $a_1^* \cdots a_n^*$, wy is in X if and only if $w'y$ is in X. If z is an arbitrary element of Σ^*, then some permutation of z is an element y in $a_1^* \cdots a_n^*$. Since X is a commutative closure, wz (or $w'z$) is in X if and only if wy (or $w'y$) is in X. Therefore $w \sim^X w'$ and the proof is complete.

Corollary. *It is decidable to determine for an arbitrary semilinear set L in N^n whether $c(\tau^{-1}(L))$ is regular.*

Proof. We use a method found in [**5**]. By Theorem 1.3 of [**3**], a Presburger formula $P(x_1, \cdots, x_n)$ over the nonnegative integers can be effectively found such that $L = \{(x_1, \cdots, x_n) \text{ in } N^n \mid P(x_1, \cdots, x_n) \text{ is true}\}$. Let Q be the following Presburger sentence:

$$(\exists k)(x_1) \cdots (x_n)(\exists x_1') \cdots (\exists x_n')$$

$$\left[\bigwedge_{i=1}^{n} (x_i' \leq k) \wedge (y_1) \cdots (y_n)(P(x_1 + y_1, \cdots, x_n + y_n) \right.$$

$$\left. \equiv P(x_1' + y_1, \cdots, x_n' + y_n)) \right].$$

Then Q is true if and only if there are a finite number of equivalence classes of N^n generated by $\equiv^L$. By Lemma 3.1, this is equivalent to the condition that $c(\tau^{-1}(L))$ is regular. Since the truth of every Presburger sentence is decidable [**4**], the corollary holds.

We now obtain the following result also due to [**5**] and [**8**].

Theorem 3.1. *It is decidable to determine for an arbitrary regular set R whether $c(R)$ is regular.*

PROOF. Let $R \subseteq \Sigma^*$ be regular and ψ the Parikh mapping of Σ^* into R. Then $L = \psi(R)$ is semilinear [7]. Since $c(R) = c(\tau^{-1}(L))$, the result follows from the preceding corollary.

REFERENCES

1. S. Ginsburg and G. F. Rose, *Operations which preserve definability in languages*, J. Assoc. Comput. Mach. **10** (1963), 175–195.

2. S. Ginsburg and E. H. Spanier, *Bounded* ALGOL-*like languages*, Trans. Amer. Math. Soc. **113** (1964), 333–368.

3. ———, *Semigroups, Presburger formulas, and languages*, Pacific J. Math. **16** (1966), 285–296.

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

5. R. Laing and J. B. Wright, *Commutative machines*, Tech. Rep., Univ. of Michigan, Ann Arbor, Mich., December 1962.

6. J. Myhill, *Finite automata and the representation of events*, WADC Tech. Rep. 57-624, 1957, pp. 112–137.

7. R. J. Parikh, *Language-generating devices*, Quarterly Progress Rep. No. 60, Research Laboratory of Electronics, January 1961, pp. 199–212, Massachusetts Institute of Technology, Cambridge, Mass.

8. V. N. Redko, *On the commutative closure of events*, Dopovidi Akad. Nauk Ukrain. RSR (1963), 1156–1159.

9. M. O. Rabin and D. Scott, *Finite automata and their decision problems*, IBM J. Res. Develop. **3** (1959), 114–125.

SYSTEM DEVELOPMENT CORPORATION AND
UNIVERSITY OF CALIFORNIA AT BERKELEY