Theoretical Computer Science 627 (2016) 71–81

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

The effect of end-markers on counter machines and commutativity

Oscar H. Ibarra [a,1], Ian McQuillan [b,*,2]

[a] Department of Computer Science, University of California, Santa Barbara, CA 93106, USA
[b] Department of Computer Science, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form 24 February 2016
Accepted 25 February 2016
Available online 2 March 2016
Communicated by D. Perrin

Keywords:
Counter machines
Commutativity
Reversal-bounds
Determinism
Finite automata

ABSTRACT

Restrictions of reversal-bounded multicounter machines are studied; in particular, those that cannot subtract from any counter until it has reached the end of the input. It is proven that this does not alter the languages accepted when the machines are nondeterministic. When the machines are deterministic, the languages (denoted by eDCM) are shown to coincide with those accepted by deterministic Parikh automata, but are strictly contained in the class of languages accepted by machines without this condition. It then follows that all commutative semilinear languages are in this restricted class. A number of decidability and complexity properties are shown, such as the ability to test, given a deterministic pushdown automaton (even if augmented by a fixed number of reversal-bounded counters), whether it is commutative. Lastly, this deterministic family, eDCM, is shown to be the smallest family of languages closed under commutative closure, right quotient with regular languages and inverse deterministic finite transductions.

1. Introduction

The *commutative closure* of a language L, $\text{comm}(L)$, is the language of all words obtained by permuting the positions of the letters of all words in L. A language L is then *commutative* if $\text{comm}(L) = L$. The *Parikh map* of a word (and a language respectively) is the vector representing the number of copies of each letter in the word (the set of Parikh vectors of all words). These provide an equivalent criterion for commutative closure; $\text{comm}(L)$ is the set of all words with the same Parikh vector as a word of L. Thus, studying the set of Parikh vectors is closely related to commutativity. It was found by Parikh [1] that every context-free language has a so-called *semilinear* (defined formally below in Section 2) Parikh map. The semilinear criteria can be equivalently expressed as, every language with a semilinear Parikh map has the same commutative closure as a regular language [2]. However, it is quite easy to create commutative languages that are not regular (nor context-free), such as $\{w \mid w$ has the same number of a's, b's and c's$\}$.

There is a model of automata that can accept every commutative semilinear language; namely the family of one-way nondeterministic reversal-bounded multicounter languages (NCM) [3,4]. In [5], it was shown that NCM is in fact the smallest trio (closed under λ-free homomorphism, inverse homomorphism and intersection with regular languages) that is also closed under taking commutative closures. NCM is equal to the family of languages accepted by another model, Parikh au-

* Corresponding author.
E-mail addresses: ibarra@cs.ucsb.edu (O.H. Ibarra), mcquillan@cs.usask.ca (I. McQuillan).
1 Supported, in part, by NSF Grant CCF-1117708 (Oscar H. Ibarra).
2 Supported, in part, by Natural Sciences and Engineering Research Council of Canada Grant 327486-2010 (Ian McQuillan).

http://dx.doi.org/10.1016/j.tcs.2016.02.034
0304-3975/

tomata [6]. It has also been shown that languages accepted by deterministic Parikh automata are closed under commutative closure.

Commutative semilinear languages (referred to as COM-SLIP in [7]) have also been studied. Since every semilinear language has the same commutative closure as a regular language, and the fact that all regular languages can be accepted by deterministic Parikh automata, it follows that COM-SLIP is contained inside the deterministic Parikh languages. The family of commutative semilinear languages has been extended to their closure under union and concatenation (referred to as COM-SLIP$^{\cdot,\cup}$ [7]), and these languages are also strictly contained in NCM, since NCM is closed under union and concatenation.

In this paper, one-way *deterministic* reversal-bounded multicounter languages (DCM) are studied, and a new restriction is introduced, eDCM, that are DCM machines that cannot subtract from any counter until hitting the end of the input. It is shown that this new family coincides with deterministic Parikh automata, and it therefore follows that both families are strictly contained inside DCM, and all commutative semilinear languages are contained in both. We then use the eDCM model to demonstrate a new language that can be accepted in DCM with only one counter that makes one counter reversal that is not in eDCM.

As these families are contained in the family of DCM languages, we explore a number of decidability and complexity properties that NCM does not have, such as decidable containment and equivalence problems. Several properties of commutative semilinear languages become easily decidable. For example, it is possible to test for either containment or equivalence between the commutative closures of any effectively semilinear languages (or between the commutative closure of any effectively semilinear language and an arbitrary DCM language). It is also shown that it is possible to decide whether an arbitrary DPCM language (a language accepted by a deterministic machine that has an unrestricted pushdown plus a fixed number of reversal-bounded counters) is commutative, and similarly for other deterministic automata models accepting semilinear languages. Also, testing membership in DCM is computable in logarithmic space on a deterministic Turing machine, and thus complexity theoretic results are presented for Turing Machines accepting semilinear languages. It is then shown that the concatenation closure of commutative semilinear languages is not always a DCM language, and therefore, the COM-SLIP$^{\cdot,\cup}$ languages are not contained in DCM; they are incomparable.

Finally, it is shown that eDCM (and hence deterministic Parikh automata) is the smallest family of languages closed under commutative closure, right quotient with regular languages, and inverse deterministic finite transductions. Such a characterization of a family of languages involving deterministic automata (that do not coincide with nondeterministic automata) using closure properties is somewhat unusual and is of interest.

2. Preliminaries

We assume familiarity with formal language and automata theory [8], and computational complexity theory [9]. We will fix the notation used in the paper. Let Σ be a finite alphabet. Then Σ^* (respectively Σ^+) is the set of all words (non-empty words) over Σ. A *word* is an element $w \in \Sigma^*$, λ is the *empty word*, and a *language* is any $L \subseteq \Sigma^*$. The *complement* of $L \subseteq \Sigma^*$ is $\overline{L} = \Sigma^* - L$. A language $L \subseteq \Sigma^*$ is *bounded* if there exists (not necessarily distinct) words $w_1, \dots, w_k$ such that $L \subseteq w_1^* \cdots w_k^*$. Further, L is *letter-bounded* if there exists (not necessarily distinct) letters $a_1, \dots, a_k$ such that $L \subseteq a_1^* \cdots a_k^*$. For $L, R \subseteq \Sigma^*$, the *right quotient* of L by R is $LR^{-1} = \{x \mid xy \in L, y \in R\}$.

Let $\mathbb{N}$ be the set of positive integers and $\mathbb{N}_0$ be the set of non-negative integers. Let $m \in \mathbb{N}_0$. Then $\pi(m)$ is 1 if $m > 0$ and 0 otherwise. Let $m \in \mathbb{N}$. Then $\mathbb{N}(m) = \{1, \dots, m\}$. A subset Q of $\mathbb{N}_0^m$ is a *linear set* if there exist vectors $\vec{v}_0, \vec{v}_1, \dots, \vec{v}_n \in \mathbb{N}_0^m$ such that $Q = \{\vec{v}_0 + i_1\vec{v}_1 + \cdots i_n\vec{v}_n \mid i_1, \dots, i_n \in \mathbb{N}_0\}$. The vectors $\vec{v}_0$ (referred to as the *constant*) and $\vec{v}_1, \dots, \vec{v}_n$ (referred to as *periods*) are called the *generators* of the linear set Q. A finite union of linear sets is called a *semilinear set*. Every finite subset of $\mathbb{N}_0^m$ (including the empty set $\emptyset$) is semilinear — it is just a finite union of linear sets with no periods. For semilinear sets $Q_1, Q_2 \subseteq \mathbb{N}^m$, $Q_1 + Q_2 = \{v \mid v = v_1 + v_2, v_1 \in Q_1, v_2 \in Q_2\}$.

Let $\Sigma = \{a_1, \dots, a_m\}$. For a word w over Σ and a letter $a \in \Sigma$, we denote by $|w|_a$ the number of occurrences of a's in w, and by $|w|$ the length of w. The *Parikh map* of w is the m-dimensional vector $\psi(w) = (|w|_{a_1}, \dots, |w|_{a_m})$. The *Parikh map* of a language $L \subseteq \Sigma^*$ is defined as $\psi(L) = \{\psi(w) \mid w \in L\}$. For $\vec{v} \in \mathbb{N}_0^m$, the inverse, $\psi^{-1}(\vec{v}) = \{w \in \Sigma^* \mid \psi(w) = \vec{v}\}$, extended to subsets of $\mathbb{N}_0^m$. A language is *semilinear* if its Parikh map is a semilinear set. The *commutative closure* of $L \subseteq \Sigma^*$ is the set

$$\mathrm{comm}(L) = \psi^{-1}(\psi(L)),$$

and a language L is said to be *commutative* if $L = \mathrm{comm}(L)$.

Of interest are languages that are both commutative and semilinear. For example, $L_1 = \{w \mid w \in \{a,b\}^*, |w|_a = 2|w|_b\}$ and $L_2 = \{w \mid w \in \{a,b,c\}^*, 2|w|_a - 5|w|_b > 4|w|_c\}$ are commutative semilinear languages.

A *one-way k-counter machine* is denoted by $M = (k, Q, \Sigma, \lhd, \delta, q_0, F)$, where $Q, \Sigma, \lhd, q_0, F$ are the set of states, input alphabet, right input end-marker not in Σ, initial state in Q, and accepting states that are a subset of Q. The transition function δ is a relation from $Q \times (\Sigma \cup \{\lhd\}) \times \{0,1\}^k$ into $Q \times \{\mathrm{S}, \mathrm{R}\} \times \{-1, 0, +1\}^k$, such that if $\delta(q, a, c_1, \dots, c_k)$ contains $(p, d, d_1, \dots, d_k)$ and $c_i = 0$ for some i, then $d_i \geq 0$ to enforce that the counters cannot store negative numbers. The symbols S and R indicate the direction that the input tape head moves, either *stay* or *right*. Further, M is *deterministic* if δ is a partial function. A configuration of M is a $k+2$-tuple $(q, w\lhd, c_1, \dots, c_k)$ representing that M is in state q, with $w \in \Sigma^*$ still to read as input, and $c_1, \dots, c_k \in \mathbb{N}_0$ being the contents of the k counters. The derivation relation $\vdash_M$ is defined between