A Lower Bound for the Nondeterministic Space Complexity of Context-free Recognition

Helmut Alt* Viliam Geffert† Kurt Mehlhorn‡

April 26, 2002

Abstract

We prove a $\log n$ lower bound on the nondeterministic space complexity of every nonregular deterministic context-free language.

1 Introduction

We show that the nondeterministic space complexity of every nonregular deterministic context-free language is at least $\log n$. Previously, this was only known for the deterministic space complexity [Alt79, AM76].

For a function $s : \mathcal{N} \to \mathcal{N}$ call a nondeterministic Turing machine $s(n)$-*space-bounded* if for all $n \in \mathcal{N}$ all computations on inputs of length n use at most $s(n)$ cells on the worktape and call it *weakly-$s(n)$-space-bounded* if for all $n \in \mathcal{N}$ and all accepted inputs of length n there is at least one accepting computation path that uses at most $s(n)$ cells on the worktape. Let $NSPACE(s(n))$ be the class of languages that are accepted by $s(n)$-space-bounded nondeterministic machines and let $WEAKNSPACE(s(n))$ be the class of languages that are accepted by weakly-$s(n)$-space-bounded nondeterministic machines. The following theorem is essential for the proof of our lower bound.

Theorem 1 *Let $s : \mathcal{N} \to \mathcal{N}$ be any function, let $\{a_1, \dots, a_k\}$ be a finite set of symbols, and let $L \subseteq a_1^* \dots a_k^*$ be a language. Then $L \in NSPACE(s(n))$ implies $\overline{L} \in NSPACE(cs(n))$ for some constant c.*

Theorem 1 is a partial extension of the Immermann-Szelepscényi result [Imm88, Sze88] to space bounds below $\log n$. The extension is partial because it only applies to languages L which are subsets of $a_1^* \dots a_k^*$ for some finite set $\{a_1, a_2, \dots, a_k\}$ of symbols.

A language L is called *deterministic context-free*, if it is accepted by a deterministic pushdown automaton. A language L is called *strictly nonregular*,

*Fachbereich Mathematik, Freie Universität Berlin, Arnimallee 2-6, W1000 Berlin 33, Germany

†Department of Computer Science, Jesenna 5, 04154 Kosice, Czechoslovakia

‡Max-Planck-Institut für Informatik and Fachbereich Informatik der Universität des Saarlandes, W 6600 Saarbrücken, Germany

if there are strings u, v, w, x, and y such that $L \cap uv^*wx^*y$ is context-free and nonregular. A context-free language L is *bounded* if there are strings $w_1, \dots, w_k$ such that $L \subseteq w_1^* w_2^* \dots w_k^*$.

For functions $f : \mathcal{N} \to \mathcal{N}$ and $g : \mathcal{N} \to \mathcal{N}$ we write $f = \Omega^*(g)$ if there is a constant $c > 0$ such that $f(n) \geq c \cdot g(n)$ for infinitely many n.

Theorem 2 *Let $s : \mathcal{N} \to \mathcal{N}$ be any function, let L be a nonregular deterministic context-free, a strictly nonregular language, or a nonregular bounded language, and let $L \in NSPACE(s(n))$. Then $s = \Omega^*(\log n)$.*

Theorem 2 generalizes the following

Theorem 3 *[Alt79] Let $s : \mathcal{N} \to \mathcal{N}$ be any function, let L be a nonregular deterministic context-free, a strictly nonregular language, or a nonregular bounded language, and let $L, \overline{L} \in WEAKNSPACE(s(n))$. Then $s = \Omega^*(\log n)$.*

We prove both theorems in section 2. Section 3 discusses the results and puts them into context.

2 Proofs

For a function $f : \mathcal{N} \to \mathcal{N}$, a language $L \subseteq \{0,1\}^*$ is called $f(n)$-*zero-bounded* if every $w \in L$ of length n contains at most $f(n)$ zeroes. A function $s : \mathcal{N} \to \mathcal{N}$ is called *fully space constructible* if there is a deterministic Turing machine which on every input of length n uses exactly $s(n)$ cells of the worktape.

In [Gef90] the following partial extension of the Immermann-Szelepcsényi result to space bounds below $\log n$ was shown:

Theorem 4 *[Gef90, Theorem 3] Let $s : \mathcal{N} \to \mathcal{N}$ be fully space constructible, let $d > 0$ be a constant, let $f : \mathcal{N} \to \mathcal{N}$ be such that $f(n) \leq d^{s(n)}$ for all n, and let $L \in \{0,1\}^* \cap NSPACE(s(n))$ be $f(n)$-zero-bounded. Then $\overline{L} \in NSPACE(cs(n))$ for some constant $c > 0$.*

The proof of Theorem 4 in [Gef90] actually works under slightly weaker hypotheses. It is only required that the value $f(n)$ can be computed using no more than $O(s(n))$ space; the full space constructibility of s is not needed. Assume that $f(n) = k$ for some constant k and all n. Then the value $f(n)$ can clearly be computed in $O(s(n))$ space, no matter what the function s is. We have thus established the following

Lemma 1 *Let $s : \mathcal{N} \to \mathcal{N}$ be any function, let $k \in \mathcal{N}$ be a constant, and let $L \in \{0,1\}^* \cap NSPACE(s(n))$ be k-zero-bounded. Then $\overline{L} \in NSPACE(cs(n))$ for some constant $c > 0$.*

Theorem 1 is a direct consequence of Lemma 1. Let $L \subseteq a_1^* a_2^* ... a_k^*$. Define

$$M_1 = \{1^{i_1} 0 1^{i_2} ... 0 1^{i_k} ; a_1^{i_1} a_2^{i_2} ... a_k^{i_k} \in L\}.$$

Then M_1 is $(k-1)$-zero-bounded and $L \in NSPACE(s(n))$ implies $M_1 \in NSPACE(s(n))$. Thus $\overline{M_1} \in NSPACE(c_1 s(n))$ according to Lemma 1 and hence $\overline{L} \in NSPACE(c_2 s(n))$ for some constants $c_1, c_2 > 0$. This proves Theorem 1.

We now turn to the proof of Theorem 2. Assume first that $L \in NSPACE(s(n))$ is a nonregular bounded context-free language, i.e., $L \subseteq w_1^* w_2^* \dots w_k^*$ for some strings $w_1, w_2, \dots, w_k$. Let $a_1, \dots, a_k$ be k distinct symbols. Define

$$M_2 = \{a_1^{i_1} \dots a_k^{i_k}; w_1^{i_1} \dots w_k^{i_k} \in L\}.$$

Then M_2 is nonregular and context-free and belongs to $NSPACE(c_1 s(n))$ for some constant $c_1 > 0$. Hence $\overline{M_2} \in NSPACE(c_2 s(n))$ for some constant $c_2 > 0$ according to Theorem 1. But then $M_2, \overline{M_2} \in WEAKSPACE(cs(n))$ for some constant c, since $NSPACE(s(n)) \subseteq WEAKNSPACE(s(n))$ for all bounds $s(n))$, and therefore $s = \Omega^*(\log n)$ according to Theorem 3.

Assume next that $L \in NSPACE(s(n))$ is strictly nonregular, i.e., there are strings u, v, w, x, y such that $L' = L \cap uv^*wx^*y$ is context-free and nonregular. Clearly, $L' \in NSPACE(s(n))$. Also, L' is bounded and hence $s = \Omega^*(\log n)$ according to the preceding paragraph.

Assume finally, that $L \in NSPACE(s(n))$ is nonregular and deterministic context-free. Then L is strictly nonregular as Stearns [Ste67] has shown. Thus $s = \Omega^*(\log n)$. This completes the proof of Theorem 2.

3 Discussion

We have shown a logarithmic lower bound on the nondeterministic space complexity of every nonregular deterministic context-free language. The lower bound does not extend to $WEAKNSPACE$. In fact, in [AM76] it was observed that the nonregular deterministic context-free language

$$L_1 = \{a^n b^m; n, m \in \mathcal{N}, n \neq m\}$$

belongs to $WEAKNSPACE(\log \log n)$.

Table 1 summarizes the current knowledge about the space complexity of context-free languages. Every nonregular deterministic context-free language requires nondeterministic space $\Omega^*(\log n)$ [this paper] and every nonregular context-free language requires nondeterministic space $\Omega^*(\log \log n)$ [LHS65, HU69]. The language

$$L_2 = \{a^n b^n; n \in \mathcal{N}\}$$

is nonregular deterministic context-free (also bounded and strictly nonregular) and belongs to $DSPACE(\log n)$. The language

$$L_3 = \{0, 1, \#\}^* \setminus \{bin(1)\#bin(2)^R\# \dots \#bin(n)^{(R)}; n \in \mathcal{N}\},$$

where $bin(i)$ is the binary representation of integer i and w^R denotes the reversal of string w, is nonregular context-free and belongs to $DSPACE(\log \log n)$ [LHS65, HU69]. Finally, every context-free language belongs to $DSPACE(\log^2 n)$ [LHS65, HU69].

Type of Bound	nonregular det. CFL	nonregular CFL
existential upper	$\boldsymbol{\log n}$	$\boldsymbol{\log\log n}$
universal upper	$\log^2 n$	$\log^2 n$
existential lower	$\log n$	$\log n$
universal lower	$\boldsymbol{\log n}$	$\boldsymbol{\log\log n}$

Table 1: Current bounds for the space complexity of context-free recognition. All upper bounds are for deterministic machines, all lower bounds are for nondeterministic machines. Entries shown in bold are known to be tight.

References

[Alt79] H. Alt. Lower bounds on space complexity for context-free recognition. *Acta Informatica*, 12:33 – 61, 1979.

[AM76] H. Alt and K. Mehlhorn. Lower bounds for the space complexity of context-free recognition. In *Proceedings 3rd ICALP*, pages 338 – 354. Edinburgh University Press, 1976.

[Gef90] V. Geffert. Tally versions of Savitch's and Immermann-Szelepcsényi theorems for sublogarithmic space. SIAM Journal of Computing, to appear.

[HU69] J.E. Hopcroft and J.D. Ullman. *Formal Languages and their Relation to Automata*. Addison-Wesley Publ. Comp., 1969.

[Imm88] N. Immermann. Nondeterministic space is closed under complementation. *Siam Journal of Computing*, 17:935 – 938, 1988.

[LHS65] P.M. Lewis, J. Hartmanis, and R.E. Stearns. Memory bounds for the recognition of context-free and context-sensitive languages. *IEEE Conference Record on Switching Circuit Theory and Logical Design*, pages 179 – 202, 1965.

[Ste67] R.E. Stearns. A regularity test for pushdown-machines. *Information and Control*, pages 323 – 340, 1967.

[Sze88] R. Szelepcsényi. The method of forced enumeration for nondeterministic automata. *Acta Informatica*, 26:279 – 284, 1988.