A Pumping Lemma for Pushdown Graphs of Any Level

Paweł Parys*

University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland, parys@mimuw.edu.pl

Abstract

We present a pumping lemma for the class of ε-contractions of pushdown graphs of level n, for each n. A pumping lemma was proposed by Blumensath, but there is an irrecoverable error in his proof; we present a new proof. Our pumping lemma also improves the bounds given in the invalid paper of Blumensath.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases pushdown graph, ε-contraction, pumping lemma

Digital Object Identifier 10.4230/LIPIcs.STACS.2012.54

1 Introduction

Higher-order pushdown systems are a very natural extension of pushdown systems. They were originally introduced by Maslov [10]. In a system of level n we have a level-n stack of level-$(n-1)$ stacks of ... of level-1 stacks. The idea is that the system operates only on the topmost level-1 stack, but additionally it can make a copy of the topmost stack of some level, or can remove the topmost stack of some level. Higher-order pushdown systems have connections with several other concepts. A result of Knapik et al. [9] shows that higher-order pushdown systems generate the same trees as *safe* higher-order recursion schemes. Carayol and Wöhrle [2] proved that the ε-contractions of graphs generated by higher-order pushdown systems are exactly the graphs in the Caucal hierarchy [3]. Thus, all these graphs have decidable monadic second-order theories.

Even though higher-order pushdown systems generate important classes of graphs, useful characterizations of their structure are still rare. We still miss techniques for disproving membership in the pushdown hierarchy. In classical automata theory, pumping lemmas provide good tools for proving that a language cannot be defined by a finite automaton or by a pushdown automaton. For indexed languages, which are the languages recognized by pushdown systems of level 2, we have a pumping lemma of Hayashi [6], and a shrinking lemma of Gilman [4]. We also have a pumping lemma of Kartzow [7] for collapsible pushdown systems of level 2. On higher levels, similar results are still missing. Blumensath [1] published a pumping lemma for all levels of the higher-order pushdown hierarchy. Unfortunately, there is an irrecoverable error in his proof (cf. [11]).

Our main theorem is the following pumping lemma applicable to every level of the higher-order pushdown graph hierarchy.

▶ **Theorem 1.1.** *Let $\mathcal{A}$ be a pushdown system of level n, and L a regular language. Let G be the ε-contraction of the pushdown graph of $\mathcal{A}$; assume that it is finitely branching. Assume that in G there exists a path of length m from the initial configuration to some configuration*

* Work partially supported by the Polish Ministry of Science grant nr N N206 567840.

29th Symposium on Theoretical Aspects of Computer Science (STACS'12).
Editors: Christoph Dürr, Thomas Wilke; pp. 54–65
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

c. Let $S_1 = (m+1) \cdot C_{\mathcal{A}L}$ *and* $S_j = 2^{S_{j-1}}$ *for* $2 \leq j \leq n$*, where* $C_{\mathcal{A}L}$ *is a constant which depends on* $\mathcal{A}$ *and on* L*. Assume also that in* G *there exists a path* p *of length at least* S_n*, which starts in* c *and belongs*[1] *to* L*. Then there are infinitely many paths in* G*, which start in* c*, belong to* L*, and end in configurations having the same state as the last configuration of* p*.*

This theorem is very similar to the pumping lemma proposed in [1]. Namely our Lemma 5.2 is an analogue of Corollary 16 from [1], and our Lemma 5.3 is an analogue of Theorem 61 from [1]; the above theorem (without the part about the regular language L) is obtained by composing these two lemmas.

Notice also that the bound S_n is $n-1$ times exponential in m, while the corresponding bound in [1] is $3n-1$ times exponential. Thus we obtain a better bound. Moreover, our bound is optimal, as explained in Section 6. The other difference is that our pumping preserves a regular property L of the paths, as well as the state of the last configuration.

2 Preliminaries

A *pushdown system* (PDS for short) of level n is given by a tuple $(A, \Gamma, \gamma_I, Q, q_I, \Delta, \lambda)$, where

- A is an input alphabet,
- Γ is a stack alphabet, and $\gamma_I \in \Gamma$ is an initial stack symbol,
- Q is a set of states, and $q_I \in Q$ is an initial state,
- $\Delta \subseteq Q \times \Gamma \times Q \times OP$ is a transition relation, where the set OP contains the operations pop^k and $\mathsf{push}^k(\alpha)$ for $1 \leq k \leq n$ and $\alpha \in \Gamma$,
- $\lambda \colon \Delta \to A \cup \{\varepsilon\}$ is a labelling of transitions.

In this paper, the letter n is always used for the level of the pushdown system.

For any alphabet Γ (of stack symbols) we define a *k-th level pushdown store* (k-pds for short) as an element of the following set Γ_*^k:

$$\Gamma_*^0 = \Gamma,$$
$$\Gamma_*^k = (\Gamma_*^{k-1})^* \qquad \text{for } 1 \leq k \leq n.$$

In other words, a 0-pds is just a single symbol, and a k-pds for $1 \leq k \leq n$ is a (possibly empty) sequence of $(k-1)$-pds's. The last element of a k-pds is also called the *topmost* one. For any $\alpha^k \in \Gamma_*^k$ and $\alpha^{k-1} \in \Gamma_*^{k-1}$ we write $\alpha^k : \alpha^{k-1}$ for the k-pds obtained from α^k by placing α^{k-1} at its end. The operator „:" is assumed to be right associative, i.e. $\alpha^2 : \alpha^1 : \alpha^0 = \alpha^2 : (\alpha^1 : \alpha^0)$. We say for $k \geq 1$ that a k-pds is *proper* if it is nonempty and every $(k-1)$-pds in it is proper; a 0-pds is always proper.

A *configuration* of $\mathcal{A}$ consists of a state and of a proper n-pds, i.e. it is an element of $Q \times \Gamma_*^n$ in which the n-pds is proper. The *initial* configuration consists of the initial state q_I and of the n-pds containing only one 0-pds, which is the initial stack symbol γ_I. For a configuration c, its state is denoted by $\mathit{state}(c)$, and its n-pds is denoted by $\pi(c)$.

Next, for configurations c, d we define when $c \vdash d$. Let α be the topmost 0-pds of $\pi(c)$. Assume that $(\mathit{state}(c), \alpha, \mathit{state}(d), \mathit{op}) \in \Delta$. We have two cases depending on op:

- if $\mathit{op} = \mathsf{pop}^k$ then $\pi(d)$ is obtained from $\pi(c)$ by replacing its topmost k-pds $\alpha^k : \alpha^{k-1}$ by α^k (i.e. we remove the topmost $(k-1)$-pds; in particular the topmost k-pds of $\pi(c)$ has to contain at least two $(k-1)$-pds's),

[1] Formally, the word consisting of labels on that path belongs to L.

- if $op = \mathsf{push}^k(\beta)$ then $\pi(d)$ is obtained from $\pi(c)$ by replacing its topmost k-pds $\alpha^k : \alpha^{k-1}$ by $(\alpha^k : \alpha^{k-1}) : \alpha^{k-1}$, and then by replacing its topmost 0-pds by β (i.e. we copy the topmost k-pds, and then we change the topmost symbol in the copy[2]).

A *run* is a function w from numbers $0, 1, \dots, l$ (for some $l \geq 0$) to configurations such that $w(i-1) \vdash w(i)$ for $1 \leq i \leq l$. The number l is called the *length* of w, and denoted by $|w|$. We say that w is a run from $w(0)$ to $w(|w|)$. For $0 \leq x \leq y \leq |w|$ we can consider the *subrun of w from x to y*; this is the run of length $y - x$ which maps i to $w(i + x)$. For two runs v, w such that $v(|v|) = w(0)$ we can consider their *composition*; this is the run of length $|v| + |w|$ which maps $i \leq |v|$ to $v(i)$, and $i > |v|$ to $w(i - |v|)$. We say that a configuration d is *reachable* from a configuration c if there exists a run w from c to d.

The *pushdown graph of* $\mathcal{A}$, denoted by $PDG(\mathcal{A})$, is the directed graph consisting of configurations of $\mathcal{A}$ reachable from the initial configuration; there is an edge from a configuration c to a configuration d when $c \vdash d$. To each edge of $PDG(\mathcal{A})$ we can assign a label from $A \cup \{\varepsilon\}$ in the following way. Let c, d be configurations such that $c \vdash d$. Notice that the transition $\delta \in \Delta$ used between c and d (in the definition of $\vdash$) is uniquely determined. We label the edge from c to d by $\lambda(\delta)$. A run of $\mathcal{A}$ can also be interpreted as a path in $PDG(\mathcal{A})$, so it makes sense to talk about edges of a run, and about labels of these edges.

We define the *ε-contraction* of $PDG(\mathcal{A})$, denoted by $PDG(\mathcal{A})/\varepsilon$, which is a directed multigraph.[3] Its vertices are the initial configuration c_I, and configurations d such that there is a run from c_I to d in which the last edge is labelled by an element of A (i.e. not by ε). In $PDG(\mathcal{A})/\varepsilon$ there is an edge from c to d labelled by $a \in A$ when in $PDG(\mathcal{A})$ there is a path from c to d whose edges except the last one are labelled by ε, and the last edge is labelled by a. We say that $PDG(\mathcal{A})/\varepsilon$ is *finitely branching* if from each of its nodes there are only finitely many outgoing edges.

A *position* is a vector $x = (x_n, x_{n-1}, \dots, x_1)$ of n positive integers. The symbol on position x in configuration c (which is an element of Γ) is defined as follows: we take the x_n-th (from the bottom) $(n-1)$-pds of $\pi(c)$, then its x_{n-1}-th $(n-2)$-pds, and so on. We say that x is a position of c, if at position x there is a symbol in c.

For $0 \leq k \leq n$, by $top^k(c)$ we denote the position of the bottommost symbol of the topmost k-pds of c. In particular $top^0(c)$ is the position of the topmost symbol in c.

For any run w, indices $0 \leq a \leq b \leq |w|$, and a position y of $w(b)$, we define a position $hist_w(b, y)(a)$. It is y when $b = a$. It is y also when $b = a + 1$, and the operation between $w(a)$ and $w(b)$ is pop^k, as well as when the operation is push^k and y is not in the topmost $(k-1)$-pds of $w(b)$. If the operation between $w(a)$ and $w(b)$ is push^k and y is in the topmost $(k-1)$-pds of $w(b)$, then $hist_w(b, y)(a)$ is the position of $w(a)$ from which a symbol was copied to y (i.e. this is y with the $(n-k+1)$-th coordinate decreased by 1). When $b > a+1$, $hist_w(b, y)(a)$ is defined (by induction) as $hist_w(a+1, hist_w(b, y)(a+1))(a)$. In other words, $hist_w(b, y)(a)$ is the (unique) position of $w(a)$, from which the symbol was copied to y in $w(b)$.

For $0 \leq k \leq n$, a run w, and an index $0 \leq b \leq |w|$ we define a set $pre^k_w(b)$ consisting of all indices a for which $0 \leq a \leq b$ and $hist_w(b, top^k(w(b)))(a) = top^k(w(a))$. Intuitively, $a \in pre^k_w(b)$ means that the topmost k-pds of $w(b)$ „comes from" the topmost k-pds of $w(a)$,

---

[2] In the classical definition the topmost symbol can be changed only when $k = 1$ (for $k \geq 2$ it has to be $\beta = \alpha$). Notice however that our theorems, true for every PDS, are in particular true for such restricted PDS's. On the other hand, it is not difficult to see that for any PDS $\mathcal{A}$ of level n there exists a PDS $\mathcal{B}$ of level n of this restricted form such that graphs $PDG(\mathcal{A})/\varepsilon$ and $PDG(\mathcal{B})/\varepsilon$ are isomorphic.

[3] In this graph, unlike in $PDG(\mathcal{A})$, we can have multiple edges between two nodes, each labeled by a different symbol.

in the sense that the topmost k-pds of $w(b)$ is a copy of the topmost k-pds of $w(a)$, but possibly some changes were done to it.

Example. Consider a PDS of level 3. Below, brackets are used in descriptions of pds's as follows: symbols taken in brackets form one 1-pds, 1-pds's taken in brackets form one 2-pds, and 2-pds's taken in brackets form one 3-pds. Consider a run w of length 6 in which $\pi(w(0)) = [[[ab]]]$ and the operations between consecutive configurations are:

$$\mathsf{push}^2(c),\ \mathsf{push}^3(d),\ \mathsf{pop}^1,\ \mathsf{push}^3(e),\ \mathsf{pop}^2,\ \mathsf{pop}^3.$$

The contents of the 3-pds's of the configurations in the run, and the *pre* sets, are presented in the table below.

i	$\pi(w(i))$	$pre^0_w(i)$	$pre^1_w(i)$	$pre^2_w(i)$	$pre^3_w(i)$
0	$[[[ab]]]$	$\{0\}$	$\{0\}$	$\{0\}$	$\{0\}$
1	$[[[ab][ac]]]$	$\{0,1\}$	$\{0,1\}$	$\{0,1\}$	$\{0,1\}$
2	$[[[ab][ac]][[ab][ad]]]$	$\{0,1,2\}$	$\{0,1,2\}$	$\{0,1,2\}$	$\{0,1,2\}$
3	$[[[ab][ac]][[ab][a]]]$	$\{3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$
4	$[[[ab][ac]][[ab][a]][[ab][e]]]$	$\{3,4\}$	$\{0,1,2,3,4\}$	$\{0,1,2,3,4\}$	$\{0,1,2,3,4\}$
5	$[[[ab][ac]][[ab][a]][[ab]]]$	$\{0,5\}$	$\{0,5\}$	$\{0,1,2,3,4,5\}$	$\{0,1,2,3,4,5\}$
6	$[[[ab][ac]][[ab][a]]]$	$\{3,6\}$	$\{0,1,2,3,6\}$	$\{0,1,2,3,6\}$	$\{0,1,2,3,4,5,6\}$

In configuration $w(0)$ symbol a is on position $(1,1,1)$ and symbol b is on position $(1,1,2)$. We have

$$hist_w(2,(2,2,1))(1) = (1,2,1) \qquad \text{and} \qquad hist_w(2,(2,2,1))(0) = (1,1,1).$$

Notice that positions y in $w(b)$ and $hist_w(b,y)(a)$ in $w(a)$ not necessarily contain the same symbol, for example on position $(1,2,2)$ in $w(1)$ we have c, and on position $(1,1,2)$ in $w(0)$ we have b, but $hist_w(1,(1,2,2))(0) = (1,1,2)$.

Easy properties. The following two propositions follow immediately from the definitions. These properties are often used implicitly later.

▶ **Proposition 2.1.** *Let w be a run, let $0 \le k \le n$, and let $0 \le a \le b \le c \le |w|$. Then*
- *$pre^k_w(b) \subseteq pre^{k+1}_w(b)$ (for $k < n$), and*
- *$a \in pre^k_w(b)$ and $b \in pre^k_w(c)$ implies $a \in pre^k_w(c)$, and*
- *$\{a,b\} \subseteq pre^k_w(c)$ implies $a \in pre^k_w(b)$.*

▶ **Proposition 2.2.** *Let w be a run, let $1 \le k \le n$, and let $0 \le a \le b \le |w|$ be such that $a \in pre^k_w(b)$. Then $a \in pre^{k-1}_w(b)$ if and only if, for all $a \le i \le b$, the size of the k-pds of $w(i)$ containing $hist_w(b, top^k(w(b)))(i)$ is not smaller than the size of the topmost k-pds of $w(a)$.*

3 Types of configurations

Let $\mathcal{A} = (A, \Gamma, \gamma_I, Q, q_I, \Delta, \lambda)$ be a PDS of level n. Below we define a function $type_{\mathcal{A}}$ which assigns to every configuration of $\mathcal{A}$ an element of a finite set $\mathcal{T}_{\mathcal{A}}$. The important properties of the $type_{\mathcal{A}}$ function are listed below, in the three facts.

▶ **Fact 3.1.** *Let $\mathcal{A}$ be a PDS of level n. Let w be a run of $\mathcal{A}$ such that $0 \in pre^0_w(|w|)$, and let c be a configuration such that $type_{\mathcal{A}}(w(0)) = type_{\mathcal{A}}(c)$. Then there exists a run v from c such that*

1. *if* $\pi(w(0)) \neq \pi(w(|w|))$ *then* $\pi(v(0)) \neq \pi(v(|v|))$*, and*
2. $0 \in pre_v^0(|v|)$*, and*
3. *all edges of* w *are labelled by* ε *if and only if all edges of* v *are labelled by* ε*, and*
4. $type_{\mathcal{A}}(w(|w|)) = type_{\mathcal{A}}(v(|v|))$*.*

▶ **Fact 3.2.** *Let* $\mathcal{A}$ *be a PDS of level* n*. Let* w *be a run of* $\mathcal{A}$ *such that at least one of its edges is not labelled by* ε*, and the position* $top^0(w(0))$ *is present in every configuration of* w*. Let* c *be a configuration such that* $type_{\mathcal{A}}(w(0)) = type_{\mathcal{A}}(c)$*. Then there exists a run* v *from* c *such that at least one of its edges is not labelled by* ε*, and the position* $top^0(c)$ *is present in every configuration of* v*.*

▶ **Fact 3.3.** *Let* $\mathcal{A}$ *be a PDS of level* n*. Let* w *be a run of* $\mathcal{A}$*, and let* c *be a configuration such that* $type_{\mathcal{A}}(w(0)) = type_{\mathcal{A}}(c)$*. Then there exists a run* v *from* c *such that* $state(v(|v|)) = state(w(|w|))$*.*

Before we define types of configurations, we define types of k-pds's, for each k. The main idea is that we have to characterize special kind of runs, called k-returns, as well as runs as described by Facts 3.2 and 3.3.

▶ **Definition 3.4.** Let $1 \leq r \leq n$, and let w be a run. We say that w is an *r-return* if

- the topmost r-pds of $w(0)$ contains at least two $(r-1)$-pds's, and
- $hist_w(|w|, top^{r-1}(w(|w|)))(0)$ is the bottommost position of the $(r-1)$-pds just below the topmost $(r-1)$-pds of $w(0)$, and
- $pre_w^{r-1}(|w|) = \{|w|\}$.

In other words, w is an r-return when the topmost r-pds of $w(|w|)$ is obtained from the topmost r-pds of $w(0)$ by removing its topmost $(r-1)$-pds (but not only in the sense of contents, but we require that really it was obtained this way). In particular we have the following proposition.

▶ **Proposition 3.5.** *Let* w *be an* r*-return. Then the topmost* r*-pds of* $w(0)$ *after removing its topmost* $(r-1)$*-pds is equal to the topmost* r*-pds of* $w(|w|)$*.*

Example. Consider a PDS of level 2, and a run w of length 6 in which $\pi(w(0)) = [[ab][cd]]$, and the operations between consecutive configurations are:

$$\mathsf{push}^2(e),\ \mathsf{pop}^1,\ \mathsf{pop}^2,\ \mathsf{pop}^1,\ \mathsf{push}^1(d),\ \mathsf{pop}^1.$$

The contents of the 2-pds's of the configurations in the run are presented in the table below.

i	0	1	2	3	4	5	6
$\pi(w(i))$	$[[ab][cd]]$	$[[ab][cd][ce]]$	$[[ab][cd][c]]$	$[[ab][cd]]$	$[[ab][c]]$	$[[ab][cd]]$	$[[ab][c]]$

The subruns of w from 0 to 2, from 0 to 4, from 1 to 2, from 3 to 4, and from 5 to 6 are 1-returns; the subruns of w from 1 to 3, and from 2 to 3 are 2-returns. These are the only subruns of w being returns, in particular w is not a 1-return because $4 \in pre_w^0(6)$.

We are going to define a type of a k-pds for each $0 \leq k \leq n$. A set of possible level-k types (types of k-pds's) will be denoted by $\mathcal{T}^k$. We also define a set $\mathcal{D}^k$; its elements correspond to kinds of runs (this correspondence is formalized in the "agrees with" notion).

▶ **Definition 3.6.** We define $\mathcal{T}^k$ (where $0 \leq k \leq n$) by induction on k, going down from $k = n$ to $k = 0$. Let $0 \leq k \leq n$. Assume we have already defined sets $\mathcal{T}^i$ for $k+1 \leq i \leq n$. We take

$$\mathcal{D}^k = Q \cup \bigcup_{r=k+1}^{n} \{r\} \times \Big(\{\textsf{non-}\varepsilon\} \cup \Big(\{0,1\} \times \mathcal{P}(\mathcal{T}^n) \times \mathcal{P}(\mathcal{T}^{n-1}) \times \dots \times \mathcal{P}(\mathcal{T}^{r+1}) \times Q \times \{0,1\}\Big)\Big),$$

$$\mathcal{T}^k = \mathcal{P}(\mathcal{T}^n) \times \mathcal{P}(\mathcal{T}^{n-1}) \times \dots \times \mathcal{P}(\mathcal{T}^{k+1}) \times Q \times \mathcal{D}^k,$$

where by $\mathcal{P}(X)$ we denote the power set of X (the set of all subsets of X).

▶ **Definition 3.7.** We define $type(\alpha^k) \subseteq \mathcal{T}^k$ for a k-pds α^k (where $0 \leq k \leq n$) by induction on k, going down from $k = n$ to $k = 0$. Let $0 \leq k \leq n$. Assume we have already defined sets *type* for i-pds's for $k+1 \leq i \leq n$.

1. Let $t = (r, f, \xi^n, \xi^{n-1}, \dots, \xi^{r+1}, q, g) \in \mathcal{D}^k$, and let w be a run. Decompose $\pi(w(|w|)) = \beta^n : \beta^{n-1} : \dots : \beta^r$. We say that w *agrees with* t if
   - w is an r-return, and
   - each edge of w is labelled by ε if and only if $f = 0$, and
   - $type(\beta^i) = \xi^i$ for $r+1 \leq i \leq n$, and
   - $q = state(w(|w|))$, and
   - $\pi(w(|w|))$ can be obtained from $\pi(w(0))$ by removing its topmost $(r-1)$-pds if and only if $g = 0$.
2. We say that a run w *agrees with* $(r, \textsf{non-}\varepsilon) \in \mathcal{D}^k$ if at least one edge of w is labelled by an element of A, and position $top^{r-1}(w(0))$ is present in every configuration of w.
3. We say that a run w *agrees with* $q \in \mathcal{D}^k \cap Q$ if $state(w(|w|)) = q$.
4. Let $t = (\rho^n, \rho^{n-1}, \dots, \rho^{k+1}, p, t') \in \mathcal{T}^k$, and let α^k be a k-pds. We say that $t \in type(\alpha^k)$ if the following is true.

   For $k+1 \leq i \leq n$, let α^i be an i-pds such that $type(\alpha^i) = \rho^i$. Then there exists a run from $(p, \alpha^n : \alpha^{n-1} : \dots : \alpha^k)$ which agrees with t'.

In point 4 of the above definition we mean that for all appropriate $\alpha^{k+1}, \alpha^{k+2}, \dots, \alpha^n$ the run exists (and not that there exist appropriate $\alpha^{k+1}, \alpha^{k+2}, \dots, \alpha^n$ such that the run exists). However in fact the „there exists" variant would be equivalent; this is described by the following lemma, which is the main technical result about types.

▶ **Lemma 3.8.** *Let $0 \leq k \leq n$, let $t \in \mathcal{D}^k$, and let w be a run which agrees with t. Decompose $\pi(w(0)) = \alpha^n : \alpha^{n-1} : \dots : \alpha^k$. Then*

$$(type(\alpha^n), type(\alpha^{n-1}), \dots, type(\alpha^{k+1}), state(w(0)), t) \in type(\alpha^k).$$

The proof of this lemma is tedious but rather straightforward. Finally, we define types of configurations.

▶ **Definition 3.9.** Let $\mathcal{T}_\mathcal{A} = \mathcal{P}(\mathcal{T}^n) \times \mathcal{P}(\mathcal{T}^{n-1}) \times \dots \times \mathcal{P}(\mathcal{T}^1) \times \Gamma \times Q$. For a configuration $c = (q, \alpha^n : \alpha^{n-1} : \dots : \alpha^0)$, let

$$type_\mathcal{A}(c) = (type(\alpha^n), type(\alpha^{n-1}), \dots, type(\alpha^1), \alpha^0, q).$$

Using Lemma 3.8 it is not difficult to show that Facts 3.1–3.3 for such definition of a type.

4 Pumping of pushdown graphs

The following technical lemma describes how pushdown graphs can be pumped.

▶ **Lemma 4.1.** *Let $\mathcal{A}$ be a PDS of level n, let $0 \leq k \leq n$, let w be a run of $\mathcal{A}$, and let $G \subseteq pre^k_w(|w|) - \{|w|\}$. Let α^k be the k-pds of $w(0)$ containing $hist_w(|w|, top^k(w(|w|)))(0)$. For $1 \leq j \leq k$, let r_j be the maximum of the sizes of the j-pds's in α^k. Define*

$$N_0 = |\mathcal{T}_\mathcal{A}| + 1 \qquad \text{and} \qquad N_j = r_j \cdot 2^{N_{j-1}} \qquad \text{for } 1 \leq j \leq k.$$

Assume that $|G| \geq N_k$. Then there exist indices $0 \leq x < y < z \leq |w|$ such that

1. *$type_\mathcal{A}(w(x)) = type_\mathcal{A}(w(y))$, and*
2. *$x \in pre^0_w(y)$ and $y \in pre^k_w(|w|)$, and*
3. *either $\pi(w(x)) \neq \pi(w(y))$, or $G \cap \{x, x+1, \ldots, y-1\} \neq \emptyset$, and*
4. *$z - 1 \in G$ and $top^0(w(y))$ is present in every configuration of the subrun of w from y to z.*

Let us comment on the statement of this lemma. The essence of the lemma is that in every appropriately long run one can find indices x, y such that $type_\mathcal{A}(w(x)) = type_\mathcal{A}(w(y))$ and $x \in pre^0_w(y)$. Notice that the notion "appropriately long" depends on the size of the stack in $w(0)$: when one starts from a bigger stack, we require a longer run. Then Fact 3.1 can be applied to the fragment of w between x and y, so this fragment can be pumped (repeated forever). The lemma is more complicated for technical reasons. The problem is that pumping any fragment of a run is not interesting enough. For example the fragment between x and y can be a loop doing nothing; we are not satisfied with finding such a loop. For this reason we have introduced the set G of "good" indices, and we assume that this set is big enough. Our goal is to have some element of G in the fragment between x and y (the second variant of condition 3). However this is not always possible, and we sometimes get the first variant of condition 3; the intuition is that then we can show (using also index z) that the graph has to be infinitely branching.

The above lemma is proved by induction on k. For $k = 0$ we have $|G| \geq |\mathcal{T}_\mathcal{A}| + 1$ and there are only $|\mathcal{T}_\mathcal{A}|$ possible values of $type_\mathcal{A}$, so there exist two indices $x, y \in G$ such that $x < y$ and $type_\mathcal{A}(w(x)) = type_\mathcal{A}(w(y))$ (we get condition 1). By assumption we know that $x, y \in pre^0_w(|w|)$; this implies that $x \in pre^0_w(y)$ (we get condition 2). We have condition 3 because $x \in G$. We take $z = y + 1$. We have $z - 1 \in G$. Because $y \in pre^0_w(|w|)$, position $top^0(w(y))$ is present in $w(z)$ (we get condition 4).

For $k > 0$ we make the induction step using the following lemma about sequences of integers. For $0 \leq i \leq |w|$ as a_i we take the size of the k-pds of $w(i)$ containing $hist_w(|w|, top^k(w(|w|)))(i)$.

▶ **Lemma 4.2.** *Let $N \geq 1$ be a natural number, let $a_0, a_1, \ldots, a_M$ be a sequence of positive integers such that $|a_i - a_{i-1}| \leq 1$ for $1 \leq i \leq M$. Let $G \subseteq \{0, 1, \ldots, M-1\}$ be such that $|G| \geq a_0 \cdot 2^N$. Then there exist two indices b, e such that $0 \leq b < e \leq M$ and $e - 1 \in G$, and*

1. *for each i such that $b \leq i \leq e$ we have $a_i \geq a_b$, and*
2. *for each i such that $0 \leq i \leq b-1$ we have $a_i \geq a_b + 1$, and*
3. *$|H_{b,e}| \geq N$, where*

$$H_{b,e} = \{i \colon b \leq i \leq e-1 \wedge \forall_j (i \leq j \leq e \Rightarrow a_j \geq a_i) \wedge \\ \wedge \exists_{g \in G}(g \geq i \wedge \forall_j (i+1 \leq j \leq g \Rightarrow a_j \geq a_i + 1))\}.$$

5 Finitely branching ε-contractions of pushdown graphs

In this section we show how finitely branching ε-contractions of pushdown graphs can be pumped; we prove Theorem 1.1. First we give an auxiliary lemma, which describes how the assumption about finite branching can be used. Then we have two lemmas, which are then composed together into Theorem 1.1. Lemma 5.2 tells us that a short run from the initial configuration cannot finish in a configuration having a big stack. Lemma 5.3 is similar to Theorem 1.1, but instead of assuming that a configuration can be reached with a short run from the initial configuration, we assume that its stack is small (and this assumption will be then satisfied thanks to Lemma 5.2).

▶ **Lemma 5.1.** *Let $\mathcal{A}$ be a PDS of level n, let w be a run of $\mathcal{A}$ such that $w(0)$ is reachable from the initial configuration, and let $0 \leq x < y \leq |w|-1$ be indices such that $type_{\mathcal{A}}(w(x)) = type_{\mathcal{A}}(w(y))$, and $x \in pre^0_w(y)$, and $\pi(w(x)) \neq \pi(w(y))$. Assume that $top^0(y)$ is present in every configuration of the subrun of w from y to $|w|$. Assume also that every edge of w between x and y is labelled by ε, and at least one edge of w between y and $|w|$ is not labelled by ε. Then $PDG(\mathcal{A})/\varepsilon$ is not finitely branching.*

Proof. Without loss of generality, we assume that w begins in the initial configuration; we can obtain such a situation by appending before w any run from the initial configuration to $w(0)$, and appropriately shifting x and y. Let g be the smallest index ($0 \leq g \leq x$) such that every edge between g and x is labelled by ε. Then $w(g)$ is a node of $PDS(\mathcal{A})/\varepsilon$.

We want to create a sequence of runs $v_1, v_2, v_3, \dots$ such that for each $i \geq 1$ we have

a) $v_1(0) = w(x)$ and $v_i(0) = v_{i-1}(|v_{i-1}|)$ for $i > 1$, and
b) $\pi(v_i(0)) \neq \pi(v_i(|v_i|))$, and
c) $0 \in pre^0_{v_i}(|v_i|)$, and
d) every edge of v_i is labelled by ε, and
e) $type_{\mathcal{A}}(v_i(0)) = type_{\mathcal{A}}(v_i(|v_i|))$.

As v_1 we can take the subrun of w from x to y. Assume that we already have v_i for some $i \geq 1$. We use Fact 3.1 for v_i (as w) and $v_i(|v_i|)$ (as c); thanks to properties c) and e) its assumptions are satisfied. We obtain a run v_{i+1} from $v_i(|v_i|)$. Conditions 1–4 of the fact immediately give us conditions b–e for v_{i+1}.

Notice, for each $i \geq 1$, that because $0 \in pre^0_{v_i}(|v_i|)$ and $\pi(v_i(0)) \neq \pi(v_i(|v_i|))$, position $top^0(v_i(|v_i|))$ (which is $top^0(v_{i+1}(0))$) is lexicographically greater than $top^0(v_i(0))$. Thus every $top^0(v_i(0))$ is different.

For every $i \geq 1$ we do the following. From condition e) and from $type_{\mathcal{A}}(w(x)) = type_{\mathcal{A}}(w(y))$ we know that $type_{\mathcal{A}}(v_i(0)) = type_{\mathcal{A}}(w(y))$. We use Fact 3.2 for the subrun of w from y to $|w|$ (as w), and for $v_i(0)$ (as c). We obtain a run u_i from $v_i(0)$ such that at least one of its edges is not labelled by ε, and position $top^0(v_i(0))$ is present in every configuration of u_i. We can assume that only the last edge of u_i is not labelled by ε (we obtain this situation by cutting u_i after the first edge not labelled by ε). Now compose the subrun of w from g to x, runs $v_1, v_2, \dots, v_{i-1}$, and run u_i. We obtain a run from $w(g)$ such that only its last edge is not labelled by ε. Thus $u_i(|u_i|)$ is a successor of $w(g)$ in $PDG(\mathcal{A})/\varepsilon$, in which position $top^0(v_i(0))$ is present. As each position $top^0(v_i(0))$ is different, they cannot be all present in only finitely many configurations, so among $u_i(|u_i|)$ there are infinitely many different configurations. This means that $PDG(\mathcal{A})/\varepsilon$ is not finitely branching. ◀

▶ **Lemma 5.2.** *Let $\mathcal{A}$ be a PDS of level n such that $PDG(\mathcal{A})/\varepsilon$ is finitely branching. Let w be a run which begins in the initial configuration, and whose last edge is not labelled by ε.*

Let m be the number of edges of w not labelled by ε. Let

$$M_1 = (m+1) \cdot (|\mathcal{T}_\mathcal{A}| + 1) \qquad \text{and} \qquad M_j = 2^{M_{j-1}} \qquad \text{for } 2 \le j \le n.$$

Then, for $1 \le k \le n$, the size of any k-pds of $w(|w|)$ is at most M_k.

Proof. Induction on m. Notice that $m \ge 1$. Define

$$M'_1 = m \cdot (|\mathcal{T}_\mathcal{A}| + 1) \qquad \text{and} \qquad M'_j = 2^{M'_{j-1}} \qquad \text{for } 2 \le j \le n.$$

Let b be the index such that the $(m-1)$-st edge of w not labelled by ε is between $w(b-1)$ and $w(b)$; if $m = 1$ we take $b = 0$. From the induction assumption, used for the subrun of w from 0 to b, we know, for $1 \le k \le n$, that the size of any k-pds of $w(b)$ is at most M'_k. This is also true for $m = 1$, as $M'_k \ge 1$.

Assume that for some k ($1 \le k \le n$) the size of some k-pds of $w(|w|)$ is greater than M_k. Let s be the bottommost position of such a k-pds. Let v be the subrun of w from b to $|w|$. For $0 \le i \le |v|$, let a_i be the size of the k-pds of $v(i)$ containing $hist_v(|v|, s)(i)$. We have $a_{|v|} \ge M_k$ and $a_0 \le M'_k$. Of course $|a_{i-1} - a_i| \le 1$ for $1 \le i \le |v|$. Let

$$G = \{i \colon 0 \le i \le |v| - 1 \wedge \forall_j (i + 1 \le j \le |v| \Rightarrow a_j \ge a_i + 1)\}.$$

Notice that $|G| \ge M_k - M'_k$, as for each j such that $M'_k \le j \le M_k - 1$ in G we have the last index i such that $a_i = j$. Let e be the greatest index such that $e - 1 \in G$; let v' be the subrun of v from 0 to e. Define

$$N_0 = |\mathcal{T}_\mathcal{A}| + 1 \qquad \text{and} \qquad N_i = M'_i \cdot 2^{N_{i-1}} \qquad \text{for } 1 \le i \le k-1.$$

We are going to use Lemma 4.1 for $k - 1$ (as k), for the run v' (as w), and for G. We have to check that its assumptions are satisfied. We need to check that $G \subseteq pre_v^{k-1}(e)$. Because only the topmost k-pds can change its size, and $a_i \ne a_{i+1}$ for $i \in G$, it follows that $hist_v(|v|, s)(i) = top^k(v(i))$ for $i \in G \cup \{e\}$, which means that $G \subseteq pre_v^k(e)$. As additionally $a_j \ge a_i$ for $i \in G$, $i \le j \le |v|$, from Proposition 2.2 we get $G \subseteq pre_v^{k-1}(e)$, as required. We also need to check that G has enough elements; this is a straightforward calculation.

From Lemma 4.1 we obtain indices $0 \le x < y < z \le e$ such that

1. $type_\mathcal{A}(v(x)) = type_\mathcal{A}(v(y))$, and
2. $x \in pre_v^0(y)$, and
3. either $\pi(v(x)) \ne \pi(v(y))$, or $G \cap \{x, x+1, \dots, y-1\} \ne \emptyset$, and
4. $z - 1 \in G$ and $top^0(v(y))$ is present in every configuration of the subrun of v from y to z.

Is it possible that $\pi(v(x)) = \pi(v(y))$? As additionally $x \in pre_v^0(y)$ (condition 2), this would mean that for every position p in $v(y)$ we have $hist_v(y, p)(x) = p$ (between $v(x)$ and $v(y)$ some new fragments of the n-pds were added and then removed; it is impossible that we have first removed something and then reproduced it). In particular a_x and a_y describe the size of the same k-pds, so $a_x = a_y$. Moreover $a_i \ge a_x$ for $x \le i \le y$. But condition 3 implies that there is some $g \in G \cap \{x, x+1, \dots, y-1\}$. This is impossible, as we have $a_y \ge a_g + 1$ (by definition of G), and $a_g \ge a_x$, which means that $a_x \ne a_y$. Thus we always have $\pi(v(x)) \ne \pi(v(y))$.

Because $z-1 \in G$, we have $a_{z-1} \ne a_z$, so since only the topmost k-pds can change its size, we know that $hist_v(|v|, s)(z) = top^k(v(z))$. Additionally $a_i \ge a_z = a_{z-1} + 1$ for $z \le i \le |v|$ (by definition of G), which means that $top^{k-1}(v(z))$ is present in every configuration of the subrun of v from z to $|v|$. Since $top^0(v(y))$ is present in $v(z-1)$ (condition 4), we know that $top^0(v(y))$ is (lexicographically) below $top^{k-1}(v(z))$, so one cannot remove $top^0(v(y))$

without removing $top^{k-1}(v(z))$. It follows that $top^0(v(y))$ is present in every configuration of the subrun of v from y to $|v|$.

Recall also that the last edge of v is not labelled by ε, and all earlier edges are labelled by ε. So every edge of v between x and y is labelled by ε, and at least one edge of v between y and $|v|$ is not labelled by ε. Thus the assumptions of Lemma 5.1 (where v is taken as w) are satisfied. We get that $PDG(\mathcal{A})/\varepsilon$ is not finitely branching, which contradicts with our assumption. ◀

▶ **Lemma 5.3.** *Let $\mathcal{A}$ be a PDS of level n such that $PDG(\mathcal{A})/\varepsilon$ is finitely branching, and let w be a run of $\mathcal{A}$ such that $w(0)$ is reachable from the initial configuration. For $1 \leq j \leq n$, let r_j be the maximum of the sizes of j-pds's of $w(0)$. Define*

$$N_0 = |\mathcal{T}_\mathcal{A}| + 1 \qquad \textit{and} \qquad N_j = r_j \cdot 2^{N_{j-1}} \qquad \textit{for } 1 \leq j \leq n.$$

Assume that at least N_n edges of w are not labelled by ε. Then for each $j \in \mathbb{N}$ there exist a run w_j from $w(0)$ which has at least j edges not labelled by ε, and such that $state(w_j(|w_j|)) = state(w(|w|))$.

Proof. Let G be the set of indices i ($0 \leq i \leq |w|-1$) such that the edge between $w(i)$ and $w(i+1)$ is not labelled by ε. We use Lemma 4.1 for n (as k), for run w, and set G. Of course $G \subseteq pre^n_w(|w|)$, as $pre^n_w(|w|)$ by definition contains all numbers from 0 to $|w|$. We also have $|G| \geq N_n$, which is the required size. From the lemma we obtain indices $0 \leq x < y < z \leq |w|$ such that

1. $type_\mathcal{A}(w(x)) = type_\mathcal{A}(w(y))$, and
2. $x \in pre^0_w(y)$, and
3. either $\pi(w(x)) \neq \pi(w(y))$, or $G \cap \{x, x+1, \ldots, y-1\} \neq \emptyset$, and
4. $z-1 \in G$ and $top^0(w(y))$ is present in every configuration of the subrun of w from y to z.

Assume first that every edge of w between x and y is labelled by ε. By condition 3 we see that $\pi(w(x)) \neq \pi(w(y))$. Notice also that at least one edge of w between y and z is not labelled by ε, namely the last edge (as $z-1 \in G$). The assumptions of Lemma 5.1 are satisfied; we get that $PDG(\mathcal{A})/\varepsilon$ is not finitely branching, which contradicts with our assumption. Thus at least one edge of w between x and y is not labelled by ε.

We want to create a sequence of runs $v_1, v_2, v_3, \ldots$ beginning at $w(x)$ such that for each $j \geq 1$ we have

a) $0 \in pre^0_{v_j}(|v_j|)$, and
b) at least j edges of v_j are not labelled by ε, and
c) $type_\mathcal{A}(v_j(0)) = type_\mathcal{A}(v_j(|v_j|))$.

As v_1 we can take the subrun of w from x to y. Assume that we already have v_j for some $j \geq 1$. We use Fact 3.1 for v_j (as w) and $v_j(|v_j|)$ (as c); thanks to properties a) and c) its assumptions are satisfied. We obtain a run v from $v_j(|v_j|)$. Let v_{j+1} be the composition of runs v_j and v. Condition 2 of the fact says that $0 \in pre^0_v(|v|)$; together with $0 \in pre^0_{v_j}(|v_j|)$ it gives us that $0 \in pre^0_{v_{j+1}}(|v_{j+1}|)$. Condition 3 of the fact says that at least one edge of v is not labelled by ε; thus at least $j+1$ edges of v_{j+1} are not labelled by ε. Condition 4 of the fact says that $type_\mathcal{A}(v(0)) = type_\mathcal{A}(v(|v|))$; thus $type_\mathcal{A}(v_{j+1}(0)) = type_\mathcal{A}(v_{j+1}(|v_{j+1}|))$.

Next, we use Fact 3.3 for the subrun of w from y to $|w|$ and for $v_j(|v_j|)$; we obtain a run v'_j from $v_j(|v_j|)$ such that $state(v'_j(|v'_j|)) = state(w(|w|))$. Finally, as w_j we take the composition of the subrun of w from 0 to x with run v_i and with run v'_i; this run satisfies the thesis of the lemma. ◀

Proof (Theorem 1.1). First we consider the following special case. Assume that the language L contains all words. Assume also that the set of states of $\mathcal{A}$ is of the form $Q \times \{0,1\}$, and a transition is labelled by ε if and only if it leads to a state with 0 on the second coordinate. Then we take $C_{\mathcal{A}L} = 3 \cdot (|\mathcal{T}_\mathcal{A}|+1) \cdot 2^{|\mathcal{T}_\mathcal{A}|+1}$. Because in $PDG(\mathcal{A})/\varepsilon$ we have a path of length m from the initial configuration to c, there exists a run w from the initial configuration to c such that exactly m of its edges are not labelled by ε, in particular the last one. Let

$$M_1 = (m+1) \cdot (|\mathcal{T}_\mathcal{A}|+1) \qquad \text{and} \qquad M_j = 2^{M_{j-1}} \qquad \text{for } 2 \leq j \leq n.$$

By Lemma 5.2 we know, for $1 \leq k \leq n$, that the size of any k-pds of c is at most M_k. Let

$$N_0 = |\mathcal{T}_\mathcal{A}|+1 \qquad \text{and} \qquad N_j = M_j \cdot 2^{N_{j-1}} \qquad \text{for } 1 \leq j \leq n.$$

A straightforward calculation proves that $S_n \geq N_n$. Because in $PDG(\mathcal{A})/\varepsilon$ we have a path of length S_n starting at c, there exists a run v starting at c such that at least $S_n \geq N_n$ of its edges are not labelled by ε. We use Lemma 5.3 for the run v (as w). It says that there exist runs w_j from c having arbitrarily many edges not labelled by ε, and such that $w_j(|w_j|)$ and $w(|w|)$ have the same state. Since one state is reached either only by ε-transitions or only by non-ε-transitions, the last edge of w_j is not labelled by ε, because the last edge of w was not labelled by ε. It means that there are arbitrarily many paths in $PDG(\mathcal{A})/\varepsilon$ starting at c, and ending in configurations with state $state(w(|w|))$.

Next, consider a situation where $\mathcal{A}$ is arbitrary, but L still contains all words. Then we convert $\mathcal{A}$ to $\mathcal{A}'$ having the above form. We simply product the states Q of $\mathcal{A}$ by $\{0,1\}$; for every transition $\delta = (q_1, \gamma, q_2, op)$ of $\mathcal{A}$, in $\mathcal{A}'$ we have, for $i = 0, 1$, transitions $((q_1, i), \gamma, (q_2, 0), op)$ if $\lambda(\delta) = \varepsilon$, or $((q_1, i), \gamma, (q_2, 1), op)$ otherwise. The initial state gets 1 on the second coordinate. Notice that only configurations having 1 on the second coordinate are in $PDG(\mathcal{A}')/\varepsilon$. Moreover there is an edge between two configurations in $PDG(\mathcal{A})/\varepsilon$ if and only if there is an edge between corresponding (obtained by putting 1 on the second coordinate of the state) configurations in $PDG(\mathcal{A}')/\varepsilon$. So the two graphs are isomorphic, thus the theorem for one of them immediately implies the theorem for the other.

For an arbitrary language L and arbitrary PDS $\mathcal{A}$ the theorem is true, because we can make a product of $\mathcal{A}$ with a finite automaton recognizing L. ◀

6 Example application

Let $\varphi\colon \mathbb{N} \to \mathbb{N}$ be an unbounded function. Let $f_1^\varphi(x) = x \cdot \varphi(x)$ and $f_{k+1}^\varphi(x) = 2^{f_k^\varphi(x)}$ for $k \geq 1$. Consider the tree T_n^φ whose nodes are

$$\{0^i 1^j : i \geq 0, j \leq f_n^\varphi(i+2)+1\},$$

and a node w is connected with a node wa by an edge labelled by a (where w is a word and $a \in \{0,1\}$ is a letter). This tree is not isomorphic to the ε-contraction of any pushdown graph of level n.

Heading for a contradiction, assume that T_n^φ is isomorphic to $PDG(\mathcal{A})/\varepsilon$ for some pushdown system $\mathcal{A}$ of level n. In this isomorphism, the empty word in T_n^φ has to correspond to the initial configuration (as it is the only configuration which can have no predecessors). Choose $i \in \mathbb{N}$ such that $\varphi(i+2) \geq C_{\mathcal{A}L}$ (where $C_{\mathcal{A}L}$ is the constant from Theorem 1.1, for $L = \{0,1\}^*$). Let c be the configuration corresponding to $0^i 1$, and d the configuration corresponding to $0^i 1^{f_n^\varphi(i+2)+1}$. We use Theorem 1.1 for the path from the initial configuration to c and for the path from c to d; their length is, respectively, $i+1$ and $f_n^\varphi(i+2)$ (which is greater or equal to S_n from the theorem). Thus we obtain infinitely many paths starting in 0^i1, which contradicts the definition of T_n^φ.

On the other hand it is known that when the function φ is constant, then tree T_n^φ is isomorphic to $PDG(\mathcal{A})/\varepsilon$ for some pushdown system $\mathcal{A}$. See e.g. [1], Example 9, where a very similar pushdown system is constructed. In this sense the length required in Theorem 1.1 is the smallest possible: S_n has to be $n-1$ times exponential in m.

7 Future work

As a continuation of this work, we have recently [8] generalized Theorem 1.1 to *collapsible* pushdown systems. Collapsible pushdown systems are an extension of higher-order pushdown systems, in which an additional operation, called *collapse*, can be performed. Trees generated by these systems correspond to all higher-order recursion schemes [5], not only to safe ones.

Our pumping lemma talks only about the length of paths, and about a regular condition on the labels on them, hence its applications are rather limited. It would be useful to show a pumping lemma which describes more precisely how the new paths (as sequences of labels) can be constructed from the original paths, similarly to the classical pumping lemma for finite automata or pushdown automata.

Acknowledgement. I would like to thank Alexander Kartzow for many useful comments.

References

1. Achim Blumensath. On the structure of graphs in the caucal hierarchy. *Theor. Comput. Sci.*, 400(1-3):19–45, 2008.
2. Arnaud Carayol and Stefan Wöhrle. The caucal hierarchy of infinite graphs in terms of logic and higher-order pushdown automata. In *FSTTCS*, pages 112–123. Springer, 2003.
3. Didier Caucal. On infinite terms having a decidable monadic theory. In Krzysztof Diks and Wojciech Rytter, editors, *MFCS*, volume 2420 of *Lecture Notes in Computer Science*, pages 165–176. Springer, 2002.
4. Robert H. Gilman. A shrinking lemma for indexed languages. *Theor. Comput. Sci.*, 163(1&2):277–281, 1996.
5. Matthew Hague, Andrzej S. Murawski, C.-H. Luke Ong, and Olivier Serre. Collapsible pushdown automata and recursion schemes. In *LICS*, pages 452–461. IEEE Computer Society, 2008.
6. Takeshi Hayashi. On derivation trees of indexed grammars. *Publ. RIMS, Kyoto Univ.*, 9:61–92, 1973.
7. Alexander Kartzow. A pumping lemma for collapsible pushdown graphs of level 2. In Marc Bezem, editor, *CSL*, volume 12 of *LIPIcs*, pages 322–336. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2011.
8. Alexander Kartzow and Paweł Parys. Strictness of the collapsible pushdown hierarchy. In preparation, 2012.
9. Teodor Knapik, Damian Niwinski, and Pawel Urzyczyn. Higher-order pushdown trees are easy. In Mogens Nielsen and Uffe Engberg, editors, *FoSSaCS*, volume 2303 of *Lecture Notes in Computer Science*, pages 205–222. Springer, 2002.
10. A. N. Maslov. The hierarchy of indexed languages of an arbitrary level. *Soviet Math. Dokl.*, 15:1170–1174, 1974.
11. Paweł Parys. The pumping lemma is incorrect? Unpublished, 2010.