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

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], see also Appendix C).

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 c. Let $S_1 = (m+1)\cdot C_{\mathcal{A}L}$ and $S_j = 2^{S_{j-1}}$ for $2 \le j \le n$, 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

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

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

[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 $state(c)$, 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 $(state(c), \alpha, state(d), op) \in \Delta$. We have two cases depending on op:

- if $op = \mathsf{pop}^k$ 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),
- 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.

---

[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.

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 $w(a)$, 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	$[[[ab][ac]][[ab][a]]]$	$\{3,6\}$	$\{0,1,2,3,6\}$	$\{0,1,2,3,6\}$	$\{0,1,2,3,4,5,6\}$

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

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 $\pi(w(0)) \ne \pi(w(|w|))$ then $\pi(v(0)) \ne \pi(v(|v|))$, and*
2. *$0 \in pre^0_v(|v|)$, 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 $1 \le r \le n$, 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
- $pre^{r-1}_w(|w|) = \{|w|\}$.

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 $4 \in pre^0_w(6)$.

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

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

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

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* $(r, \mathsf{non}\text{-}\varepsilon) \in \mathcal{D}^k$ 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, and can be found in Appendix B. Finally, we define types of configurations.

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

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

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 $0 \le k \le n$, 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 $1 \le j \le k$, let r_j be the maximum of the sizes of the j-pds's in α^k. Define*

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

Assume that $|G| \ge N_k$. Then there exist indices $0 \le x < y < z \le |w|$ 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.

For the proof, we need a lemma about sequences of integers (proven in Appendix A).

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

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

Proof (Lemma 4.1). Induction on k. Consider first the case $k = 0$. We have $|G| \ge |\mathcal{T}_\mathcal{A}| + 1$ 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).

Now assume that $k \ge 1$. For $0 \le i \le |w|$, let a_i be the size of the k-pds of $w(i)$ containing $hist_w(|w|, top^k(w(|w|)))(i)$. By definition $a_0 = r_k$ (it is the size of α^k). Notice that $|G| \ge N_k = a_0 \cdot 2^{N_{k-1}}$. We apply Lemma 4.2 to the sequence $a_0, a_1, \ldots, a_{|w|}$, to N_{k-1} (as N) and G. Of course consecutive elements of the sequence differ by at most one, because operations of a pushdown system can change the size of a k-pds by at most one, so the assumption of the corollary is satisfied. From the corollary we obtain indices b, e. By condition 3 of the corollary we know that $|H_{b,e}| \geq N_{k-1}$.

Now observe that $H_{b,e} \cup \{e-1\} \subseteq pre^k_w(|w|)$. Indeed, consider any $i \in H_{b,e}$. If $i \in G \subseteq pre^k_w(|w|)$, we are done. Otherwise, by definition of $H_{b,e}$, it has to be $a_{i+1} \geq a_i + 1$. But only the topmost k-pds can change its size, so $hist_w(|w|, top^k(w(|w|)))(i)$ is in the topmost k-pds of $w(i)$; this by definition means that $i \in pre^k_w(|w|)$. Moreover $e-1 \in G \subseteq pre^k_w(|w|)$.

Because $H_{b,e} \cup \{e-1\} \subseteq pre^k_w(|w|)$ (see the above paragraph), we know that a_i is the size of the topmost k-pds of $w(i)$ for $i \in H_{b,e} \cup \{e-1\}$. Notice that $a_i \leq a_j$ for $i \in H_{b,e}$, $i \leq j \leq e$ (follows from the definition of $H_{b,e}$). This implies that $H_{b,e} \subseteq pre^{k-1}_w(e-1)$ (Proposition 2.2).

What operation is performed between $w(e-1)$ and $w(e)$? Definitely this is not a pop^r operation for $r \geq k+1$, as it would remove the topmost k-pds of $w(e-1)$, which is impossible because $e-1 \in G \subseteq pre^k_w(|w|)$. If this operation is pop^k, then $e \in pre^k_w(|w|)$, and a_e is the size of the topmost k-pds of $w(e)$; because $a_i \leq a_e$ for $i \in H_{b,e}$, we have $H_{b,e} \subseteq pre^{k-1}_w(e)$. If this is any other operation (i.e. pop^r for $r \leq k-1$, or push^r for any r), we have $e-1 \in pre^{k-1}_w(e)$, so also $H_{b,e} \subseteq pre^{k-1}_w(e)$. Let v be the subrun of w from b to e, and let $H' = \{i - b \colon i \in H_{b,e}\}$. It follows that $H' \subseteq pre^{k-1}_v(|v|)$.

Let α^{k-1} be the $(k-1)$-pds of $v(0)$ containing $hist_v(|v|, top^{k-1}(v(|v|)))(0)$; equivalently: the $(k-1)$-pds of $w(b)$ containing $hist_w(e, top^{k-1}(w(e)))(b)$. Notice that α^{k-1} is one of $(k-1)$-pds's of the k-pds of $w(b)$ containing $hist_w(e, top^k(w(e)))(b)$. Because $e-1 \in pre^k_w(e)$ (as the operation between $w(e-1)$ and $w(e)$ cannot be pop^r for $r \geq k+1$) and $e-1 \in G \subseteq pre^k_w(|w|)$, we have $hist_w(e, top^k(w(e)))(b) = hist_w(|w|, top^k(w(|w|)))(b)$. Recall that $a_i \geq a_b + 1$ for $0 \leq i \leq b-1$ (condition 2 of Lemma 4.2). This implies that the topmost $(k-1)$-pds (and obviously all the deeper $(k-1)$-pds's) of the k-pds of $w(b)$ containing $hist_w(|w|, top^k(w(|w|)))(b)$ was never modified between $w(0)$ and $w(b)$ (as it was never a topmost $(k-1)$-pds). It follows that α^{k-1} is one of the $(k-1)$-pds's of α^k. Thus, for $1 \leq j \leq k-1$, the maximum of the sizes of j-pds's in α^{k-1} is at most r_j.

Next, we apply the induction assumption for $k-1$ (as k), v (as w), H' (as G). The above analysis shows that its assumptions are satisfied. We obtain three indices, call them x', y', z'. Let $x = x' + b$ and $y = y' + b$. Let z be the smallest index such that $z \geq z' + b$ and $z - 1 \in G$ (it exists because $z' + b \leq e$ and $e - 1 \in G$). We claim that x, y, z satisfy conditions 1–4. From the induction assumption we have the following properties:

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

We immediately get condition 1 and the first part of condition 2. From $y \in pre^{k-1}_w(e)$ we know that $y \in pre^k_w(e)$; we also know that $y \leq e-1$. As already observed, the operation performed between $w(e-1)$ and $w(e)$ cannot be pop^r for $r \geq k+1$, which implies that $e-1 \in pre^k_w(e)$, so $y \in pre^k_w(e-1)$. Because $e-1 \in G \subseteq pre^k_w(|w|)$ we get $y \in pre^k_w(|w|)$ (the second part of condition 2).

Because $y \in pre^k_w(|w|)$ and $x \in pre^0_w(y)$, we know that the size of the topmost k-pds of x and of y is a_x and a_y, respectively. Assume that $\pi(w(x)) = \pi(w(y))$; this in particular means that $a_x = a_y$, and $H_{b,e} \cap \{x, x+1, \ldots, y-1\} \neq \emptyset$. Let i be any element of $H_{b,e} \cap \{x, x+1, \ldots, y-1\}$. By the definition of the set $H_{b,e}$, there exists $g \in G$ such that $g \geq i$ and $a_j \geq a_i + 1$ for $i+1 \leq j \leq g$. Because $x \in pre^0_w(y)$ (hence also $x \in pre^{k-1}_w(y)$ and $x \in pre^k_w(y)$), from Proposition 2.2 we know that $a_i \geq a_x$, hence $a_y < a_i + 1$. This implies that $g \leq y-1$. We get that $g \in G \cap \{x, x+1, \ldots, y-1\}$ (condition 3).

Because $z' + b - 1 \in H_{b,e}$, we know that $a_j \geq a_{z'+b-1} + 1$ for $z' + b \leq j \leq z - 1$ (as $z - 1$ is the first element of G not smaller than $z' + b - 1$). As $a_{z'+b-1} \neq a_{z'+b}$, we know that the size of the topmost k-pds of $w(z' + b - 1)$ is $a_{z'+b-1}$ (only the topmost k-pds can change its size). We get that the topmost $(k-1)$-pds of $w(z' + b - 1)$ becomes covered after the next operation

$(a_{z'+b} > a_{z'+b-1})$, and is not the topmost one until z. Thus $top^0(w(z'+b-1))$ is present in every configuration of the subrun of w from $z'+b-1$ to z. As additionally $top^0(w(y))$ is present in every configuration of the subrun of w from y to $z'+b-1$, we know that it is also present in every configuration of the subrun of w from y to z (we get condition 4). ◀

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*

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

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, performed in Appendix A.

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 $1 \le j \le n$, let r_j be the maximum of the sizes of j-pds's of $w(0)$. Define*

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

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 $G \subseteq pre_w^n(|w|)$, as $pre_w^n(|w|)$ 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. $x \in pre_w^0(y)$, and
3. either $\pi(w(x)) \neq \pi(w(y))$, or $G \cap \{x, x+1, \dots, y-1\} \neq \emptyset$, 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 $v_1, v_2, v_3, \dots$ beginning at $w(x)$ such that for each $j \geq 1$ we have

a) $0 \in pre_{v_j}^0(|v_j|)$, 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 $0 \in pre_v^0(|v|)$; together with $0 \in pre_{v_j}^0(|v_j|)$ it gives us that $0 \in pre_{v_{j+1}}^0(|v_{j+1}|)$. 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 $S_n \geq N_n$ (see Appendix A). 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.

Finally, consider an arbitrary language L and arbitrary PDS $\mathcal{A}$. Roughly speaking, the theorem is then true, because we can make a product of $\mathcal{A}$ with a finite automaton recognizing L. More precisely, we proceed as follows. Let $\mathcal{A} = (A, \Gamma, \gamma_I, Q, q_I, \Delta, \lambda)$, and let $A' = A \cup \{\$\}$, where $\$ \not\in A$ is a fresh symbol. Let $\mathcal{A}_L = (Q_L, A', q_L^I, F, \Delta_L)$ be a deterministic finite automaton recognizing language $A^*\$L$ (where Q_L is its set of states, A' its input alphabet, $q_L^I \in Q$ an initial state, $F \subseteq Q$ a set of accepting states, and $\Delta_L \colon Q_L \times A' \to Q_L$ a transition function). We assume that $\mathcal{A}_L$ has the additional property: there is a state $q_L^\$$ which is reached if and only if the input word read so far is from $A^*\$$. We create a new pushdown system $\mathcal{A}' = (A', \Gamma, \gamma_I, Q \times Q_L, (q_I, q_L^I), \Delta', \lambda')$. For each transition $\delta = (q_1, \gamma, q_2, op) \in \Delta$ and every $p \in Q_L$ we add to Δ'

- a transition $\delta' = ((q_1, p), \gamma, (q_2, p'), op)$, where $p' = p$ if $\lambda(\delta) = \varepsilon$ and $p' = \Delta_L(p, \lambda(\delta))$ otherwise; let the label of δ' be the label of δ;
- a transition $\delta' = ((q_1, p), \gamma, (q_2, \Delta_L(p, \$)), op)$ if $\lambda(\delta) \neq \varepsilon$; let the label of δ' be $\$$.

We also create a pushdown system $\mathcal{A}''$ which is almost identical to $\mathcal{A}'$, but its initial state is $(q_I, q_L^\$)$.

As $C_{\mathcal{A}L}$ we take the greater[4] of the constants for $\mathcal{A}'$ and $\mathcal{A}''$, and language A'^*, from the above special case. Let c be the configuration of $\mathcal{A}$ for which we want to do the pumping; in $PDG(\mathcal{A})/\varepsilon$ there is a path of length m from the initial configuration to c. Assume first that $m > 0$. Then there is a run w of $\mathcal{A}$ from the initial configuration to c such that exactly m of its edges are not labelled by ε, in particular the last edge. Let c' be the configuration with state $(state(c), q_L^\$)$ and the same pds as c. In $\mathcal{A}'$ we have a run w' from the initial configuration to c' with the same property (so there is a path in $PDG(\mathcal{A}')/\varepsilon$ of length m). Indeed, we replace each transition of w by a corresponding transition of $\mathcal{A}'$ of the first kind, except the last transition, which we replace by a transition of the second kind. The labels of this run, after omitting ε, form a word from $A^*\$$, so the second coordinate of $state(w'(|w'|))$ is $q_L^\$$. Similarly, because in $PDG(\mathcal{A})/\varepsilon$ there is a path p of length at least S_n which begins in c and belongs to L, in $PDG(\mathcal{A})/\varepsilon$ there is a path of length at least S_n which begins in c' and ends in a configuration having an accepting state of $\mathcal{A}_L$ on the second coordinate of the state. We use the above special case for $\mathcal{A}'$ (for language containing all words). We obtain infinitely many paths in $PDG(\mathcal{A}')/\varepsilon$, which start in c', and end in configurations having the same state as the original path. Because the second coordinate of this state is an accepting state of $\mathcal{A}_L$, the paths belong to L. Thus there are infinitely many paths in $PDG(\mathcal{A})/\varepsilon$, which start in c, belong to L, and end in configurations having the same state as p. The case $m = 0$ is solved analogously, but as c' we take the initial configuration of $\mathcal{A}''$. ◀

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.

[4] In fact these two constants are the same, as they do not depend on the initial state.

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 $0^i 1$, 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.

A Combinatorial lemmas

In this appendix we prove combinatorial facts used throughout the paper. We begin by a proof of Lemma 4.2. In its proof we use the following auxiliary lemma (where the set $H_{b,e}$ is defined as in the statement of Lemma 4.2).

▶ **Lemma A.1.** *Let N be a positive integer, let $a_B, a_{B+1}, \dots, a_E$ be a sequence of positive integers such that $|a_i - a_{i-1}| \leq 1$ and $a_i \geq a_B$ for $B+1 \leq i \leq E$. Let $G \subseteq \{B, B+1, \dots, E-1\}$ be such that $|G| \geq 2^N - 1$. Then there exists an index e such that $e-1 \in G$ and $|H_{B,e}| \geq N$.*

Proof. Induction on $E-B$. Notice that for $B \leq b \leq e$ we always have $H_{b,e} \subseteq H_{B,e}$. We have the following cases.

Case 0. Assume $N=1$. Then as e we can take any index such that $e-1 \in G$ (notice that $|G| \geq 2^1 - 1 = 1$). Let i be the greatest index such that $i \leq e-1$ and $a_i = a_B$ (it exists as B can be always taken as i). Notice that $i \in H_{B,e}$, because $a_j \geq a_i$ for $B \leq j \leq E$, and $a_j \geq a_i + 1$ for $i+1 \leq j \leq e-1$ ($e-1$ can be taken as g from the definition of $H_{B,e}$).

Case 1. Assume for some b such that $B+1 \leq b \leq E$, we have $a_b = a_B$ and $G \subseteq \{b, b+1, \dots, E-1\}$. We use the induction assumption for the shorter sequence $a_b, a_{b+1}, \dots, a_E$, and for N and G. The assumption that $a_i \geq a_b$ for $b+1 \leq i \leq E$ is satisfied because $a_b = a_B$. We obtain e such that $e-1 \in G$ and $|H_{b,e}| \geq N$. Since $H_{b,e} \subseteq H_{B,e}$ (in fact even $H_{b,e} = H_{B,e}$), this e is good for the original sequence.

Case 2. Assume that $N \geq 2$ and that for some M such that $B+1 \leq M \leq E$ we have $a_i \geq a_B + 1$ for $B+1 \leq i \leq M$, and $|G \cap \{B+1, B+2, \dots, M-1\}| \geq 2^{N-1} - 1$. Let $G' = G \cap \{B+1, B+2, \dots, M-1\}$. We use the induction assumption for the shorter sequence $a_{B+1}, a_{B+2}, \dots, a_M$, and for $N-1$ (as N) and G' (as G). The assumption that $a_i \geq a_{B+1}$ for $B+2 \leq i \leq M$ is satisfied because $a_{B+1} = a_B + 1$. We obtain e such that $e-1 \in G' \subseteq G$ and $|H_{B+1,e}| \geq N-1$. This e is good for the original sequence: because $e \leq M$ we see that $B \in H_{B,e}$ (as g in the definition of $H_{B,e}$ we can take $e-1$); we have $H_{B,e} = \{B\} \cup H_{B+1,e}$, so $|H_{B,e}| = |H_{B+1,e}| + 1 \geq N$.

Case 3. Assume that none of the above cases hold. Then in particular $N \geq 2$. Let $b \geq B+1$ be the smallest index such that $a_b = a_B$. If such a number b would not exist, then $a_i \geq a_B + 1$ for $B+1 \leq i \leq E$ and $|G \cap \{B+1, B+2, \dots, E-1\}| \geq |G| - 1 \geq 2^N - 2 \geq 2^{N-1} - 1$, so we could take $M = E$ in case 2; thus such a number b exists. Let $G' = G \cap \{b, b+1, \dots, E-1\}$. We have $a_i \geq a_B + 1$ for $B+1 \leq i \leq b-1$. Because $b-1$ can not be taken as M in case 2, we either have $b-1 = B$, or $|G \cap \{B+1, B+2, \dots, b-2\}| \leq 2^{N-1} - 2$. We have $|G'| \geq (2^N - 1) - 2 - (2^{N-1} - 2) = 2^{N-1} - 1$, as $G' = G - \{B, b-1\} - (G \cap \{B+1, B+2, \dots, b-2\})$. We use the induction assumption for the shorter sequence $a_b, a_{b+1}, \dots, a_E$, and for $N-1$ (as N) and G' (as G). The assumption that $a_i \geq a_b$ for $b+1 \leq i \leq E$ is satisfied because $a_b = a_B$. We obtain an index e such that $e-1 \in G' \subseteq G$ and $|H_{b,e}| \geq N-1$. Observe that this e is good also for the original sequence. Because case 1 does not hold, we have $G \not\subseteq \{b, b+1, \dots, E-1\}$, so for some $g \in G$ we have $B \leq g \leq b-1$, which means that $B \in H_{B,e}$; of course $B \notin H_{b,e}$. Thus $|H_{B,e}| \geq |H_{b,e}| + 1 \geq N$. ◀

Proof (Lemma 4.2). For each $0 \leq j \leq M$ denote

$$m_j = \min_{0 \leq i \leq j} a_i.$$

Notice that $m_0, m_1, \dots, m_M$ is a non-increasing sequence of positive integers such that $m_0 = a_0$. Because we have at least $a_0 \cdot 2^N$ elements of G, for some m there are at least 2^N elements of $G \cap \{i : m_i = m\}$. Choose such an m; let b be the first index such that $m_b = m$ and E the last such index. Let $G' = G \cap \{b, b+1, \dots, E-1\}$. We have $a_i \geq a_b = m$ for $b+1 \leq i \leq E$, and $|G'| \geq 2^N - 1$, so we can use Lemma A.1 for sequence $a_b, a_{b+1}, \dots, a_E$, for N and G' (as G). We obtain e such that $e \in G' \subseteq G$ and $|H_{b,e}| \geq N$ (we get condition 3). Because $e \leq E$ we get condition 1, and by minimality of b we get condition 2. ◀

Addendum to the proof of Lemma 5.2. Below we check that the set G has enough elements, so that Lemma 4.1 can be applied. Recall that any i-pds ($1 \leq i \leq k-1$) of $v(0)$ has size at most M_i'.

Thus it is enough to show that $M_k - M'_k \geq N_{k-1}$ (we have $|G| \geq M_k - M'_k$). We will show by induction on i that $M_i - M'_i \geq N_{i-1}$ for $1 \leq i \leq k$. For $i = 1$ we just have $M_1 - M'_1 = |\mathcal{T}_\mathcal{A}|+1 = N_0$. For $i \geq 2$ we have

$$M_i - M'_i = 2^{M_{i-1}} - 2^{M'_{i-1}} = 2^{M'_{i-1}}(2^{M_{i-1}-M'_{i-1}} - 1) \geq 2^{M'_{i-1}}(2^{N_{i-2}} - 1) \geq$$
$$\geq 2 \cdot M'_{i-1}(2^{N_{i-2}} - 1) \geq M'_{i-1} \cdot 2^{N_{i-2}} = N_{i-1}.$$

Above, the first inequality is true because of the induction assumption, the second inequality because for any natural number m we have $2^m \geq 2m$, and the last inequality because $N_{i-2} \geq 1$, so $2^{N_{i-2}} - 1 \geq 2^{N_{i-2}-1}$. ◀

Addendum to the proof of Theorem 1.1. Below we prove that $S_n \geq N_n$. In fact we will prove that $S_j \geq 3N_j$ by induction on j (for $1 \leq j \leq n$). For $j = 1$ we have

$$S_1 = (m+1) \cdot 3 \cdot (|\mathcal{T}_\mathcal{A}| + 1) \cdot 2^{|\mathcal{T}_\mathcal{A}|+1} = 3 \cdot M_1 \cdot 2^{N_0} = 3N_1.$$

Let now $j \geq 2$. It is easy to see that $N_{j-1} \geq N_0 \geq 2$, so $2^{N_{j-1}} \geq 3$. We also have $N_{j-1} = M_{j-1} \cdot 2^{N_{j-2}} \geq M_{j-1}$, so $2^{N_{j-1}} \geq 2^{M_{j-1}} = M_j$. From the induction assumption we get that

$$S_j = 2^{S_{j-1}} \geq 2^{3N_{j-1}} = 2^{N_{j-1}} \cdot 2^{N_{j-1}} \cdot 2^{N_{j-1}} \geq 3 \cdot M_j \cdot 2^{N_{j-1}} = 3N_j.$$

◀

B Types and sequence equivalence

B.1 Types of pds's

In this section we prove Lemma 3.8. This proof is by induction on k (from n to 0). To prove this lemma, we need to define a relation $\sim_l$ over k-pds's for $l \leq k \leq n$, and a relation $\approx_l$ over configurations; then we prove some auxiliary lemmas.

▶ **Definition B.1.** Let $0 \leq l \leq n$. The relation $\sim_l$ is the smallest relation satisfying the following:

- for any k-pds α^k, where $l \leq k \leq n$, we have $\alpha^k \sim_l \alpha^k$, and
- for any k-pds's α^k, β^k such that $type(\alpha^k) = type(\beta^k)$, where $l+1 \leq k \leq n$ (but not $k = l$), we have $\alpha^k \sim_l \beta^k$, and
- for any k-pds's α^k, β^k and $(k-1)$-pds's α^{k-1}, β^{k-1} such that $\alpha^k \sim_l \beta^k$ and $\alpha^{k-1} \sim_l \beta^{k-1}$, where $l+1 \leq k \leq n$, we have $\alpha^k : \alpha^{k-1} \sim_l \beta^k : \beta^{k-1}$.

▶ **Definition B.2.** Let $0 \leq l \leq n$, and let c and c' be configurations. Decompose

$$\pi(c) = \alpha^n : \alpha^{n-1} : \cdots : \alpha^l \qquad \text{and} \qquad \pi(c') = \alpha'^n : \alpha'^{n-1} : \cdots : \alpha'^l.$$

We say that $c \approx_l c'$ if $state(c) = state(c')$ and $\alpha^k \sim_l \alpha'^k$ for $l \leq k \leq n$.

At the end of our proofs it will be easy to see that the above relations are equivalence relations (transitivity follows from Proposition B.8). However there is no direct argument showing that, so in the proofs we cannot assume transitivity of these relations.

The following propositions are immediate consequences of the definitions.

▶ **Proposition B.3.** *Let $0 \leq l \leq k \leq n$ and let α^k, β^k be k-pds's such that $\alpha^k \sim_l \beta^k$. Decompose $\alpha^k = \alpha'^k : \alpha'^{k-1} : \cdots : \alpha'^l$ and $\beta^k = \beta'^k : \beta'^{k-1} : \cdots : \beta'^l$. Then either*

- *$\alpha'^i \sim_l \beta'^i$ for $l \leq i \leq k$, or*
- *there exists a number m such that $l+1 \leq m \leq k$, and $\alpha'^i \sim_l \beta'^i$ for $m+1 \leq i \leq k$, and $type(\alpha'^m : \alpha'^{m-1} : \cdots : \alpha'^l) = type(\beta'^m : \beta'^{m-1} : \cdots : \beta'^l)$.*

Indeed, from the definition of $\sim_l$, either the types match, or we have the last case of the definition, which inductively gives the property.

▶ **Proposition B.4.** *Let $0 \leq l \leq n$, and let w and v be runs of length 1 such that $w(0) \approx_l v(0)$. Assume that these runs perform the same transition, which uses a push^k operation for any k, or a pop^k operation for $1 \leq k \leq l$. Then $w(1) \approx_l v(1)$.*

▶ **Proposition B.5.** *Let $1 \leq k \leq n$, and let α^k be a k-pds. Then type$(\alpha^k) = \emptyset$ if and only if α^k is empty.*

The next proposition has in its assumptions the induction assumption of Lemma 3.8, and the induction assumption of Lemma B.7.

▶ **Proposition B.6.** *Let $0 \leq l \leq k \leq n$ and let α^k, β^k be k-pds's such that $\alpha^k \sim_l \beta^k$. Assume that the statement of Lemma 3.8 holds for $k \geq l+1$. Assume also that for each $t \in \mathcal{D}^k$, each run w which agrees with t, and each configuration c such that $w(0) \approx_l c$ there exists a run from c which agrees with t. Then type$(\alpha^k) =$ type(β^k).*

Proof. Decompose $\alpha^k = \alpha'^k : \alpha'^{k-1} : \cdots : \alpha'^l$ and $\beta^k = \beta'^k : \beta'^{k-1} : \cdots : \beta'^l$. Take any $t' = (\rho^n, \rho^{n-1}, \ldots, \rho^{k+1}, q, t) \in type(\alpha^k)$. To prove that $t \in type(\beta^k)$, for $k+1 \leq i \leq n$ take any β'^i such that $type(\beta'^i) = \rho^i$. Let $c = (q, \beta'^n : \beta'^{n-1} : \cdots : \beta'^{k+1} : \beta^k)$. We have to find a run from c which agrees with t. By definition of *type*, because $t \in type(\alpha^k)$, there is a run w from $(q, \beta'^n : \beta'^{n-1} : \cdots : \beta'^{k+1} : \alpha^k)$ which agrees with t.

If $\alpha'^i \sim_l \beta'^i$ for $l \leq i \leq k$, we have $w(0) \approx_l c$. Then (by the second assumption of the proposition), there exists a run from c which agrees with t, as required. Otherwise (Proposition B.3), there exists a number m such that $l+1 \leq m \leq k$, and $\alpha'^i \sim_l \beta'^i$ for $m+1 \leq i \leq n$, and $type(\alpha'^m : \alpha'^{m-1} : \cdots : \alpha'^l) = type(\beta'^m : \beta'^{m-1} : \cdots : \beta'^l)$. Let

$$d = (q, \beta'^n : \beta'^{n-1} : \cdots : \beta'^{m+1} : \alpha'^m : \alpha'^{m-1} : \cdots : \alpha'^l).$$

We have $w(0) \approx_l d$. Then (by the second assumption of the proposition), there exists a run v from d which agrees with t. Next, we use Lemma 3.8 for m (which is $\geq l+1$) as k, for t, and for v as w; we have assumed that it is true. It says that

$$(type(\beta'^n), type(\beta'^{n-1}), \ldots, type(\beta'^{m+1}), q, t) \in type(\alpha'^m : \alpha'^{m-1} : \cdots : \alpha'^l).$$

By the definition of *type*, and by the equality of *type* for $\alpha'^m : \alpha'^{m-1} : \cdots : \alpha'^l$ and $\beta'^m : \beta'^{m-1} : \cdots : \beta'^l$ we get a run from c which agrees with t, as required. ◀

The next lemma will be used in the proof of Lemma 3.8. The induction assumption for Lemma 3.8 is assumed in Lemma B.7.

▶ **Lemma B.7.** *Let $0 \leq l \leq r-1 \leq n$. Assume that the statement of Lemma 3.8 holds for $k \geq l+1$. Let $t \in \mathcal{D}^{r-1}$, let w be a run which agrees with t, and let c be a configuration such that $w(0) \approx c$. Then there exists a run from c which agrees with t.*

Proof. We make an external induction on r (from $n+1$ to $l+1$), and an internal induction on the length of w. To avoid special case for $r = n+1$, we assume that $\mathcal{D}^{n+1} = \emptyset$. If $t \in \mathcal{D}^r$, we can directly use the external induction assumption. Thus we assume that $t \in \mathcal{D}^{r-1} - \mathcal{D}^r$.

Let $\pi(w(0)) = \alpha^n : \alpha^{n-1} : \cdots : \alpha^0$ and $\pi(c) = \alpha'^n : \alpha'^{n-1} : \cdots : \alpha'^0$. Notice (by definition of $\approx$ and $\sim$) that $w(0) \approx c$ implies that $\alpha^0 = \alpha'^0$, and $state(w(0)) = state(c)$, and α^i is empty if and only if α'^i is empty for $1 \leq i \leq n$. We have a trivial case when $|w| = 0$. A run of length 0 can agree only with $t = state(w(0))$; then the run of length 0 from c also agrees with t. For the rest of the proof we assume that $|w| \geq 1$. Let $\delta \in \Delta$ be the first transition used in w, between $w(0)$ and $w(1)$. Notice that the same transition can be performed from c; let c_1 be the resulting configuration. We have $state(w(1)) = state(c_1)$.

First we prove the following claim, denoted by (*).

> Let $u \in \mathcal{D}^{r-1}$ be such that the subrun of w from 1 to $|w|$ agrees with u. Assume also that the operation in δ is pop^k for $1 \leq k \leq r-1$ or push^k for $1 \leq k \leq r-1$. Then there exists a run from c_1 which agrees with u.

To prove the claim, assume first that the operation in δ is not pop^k for $k \geq l+1$. Then, by Proposition B.4, $w(1) \approx_l c_1$. Thus we can use the internal induction assumption for u (as t), the subrun of w from 1 to $|w|$ (as w), and c_1 (as c); we obtain a run from c_1 which agrees with u, as required by the claim.

Next assume that the operation in δ is pop^k for $k \geq l+1$. Decompose

$$\pi(w(1)) = \rho^n : \rho^{n-1} : \dots : \rho^l \qquad \text{and} \qquad \pi(c_1) = \rho'^n : \rho'^{n-1} : \dots : \rho'^l.$$

Notice that $\rho^i = \alpha^i$ and $\rho'^i = \alpha'^i$ for $k+1 \leq i \leq n$, and $\rho^k : \rho^{k-1} : \dots : \rho^l = \alpha^k$ and $\rho'^k : \rho'^{k-1} : \dots : \rho'^l = \alpha'^k$. We have $state(w(1)) = state(c_1)$ and $\rho^i \sim_l \rho'^i$ for $k+1 \leq i \leq n$, and $\rho^k : \rho^{k-1} : \dots : \rho^l \sim_l \rho'^k : \rho'^{k-1} : \dots : \rho'^l$. When also $\rho^i \sim_l \rho'^i$ for $l \leq i \leq k$, we have $w(1) \approx_l c_1$ and we are done (see case 1a). Otherwise (by Proposition B.3) there exists a number m such that $l+1 \leq m \leq k$, and $\rho^i \sim_l \rho'^i$ for $m+1 \leq i \leq n$, and $type(\rho^m : \rho^{m-1} : \dots : \rho^l) = type(\rho'^m : \rho'^{m-1} : \dots : \rho'^l)$. Let

$$d = (state(w(1)), \rho'^n : \rho'^{n-1} : \dots : \rho'^{m+1} : \rho^m : \rho^{m-1} : \dots : \rho^l).$$

We have $w(1) \approx_l d$. From the internal induction assumption for u (as t), the subrun of w from 1 to $|w|$ (as w), and d (as c) we obtain a run w' from d which agrees with u. Next, we use Lemma 3.8 for m (which is $\geq l+1$) as k, for u as t, and for w' as w; we have assumed the it is true. It says that

$$(type(\rho'^n), type(\rho'^{n-1}), \dots, type(\rho'^{m+1}), state(d), u) \in type(\rho^m : \rho^{m-1} : \dots : \rho^l).$$

By the definition of $type$, and by the equality of $type$ for $\rho^m : \rho^{m-1} : \dots : \rho^l$ and $\rho'^m : \rho'^{m-1} : \dots : \rho'^l$, we obtain a run v' from c_1 which agrees with u. This finishes the proof of the claim.

We have the following cases depending on the form of t.

Case 1. Consider first the case when $t \in Q$. Then $r = n+1$. The subrun w' of w from 1 to $|w|$ also agrees with t. From claim (*) follows that there exists a run v' from c_1 which agrees with t. Then as v we take the composition of the one-step run from c to c_1 with run v'; of course v also agrees with t.

Case 2. Next, consider the case when t is of the form $t = (r, \mathsf{non}\text{-}\varepsilon)$ (recall that $t \in \mathcal{D}^{r-1} - \mathcal{D}^r$, so $t = (r', \mathsf{non}\text{-}\varepsilon)$ implies that $r' = r$). We have $r \geq l$, as $t \in \mathcal{D}^l$. One possibility is that at least one of the edges of w before the last edge is labelled by an element of A. Then the subrun w' of w from 0 to $|w|-1$ also agrees with t. From the internal induction assumption, used for t, w' (as w), and c, we obtain a run from c which agrees with t. Thus we can assume that only the last edge of w is labelled by an element of A, and all earlier edges are labelled by ε.

Next, observe that the operation in δ cannot be pop^k for $k \geq r$, as then position $top^{r-1}(w(0))$ would not be present in $w(1)$. We have several subcases.

Case 2a. Assume first that $|w| = 1$. Then $\lambda(\delta) \neq \varepsilon$, thus the one-step run from c to c_1 agrees with t (since the operation in δ is not pop^k for $k \geq r$, position $top^{r-1}(c)$ is present in c_1).

Case 2b. Assume that $|w| \geq 2$, and the operation in δ is pop^k for $1 \leq k \leq r-1$, or push^k for $1 \leq k \leq r-1$. Then $top^{r-1}(w(0)) = top^{r-1}(w(1))$, so the subrun w' of w from 1 to $|w|$ agrees with t (as position $top^{r-1}(w'(0))$ is present in every configuration of w'). From claim (*) (used for t as u) we obtain a run v' from c_1 which agrees with t. Let v be the composition of the one-step run from c to c_1 with run v'. Again because $top^{r-1}(v(0)) = top^{r-1}(v(1))$, we obtain that v agrees with t.

Case 2c. Assume that $|w| \geq 2$, and the operation in δ is push^k for $r \leq k \leq n$, and position $top^{k-1}(w(1))$ is present in every configuration of w, starting from $w(1)$. This means that the subrun w' of w from 1 to $|w|$ agrees with $u = (k, \mathsf{non}\text{-}\varepsilon)$. From the internal induction assumption, used for u (as t), w' (as w), and c_1 (as c), we obtain a run v' from c_1 which agrees with u. Let v be the composition of the one-step run from c to c_1 with run v'. Of course position $top^{r-1}(v(0))$ cannot be removed if position $top^{k-1}(v(1))$ is not removed (as $top^{r-1}(v(0))$ is deeper than $top^{k-1}(v(1))$), so v agrees with t.

Case 2d. Finally, assume that $|w| \geq 2$, the operation in δ is push^k for $r \leq k \leq n$, and position $top^{k-1}(w(1))$ is not present in some configuration of w after $w(1)$; let $w(j)$ be the first such configuration. Since position $top^{r-1}(w(0))$ is still present in $w(j)$, and $r \leq k$, it has to be $\pi(w(0)) = \pi(w(j))$. If for some i such that $1 \leq i \leq j-1$ we have $i \in pre_w^{k-1}(j)$, it would mean that $top^{k-1}(w(1))$ is not present also in $w(i)$; this contradicts with our choice of j, which was the smallest such index. It follows that the subrun w_1 of w from 1 to j is a k-return. Run w_1 agrees with $u = (k, f, \mathit{type}(\alpha^n), \mathit{type}(\alpha^{n-1}), \dots, \mathit{type}(\alpha^{k+1}), \mathit{state}(w(j)), 0)$, where $f = 0$ if and only if all edges of w_1 are labelled ε (hence if and only if $j < |w|$). From the internal induction assumption, used for u (as t), w_1 (as w), and c_1 (as c), we obtain a run v_1 from c_1 which agrees with u. This means in particular that $\mathit{state}(v_1(|v_1|)) = \mathit{state}(w(j))$ and $\pi(v_1(|v_1|)) = \pi(c)$ (because the last coordinate of u is 0). Notice that, by definition of a k-return, position $\mathit{top}^{r-1}(c) = \mathit{top}^{r-1}(v_1(|v_1|))$ is present in every configuration of v_1. If $f = 1$ we are done: the composition of the one-step run from c to c_1 with run v_1 agrees with t. Otherwise, let w_2 be the subrun of w from j to $|w|$. Because $\mathit{top}^{r-1}(w(0)) = \mathit{top}^{r-1}(w(j))$, run w_2 agrees with t. From the internal induction assumption, used for t, w_2 (as w), and $v_1(|v_1|)$ (as c), we obtain a run v_2 from $v_1(|v_1|)$ which agrees with t. The composition of the one-step run from c to c_1 with run v_1 and with run v_2 is a run from c which agrees with t.

Case 3. Next, consider the case when t is of the form $t = (r, f, \xi^n, \xi^{n-1}, \dots, \xi^{r+1}, q, g)$ (again, $t = (r', f, \xi^n, \xi^{n-1}, \dots, \xi^{r'+1}, q, g)$ implies that $r' = r$, because $t \in \mathcal{D}^{r-1} - \mathcal{D}^r$). Decompose $\pi(w(|w|)) = \beta^n : \beta^{n-1} : \dots : \beta^r$. Because w is an r-return, we have $\beta^r = \alpha^r$. Additionally $\xi^i = \mathit{type}(\beta^i)$ for $r+1 \le i \le n$, and $q = \mathit{state}(w(|w|))$. Observe that the operation in δ cannot be pop^k for $k > r$. Indeed, then the bottommost position y of the $(r-1)$-pds just below the topmost $(r-1)$-pds of $w(0)$ would be removed in $w(1)$ by this operation, so we cannot have $\mathit{hist}_w(i, x)(0) = y$ for any $1 \le i \le |w|$ and any position x of $w(i)$; in particular w cannot be an r-return. We have several subcases depending on the operation in δ.

Case 3a. Assume first that the operation in δ is push^k for $1 \le k \le r-1$, or pop^k for $1 \le k \le r-1$. Let w' be the subrun of w from 1 to $|w|$. We see that w' is an r-return, because w is an r-return and between $w(0)$ and $w(1)$ we make changes only inside the topmost $(r-1)$-pds. We also see that $\pi(w'(|w'|))$ can be obtained from $\pi(w'(0))$ by removing its topmost $(r-1)$-pds if and only if $\pi(w(|w|))$ can be obtained from $\pi(w(0))$ by removing its topmost $(r-1)$-pds. Let $u = (r, f', \xi^n, \xi^{n-1}, \dots, \xi^{r+1}, q, g)$, where $f' = 0$ if and only if every edge of w' is labelled by ε. We get that w' agrees with u. From claim (*) we have a run v' from c_1 which agrees with u. Let v be the composition of the one-step run from c to c_1 with run v'; we obtain a run from c. Because v' is an r-return, also v is an r-return. Moreover $\pi(v(|v|))$ can be obtained from $\pi(v(0))$ by removing its topmost $(r-1)$-pds if and only if $\pi(v'(|v'|))$ can be obtained from $\pi(v'(0))$ by removing its topmost $(r-1)$-pds. Additionally, because $f = 0$ if and only if $f' = 0$ and $\lambda(\delta) = \varepsilon$, we see that $f = 0$ if and only if every edge of v is labelled by ε. It follows that v agrees with t. Thus it remains to show that there exists a run from c_1 which agrees with u.

Case 3b. Next, assume that the operation in δ is pop^r. We have $\pi(w(1)) = \alpha^n : \alpha^{n-1} : \dots : \alpha^r$ and $\pi(c_1) = \alpha'^n : \alpha'^{n-1} : \dots : \alpha'^r$. Let y be the bottommost position of the $(r-1)$-pds just below the topmost $(r-1)$-pds of $w(0)$ (like in Definition 3.4). We have $\mathit{hist}_w(1, \mathit{top}^{r-1}(w(1)))(0) = y$ and no other copy of y is present in $w(1)$; thus $\mathit{hist}_w(|w|, \mathit{top}^{r-1}(w(|w|)))(0) = y$ implies $1 \in \mathit{pre}_w^{r-1}(|w|)$. From Definition 3.4 it follows that $|w| = 1$. We have $q = \mathit{state}(w(1))$, and $g = 0$, and $\beta^i = \alpha^i \sim_l \alpha'^i$ for $r+1 \le i \le n$ (also for $i = r$, but this is useless). To conclude that the one-step run from c to c_1 agrees with t, it remains to show that $\mathit{type}(\beta^i) = \mathit{type}(\alpha'^i)$ for $r+1 \le i \le n$. The external induction assumption says that for $i \ge r+1$, for each $t' \in \mathcal{D}^i$, each run w' which agrees with t', and each configuration c' such that $w'(0) \approx_l c'$ there exists a run from c' which agrees with t'. Exactly this, plus the assumption about Lemma 3.8 for $k \ge l+1$, allows us to use Proposition B.6 for β^i and α'^i; we get that $\mathit{type}(\beta^i) = \mathit{type}(\alpha'^i)$ as required.

Case 3c. Finally, assume the operation in δ is push^k for $r \le k \le n$. Consider first the situation when the subrun of w from 1 to $|w|$ is an r-return, and $k > r$; denote this subrun by w'. Notice that $\pi(w(|w|))$ cannot be obtained from $\pi(w(0))$ by removing its topmost $(r-1)$-pds, since even the $(r-1)$-pds of $w(1)$ just below its topmost $(r-1)$-pds is not removed in $w(|w|)$, as w' is an r-return; so $g = 0$. Let $u = (r, f', \xi^n, \xi^{n-1}, \dots, \xi^{r+1}, q, g')$, where $f' = 0$ if and only if every edge of w' is labelled by ε, and $g' = 0$ if and only if $\pi(w'(|w'|))$ can be obtained from $\pi(w'(0))$ by removing its topmost $(r-1)$-pds. We get that w' agrees with u. We use the internal induction assumption for u (as t), w' (as w), and c_1 (as c); we obtain a run v' from c_1 which agrees with u. Let v be the composition of the one-step run from c to c_1 with run v'; we obtain a run from c. Because v' is an r-return and $k \neq r$, also v is an r-return. Like above, $\pi(v(|v|))$ cannot be obtained from $\pi(v(0))$ by removing its topmost $(r-1)$-pds, because v' is an r-return. Additionally, because $f = 0$ if and only if $f' = 0$ and $\lambda(\delta) = \varepsilon$, we see that $f = 0$ if and only if every edge of v is labelled by ε. Thus v agrees with t.

Next, consider the situation when there exists j $(2 \leq j \leq |w| - 1)$ such that the subrun of w from 1 to j is a k-return, and the subrun of w from j to $|w|$ is an r-return. Denote this subruns w_1 and w_2. Notice that the topmost k-pds of $w(0)$ and of $w(j)$ are the same. Decompose $\pi(w(j)) = \rho^n : \rho^{n-1} : \cdots : \rho^{k+1} : \alpha^k : \alpha^{k-1} : \cdots : \alpha^l$. Let $u_1 = (k, f_1, type(\rho^n), type(\rho^{n-1}), \ldots, type(\rho^{k+1}), state(w(j)), g_1)$, where $f_1 = 0$ if and only if every edge of w_1 is labelled by ε, and $g_1 = 0$ if and only if $\pi(w(j))$ can be obtained from $\pi(w(1))$ by removing its topmost $(k-1)$-pds. Let $u_2 = (r, f_2, \xi^n, \xi^{n-1}, \ldots, \xi^{r+1}, q, g_2)$, where $f_2 = 0$ if and only if every edge of w_2 is labelled by ε, and $g_2 = 0$ if and only if $\pi(w(|w|))$ can be obtained from $\pi(w(j))$ by removing its topmost $(r-1)$-pds. We see that w_1 agrees with u_1 and w_2 agrees with u_2. We use the internal induction assumption for u_1 (as t), w_1 (as w), and c_1 (as c); we obtain a run v_1 from c_1 which agrees with u_1. Notice that the topmost k-pds of c and of $v_1(|v_1|)$ are the same. Decompose $\pi(v_1(|v_1|)) = \rho'^n : \rho'^{n-1} : \cdots : \rho'^{k+1} : \alpha'^k : \alpha'^{k-1} : \cdots : \alpha'^l$. Because $type(\rho^i) = type(\rho'^i)$ for $k+1 \leq i \leq n$, we get that $w(j) \approx_l v_1(|v_1|)$. We use the internal induction assumption for u_2 (as t), w_2 (as w), and $v_1(|v_1|)$ (as c); we obtain a run v_2 from $v_1(|v_1|)$ which agrees with u_2. Let v be the composition of the one-step run from c to c_1 with run v_1 and with run v_2. Because v_1 is a k-return and v_2 is an r-return, and $r \leq k$, we get that v is an r-return. Because $f = 0$ if and only if $f_1 = 0$ and $f_2 = 0$ and $\lambda(\delta) = \varepsilon$, we see that $f = 0$ if and only if every edge of v is labelled by ε. We see that $\pi(w(|w|))$ can be obtained from $\pi(w(0))$ by removing its topmost $(r-1)$-pds if and only if $\pi(w_1(|w_1|))$ can be obtained from $\pi(w_1(0))$ by removing its topmost $(k-1)$-pds and $\pi(w_2(|w_2|))$ can be obtained from $\pi(w_2(0))$ by removing its topmost $(r-1)$-pds; similarly for v and v_1, v_2. If follows that v agrees with t.

It remains to show that, when the operation in δ is push^k for $r \leq k \leq n$, we always have one of the above situations. Let s be the size of the topmost k-pds of $w(0)$. Let y be the bottommost position of the $(r-1)$-pds just below the topmost $(r-1)$-pds of $w(0)$. For $1 \leq j \leq |w|$, we look at the size of the k-pds of $w(j)$ containing $hist_w(|w|, top^{r-1}(w(|w|)))(j)$ (equivalently: containing $hist_w(|w|, top^k(w(|w|)))(j)$). Recall that $hist_w(|w|, top^{r-1}(w(|w|)))(0) = y$ (and y is in the topmost k-pds of $w(0)$, and push^k makes changes only inside the topmost k-pds), so for $j = 1$ this is the topmost k-pds (in other words $1 \in pre^k_w(|w|)$) and its size is $s + 1$. Assume first that this size is at least $s + 1$ for each j. Then $1 \in pre^{k-1}_w(|w|)$ (by Proposition 2.2). Because w is an r-return, we know that $1 \notin pre^{r-1}_w(|w|)$, so $r \neq k$ $(r < k)$. We see that $hist_w(|w|, top^{r-1}(w(|w|)))(1)$ is the copy of y in the topmost $(k-1)$-pds, so it is the bottommost position of the $(r-1)$-pds just below the topmost $(r-1)$-pds of $w(1)$. We get that the subrun of w from 1 to $|w|$ is an r-return, and $r \neq k$. The opposite possibility is that for some j $(1 \leq j \leq |w|)$ the size of the k-pds containing $hist_w(|w|, top^{r-1}(w(|w|)))(j)$ becomes s. Fix the first such j. It is easy to see that the subrun of w from 1 to j is a k-return, and the subrun of w from j to $|w|$ is an r-return. ◀

Proof (Lemma 3.8). This is induction on k. Denote

$$t' = (type(\alpha^n), type(\alpha^{n-1}), \ldots, type(\alpha^{k+1}), state(w(0)), t).$$

To prove that $t' \in type(\alpha^k)$, for $k+1 \leq i \leq n$ take any i-pds β^i such that $type(\beta^i) = type(\alpha^i)$. Let $c = (state(w(0)), \beta^n : \beta^{n-1} : \cdots : \beta^{k+1} : \alpha^k)$. We have to prove that there exists a run from c which agrees with t. Observe that $w(0) \approx_k c$. We use Lemma B.7 for k (as l), w, t, and c. Its assumptions are satisfied thanks to our induction assumption. We obtain a run from c which agrees with t, as required. ◀

▶ **Proposition B.8.** *Let $1 \leq k \leq n$, let α^k, β^k be k-pds's, and let $\alpha^{k-1}, \beta^{k-1}$ be $(k-1)$-pds's. Assume that $type(\alpha^k) = type(\beta^k)$ and $type(\alpha^{k-1}) = type(\beta^{k-1})$. Then $type(\alpha^k : \alpha^{k-1}) = type(\beta^k : \beta^{k-1})$.*

Proof. Take any $t = (\rho^n, \rho^{n-1}, \ldots, \rho^{k+1}, q, t') \in type(\alpha^k : \alpha^{k-1})$. To prove that $t \in type(\beta^n : \beta^{n-1})$, for $k+1 \leq i \leq n$ take any β^i such that $type(\beta^i) = \rho^i$. Let $c = (q, \beta^n : \beta^{n-1} : \cdots : \beta^{k-1})$. We have to prove that there exists a run from c which agrees with t'. By definition of *type*, because $t \in type(\alpha^k : \alpha^{k-1})$, there is a run from $(q, \beta^n : \beta^{n-1} : \cdots : \beta^{k+1} : \alpha^k : \alpha^{k-1})$ which agrees with t'. From Lemma 3.8 it follows that $(\rho^n, \rho^{n-1}, \dots, \rho^{k+1}, type(\alpha^k), q, t') \in type(\alpha^{k-1})$; which by assumptions of the proposition means that $(\rho^n, \rho^{n-1}, \dots, \rho^{k+1}, type(\beta^k), q, t') \in type(\beta^{k-1})$. From the definition of *type* we get that there exists a run from c which agrees with t'. ◀

B.2 Types of configurations

In this section we prove Facts 3.1–3.3. To simplify the notation we have the following definition.

▶ **Definition B.9.** For a configuration c and for $1 \le k \le n$ we define $pds^k(c) \in \Gamma^k_*$ as the topmost k-pds of c with its topmost $(k-1)$-pds removed. Additionally $pds^0(c)$ is the topmost 0-pds of c.

In other words, we always have $\pi(c) = pds^n(c) : pds^{n-1}(c) : \dots : pds^0(c)$. Recall that

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

Proof (Fact 3.1). If $|w| = 0$, as v we can take the trivial run consisting of c. So assume that $|w| \ge 1$. Let $\delta \in \Delta$ be the first transition used in w, between $w(0)$ and $w(1)$. Notice that the same transition can be performed from c; let c_1 be the resulting configuration. We have $state(w(1)) = state(c_1)$. Observe that the operation in δ cannot be pop^k for any k, as such operation removes the topmost symbol of $w(0)$, which contradicts with the assumption that $0 \in pre^0_w(|w|)$. So the operation in δ is push^k for some k.

Case 1. First, consider the special case when $|w| = 1$. As v we take the one-step run from c to c_1. We have $\pi(c) \ne \pi(c_1)$ (condition 1) and $0 \in pre^0_w(1)$ (condition 2). The only edge of w and of v is labelled by the same, i.e. by $\lambda(\delta)$ (condition 3). Thanks to Proposition B.8 we get $type_{\mathcal{A}}(w(1)) = type_{\mathcal{A}}(c_1)$ (condition 4).

Case 2. Next, consider the special case when the subrun w' of w from 1 to $|w|$ is a k-return (where the operation in δ is push^k for the same k). Notice that w' agrees with

$$t = (k, f, type(pds^n(w(|w|))), type(pds^{n-1}(w(|w|))), \dots, type(pds^{k+1}(w(|w|))), state(w(|w|)), g),$$

where $f = 0$ if and only if every edge of w' is labelled by ε, and $g = 0$ if and only if $\pi(w(|w|))$ can be obtained from $\pi(w(1))$ by removing its topmost $(k-1)$-pds. In other words $g = 0$ if and only if $\pi(w(|w|)) = \pi(w(0))$. From Lemma 3.8 (used for w, t, and 0 as k) we get that

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

Thanks to Proposition B.8 we know that $type_{\mathcal{A}}(w(1)) = type_{\mathcal{A}}(c_1)$, so

$$(type(pds^n(c_1)), type(pds^{n-1}(c_1)), \dots, type(pds^1(c_1)), state(c_1), t) \in type(pds^0(c_1)).$$

By definition of *type*, there exists a run v' from c_1 which agrees with t. Let v be the composition of the one-step run from c to c_1 with run v'. Notice that $\pi(w(0)) \ne \pi(w(|w|))$ implies that $g = 1$ which implies that $\pi(v'(|v'|))$ cannot be obtained from $\pi(c_1)$ by removing the topmost $(k-1)$-pds, so $\pi(v(0)) \ne \pi(v(|v|))$ (we get condition 1). By definition of a k-return, we have $hist_v(|v|, top^{k-1}(v(|v|)))(1) = top^{k-1}(v(0))$, as $top^{k-1}(v(0))$ is the bottommost symbol of the $(k-1)$-pds of $v(1)$ just below its topmost $(k-1)$-pds. This means that $0 \in pre^{k-1}_v(|v|)$ (we get condition 2). All edges of w are labelled by ε if and only if $\lambda(\delta) = \varepsilon$ and $f = 0$, hence if and only if all edges of v are labelled by ε (we get condition 3). Because w' and v' agree with t, we have $state(w(|w|)) = state(v(|v|))$ and $type(pds^i(w(|w|))) = type(pds^i(v(|v|)))$ for $k+1 \le i \le n$. Moreover, the topmost k-pds of $w(|w|)$ and of $w(0)$ are the same, and the topmost k-pds of $v(|v|)$ and of $v(0)$ are the same. This way we get $type_{\mathcal{A}}(w(|w|)) = type_{\mathcal{A}}(v(|v|))$ (condition 4).

The general case. We make an induction on the length of w. Assume first that for some j such that $1 \le j \le |w| - 1$ we have $j \in pre^0_w(|w|)$. Since $0 \in pre^0_w(|w|)$, then also $0 \in pre^0_w(j)$. We use the induction assumption for the subrun of w from 0 to j (as w), and for c; we obtain a run v_1 from c. We have $type_{\mathcal{A}}(w(j)) = type_{\mathcal{A}}(v_1(|v_1|))$. Next, we use the induction assumption for the subrun of w from j to $|w|$ (as w), and for $v_1(|v_1|)$ (as c); we obtain a run v_2 from $v_1(|v_1|)$. Let v be the composition of runs v_1 and v_2. We have $type_{\mathcal{A}}(w(|w|)) = type_{\mathcal{A}}(v(|v|))$ (condition 4). Because

$0 \in pre^0_{v_1}(|v_1|)$ and $0 \in pre^0_{v_2}(|v_2|)$, we get $0 \in pre^0_v(|v|)$ (condition 2). Because $0 \in pre^0_v(|v_1|)$ and $|v_1| \in pre^0_v(|v|)$, if $\pi(v(0))$ and $\pi(v(|v|))$ are equal, then also $\pi(v(|v_1|))$ is equal to them. In the light of condition 1 from the induction assumption, this implies that $\pi(w(0)) = \pi(w(j)) = \pi(w(|w|))$ (we get condition 1). From condition 3 of the induction assumptions follows that every edge of w is labelled by ε if and only if every edge of v is labelled by ε (we get condition 3).

Thus in general it is enough to show that either $|w| = 0$ (in which case the thesis is trivial), or we have some of the two special cases described above, or there exists j such that $1 \leq j \leq |w|-1$ and $j \in pre^0_w(|w|)$. So assume that $|w| \geq 2$ and $1 \notin pre^0_w(|w|)$. Recall that the operation performed between $w(0)$ and $w(1)$ is push^k. Because $0 \in pre^0_w(|w|)$ (which means $hist_w(|w|, top^0(w(|w|)))(0) = top^0(w(0))$) and $1 \notin pre^0_w(|w|)$, it has to be $hist_w(|w|, top^0(w(|w|)))(1) = top^0(w(0))$ (as $top^0(w(0))$ has two copies in $w(1)$; these are $top^0(w(0))$ and $top^0(w(1))$). Let $j \geq 1$ be the smallest positive index for which $j \in pre^{k-1}_w(|w|)$; such j exists since $|w|$ always can be taken as j. If $j = |w|$, we see that the subrun of w from 1 to $|w|$ is a k-return, so we have case 2. So assume that $j < |w|$. Concentrate on the $(k-1)$-pds of $w(i)$ containing $hist_w(|w|, top^{k-1}(w(|w|)))(i)$ for $1 \leq i \leq j$. This is not the topmost $(k-1)$-pds for $i < j$ (because $i \notin pre^{k-1}_w(|w|)$, so it is never modified: either it remains unchanged (and everything in the $(k-1)$-pds's below it in the same k-pds), or we copy the whole k-pds. Because $hist_w(|w|, top^0(w(|w|)))(1) = top^0(w(0))$ is the topmost symbol in this $(k-1)$-pds, the same is true for $w(j)$: we have $hist_w(|w|, top^0(w(|w|)))(j) = top^0(w(j))$. It means that $j \in pre^0_w(|w|)$, thus we can take j to the inductive case. ◀

Proof (Facts 3.2, 3.3). For Fact 3.2 let $t = (1, \mathsf{non}\text{-}\varepsilon)$, and for Fact 3.3 let $t = state(w(|w|))$. In both cases w agrees with t. Also in both cases we want to find a run v from c which agrees with t. From Lemma 3.8 used for $k = 0$, for w, and for t, we know that

$$(type(pds^n(w(0))), type(pds^{n-1}(w(0))), \ldots, type(pds^1(w(0))), state(w(0)), t) \in \\ \in type(pds^0(w(0))).$$

Because $type_{\mathcal{A}}(w(0)) = type_{\mathcal{A}}(c)$ it means that

$$(type(pds^n(c)), type(pds^{n-1}(c)), \ldots, type(pds^1(c)), state(c), t) \in type(pds^0(c)).$$

From point 4 of Definition 3.7 it follows that there exists a run from c which agrees with t, as required. ◀

C The error in the Blumensath pumping lemma

We will show that Lemma 60 in [1] is false; this is a key lemma on which the pumping lemma is based. This lemma says that in each long enough run w of any pushdown system there exists a *pumping pair* of configurations $w(x), w(y)$ $(x < y)$. From the definition of a pumping pair we use only the following:

- $state(w(x)) = state(w(y))$ (the state in these configurations is the same),
- $\pi(w(x)) \lhd_1 \pi(w(y))$.

For two k-pds's α^k, β^k (for $0 \leq k \leq n$), the relation $\alpha^k \lhd_1 \beta^k$ is defined as follows (Definition 20 in [1]). If $k = 0$, then $\alpha^k \lhd_1 \beta^k$ always holds. For $k \geq 1$, suppose that $\alpha^k = a_1 a_2 \ldots a_r$ and $\beta^k = b_1 b_2 \ldots b_s$, where a_i, b_i are $(k-1)$-pds's. Then $\alpha^k \lhd_1 \beta^k$ holds if

$$r \leq s, \qquad a_i = b_i, \text{ for } 1 \leq i \leq r-1, \qquad \text{and} \qquad a_r \lhd_1 b_i, \text{ for } r \leq i \leq s.$$

Consider the following pushdown system $\mathcal{A}$ of level 3. The stack alphabet contains only the a symbol. The system first executes the $\mathsf{push}^1(a)$ operation, and then repeats the following sequence of operations:

$$\mathsf{push}^2(a), \quad \mathsf{push}^3(a), \quad \mathsf{pop}^1, \quad \mathsf{push}^3(a), \quad \mathsf{pop}^2, \quad \mathsf{push}^3(a).$$

Thus it has 7 states: an initial state and a loop of 6 states.

This pushdown system has only one infinite run starting in the initial configuration. Consider its subrun w starting from the second configuration. The n-pds of $w(0)$ is $[[[aa]]]$ (one 1-pds with two symbols). Our PDS works as follows. First observe that it never makes any pop^3 operation. Hence only the topmost 2-pds is accessed. By making a push^3 operation we keep the history of the current contents of the topmost 2-pds.

Now observe how the topmost 2-pds changes. It has three possible contents, between which we loop:

$$\alpha = [[aa]],$$
$$\beta = [[aa][aa]],$$
$$\gamma = [[aa][a]].$$

We have $\alpha \lhd_1 \beta$ and $\gamma \lhd_1 \beta$, but α and γ are $\lhd_1$-incomparable.

Lemma 60 says that in w there is a pumping pair $w(x), w(y)$ $(x < y)$. We will show that this is not true. Fix some such pair. Let $\pi(w(x)) = \xi_1 \dots \xi_k$ and $\pi(w(y)) = \zeta_1 \dots \zeta_l$. The configurations $w(x), w(y)$ have the same state, which means that $\xi_k = \zeta_l$. As $x < y$, it necessarily holds $l \geq k+3$. Because $\pi(w(x)) \lhd_1 \pi(w(y))$, it has to be

$$\xi_k \lhd_1 \zeta_k \qquad \text{and} \qquad \xi_k \lhd_1 \zeta_{k+1} \qquad \text{and} \qquad \xi_k \lhd_1 \zeta_{k+2}.$$

We know that $\zeta_1, \zeta_2, \dots, \zeta_{l-1} = \alpha, \beta, \gamma, \alpha, \beta, \gamma, \dots$ (and ζ_l is either equal to ζ_{l-1}, or is the next symbol). This means that

$$\xi_k \lhd_1 \alpha \qquad \text{and} \qquad \xi_k \lhd_1 \beta \qquad \text{and} \qquad \xi_k \lhd_1 \gamma.$$

But none of α, β, γ satisfies this. Hence there is no pumping pair.

Other arguments against the proof. After reading the above counterexample, one may think that maybe there is some small mistake in the definitions or in the statement of the lemma, and that possibly it can be corrected by improving some details. However the problem in the proof is much deeper. For simplicity assume that we consider a PDS of level 3 which has only one infinite run from the initial configuration. Lemma 60 says that there exists a pumping pair arbitrarily far in this run (as we can consider its subrun starting from any moment). From Lemma 57, and the definition of the substitution used there, it follows that the size of the 3-pds, and the maximal size of 1-pds's has to grow with similar speed (the dependence between them has to be linear). However it is not difficult to show a PDS in which the size of the 3-pds is exponential in the maximal size of 1-pds's.