Equivalence of pushdown automata via first-order grammars

Petr Jančar

Dept of Computer Science, Faculty of Science, Palacký University in Olomouc, Czechia

petr.jancar@upol.cz

Abstract

A decidability proof for bisimulation equivalence of first-order grammars is given. It is an alternative proof for a result by Sénizergues (1998, 2005) that subsumes his affirmative solution of the famous decidability question for deterministic pushdown automata.

The presented proof is conceptually simpler, and a particular novelty is that it is not given as two semidecision procedures but it provides an explicit algorithm that might be amenable to a complexity analysis.

1 Introduction

Decision problems for semantic equivalences have been a frequent topic in computer science. For pushdown automata (PDA) *language equivalence* was quickly shown undecidable, while the decidability in the case of deterministic PDA (DPDA) is a famous result by Sénizergues [1]. A finer equivalence, called *bisimulation equivalence* or *bisimilarity*, has emerged as another fundamental behavioural equivalence [2]; for deterministic systems it essentially coincides with language equivalence. By [3] we can exemplify the first decidability results for infinite-state systems (a subclass of PDA, in fact), and refer to [4] for a survey of results in a relevant area.

One of the most involved results in this area shows the decidability of bisimilarity of equational graphs with finite out-degree, which are equivalent to PDA with alternative-free ε-steps (if an ε-step is enabled, then it has no alternative); Sénizergues [5] has thus generalized his decidability result for DPDA.

We recall that the complexity of the DPDA problem remains far from clear, the problem is known to be PTIME-hard and to be in TOWER (i.e., in the first complexity class beyond elementary in the terminology of [6]); the upper bound was shown by Stirling [7] (and formulated more explicitly in [8]). For PDA the bisimulation equivalence problem is known to be nonelementary [9] (in fact, TOWER-hard), even for real-time PDA, i.e. PDA with no ε-steps. For the above mentioned PDA with alternative-free ε-steps the problem is even not primitive recursive; its Ackermann-hardness was shown in [8].

The decidability proofs, both for DPDA and PDA, are involved and hard to understand. This paper aims to contribute to a clarification of the more general decidability proof, showing an algorithm deciding bisimilarity of PDA with alternative-free ε-steps.

The proof is shown in the framework of labelled transition systems generated by *first-order grammars* (FO-grammars), which seems to be a particularly convenient formalism; it is called *term context-free grammars* in [10]. Here the states (or configurations) are first-order terms over a specified finite set of function symbols (or "nonterminals"); the transitions are induced by a *first-order grammar*, which is a finite set of labelled rules for rewriting the

1

roots of terms. This framework is equivalent to the framework of [5]; cf., e.g., [11, 10] and the references therein, or also [12] for a concrete transformation of PDA to FO-grammars. The proof here is in principle based on the high-level ideas from the proof in [5] but with various simplifications and new modifications. The presented proof has resulted by a thorough reworking of the conference paper [13], <mark>aiming to get an algorithm that might be amenable to a complexity analysis.</mark>

Proof overview. We give a flavour of the process that is formally realized in the paper. It is standard to characterize bisimulation equivalence (also called bisimilarity) in terms of a turn-based game between Attacker and Defender, say. If two PDA-configurations, modelled by first-order terms E, F in our framework, are non-bisimilar, then Attacker can force his win within k rounds of the game, for some number $k \in \mathbb{N}$; in this case $k-1$ for the least such k can be viewed as the *equivalence-level* $\text{EL}(E, F)$ of terms E, F: we write $E \sim_{k-1} F$ and $E \not\sim_k F$. If E, F are bisimilar, i.e. $E \sim F$, then Defender has a winning strategy and we put $\text{EL}(E, F) = \omega$. A natural idea is to search for a computable function f attaching a number $f(\mathcal{G}, E, F) \in \mathbb{N}$ to terms E, F and a grammar $\mathcal{G}$ so that it is guaranteed that $\text{EL}(E, F) \leq f(\mathcal{G}, E, F)$ or $\text{EL}(E, F) = \omega$; this immediately yields an algorithm that computes $\text{EL}(E, F)$ (concluding that $\text{EL}(E, F) = \omega$ when finding that $\text{EL}(E, F) > f(\mathcal{G}, E, F)$).

We will show such a computable function f by analysing optimal plays from $E_0 \not\sim F_0$; such an optimal play gives rise to a sequence $(E_0, F_0), (E_1, F_1), \ldots, (E_k, F_k)$ of pairs of terms where $\text{EL}(E_i, F_i) = \text{EL}(E_{i-1}, F_{i-1}) - 1$ for $i = 1, 2, \ldots, k$, and $\text{EL}(E_k, F_k) = 0$ (hence $\text{EL}(E_0, F_0) = k$). This sequence is then suitably modified to yield a certain sequence

$$(E_0', F_0'), (E_1', F_1'), \ldots, (E_k', F_k') \tag{1}$$

such that $(E_0', F_0') = (E_0, F_0)$ and $\text{EL}(E_i', F_i') = \text{EL}(E_i, F_i)$ for all $i = 1, 2, \ldots, k$; here we use simple congruence properties (if E' arises from E by replacing a subterm H with H' such that $H \sim_k H'$, then $E \sim_k E'$). Doing this modification carefully, adhering to a sort of "balancing policy" (inspired by one crucial ingredient in [1, 5], used also in [14]) we derive that if k is "large", then the sequence (1) contains a "long" subsequence

$$(\overline{E}_1\sigma, \overline{F}_1\sigma), (\overline{E}_2\sigma, \overline{F}_2\sigma), \ldots, (\overline{E}_z\sigma, \overline{F}_z\sigma), \tag{2}$$

called an (n, s, g)-sequence, where the variables in all "tops" $\overline{E}_j$, $\overline{F}_j$ are from the set $\{x_1, \ldots, x_n\}$, σ is the common "tail" substitution (maybe with "large" terms $x_i\sigma$), and the size-growth of the tops is bounded: $\text{SIZE}(\overline{E}_j, \overline{F}_j) \leq s + g \cdot (j-1)$ for $j = 1, 2, \ldots, z$. The numbers n, s, g are elementary in the size of the grammar $\mathcal{G}$. Then another fact is used (whose analogues in different frameworks could be traced back to [1, 5] and other related works): if $\text{EL}(\overline{E}_1, \overline{F}_1) = e < \ell = \text{EL}(\overline{E}_1\sigma, \overline{F}_1\sigma)$, then there is $i \in \{1, 2, \ldots, n\}$ and a term $H \neq x_i$ reachable from $\overline{E}_1$ or $\overline{F}_1$ within e moves (i.e. root-rewriting steps) such that $x_i\sigma \sim_{\ell-e} H\sigma$. This entails that for $j = e+2, e+3, \ldots, z$ the tops $(\overline{E}_j, \overline{F}_j)$ in (2) can be replaced with $(\overline{E}_j[x_i/H'], \overline{F}_j[x_i/H'])$, where H' is the regular term $H[x_i/H][x_i/H][x_i/H] \cdots$, without changing the equivalence-level; hence $\text{EL}(\overline{E}_j\sigma, \overline{F}_j\sigma) = \text{EL}(\overline{E}_j[x_i/H']\sigma, \overline{F}_j[x_i/H']\sigma)$. Though H' might be an infinite regular term, its natural graph presentation is not larger than the presentation of H. Moreover, x_i does not occur in H', and thus the term $x_i\sigma$ ceases to play any role in the pairs $(\overline{E}_j[x_i/H']\sigma, \overline{F}_j[x_i/H']\sigma)$ $(j = e+2, e+3, \ldots, z)$.

By continuing this reasoning inductively ("removing" one $x_i\sigma$ in each of at most n phases), we note that the length of (n, s, g)-sequences (2) is bounded by a (maybe large) constant

2

determined by the grammar $\mathcal{G}$. By a careful analysis we then show that such a constant is, in fact, computable when a grammar is given.

Further remarks on related research. Further work is needed to fully understand the bisimulation problems on PDA and their subclasses, also regarding their computational complexity. E.g., even the case of BPA processes, generated by real-time PDA with a single control-state, is not quite clear. Here the bisimilarity problem is EXPTIME-hard [15] and in 2-EXPTIME [16] (proven explicitly in [17]); for the subclass of normed BPA the problem is polynomial [18] (see [19] for the best published upper bound). Another issue is the precise decidability border. This was also studied in [20]; allowing that ε-steps can have alternatives (though they are restricted to be stack-popping) leads to undecidability of bisimilarity. This aspect has been also refined, for branching bisimilarity [21]. For second-order PDA the undecidability is established without ε-steps [22]. We can refer to the survey papers [23, 24] for the work on higher-order PDA, and in particular mention that the decidability of equivalence of *deterministic* higher-order PDA remains open; some progress in this direction was made by Stirling in [25].

Finally we remark that recently (while this paper was under review) the author cooperated with Sylvain Schmitz on developing a concrete version of the algorithm suggested here, and its complexity analysis has revealed an Ackermannian upper bound; with the lower bound from [8] this yields the Ackermann-completeness of the studied equivalence problem [26].

Organization of the paper. After the preliminaries in Section 2 we state the main theorem in Section 3. The theorem is proven in Section 7, using the notions and results discussed in Sections 4, 5, and 6; each of these sections starts with an informal summary.

2 Basic Notions and Facts

In this section we define basic notions and observe their simple properties. Some standard definitions are restricted when we do not need full generality.

By $\mathbb{N}$ and $\mathbb{N}_+$ we denote the sets of nonnegative integers and of positive integers, respectively. By $[i, j]$, for $i, j \in \mathbb{N}$, we denote the set $\{i, i+1, \ldots, j\}$. For a set $\mathcal{A}$, by $\mathcal{A}^*$ we denote the set of finite sequences of elements of $\mathcal{A}$, which are also called *words* (over $\mathcal{A}$). By $|w|$ we denote the *length* of $w \in \mathcal{A}^*$, and by ε the *empty sequence*; hence $|\varepsilon| = 0$. We put $\mathcal{A}^+ = \mathcal{A}^* \smallsetminus \{\varepsilon\}$.

Labelled transition systems. A *labelled transition system*, an *LTS* for short, is a tuple $\mathcal{L} = (\mathcal{S}, \Sigma, (\xrightarrow{a})_{a\in\Sigma})$ where $\mathcal{S}$ is a finite or countable set of *states*, Σ is a finite or countable set of *actions* and $\xrightarrow{a} \subseteq \mathcal{S} \times \mathcal{S}$ is a set of *a-transitions* (for each $a \in \Sigma$). We say that $\mathcal{L}$ is a *deterministic LTS* if for each pair $s \in \mathcal{S}$, $a \in \Sigma$ there is at most one s' such that $s \xrightarrow{a} s'$ (which stands for $(s, s') \in \xrightarrow{a}$). By $s \xrightarrow{w} s'$, where $w = a_1 a_2 \ldots a_n \in \Sigma^*$, we denote that there is a *path* $s = s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} s_2 \cdots \xrightarrow{a_n} s_n = s'$; the length of such a path is n, which is zero for the (trivial) path $s \xrightarrow{\varepsilon} s$. If $s \xrightarrow{w} s'$, then s' is *reachable from* s. By $s \xrightarrow{w}$ we denote that w is *enabled in* s, or w is *performable from* s, i.e., $s \xrightarrow{w} s'$ for some s'. If $\mathcal{L}$ is deterministic, then the expressions $s \xrightarrow{w} s'$ and $s \xrightarrow{w}$ also denote a unique path.

Bisimilarity, eq-levels. Given $\mathcal{L} = (\mathcal{S}, \Sigma, (\xrightarrow{a})_{a\in\Sigma})$, a *set* $\mathcal{D} \subseteq \mathcal{S} \times \mathcal{S}$ *covers* $(s, t) \in \mathcal{S} \times \mathcal{S}$ if for any $s \xrightarrow{a} s'$ there is $t \xrightarrow{a} t'$ such that $(s', t') \in \mathcal{D}$, and for any $t \xrightarrow{a} t'$ there is $s \xrightarrow{a} s'$ such that $(s', t') \in \mathcal{D}$. For $\mathcal{D}, \mathcal{D}' \subseteq \mathcal{S} \times \mathcal{S}$ we say that $\mathcal{D}'$ *covers* $\mathcal{D}$ if $\mathcal{D}'$ covers each $(s, t) \in \mathcal{D}$. A

set $\mathcal{D} \subseteq \mathcal{S} \times \mathcal{S}$ is a *bisimulation* if $\mathcal{D}$ covers $\mathcal{D}$. States $s, t \in \mathcal{S}$ are *bisimilar*, written $s \sim t$, if there is a bisimulation $\mathcal{D}$ containing (s, t). A standard fact is that $\sim \subseteq \mathcal{S} \times \mathcal{S}$ is an equivalence relation, and it is the largest bisimulation, namely the union of all bisimulations.

We also put $\sim_0 = \mathcal{S} \times \mathcal{S}$, and define $\sim_{k+1} \subseteq \mathcal{S} \times \mathcal{S}$ (for $k \in \mathbb{N}$) as the set of pairs covered by $\sim_k$. It is obvious that $\sim_k$ are equivalence relations, and that $\sim_0 \supseteq \sim_1 \supseteq \sim_2 \supseteq \cdots\cdots \supseteq \sim$. For the (first limit) ordinal ω we put $s \sim_\omega t$ if $s \sim_k t$ for all $k \in \mathbb{N}$; hence $\sim_\omega = \bigcap_{k \in \mathbb{N}} \sim_k$. We will only consider *image-finite* LTSs, where the set $\{s' \mid s \xrightarrow{a} s'\}$ is finite for each pair $s \in \mathcal{S}$, $a \in \Sigma$. In this case $\bigcap_{k \in \mathbb{N}} \sim_k$ is a bisimulation (for each $(s, t) \in \bigcap_{k \in \mathbb{N}} \sim_k$ and $s \xrightarrow{a} s'$, in the finite set $\{t' \mid t \xrightarrow{a} t'\}$ there must be one t' such that $s' \sim_k t'$ for infinitely many k, which entails $(s', t') \in \bigcap_{k \in \mathbb{N}} \sim_k$), and thus $\sim = \bigcap_{k \in \mathbb{N}} \sim_k = \sim_\omega$.

To each pair of states s, t we attach their *equivalence level* (eq-level):

$$\mathrm{EL}(s, t) = \max \{k \in \mathbb{N} \cup \{\omega\} \mid s \sim_k t\}.$$

Hence $\mathrm{EL}(s, t) = 0$ iff $\{a \in \Sigma \mid s \xrightarrow{a}\} \neq \{a \in \Sigma \mid t \xrightarrow{a}\}$ (i.e., s and t enable different sets of actions). The next proposition captures a few additional simple facts; we should add that we handle ω as an infinite amount, stipulating $\omega > n$ and $\omega + n = \omega - n = \omega$ for all $n \in \mathbb{N}$.

Proposition 1.   *1. If $\mathrm{EL}(t, t') > \mathrm{EL}(s, t)$, then $\mathrm{EL}(s, t) = \mathrm{EL}(s, t')$.*

2. *If $\omega > \mathrm{EL}(s, t) > 0$, then there is either a transition $s \xrightarrow{a} s'$ such that for all transitions $t \xrightarrow{a} t'$ we have $\mathrm{EL}(s', t') \leq \mathrm{EL}(s, t) - 1$, or a transition $t \xrightarrow{a} t'$ such that for all transitions $s \xrightarrow{a} s'$ we have $\mathrm{EL}(s', t') \leq \mathrm{EL}(s, t) - 1$.*

3. *If $|w| \leq \mathrm{EL}(s, t)$ and $s \xrightarrow{w} s'$, then $t \xrightarrow{w} t'$ for t' such that $\mathrm{EL}(s', t') \geq \mathrm{EL}(s, t) - |w|$.*

Proof. 1. If $s \sim_k t$, $s \not\sim_{k+1} t$, and $t \sim_{k+1} t'$, then $s \sim_k t'$ and $s \not\sim_{k+1} t'$.

The points 2 and 3 trivially follow from the definition of $\sim_k$ (for $k \in \mathbb{N} \cup \{\omega\}$). $\qquad\square$

First-order terms, regular terms, finite graph presentations.   We will consider LTSs in which the states are first-order regular terms.

The terms are built from *variables* taken from a fixed countable set

$$\mathrm{VAR} = \{x_1, x_2, x_3, \dots\}$$

and from *function symbols*, also called *(ranked) nonterminals*, from some specified finite set $\mathcal{N}$; each $A \in \mathcal{N}$ has $arity(A) \in \mathbb{N}$. We reserve symbols A, B, C, D to range over nonterminals, and E, F, G, H, T, U, V, W to range over terms. An example of a finite term is $E_1 = A(D(x_5, C(x_2, B)), x_5, B)$, where the arities of nonterminals A, B, C, D are $3, 0, 2, 2$, respectively. Its syntactic tree is depicted on the left of Fig.1.

We identify terms with their syntactic trees. Thus a *term over* $\mathcal{N}$ is (viewed as) a rooted, ordered, finite or infinite tree where each node has a label from $\mathcal{N} \cup \mathrm{VAR}$; if the label of a node is $x \in \mathrm{VAR}$, then the node has no successors, and if the label is $A \in \mathcal{N}$, then it has m (immediate) successor-nodes where $m = arity(A)$. A subtree of a term E is also called a *subterm* of E. We make no difference between isomorphic (sub)trees, and thus a subterm can have more (maybe infinitely many) *occurrences* in E. Each *subterm-occurrence* has its (nesting) *depth in E*, which is its (naturally defined) distance from the root of E. E.g., $C(x_2, B)$ is a depth-2 subterm of E_1; x_5 is a subterm with a depth-1 and a depth-2 occurrences.
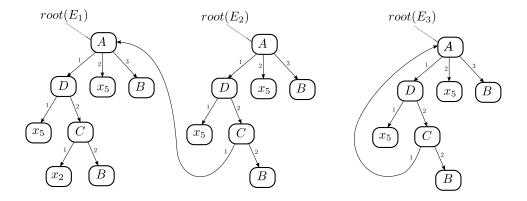
4

Figure 1: Finite terms E_1, E_2, and a graph presenting a regular infinite term E_3

We also use the standard notation for terms: we write $E = x_i$ or $E = A(G_1, \ldots, G_m)$ with the obvious meaning; in the latter case $\text{ROOT}(E) = A \in \mathcal{N}$, $m = arity(A)$, and $G_1, \ldots, G_m$ are the ordered depth-1 occurrences of subterms of E, which are also called the *root-successors* in E.

A *term* is *finite* if the respective tree is finite. A (possibly infinite) *term* is *regular* if it has only finitely many subterms (though the subterms may be infinite and may have infinitely many occurrences). We note that any regular term has at least one *graph presentation*, i.e. a finite directed graph with a designated root, where each node has a label from $\mathcal{N} \cup \text{VAR}$; if the label of a node is $x \in \text{VAR}$, then the node has no outgoing arcs, if the label is $A \in \mathcal{N}$, then it has m ordered outgoing arcs where $m = arity(A)$. We can see an example of such a graph presenting a term E_3 on the right in Fig. 1. The standard tree-unfolding of the graph is the respective term, which is infinite if there are cycles in the graph. There is a bijection between the nodes in the *least* graph presentation of E and (the roots of) the subterms of E.

Sizes, heights, and variables of terms. By $\text{TERMS}_\mathcal{N}$ we denote the set of all regular terms over $\mathcal{N}$ (and VAR); we do not consider non-regular terms. By a "term" we mean a general regular term unless the context makes clear that the term is finite.

By $\text{SIZE}(E)$ we mean the number of nodes in the least graph presentation of E. E.g., in Fig.1 $\text{SIZE}(E_1) = 6$ (E_1 has six subterms) and $\text{SIZE}(E_3) = 5$. By $\text{SIZE}(\{E_1, E_2, \ldots, E_n\})$ we mean the number of nodes in the least graph presentation in which a distinguished node R_i corresponds to the (root of the) term E_i, for each $i \in [1, n]$. (Since $E_1, E_2, \ldots, E_n$ can share some subterms, $\text{SIZE}(\{E_1, E_2, \ldots, E_n\})$ can be smaller than $\sum_{i \in [1,n]} \text{SIZE}(E_i)$.) We usually write $\text{SIZE}(E, F)$ instead of $\text{SIZE}(\{E, F\})$. E.g., $\text{SIZE}(E_1, E_2) = 9$ in Fig. 1.

For a finite term E we define $\text{HEIGHT}(E)$ as the maximal depth of a subterm; e.g., $\text{HEIGHT}(E_1) = 3$ in Fig.1.

We put $\text{VAR}(E) = \{x \in \text{VAR} \mid x \text{ occurs in } E\}$ and $\text{VAR}(E, F) = \{x \in \text{VAR} \mid x \text{ occurs in } E \text{ or } F\}$. E.g., $\text{VAR}(E_1, E_2) = \{x_2, x_5\}$ in Fig.1.

Substitutions, associative composition, iterated substitutions. A *substitution* σ is a mapping $\sigma : \text{VAR} \to \text{TERMS}_\mathcal{N}$ whose *support*

$$\text{SUPP}(\sigma) = \{x \in \text{VAR} \mid \sigma(x) \neq x\}$$

5

is *finite*; we reserve the symbol σ for substitutions. By *applying a substitution* σ to a term E we get the term $E\sigma$ that arises from E by replacing each occurrence of $x \in \mathrm{VAR}$ with $\sigma(x)$; given graph presentations, in the graph of E we just redirect each arc leading to a node labelled with x towards the root of $\sigma(x)$ (which includes the special "root-designating arc" when $E = x$). Hence $E = x$ implies $E\sigma = x\sigma = \sigma(x)$. The natural *composition of substitutions*, where $\sigma = \sigma_1\sigma_2$ is defined by $x\sigma = (x\sigma_1)\sigma_2$, can be easily verified to be associative. We thus write $E\sigma_1\sigma_2$ instead of $(E\sigma_1)\sigma_2$ or $E(\sigma_1\sigma_2)$. For $i \in \mathbb{N}$ we define σ^i inductively: σ^0 is the empty-support substitution, and $\sigma^{i+1} = \sigma\sigma^i$.

By $[x_{i_1}/H_1, x_{i_2}/H_2, \ldots, x_{i_k}/H_k]$, where $i_j \neq i_{j'}$ for $j \neq j'$, we denote the substitution σ such that $x_{i_j}\sigma = H_j$ for all $j \in [1, k]$ and $x\sigma = x$ for all $x \in \mathrm{VAR} \smallsetminus \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$. We will use $\sigma^\omega = \sigma\sigma\sigma\cdots$ just for the special case $\sigma = [x_i/H]$, where σ^ω is clearly well-defined; a graph presentation of the term $x_i\sigma^\omega$ arises from a graph presentation of H by redirecting each arc leading to x_i (if any exists) towards the root; we have $x_i\sigma^\omega = H$ if $x_i \notin \mathrm{VAR}(H)$, or if $H = x_i$. In Fig.1, for $\sigma = [x_2/E_1]$ we have $E_2 = E_1\sigma$ and $E_3 = E_1\sigma^\omega$.

By $\sigma_{[-x_i]}$ we denote the substitution arising from σ by removing x_i from its support (if it is there): hence $x_i\sigma_{[-x_i]} = x_i$ and $x\sigma_{[-x_i]} = x\sigma$ for all $x \in \mathrm{VAR} \smallsetminus \{x_i\}$.

We note a trivial fact (for later use):

Proposition 2. *If $H \neq x_i$, then for the term $H' = H[x_i/H][x_i/H][x_i/H]\cdots$ we have $x_i \notin \mathrm{VAR}(H')$, and thus $H'\sigma = H'\sigma_{[-x_i]}$ for any σ. We also have $\mathrm{SIZE}(H') \leq \mathrm{SIZE}(H)$.*

First-order grammars. A *first-order grammar*, or just a *grammar* for short, is a tuple $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$ where $\mathcal{N}$ is a finite nonempty set of ranked *nonterminals*, viewed as function symbols with arities, Σ is a finite nonempty set of *actions* (or "letters"), and $\mathcal{R}$ is a finite nonempty set of *rules* of the form

$$A(x_1, x_2, \ldots, x_m) \xrightarrow{a} E \tag{3}$$

where $A \in \mathcal{N}$, $arity(A) = m$, $a \in \Sigma$, and E is a finite term over $\mathcal{N}$ in which each occurring variable is from the set $\{x_1, x_2, \ldots, x_m\}$; we can have $E = x_i$ for some $i \in [1, m]$.

LTSs generated by rules, and by actions, of grammars. Given $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, by $\mathcal{L}_\mathcal{G}^\mathrm{R}$ we denote the (*rule-based*) LTS $\mathcal{L}_\mathcal{G}^\mathrm{R} = (\mathrm{TERMS}_\mathcal{N}, \mathcal{R}, (\xrightarrow{r})_{r \in \mathcal{R}})$ where each rule r of the form $A(x_1, x_2, \ldots, x_m) \xrightarrow{a} E$ induces transitions $A(x_1, \ldots, x_m)\sigma \xrightarrow{r} E\sigma$ for all substitutions σ. The transition induced by σ with $\mathrm{SUPP}(\sigma) = \emptyset$ is $A(x_1, \ldots, x_m) \xrightarrow{r} E$.

Using terms from Fig.1 as examples, if a rule r_1 is $A(x_1, x_2, x_3) \xrightarrow{b} x_2$, then we have $E_3 \xrightarrow{r_1} x_5$ (since E_3 can be written as $A(x_1, x_2, x_3)\sigma$ where $x_2\sigma = x_5$); the action b only plays a role in the LTS $\mathcal{L}_\mathcal{G}^\mathrm{A}$ defined below (where we have $E_3 \xrightarrow{b} x_5$). For a rule $r_2 : A(x_1, x_2, x_3) \xrightarrow{a} C(x_2, D(x_2, x_1))$ we deduce $E_1 \xrightarrow{r_2} C(x_5, D(x_5, D(x_5, C(x_2, B))))$; we note that the third root-successor in E_1 thus "disappears" since $x_3 \notin \mathrm{VAR}(C(x_2, D(x_2, x_1)))$.

By definition, the LTS $\mathcal{L}_\mathcal{G}^\mathrm{R}$ is deterministic (for each F and r there is at most one H such that $F \xrightarrow{r} H$). We note that *variables* are *dead* (have no outgoing transitions). We also note that $F \xrightarrow{w} H$ implies $\mathrm{VAR}(H) \subseteq \mathrm{VAR}(F)$ (each variable occurring in H also occurs in F) but not $\mathrm{VAR}(F) \subseteq \mathrm{VAR}(H)$ in general.

Remark. Since the rhs (right-hand sides) E in the rules (3) are finite, all terms reachable from a finite term are finite. The "finite-rhs version" with general regular terms in LTSs has
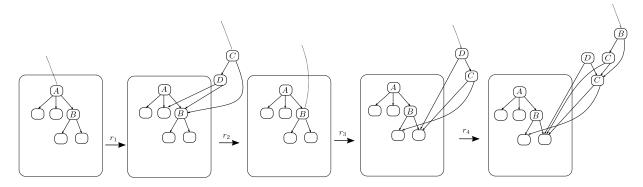
Figure 2: Path $A(T_1, T_2, B(T_3, T_4)) \xrightarrow{r_1} C(D(T_2, B(T_3, T_4)), B(T_3, T_4)) \xrightarrow{r_2} B(T_3, T_4) \xrightarrow{r_3 r_4}$ in $\mathcal{L}_{\mathcal{G}}^{\mathrm{R}}$

been chosen for technical convenience. This is not crucial, since the equivalence problem for the "regular-rhs version" can be easily reduced to the problem for our finite-rhs version.

The deterministic rule-based LTS $\mathcal{L}_{\mathcal{G}}^{\mathrm{R}}$ is helpful technically, but we are primarily interested in the (image-finite nondeterministic) *action-based* LTS $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}} = (\mathrm{TERMS}_{\mathcal{N}}, \Sigma, (\xrightarrow{a})_{a \in \Sigma})$ where each rule $A(x_1, \ldots, x_m) \xrightarrow{a} E$ induces the transitions $A(x_1, \ldots, x_m)\sigma \xrightarrow{a} E\sigma$ for all substitutions σ. (Hence the rules r_1 and r_2 in the above examples induce $E_3 \xrightarrow{b} x_5$ and $E_1 \xrightarrow{a} C(x_5, D(x_5, D(x_5, C(x_2, B))))$.)

Fig.2 sketches a path in some LTS $\mathcal{L}_{\mathcal{G}}^{\mathrm{R}}$ where we have, e.g., $r_1 : A(x_1, x_2, x_3) \xrightarrow{a_1} C(D(x_2, x_3), x_3)$ and $r_2 : C(x_1, x_2) \xrightarrow{a_2} x_2$ for some actions a_1, a_2 (which would replace r_1, r_2 in the LTS $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}}$). In the rectangle just a part of a regular-term presentation is sketched. Hence the initial root-node A might be accessible from later roots due to its possible undepicted ingoing arcs. On the other hand, the root-node D after the steps $r_1 r_2 r_3$ is not accessible (and can be omitted) in the presentation of the final term.

Eq-levels of pairs of terms. Given a grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, by $\mathrm{EL}(E, F)$ we refer to the equivalence level of (regular) terms E, F in $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}}$, with the following adjustment: though variables x_i are handled as dead also in $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}}$, we stipulate $\mathrm{EL}(x_i, H) = 0$ if $H \neq x_i$ (while $\mathrm{EL}(x_i, x_i) = \omega$); this would be achieved automatically if we enriched $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}}$ with transitions $x \xrightarrow{a_x} x$ where a_x is a special action added to each variable $x \in \mathrm{VAR}$. This adjustment gives us the point 1 in the next proposition on compositionality.

We put $\sigma \sim_k \sigma'$ if $x\sigma \sim_k x\sigma'$ for all $x \in \mathrm{VAR}$, and define

$$\mathrm{EL}(\sigma, \sigma') = \max\left\{ k \in \mathbb{N} \cup \{\omega\} \mid \sigma \sim_k \sigma' \right\}.$$

Proposition 3. *For all $\sigma, \sigma', \sigma'', E, F$, and $k \in \mathbb{N} \cup \{\omega\}$ the following conditions hold:*

1. *If $\sigma' \sim_k \sigma''$, then $\sigma'\sigma \sim_k \sigma''\sigma$. Hence $\mathrm{EL}(\sigma', \sigma'') \leq \mathrm{EL}(\sigma'\sigma, \sigma''\sigma)$.*

   In particular, $\mathrm{EL}(E, F) \leq \mathrm{EL}(E\sigma, F\sigma)$.

2. *If $\sigma' \sim_k \sigma''$, then $\sigma\sigma' \sim_k \sigma\sigma''$. Hence $\mathrm{EL}(\sigma', \sigma'') \leq \mathrm{EL}(\sigma\sigma', \sigma\sigma'')$.*

   In particular, $\mathrm{EL}(\sigma', \sigma'') \leq \mathrm{EL}(E\sigma', E\sigma'')$.

Proof. It suffices to prove the claims for $k \in \mathbb{N}$, since $\sim_\omega = \bigcap_{k \in \mathbb{N}} \sim_k$. We use an induction on k, noting that for $k = 0$ the claims are trivial.

Assuming $k > 0$ and $E \sim_k F$, we show that $E\sigma \sim_k F\sigma$: We cannot have $\{E, F\} = \{x_i, H\}$ for some $H \neq x_i$ (since then $\mathrm{EL}(E, F) = 0$ by our definition). Hence either $E = F = x$ for some $x \in \mathrm{VAR}$, in which case $E\sigma = F\sigma$, or $E \notin \mathrm{VAR}$ and $F \notin \mathrm{VAR}$. In the latter case every transition $E\sigma \xrightarrow{a} G$ ($F\sigma \xrightarrow{a} G$) is, in fact, $E\sigma \xrightarrow{a} E'\sigma$ ($F\sigma \xrightarrow{a} F'\sigma$) where $E \xrightarrow{a} E'$ ($F \xrightarrow{a} F'$), and there must be a corresponding transition $F \xrightarrow{a} F'$ ($E \xrightarrow{a} E'$) such that $E' \sim_{k-1} F'$ (by Proposition 1(3)); by the induction hypothesis $E'\sigma \sim_{k-1} F'\sigma$, which shows that $E\sigma \sim_k F\sigma$ (since $(E\sigma, F\sigma)$ is covered by $\sim_{k-1}$).

This gives us the point 1. For the point 2 we note that $\sigma' \sim_k \sigma''$ implies $E\sigma' \sim_k E\sigma''$, which is even more straightforward to verify. $\qquad\square$

The next lemma shows a simple but important fact (whose analogues in different frameworks could be traced back to [1, 5] and other related works). Its claim is sketched in a part of Figure 3. (We recall that E, F denote general regular terms when we do not say that they are finite.)

Lemma 4. *If* $\mathrm{EL}(E, F) = k < \ell = \mathrm{EL}(E\sigma, F\sigma)$, *then there are* $x_i \in \mathrm{SUPP}(\sigma)$, $H \neq x_i$, *and* $w \in \Sigma^*$, $|w| \leq k$, *such that* $E \xrightarrow{w} x_i$, $F \xrightarrow{w} H$ *or* $E \xrightarrow{w} H$, $F \xrightarrow{w} x_i$, *and* $x_i\sigma \sim_{\ell-k} H\sigma$.

Proof. We assume $\mathrm{EL}(E, F) = k < \ell = \mathrm{EL}(E\sigma, F\sigma)$ and use an induction on k. If $k = 0$, then necessarily $\{E, F\} = \{x_i, H\}$ for some $x_i \neq H$ (since $E \notin \mathrm{VAR}$, $F \notin \mathrm{VAR}$ would imply $\mathrm{EL}(E\sigma, F\sigma) = 0$ as well); the claim is thus trivial (if $x_i \notin \mathrm{SUPP}(\sigma)$, i.e. $x_i\sigma = x_i$, then $H = x_j$ and $x_j\sigma = x_i$, which entails that $x_j \in \mathrm{SUPP}(\sigma)$).

For $k > 0$ we must have $E \notin \mathrm{VAR}$, $F \notin \mathrm{VAR}$. There must be a transition $E \xrightarrow{a} E'$ (or $F \xrightarrow{a} F'$) such that for all $F \xrightarrow{a} F'$ (for all $E \xrightarrow{a} E'$) we have $\mathrm{EL}(E', F') \leq k-1$ (by Proposition 1(2)). On the other hand, for each $E\sigma \xrightarrow{a} G_1$ (and each $F\sigma \xrightarrow{a} G_2$) there is $F\sigma \xrightarrow{a} G_2$ ($E\sigma \xrightarrow{a} G_1$) such that $\mathrm{EL}(G_1, G_2) \geq \ell-1$ (by Proposition 1(3)); since $E \notin \mathrm{VAR}$ and $F \notin \mathrm{VAR}$, the transitions $E\sigma \xrightarrow{a} G_1$, $F\sigma \xrightarrow{a} G_2$ can be written $E\sigma \xrightarrow{a} E'\sigma$, $F\sigma \xrightarrow{a} F'\sigma$, respectively, where $E \xrightarrow{a} E'$, $F \xrightarrow{a} F'$. Hence there is a pair of transitions $E \xrightarrow{a} E'$, $F \xrightarrow{a} F'$ such that $\mathrm{EL}(E', F') = k' \leq k-1$ and $\mathrm{EL}(E'\sigma, F'\sigma) = \ell' \geq \ell-1$. We apply the induction hypothesis and deduce that there are $x_i \in \mathrm{SUPP}(\sigma)$, $H \neq x_i$, and $w \in \Sigma^*$, $|w| \leq k'$, such that $E' \xrightarrow{w} x_i$, $F' \xrightarrow{w} H$ or $E' \xrightarrow{w} H$, $F' \xrightarrow{w} x_i$, and $x_i\sigma \sim_{\ell'-k'} H\sigma$, which entails $x_i\sigma \sim_{\ell-k} H\sigma$ (since $\ell - k = (\ell - 1) - (k - 1) \leq \ell' - k'$). Since $E \xrightarrow{aw} x_i$, $F \xrightarrow{aw} H$ or $E \xrightarrow{aw} H$, $F \xrightarrow{aw} x_i$, we are done. $\qquad\square$

Bounded growth of sizes and heights. We fix a grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, and note a few simple facts to aid later analysis; we also introduce the constants SINC (size increase), HINC (height increase) related to $\mathcal{G}$. We recall that the rhs-terms E in the rules (3) are finite, and we put

$$\mathrm{HINC} = \max \{\mathrm{HEIGHT}(E) - 1 \mid E \text{ is the rhs of a rule in } \mathcal{R}\}. \tag{4}$$

We add that in this paper we stipulate $\max \emptyset = 0$.

By $\mathrm{NTSIZE}(E)$ we mean the number of nonterminal nodes in the least graph presentation of E (hence the number of non-variable subterms of E). We put

$$\mathrm{SINC} = \max \{\mathrm{NTSIZE}(E) \mid E \text{ is the rhs of a rule in } \mathcal{R}\}. \tag{5}$$

The next proposition shows (generous) upper bounds on the size and height increase caused by (sets of) transition sequences. (It is helpful to recall Fig. 2, assuming that the rectangle contains a presentation of G.)

Proposition 5.

1. *If* $G \xrightarrow{w} F$, *then* $\mathrm{SIZE}(F) \leq \mathrm{SIZE}(G) + |w| \cdot \mathrm{SINC}$.

2. *If* $G \xrightarrow{w} F$ *where* G *is a finite term, then* $\mathrm{HEIGHT}(F) \leq \mathrm{HEIGHT}(G) + |w| \cdot \mathrm{HINC}$.

3. *If* $G \xrightarrow{v_1} F_1$, $G \xrightarrow{v_2} F_2$, $\cdots$, $G \xrightarrow{v_p} F_p$, *where* $|v_i| \leq d$ *for all* $i \in [1, p]$, *then* $\mathrm{SIZE}(\{F_1, F_2, \ldots, F_p\}) \leq \mathrm{SIZE}(G) + p \cdot d \cdot \mathrm{SINC}$.

Proof. The points 1 and 2 are immediate. A "blind" use of 1 in the point 3 would yield $\mathrm{SIZE}(\{F_1, F_2, \ldots, F_p\}) \leq p \cdot \big(\mathrm{SIZE}(G) + d \cdot \mathrm{SINC}\big)$. But since the terms F_i can share subterms of G, we get the stronger bound $\mathrm{SIZE}(G) + p \cdot d \cdot \mathrm{SINC}$. $\square$

Shortest sink words. If $A(x_1, \ldots, x_{arity(A)}) \xrightarrow{w} x_i$ in $\mathcal{L}_{\mathcal{G}}^{\mathrm{R}}$ (hence $w \in \mathcal{R}^+$), then we call w an (A, i)-*sink word*. We note that such w can be written rw' where $A(x_1, \ldots, x_{arity(A)}) \xrightarrow{r} E \xrightarrow{w'} x_i$; hence w' "sinks" along a branch of E to x_i, or $w' = \varepsilon$ when $E = x_i$. This suggests a standard dynamic programming approach to find and fix some shortest (A, i)-sink words $w_{[A,i]}$ for all elements (A, i) of the set $\mathrm{NA} = \{(B, j) \mid B \in \mathcal{N}, j \in [1, arity(B)]\}$ for which such words exist. We can clearly (generously) bound the lengths of $w_{[A,i]}$ by $h^{|\mathrm{NA}|}$ where $h = 2 + \mathrm{HINC}$ (i.e., $h = 1 + \max\big\{\mathrm{HEIGHT}(E) \mid E \text{ is the rhs of a rule in } \mathcal{R}\big\}$). We put

$$d_0 = 1 + \max\{\, |w_{[A,i]}|; A \in \mathcal{N}, i \in [1, arity(A)] \,\}. \tag{6}$$

The above discussion entails that d_0 is a (quickly) computable number, whose value is at most exponential in the size of the given grammar $\mathcal{G}$.

Remark. For any grammar $\mathcal{G}$ we can construct a "normalized" grammar $\mathcal{G}'$ in which $w_{[A,i]}$ exists for each $(A, i) \in \mathrm{NA}$, while the LTSs $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}}$ and $\mathcal{L}_{\mathcal{G}'}^{\mathrm{A}}$ are isomorphic. (We can refer to [27] for more details.) We do not need such normalization in this paper.

Convention. When having a fixed grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, we also put

$$m = \max\big\{arity(A) \mid A \in \mathcal{N}\big\} \tag{7}$$

but we will often write $A(x_1, \ldots, x_m)$ even if $arity(A)$ might not be maximal. This is harmless since such m could be always replaced with $arity(A)$ if we wanted to be pedantic. (In fact, the grammar could be also normalized so that the arities of nonterminals are the same [27] but this is a superfluous technical issue here.)

3   Main Result (Computability of Equivalence Levels)

Small numbers. We use the notion of "small" numbers determined by a grammar $\mathcal{G}$; by saying that a *number* $d \in \mathbb{N}$ is *small* we mean that it is a computable number (for a given grammar $\mathcal{G}$) that is elementary in the size of $\mathcal{G}$.

E.g., the numbers m, HINC, SINC (defined by (7), (4), (5)) are trivially small, and we have also shown that d_0 (defined by (6)) is small. In what follows we will also introduce further specific small numbers, summarized in Table 1 at the end of the paper.
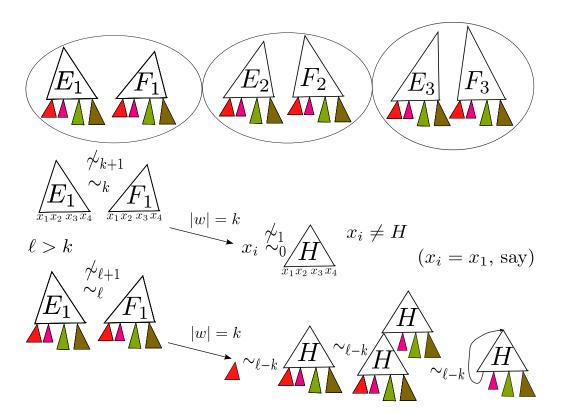
Figure 3: In an (n, s, g)-sequence, (E_1, F_1) helps to get rid of one term in the tail-substitution σ

Main theorem. We first note a fact that is obvious (by induction on k):

Proposition 6. *There is an algorithm that, given a grammar $\mathcal{G}$, terms T, U, and $k \in \mathbb{N}$, decides if $T \sim_k U$ in the LTS $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}}$.*

Hence the next theorem adds the decidability of $\sim$ (i.e., of $\sim_k$ for $k = \omega$).

Theorem 7. *For any grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$ there is a small number c and a computable (not necessarily small) number $\mathcal{E}$ such that for all $T, U \in \mathrm{TERMS}_{\mathcal{N}}$ we have:*

$$\text{if } T \not\sim U \text{ then } \mathrm{EL}(T, U) \leq c \cdot \big(\mathcal{E} \cdot \mathrm{SIZE}(T, U) + (\mathrm{SIZE}(T, U))^2\big). \tag{8}$$

Corollary 8. *It is decidable, given $\mathcal{G}$, T, U, if $T \sim U$ in $\mathcal{L}_{\mathcal{G}}^{\mathrm{A}}$.*

Theorem 7 is proven in Section 7; the proof uses the notions and results from Sections 4, 5, and 6. Each section starts with an informal summary, and the collection of these summaries yields a more detailed informal overview of the proof than that given in the introduction.

4 Bounding the Lengths of "(n, s, g)-sequences"

The top of Figure 3 depicts (a prefix of) a sequence of the form

$$(E_1\sigma, F_1\sigma), (E_2\sigma, F_2\sigma), \ldots, (E_z\sigma, F_z\sigma)$$
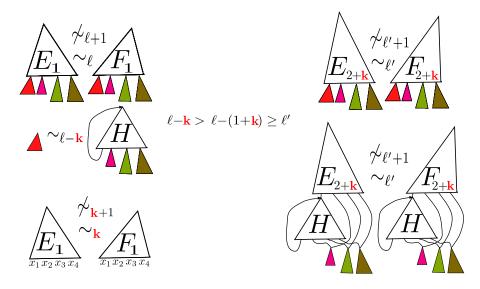
Figure 4: Support of σ can be safely decreased after the eq-level drops sufficiently

(E_i, F_i being regular terms) where we assume that the eq-levels are finite and decreasing:

$$\omega > \text{EL}(E_1\sigma, F_1\sigma) > \text{EL}(E_2\sigma, F_2\sigma) > \cdots > \text{EL}(E_z\sigma, F_z\sigma).$$

We then have $\text{EL}(E_1, F_1) = k \leq \ell = \text{EL}(E_1\sigma, F_1\sigma)$ (for some $k, \ell \in \mathbb{N}$), by Proposition 3. If σ is the empty-support substitution, then $k = \ell$ and the sequence length z is bounded by $1 + k$. If $k < \ell$, then Lemma 4 yields some x_i and $H \neq x_i$ ($x_i = x_1$ in Figure 3) where $x_i\sigma \sim_{\ell-k} H\sigma$; hence in each pair $(E_j\sigma, F_j\sigma)$ where $j \in [2 + k, z]$ we can (repeatedly) replace $x_i\sigma$ with $H\sigma$ without changing the eq-level of the pair. This is depicted in Figure 4; since $\text{EL}(E_{2+k}\sigma, F_{2+k}\sigma) = \ell' < \ell - k$, the respective eq-levels do not change due to Propositions 3 and 1.

If, moreover, we are guaranteed that the size growth of (E_j, F_j) is controlled, i.e.,

$$\text{SIZE}(E_j, F_j) \leq s + g \cdot (j-1)$$

for some fixed constants s and g (and $j \in [1, z]$), and $\text{VAR}(E_j, F_j) \subseteq \{x_1, \ldots, x_n\}$ for some fixed n (which bounds the support of σ), then a bound on the lengths z of such (n, s, g)-sequences is determined by the respective grammar $\mathcal{G}$ (independently of the sizes of terms $x_i\sigma$). This is straightforward, as we now show.

Given n, s, g, the number of respective pairs (E_1, F_1) is bounded, and there is thus $e \in \mathbb{N}$ that is the largest $\text{EL}(E_1, F_1)$ for such pairs (we recall that $\omega > \text{EL}(E_1\sigma, F_1\sigma) \geq \text{EL}(E_1, F_1)$); at this moment we do not claim that e is computable. For each (n, s, g)-sequence $(E_1\sigma, F_1\sigma)$, $(E_2\sigma, F_2\sigma)$, $\ldots$, $(E_z\sigma, F_z\sigma)$ where $z > 1+e$ we have $(\text{EL}(E_1, F_1) < \text{EL}(E_1\sigma, F_1\sigma)$ and) either $E_1 \xrightarrow{w} x_i$ and $F_1 \xrightarrow{w} H$, or $E_1 \xrightarrow{w} H$ and $F_1 \xrightarrow{w} x_i$, $|w| \leq e$, for the respective x_i, H discussed above and illustrated in Figures 3 and 4; hence $\text{SIZE}(H) \leq \text{SIZE}(E_1, F_1) + e \cdot \text{SINC}$ (by Proposition 5(1)). This entails that replacing $x_i\sigma$ with $H'\sigma_{[-x_i]}$ where $H' = H[x_i/H][x_i/H][x_i/H]\cdots$ (recall Proposition 2) in the pairs $(E_j\sigma, F_j\sigma)$ for $j = 1+e+1, 1+e+2, \ldots, z$ gives us an $(n-1, s', g)$-sequence of length $z-(1+e)$, where $s' = s + g \cdot (1+e) + s + e \cdot \text{SINC}$ (which bounds the size of terms E_{2+e}, F_{2+e} extended by a shared subterm H'). To be precise, for the terms $E'_j = E_j[x_i/H']$ and $F'_j = F_j[x_i/H']$ we only have $\text{VAR}(E'_j, F'_j) \subseteq \{x_1, \ldots, x_n\} \smallsetminus \{x_i\}$,

and x_n can occur in them (when $x_i \neq x_n$). In this case we just replace x_n with x_i in all E'_j, F'_j ($j = 1+e+1, 1+e+2, \ldots, z$) and use the tail-substitution σ' that arises from σ by putting $x_i\sigma' = x_n\sigma$ and $x_n\sigma' = x_i$. An inductive argument thus establishes that there is indeed a claimed bound (on the lengths of (n, s, g)-sequences) determined by the grammar.

We will later show that such a bound is even computable when $\mathcal{G}, n, s, g$ are given. Moreover, we will also show how to compute small n, s, g to a given $\mathcal{G}$ so that the computable bound on the length of (n, s, g)-sequences gives us the number $\mathcal{E}$ in Theorem 7.

In the rest of this section we formalize the above ideas showing that (n, s, g)-sequences are bounded. In this formalization we also define the notion of (n, s, g)-*candidates*, candidates for "non-equivalence bases"; intuitively, the *base* $\mathcal{B}_{n,s,g}$ is intended to collect all possible "tops" (E_j, F_j), $(E_j[x_i/H'], F_j[x_i/H'])$, ... from all (eqlevel-decreasing) (n, s, g)-sequences that undergo the above described inductive transformation.

Eqlevel-decreasing (n, s, g)-sequences. We fix a grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$. By an *eqlevel-decreasing sequence* we mean a sequence $(T_1, U_1), (T_2, U_2), \ldots, (T_z, U_z)$ of pairs of terms (where $z \in \mathbb{N}_+$) such that $\omega > \mathrm{EL}(T_1, U_1) > \mathrm{EL}(T_2, U_2) > \cdots > \mathrm{EL}(T_z, U_z)$. The length z of such a sequence is obviously at most $1 + \mathrm{EL}(T_1, U_1)$.

For $n, s, g \in \mathbb{N}$ we say that an eqlevel-decreasing sequence in the form

$$(E_1\sigma, F_1\sigma), (E_2\sigma, F_2\sigma), \ldots, (E_z\sigma, F_z\sigma) \tag{9}$$

is an (n, s, g)-*sequence* if $\mathrm{VAR}(E_j, F_j) \subseteq \{x_1, \ldots, x_n\}$ and $\mathrm{SIZE}(E_j, F_j) \leq s + g \cdot (j-1)$ for all $j \in [1, z]$. (The size of "tops" (E_j, F_j) is at most s at the start, and g bounds the "growth-rate" of tops; the terms $x_i\sigma$, $i \in [1, n]$, might be large but the "tail substitution" σ is the same in all elements of the sequence.)

Candidates for (non-equivalence) bases. To show a bound on the lengths of (n, s, g)-sequences in a convenient form (in Lemma 10), we introduce further notions; we start with a piece of notation. For any $n, s \in \mathbb{N}$ we put

- $\mathrm{PAIRS}_{\mathrm{VAR}:n} = \{(E, F) \in \mathrm{TERMS}_{\mathcal{N}} \times \mathrm{TERMS}_{\mathcal{N}} \mid \mathrm{VAR}(E, F) = \{x_1, \ldots, x_n\}\}$,

- $\mathrm{PAIRS}_{\mathrm{SIZE} \leq s} = \{(E, F) \in \mathrm{TERMS}_{\mathcal{N}} \times \mathrm{TERMS}_{\mathcal{N}} \mid \mathrm{SIZE}(E, F) \leq s\}$,

- $\mathrm{PAIRS}_{n,s} = \mathrm{PAIRS}_{\mathrm{VAR}:n} \cap \mathrm{PAIRS}_{\mathrm{SIZE} \leq s}$.

Given $n, s, g \in \mathbb{N}$, we say that $\mathcal{B} \subseteq \mathrm{TERMS}_{\mathcal{N}} \times \mathrm{TERMS}_{\mathcal{N}}$ is an (n, s, g)-*candidate* (intended to collect the tops of $(n, s, g,)$-sequences that undergo the above described inductive transformation) if the following conditions 1–3 hold (in which an implicit induction on n is used):

1. $\mathcal{B} \subseteq \left(\mathrm{PAIRS}_{\mathrm{VAR}:0} \cup \mathrm{PAIRS}_{\mathrm{VAR}:1} \cup \cdots \cup \mathrm{PAIRS}_{\mathrm{VAR}:n} \right) \cap \not\sim$.

2. $(\mathcal{B} \cap \mathrm{PAIRS}_{\mathrm{VAR}:n}) \subseteq \mathrm{PAIRS}_{\mathrm{SIZE} \leq s}$.

3. If $n > 0$, then the set $\mathcal{B}' = \mathcal{B} \smallsetminus \mathrm{PAIRS}_{\mathrm{VAR}:n}$ is an $(n-1, s', g)$-candidate where

$$s' = 2s + g \cdot (1+e) + e \cdot \mathrm{SINC} \text{ for } e = \max\{\mathrm{EL}(E, F) \mid (E, F) \in \mathcal{B} \cap \mathrm{PAIRS}_{\mathrm{SIZE} \leq s}\}. \tag{10}$$

Every (n, s, g)-candidate $\mathcal{B}$ yields a *bound* $\mathcal{E}_{\mathcal{B}}^{n,s,g} \in \mathbb{N}_+$, denoted just $\mathcal{E}_{\mathcal{B}}$ when n, s, g are clear from the context; in the above notation (around (10)) we define $\mathcal{E}_{\mathcal{B}}^{n,s,g}$ as follows:

$$\text{if } n = 0, \text{ then } \mathcal{E}_{\mathcal{B}}^{n,s,g} = 1 + e; \text{ if } n > 0, \text{ then } \mathcal{E}_{\mathcal{B}}^{n,s,g} = 1 + e + \mathcal{E}_{\mathcal{B}'}^{n-1,s',g}.$$

An (n, s, g)-*candidate* $\mathcal{B}$ is *full below* an *eq-level* $\overline{e} \in \mathbb{N} \cup \{\omega\}$ if each pair $(E, F) \in \big(\text{PAIRS}_{\text{VAR}:0} \cup \text{PAIRS}_{\text{VAR}:1} \cup \cdots \cup \text{PAIRS}_{\text{VAR}:n}\big) \cap \text{PAIRS}_{\text{SIZE} \leq s}$ such that $\text{EL}(E, F) < \overline{e}$ belongs to $\mathcal{B}$, and, moreover, in the case $n > 0$ the $(n-1, s', g)$-candidate $\mathcal{B}'$ is full below $\overline{e}$. We say that $\mathcal{B}$ is *full* if it is full below ω (in which case $\mathcal{B}$ contains *all* relevant non-equivalent pairs).

Proposition 9. *For any n, s, g there is the unique full (n, s, g)-candidate, denoted $\mathcal{B}_{n,s,g}$.*

Proof. Given n, s, g, the full (n, s, g)-candidate $\mathcal{B} = \mathcal{B}_{n,s,g}$ is defined as follows: $\mathcal{B} \cap \text{PAIRS}_{\text{SIZE} \leq s} = \big(\text{PAIRS}_{\text{VAR}:0} \cup \text{PAIRS}_{\text{VAR}:1} \cup \cdots \cup \text{PAIRS}_{\text{VAR}:n}\big) \cap \text{PAIRS}_{\text{SIZE} \leq s} \cap \not\sim$ and, moreover, in the case $n > 0$ the set $\mathcal{B}' = \mathcal{B} \smallsetminus \text{PAIRS}_{n,s}$ is the full $(n-1, s', g)$-candidate (where s' is defined as in (10)). $\square$

The unique full (n, s, g)-candidate $\mathcal{B}_{n,s,g}$ will be also called the (n, s, g)-*base*.

The (n, s, g)-sequences have bounded lengths. We show the announced bound.

Lemma 10. *If $(E_1\sigma, F_1\sigma), (E_2\sigma, F_2\sigma), \ldots, (E_z\sigma, F_z\sigma)$ is an (n, s, g)-sequence and $\mathcal{B}$ is an (n, s, g)-candidate that is full below $1 + \text{EL}(E_1\sigma, F_1\sigma)$, then $z \leq \mathcal{E}_{\mathcal{B}}$; in particular, $z \leq \mathcal{E}_{\mathcal{B}_{n,s,g}}$.*

Proof. We consider an (n, s, g)-sequence $(E_1\sigma, F_1\sigma), (E_2\sigma, F_2\sigma), \ldots, (E_z\sigma, F_z\sigma)$ as in (9), and an (n, s, g)-candidate $\mathcal{B}$ that is full below $1 + \text{EL}(E_1\sigma, F_1\sigma)$. Since $\omega > \text{EL}(E_1\sigma, F_1\sigma) \geq \text{EL}(E_1, F_1)$ (by Proposition 3(1)), we have $(E_1, F_1) \in \mathcal{B} \cap \text{PAIRS}_{\text{SIZE} \leq s}$. This entails

$$\text{EL}(E_1, F_1) = k \leq e = \max \big\{ \text{EL}(E, F) \mid (E, F) \in \mathcal{B} \cap \text{PAIRS}_{\text{SIZE} \leq s} \big\}.$$

If $\text{EL}(E_1\sigma, F_1\sigma) = \text{EL}(E_1, F_1) = k$, which is surely the case when $n = 0$ (in this case $(E_1\sigma, F_1\sigma) = (E_1, F_1)$), then $z \leq 1 + k$, due to the required eqlevel-decreasing property of (n, s, g)-sequences; in this case $z \leq 1 + e \leq \mathcal{E}_{\mathcal{B}}$.

We proceed inductively (on n), assuming $n > 0$ and $\text{EL}(E_1\sigma, F_1\sigma) = \ell > k = \text{EL}(E_1, F_1)$. By Lemma 4 there is $x_i \in \text{SUPP}(\sigma)$, $i \in [1, n]$, and $H \neq x_i$ such that $E_1 \xrightarrow{w} x_i$ and $F_1 \xrightarrow{w} H$, or $E_1 \xrightarrow{w} H$ and $F_1 \xrightarrow{w} x_i$, for some $w \in \Sigma^*$ with $|w| \leq k$, where $x_i\sigma \sim_{\ell-k} H\sigma$. Hence $\sigma \sim_{\ell-k} [x_i/H]\sigma$, which entails that $\sigma \sim_{\ell-k} [x_i/H]^j\sigma$ for all $j \in \mathbb{N}$ (by applying Proposition 3(2) repeatedly). We can also easily check that $[x_i/H]^{\ell-k}\sigma \sim_{\ell-k} [x_i/H]^\omega\sigma$ (by induction on $\ell - k$), hence

$$x_i\sigma \sim_{\ell-k} H'\sigma_{[-x_i]} \text{ where } H' = H[x_i/H][x_i/H][x_i/H] \cdots$$

(We also recall Proposition 2.) We note that

$$\text{SIZE}(H') \leq \text{SIZE}(H) \leq \max\{\text{SIZE}(E_1), \text{SIZE}(F_1)\} + k \cdot \text{SINC} \leq s + e \cdot \text{SINC}$$

(by using Proposition 5(1)). For each $j \in [k+2, z]$ we now put

$$(E_j', F_j') = (E_j[x_i/H'], F_j[x_i/H']), \text{ hence } (E_j'\sigma, F_j'\sigma) = (E_j'\sigma_{[-x_i]}, F_j'\sigma_{[-x_i]}),$$

and note that $\text{EL}(E_j\sigma, F_j\sigma) = \text{EL}(E_j'\sigma_{[-x_i]}, F_j'\sigma_{[-x_i]})$, since $\text{EL}(E_j\sigma, F_j\sigma) < \ell-k$ (for each $j \geq k+2$); here we use that $\text{EL}(E_j\sigma, E_j[x_i/H']\sigma) \geq \ell-k$ and $\text{EL}(F_j\sigma, F_j[x_i/H']\sigma) \geq \ell-k$, and we recall Proposition 1(1). We also note that for each $j \in [k+2, z]$ we have

$\text{SIZE}(E'_j, F'_j) \le \text{SIZE}(E_j, F_j) + \text{SIZE}(H') \le s + g \cdot (j-1) + s + e \cdot \text{SINC} = 2s + g \cdot (j-1) + e \cdot \text{SINC}.$

Hence $\text{SIZE}(E'_{k+2}, F'_{k+2}) \le 2s + g \cdot (1+k) + e \cdot \text{SINC} \le 2s + g \cdot (1+e) + e \cdot \text{SINC} = s'$ (recall s' from (10)). Thus the sequence

$$(E'_{k+2}\sigma_{[-x_i]}, F'_{k+2}\sigma_{[-x_i]}), \ (E'_{k+3}\sigma_{[-x_i]}, F'_{k+3}\sigma_{[-x_i]}) \ldots, \ (E'_z\sigma_{[-x_i]}, F'_z\sigma_{[-x_i]})$$

is "almost" an $(n-1, s', g)$-sequence. The only problem is that x_n can occur in E'_j, F'_j. But we use the fact that x_i does not occur in E'_j, F'_j, and we replace x_n with x_i, while replacing $\sigma_{[-x_i]}$ with σ' where $x_n\sigma' = x_n$, $x_i\sigma' = x_n\sigma_{[-x_i]}$, and $x\sigma' = x\sigma_{[-x_i]}$ for all $x \in \text{VAR} \smallsetminus \{x_i, x_n\}$.

We note that the $(n-1, s', g)$-candidate $\mathcal{B}' = \mathcal{B} \smallsetminus \text{PAIRS}_{\text{VAR}:n}$ is full below $1 + \text{EL}(E'_{k+2}\sigma_{[-x_i]}, F'_{k+2}\sigma_{[-x_i]})$ (since $\mathcal{B}'$ is full below $1 + \text{EL}(E_1\sigma, F_1\sigma)$, and $\text{EL}(E'_{k+2}\sigma_{[-x_i]}, F'_{k+2}\sigma_{[-x_i]}) = \text{EL}(E_{k+2}\sigma, F_{k+2}\sigma) < \text{EL}(E_1\sigma, F_1\sigma))$. By the induction hypothesis $z - (k+1) \le \mathcal{E}_{\mathcal{B}'}$, and thus $z \le 1 + k + \mathcal{E}_{\mathcal{B}'} \le 1 + e + \mathcal{E}_{\mathcal{B}'} = \mathcal{E}_{\mathcal{B}}$. $\qquad\square$

In the final argument of the proof of Theorem 7 (in Section 7) we will use $\mathcal{E}_{\mathcal{B}_{n,s,g}}$ as $\mathcal{E}$ in (8), for some specific small n, s, g. Though we have defined the (n, s, g)-base $\mathcal{B}_{n,s,g}$ only semantically, it will turn out that it coincides with an effectively constructible "sound" (n, s, g)-candidate. But we first need some further technicalities to clarify the specific n, s, g (as well as c in (8)).

5 Plays (of Bisimulation Game) and their Balancing

In Section 1 we discussed the notion of optimal plays, which we make more precise now. We assume a given grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$; for $r \in \mathcal{R}$ of the form $A(x_1, \ldots, x_m) \xrightarrow{a} E$ we put $\text{LAB}(r) = a$. For technical convenience, by a *play* we only mean an optimal play from a non-equivalent pair, i.e., a sequence

$$\begin{array}{ccccccc} T_0 & \xrightarrow{r_1} & T_1 & \xrightarrow{r_2} & T_2 \ldots & \xrightarrow{r_k} & T_k \\ U_0 & r'_1 & U_1 & r'_2 & U_2 & r'_k & U_k \end{array} \tag{11}$$

where for each $i \in [1, k]$ we have $r_i, r'_i \in \mathcal{R}$, $\text{LAB}(r_i) = \text{LAB}(r'_i)$, $T_{i-1} \xrightarrow{r_i} T_i$, $U_{i-1} \xrightarrow{r'_i} U_i$ (in the LTS $\mathcal{L}^{\text{R}}_{\mathcal{G}}$); moreover, $\omega > \text{EL}(T_0, U_0)$ and $\text{EL}(T_i, U_i) = \text{EL}(T_{i-1}, U_{i-1}) - 1$ for each $i \in [1, k]$ (in the LTS $\mathcal{L}^{\text{A}}_{\mathcal{G}}$). If $\text{EL}(T_k, E_k) = 0$, then it is a *completed play*, in which case $\text{EL}(T_0, U_0) = k$. (We recall that T_0, U_0 can be regular terms of a large size.) The length of the play (11) is (defined to be) k, and another presentation of the play is $\begin{smallmatrix} T_0 & \xrightarrow{u} & T_k \\ U_0 & u' & U_k \end{smallmatrix}$, or also just $\begin{smallmatrix} T_0 & \xrightarrow{u} \\ U_0 & u' \end{smallmatrix}$, where $u = r_1 r_2 \cdots r_k$ and $u' = r'_1 r'_2 \cdots r'_k$.

Our aim is to bound the lengths of completed plays in the way stated in Theorem 7. To facilitate this task, in this section we show a particular transformation of a completed play (11) into a sequence of plays of the same overall length (i.e., the sum of lengths) that are connected by so-called *eqlevel-concatenation* $\odot$; such concatenation

$$\begin{bmatrix} T & \xrightarrow{u_1} & T' \\ U & u'_1 & U' \end{bmatrix} \odot \begin{bmatrix} T'' & \xrightarrow{u_2} & T''' \\ U'' & u'_2 & U''' \end{bmatrix}$$

is defined if (and only if) $\text{EL}(T', U') = \text{EL}(T'', U'')$, though the pairs (T', U') and (T'', U'') can differ. The overall length of this concatenation is $|u_1| + |u_2|$; if $\begin{smallmatrix} T'' & \xrightarrow{u_2} & T''' \\ U'' & u'_2 & U''' \end{smallmatrix}$ is a completed play, then this length ($|u_1| + |u_2|$) is obviously the same as the length of any completed play starting with (T, U).

In the first phase of the mentioned transformation of a completed play (11) we will replace it with the concatenation of two plays in the form

$$\begin{bmatrix} T_0 & \xrightarrow{v_0 u_1} & T_1' \\ U_0 & \xrightarrow{v_0' u_1'} & U_1' \end{bmatrix} \odot \begin{bmatrix} T_1'' & \xrightarrow{v} \\ U_1'' & \xrightarrow{v'} \end{bmatrix}$$

where $v_0 u_1$ is a certain prefix of $u = r_1 r_2 \cdots r_k$, $v_0' u_1'$ is a prefix of $u' = r_1' r_2' \cdots r_k'$ (of the same length as $v_0 u_1$), and $\begin{bmatrix} T_1'' & \xrightarrow{v} \\ U_1'' & \xrightarrow{v'} \end{bmatrix}$ is a completed play (while v (v') is generally *not* a suffix of u (u')). Further we replace $\begin{bmatrix} T_1'' & \xrightarrow{v} \\ U_1'' & \xrightarrow{v'} \end{bmatrix}$ with $\begin{bmatrix} T_1'' & \xrightarrow{v_1 u_2} & T_2' \\ U_1'' & \xrightarrow{v_1' u_2'} & U_2' \end{bmatrix} \odot \begin{bmatrix} T_2'' & \xrightarrow{\bar{v}} \\ U_2'' & \xrightarrow{\bar{v}'} \end{bmatrix}$ where $v_1 u_2$ is a certain prefix of v, $v_1' u_2'$ is a prefix of v', and $\begin{bmatrix} T_2'' & \xrightarrow{\bar{v}} \\ U_2'' & \xrightarrow{\bar{v}'} \end{bmatrix}$ is a completed play; we continue in this way, doing ℓ phases for a certain number ℓ, until finally getting

$$\begin{bmatrix} T_0 & \xrightarrow{v_0} & \bar{T}_1 & \xrightarrow{u_1} & T_1' \\ U_0 & \xrightarrow{v_0'} & \bar{U}_1 & \xrightarrow{u_1'} & U_1' \end{bmatrix} \odot \begin{bmatrix} T_1'' & \xrightarrow{v_1} & \bar{T}_2 & \xrightarrow{u_2} & T_2' \\ U_1'' & \xrightarrow{v_1'} & \bar{U}_2 & \xrightarrow{u_2'} & U_2' \end{bmatrix} \odot \begin{bmatrix} T_2'' & \xrightarrow{v_2} & \\ U_2'' & \xrightarrow{v_2'} & \end{bmatrix} \cdots \cdots \begin{bmatrix} \bar{T}_\ell & \xrightarrow{u_\ell} & T_\ell' \\ \bar{U}_\ell & \xrightarrow{u_\ell'} & U_\ell' \end{bmatrix} \odot \begin{bmatrix} T_\ell'' & \xrightarrow{v_\ell} & \bar{T}_{\ell+1} \\ U_\ell'' & \xrightarrow{v_\ell'} & \bar{U}_{\ell+1} \end{bmatrix}$$
(12)

where $\begin{bmatrix} T_\ell'' & \xrightarrow{v_\ell} & \bar{T}_{\ell+1} \\ U_\ell'' & \xrightarrow{v_\ell'} & \bar{U}_{\ell+1} \end{bmatrix}$ is completed and "non-transformable"; the overall length of (12) is thus equal to $k = \mathrm{EL}(T_0, U_0)$. In fact, we have $\ell = 0$ when already (11) is non-transformable; we thus put $(T_0'', U_0'') = (T_0, U_0)$ for convenience. (Later we repeat (12) as (17) without the bars in the notation $\bar{T}_j, \bar{U}_j$; now the bars are added to avoid the confusion with T_j, U_j in (11).)

More concretely, we will perform the transformation so that for each phase $j \in [1, \ell]$ we have $|u_j| = d_0$ (for d_0 defined by (6)), one of the terms $\bar{T}_j, \bar{U}_j$ is the *pivot* W_j, and the pair (T_j'', U_j'') is the *balancing result*, or the *bal-result* for short, *related to the pivot* W_j.

In fact, if $W_j = \bar{U}_j$, then we have $U_j'' = U_j'$ (and $T_j'' \neq T_j'$); in this case the j-th phase consists in replacing the completed play $\begin{bmatrix} T_{j-1}'' & \xrightarrow{v} \\ U_{j-1}'' & \xrightarrow{v'} \end{bmatrix}$ with the eqlevel-concatenation

$\begin{bmatrix} T_{j-1}'' & \xrightarrow{v_{j-1}} & \bar{T}_j & \xrightarrow{u_j} & T_j' \\ U_{j-1}'' & \xrightarrow{v_{j-1}'} & \bar{U}_j & \xrightarrow{u_j'} & U_j' \end{bmatrix} \odot \begin{bmatrix} T_j'' & \xrightarrow{\bar{v}} \\ U_j' & \xrightarrow{\bar{v}'} \end{bmatrix}$ (where $\begin{bmatrix} T_{j-1}'' & \xrightarrow{v_{j-1}} & \bar{T}_j & \xrightarrow{u_j} & T_j' \\ U_{j-1}'' & \xrightarrow{v_{j-1}'} & \bar{U}_j & \xrightarrow{u_j'} & U_j' \end{bmatrix}$ is a prefix of $\begin{bmatrix} T_{j-1}'' & \xrightarrow{v} \\ U_{j-1}'' & \xrightarrow{v'} \end{bmatrix}$); this is called a *left balancing step* (the *left* term in (T_j', U_j') has been replaced with T_j'' so that $\mathrm{EL}(T_j', U_j') = \mathrm{EL}(T_j'', U_j')$). Similarly, if $W_j = \bar{T}_j$, then we have $T_j'' = T_j'$, and we have performed a *right balancing step*, replacing U_j' with U_j''.

We thus have pivots $W_1, W_2, \ldots, W_\ell$, each having its related bal-result. Since the sequence

$$(T_1'', U_1''), (T_2'', U_2''), \ldots, (T_\ell'', U_\ell'')$$

of bal-results is eqlevel-decreasing, no pair can repeat in the sequence.

We will "balance" in a way that will also yield a *pivot path*

$$W_0 \xrightarrow{w_0} W_1 \xrightarrow{w_1} W_2 \xrightarrow{w_2} \cdots W_\ell \xrightarrow{w_\ell} W_{\ell+1}$$
(13)

where $w_0 \in \mathcal{R}^*$, $w_j \in \mathcal{R}^+$ for $j \in [1, \ell]$, $W_0 \in \{T_0, U_0\}$, $W_{\ell+1} \in \{\bar{T}_{\ell+1}, \bar{U}_{\ell+1}\}$, and we will guarantee the following properties:

1. There is some small n such that for each $j \in [1, \ell]$ there are small finite terms G, E, F, with $\mathrm{VAR}(E, F) \subseteq \mathrm{VAR}(G) \subseteq \{x_1, \ldots, x_n\}$, such that

15

$$W_j = G\sigma \text{ and } (T_j'', U_j'') = (E\sigma, F\sigma),$$

for a substitution σ (with $\text{SUPP}(\sigma) \subseteq \{x_1, \ldots, x_n\}$). (Hence the terms in the bal-result arise from the pivot W by replacing a small top of W by other small tops.) This is depicted in Figure 5 for some W_j and W_{j+1} (and in more detail in Figure 6).

2. Each pivot-path segment $W_j \xrightarrow{w_j} W_{j+1}$ (for $j \in [0, \ell]$) is either short (i.e., its length is small), or it has a short prefix and a short suffix while the middle part is "quickly sinking" (to a deep subterm of W_j if this part is long).

We note that we do not exclude that a pivot W occurs more than once in the pivot path ($W = W_j$ and $W = W_{j'}$ for $j \neq j'$), but the number of its occurrences must be small; this follows from the point 1 which entails that there is only a small number of possible bal-results related to one pivot, and from the fact that the bal-results cannot repeat.

Figure 5 depicts a "non-sinking segment" on the pivot path. (In such a segment, no root-successor of the starting term is exposed.) By the above point 2 it is intuitively clear that any long non-sinking segment must contain a large number of pivots, and that the possible increase of (the tops of) the pivots is controlled. Hence any long non-sinking segment of the pivot path gives rise to a long (n, s, g)-sequence, for some small n, s, g; here we use the point 1 (and recall Figure 5). This is a crucial fact for our proof of Theorem 7.

In this section, our task is to show a transformation that guarantees a suitable pivot path (13) and the above properties 1 and 2.

A concrete way how we do a *left balancing step* is captured by Figure 6. Informally speaking, if the left-hand side does not sink to a root-successor within less than d_0 moves (for d_0 defined by (6)), which is the case in Figure 6 due to $A(x_1, \ldots, x_m) \xrightarrow{u} E'$, then the other side ($U = G\sigma$ in Figure 6) can become a pivot, and the bal-result can be created as depicted; the original root-successors in the left-hand side are replaced by suitable terms that are shortly reachable from the pivot, so that the respective eq-level does not change ($\text{EL}(E'\sigma', U') = \text{EL}(E'\sigma'', U')$ in Figure 6). The existence of such a transformation (we claim nothing about its effectiveness) is clear by Propositions 1 and 3. *Right balancing steps* are analogous (they are elligible when the right-hand side does not sink within less than d_0 moves).

We observe that any path $W \xrightarrow{v} W'$ can sink to some depth-$|v|$ subterm of W at most, surely not deeper; hence W' arises from W by replacing its "$|v|$-top" with another top; the size of these tops is small when v is short. This observation now easily entails the above property 1, guaranteed by our transformation.

To guarantee a suitable pivot path (13) and its property 2, as a first attempt we consider the following procedure in the j-th phase (of the transformation of (11) into (12)): when we are about to replace the completed play $\begin{bmatrix} T_{j-1}'' & v \\ U_{j-1}'' & v' \end{bmatrix} \xrightarrow{\quad}$, we use its shortest prefix of the form

$\begin{bmatrix} T_{j-1}'' & \xrightarrow{v_{j-1}} & \bar{T}_j & \xrightarrow{u_j} & T_j' \\ U_{j-1}'' & \xrightarrow{v_{j-1}'} & \bar{U}_j & \xrightarrow{u_j'} & U_j' \end{bmatrix}$ where $\begin{bmatrix} \bar{T}_j & \xrightarrow{u_j} & T_j' \\ \bar{U}_j & \xrightarrow{u_j'} & U_j' \end{bmatrix}$ enables a (left or right) balancing step (i.e., some side does not sink to a root successor within less than d_0 moves).

But doing this balancing as suggested would complicate our task of creating a suitable pivot path (13), as we now discuss. First we note that we can smoothly define $W_0 \xrightarrow{w_0} W_1$: it is $U_0 \xrightarrow{v_0} U_1$ if $W_1 = U_1$, and $T_0 \xrightarrow{v_0} T_1$ if $W_1 = T_1$. Similarly we define $W_\ell \xrightarrow{w_\ell} W_{\ell+1}$ as $\bar{U}_\ell \xrightarrow{u_\ell' v_\ell'} \bar{U}_{\ell+1}$ if $W_\ell = \bar{U}_\ell$, and as $\bar{T}_\ell \xrightarrow{u_\ell v_\ell} \bar{T}_{\ell+1}$ if $W_\ell = \bar{T}_\ell$. If in the consecutive phases

j and $j{+}1$ we have the pivot on the same side, say $W_j = \bar{U}_j$ and $W_{j+1} = \bar{U}_{j+1}$, then we have no problem either: we define $W_j \xrightarrow{w_j} W_{j+1}$ simply as $\bar{U}_j \xrightarrow{u'_j v'_j} \bar{U}_{j+1}$ (which is legal since $U'_j = U''_j$).

A problem to define $W_j \xrightarrow{w_j} W_{j+1}$ arises when there is a switch of balancing sides. Hence we add a simple condition to be satisfied when such a switch is allowed to occur. Suppose $W_j = U_j$, and let Figure 6 describe the respective left balancing step. In the $(j{+}1)$-th phase of the transformation we have

$$\begin{bmatrix} T_0 & \xrightarrow{v_0} & \bar{T}_1 & \xrightarrow{u_1} & T'_1 \\ U_0 & v'_0 & \bar{U}_1 & u'_1 & U'_1 \end{bmatrix} \odot \cdots \odot \begin{bmatrix} T''_{j-1} & \xrightarrow{v_{j-1}} & \bar{T}_j & \xrightarrow{u_j} & E'\sigma' \\ U''_{j-1} & v'_{j-1} & W_j & u'_j & U'_j \end{bmatrix} \odot \begin{bmatrix} E'\sigma'' & \xrightarrow{v} \\ U'_j & v' \end{bmatrix}$$

and we are about to replace the (current) completed play $\begin{bmatrix} E'\sigma'' & \xrightarrow{v} \\ U'_j & v' \end{bmatrix}$. We would prefer to do another left balancing, ideally for a short prefix of this completed play. This is not possible only if the path $E'\sigma'' \xrightarrow{v}$ is quickly sinking in the beginning, i.e., within each segment of length d_0 a root-successor of the term starting the segment is exposed; the path $E'\sigma'' \xrightarrow{v}$ thus has a short prefix $E'\sigma'' \xrightarrow{v_{j1}} x_i\sigma''$ for some x_i (since E' is a small finite term and the path sinks along one of its branches). But $x_i\sigma''$ is reachable from the last pivot W_j ($W_j = U_j$) by a short word $\bar{v}$ (e.g., if $x_i\sigma'' = V_2$ in Figure 6, then we use the path $U \xrightarrow{\bar{v}_2} V_2$). Hence only after such a short prefix $\begin{bmatrix} E'\sigma'' & \xrightarrow{v_{j1}} & x_i\sigma'' \\ U'_j & v'_{j1} & \bar{U} \end{bmatrix}$ we allow to balance on both sides (if a left balancing is not possible earlier).

If a switch of balancing sides indeed happens in our discussed case, then we can write the j-th and the $(j{+}1)$-th play in the sequence (12) in the form

$$\begin{bmatrix} T''_{j-1} & \xrightarrow{v_{j-1}} & \bar{T}_j & \xrightarrow{u_j} & E'\sigma' \\ U''_{j-1} & v'_{j-1} & W_j & u'_j & U'_j \end{bmatrix} \odot \begin{bmatrix} E'\sigma'' & \xrightarrow{v_{j1}} & x_i\sigma'' & \xrightarrow{v_{j2}} & W_{j+1} & \xrightarrow{u_{j+1}} & T'_{j+1} \\ U'_j & v'_{j1} & \bar{U} & v'_{j2} & \bar{U}_{j+1} & u'_{j+1} & U'_{j+1} \end{bmatrix}$$

and define $W_j \xrightarrow{w_j} W_{j+1}$ as $W_j \xrightarrow{\bar{v}} x_i\sigma'' \xrightarrow{v_{j2}} W_{j+1}$ (where $\bar{v}$ is shorter than the short word $u_j v_{j1}$ but this does not matter). To summarize: for the consecutive phases j and $j{+}1$ where the j-th phase is a left-balancing step captured by Figure 6, we get

$$\begin{bmatrix} T''_{j-1} & \xrightarrow{v_{j-1}} & \bar{T}_j & \xrightarrow{u_j} & E'\sigma' \\ U''_{j-1} & v'_{j-1} & W_j & u'_j & U'_j \end{bmatrix} \odot \begin{bmatrix} E'\sigma'' & \xrightarrow{v_j} & \bar{T}_{j+1} & \xrightarrow{u_{j+1}} & T'_{j+1} \\ U'_j & v'_j & \bar{U}_{j+1} & u'_{j+1} & U'_{j+1} \end{bmatrix}$$

where either v_j is short and $W_{j+1} = \bar{U}_{j+1}$ or we can write $v_j = v_{j1} v_{j2}$ where v_{j1} is short and we have $E'\sigma'' \xrightarrow{v_{j1}} x_i\sigma''$. In the latter case we can write

$$\begin{bmatrix} E'\sigma'' & \xrightarrow{v_j} & \bar{T}_{j+1} & \xrightarrow{u_{j+1}} & T'_{j+1} \\ U'_j & v'_j & \bar{U}_{j+1} & u'_{j+1} & U'_{j+1} \end{bmatrix} = \begin{bmatrix} E'\sigma'' & \xrightarrow{v_{j1}} & x_i\sigma'' & \xrightarrow{v_{j2}} & \bar{T}_{j+1} & \xrightarrow{u_{j+1}} & T'_{j+1} \\ U'_j & v'_{j1} & \bar{U} & v'_{j2} & \bar{U}_{j+1} & u'_{j+1} & U'_{j+1} \end{bmatrix} \qquad (14)$$

where $W_{j+1} \in \{\bar{T}_{j+1}, \bar{U}_{j+1}\}$, and both paths $x_i\sigma'' \xrightarrow{v_{j2}} \bar{T}_{j+1}$ and $\bar{U} \xrightarrow{v'_{j2}} \bar{U}_{j+1}$ (that may be long) are quickly sinking (since there is no balancing possibility there).

Hence the above property 2 of the pivot path is also clear (including the case $W_{j+1} = \bar{U}_{j+1}$, where $W_j \xrightarrow{w_j} W_{j+1}$ is $\bar{U}_j \xrightarrow{u'_j v'_{j1} v'_{j2}} \bar{U}_{j+1}$).

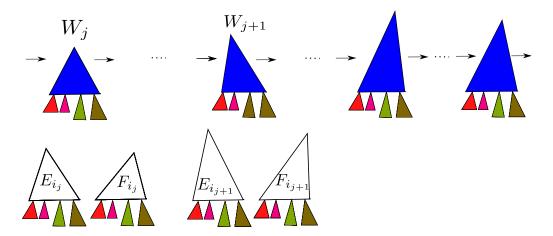Now we define the described transformation in a more formal way.

Figure 5: Non-sinking segment on the pivot path gives rise to an (n, s, g)-sequence

Modified optimal plays, and their eqlevel-concatenation. We still assume a fixed grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$. Now we let u, v, w (with subscripts etc.) range over $\mathcal{R}^*$ (not over Σ^*); hence $E \xrightarrow{w} F$ determines one path in the LTS $\mathcal{L}_{\mathcal{G}}^{\mathrm{R}}$. For $r \in \mathcal{R}$ of the form $A(x_1, \ldots, x_m) \xrightarrow{a} E$ we put $\mathrm{LAB}(r) = a$; this is extended to the respective homomorphism $\mathrm{LAB} : \mathcal{R}^* \to \Sigma^*$.

An *optimal play*, or just a *play* for short, is a sequence

$$(T_0, U_0)(r_1, r_1')(T_1, U_1)(r_2, r_2')(T_2, U_2) \cdots (r_k, r_k')(T_k, U_k),$$

denoted as

$$
\begin{array}{cccccccc}
T_0 & \xrightarrow{r_1} & T_1 & \xrightarrow{r_2} & T_2 & \cdots & \xrightarrow{r_k} & T_k \\
U_0 & r_1' & U_1 & r_2' & U_2 & & r_k' & U_k
\end{array},
\tag{15}
$$

where $T_0 \not\sim U_0$ and for each $j \in [1, k]$ we have $T_{j-1} \xrightarrow{r_j} T_j$, $U_{j-1} \xrightarrow{r_j'} U_j$, $\mathrm{LAB}(r_j) = \mathrm{LAB}(r_j')$, and $\mathrm{EL}(T_j, U_j) = \mathrm{EL}(T_{j-1}, U_{j-1}) - 1$. It is clear (by Proposition 1(2,3)) that for any $T_0 \not\sim U_0$ there is a play of the form (15) such that $k = \mathrm{EL}(T_0, U_0)$ (and $\mathrm{EL}(T_k, U_k) = 0$).

A play μ of the form (15) is a *play from* $\mathrm{START}(\mu) = (T_0, U_0)$ *to* $\mathrm{END}(\mu) = (T_k, U_k)$, and is also written as $\begin{smallmatrix} T_0 \\ U_0 \end{smallmatrix} \xrightarrow[u']{u} \begin{smallmatrix} T_k \\ U_k \end{smallmatrix}$, or just as $\begin{smallmatrix} T_0 \\ U_0 \end{smallmatrix} \xrightarrow[u']{u}$, where $u = r_1 r_2 \cdots r_k$ and $u' = r_1' r_2' \cdots r_k'$; we put $\mathrm{LENGTH}(\mu) = k$ and $\mathrm{PAIRS}(\mu) = \{(T_i, U_i) \mid i \in [0, k]\}$. We also consider the trivial plays of the form (T_0, U_0) with the length $k = 0$ (for $T_0 \not\sim U_0$). A play (15) is a *completed play* if $\mathrm{EL}(T_k, U_k) = 0$.

The *standard concatenation* $\mu\nu$ of plays $\mu = \begin{smallmatrix} T \\ U \end{smallmatrix} \xrightarrow[u']{u} \begin{smallmatrix} T' \\ U' \end{smallmatrix}$ and $\nu = \begin{smallmatrix} T'' \\ U'' \end{smallmatrix} \xrightarrow[v']{v} \begin{smallmatrix} T''' \\ U''' \end{smallmatrix}$ is defined if (and only if) $(T', U') = (T'', U'')$; in this case $\mu\nu$ is the play $\begin{smallmatrix} T \\ U \end{smallmatrix} \xrightarrow[u'v']{uv} \begin{smallmatrix} T''' \\ U''' \end{smallmatrix}$ (hence $\mathrm{END}(\mu)$ and $\mathrm{START}(\nu)$ get merged).

We aim to show a bound of the form (8) on the lengths of completed plays from (T, U). The use of (n, s, g)-sequences, bounded by Lemma 10, will become clear after we introduce a special modification of plays. Generally,

a *modified play* π is a sequence of plays $\mu_1, \mu_2, \ldots, \mu_\ell$ $(\ell \geq 1)$

where for each $j \in [1, \ell-1]$ we have $\mathrm{EL}(\mathrm{END}(\mu_j)) = \mathrm{EL}(\mathrm{START}(\mu_{j+1}))$ but $\mathrm{END}(\mu_j) \neq \mathrm{START}(\mu_{j+1})$; it is a *modified play from* $\mathrm{START}(\pi) = \mathrm{START}(\mu_1)$ *to* $\mathrm{END}(\pi) = \mathrm{END}(\mu_\ell)$, and it is a *completed modified play* if $\mathrm{EL}(\mathrm{END}(\mu_\ell)) = 0$. (As expected, if $\mathrm{END}(\mu) = (T, U)$, then by $\mathrm{EL}(\mathrm{END}(\mu))$ we refer to the eq-level $\mathrm{EL}(T, U)$; similarly in the other cases.)

We put $\text{LENGTH}(\pi) = \sum_{j\in[1,\ell]} \text{LENGTH}(\mu_j)$, and $\text{PAIRS}(\pi) = \bigcup_{j\in[1,\ell]} \text{PAIRS}(\mu_j)$. We do not consider peculiar modified plays where $\text{END}(\mu_j) = \text{START}(\mu_{j+p})$ for $p \geq 2$, in which case $\mu_{j+1}, \mu_{j+2}, \cdots, \mu_{j+p-1}$ are zero-length plays; we implicitly deem the modified plays to be *normalized* by (repeated) replacing such segments $\mu_j, \mu_{j+1}, \cdots, \mu_{j+p-1}, \mu_{j+p}$ with $\mu_j\mu_{j+p}$. E.g., a modified play μ_1, μ_2, μ_3 of the form $T_0 \xrightarrow[U_0]{u_1} T, T', T \xrightarrow[U]{u_2} T''$ (where $\text{EL}(T,U) = \text{EL}(T',U')$) is

replaced with $\mu_1\mu_3 = T_0 \xrightarrow[U_0]{u_1 u_2} T''$.

Proposition 11. *For any $T \not\sim U$ there is a completed play from (T,U), and we have* $\text{LENGTH}(\pi) = \text{EL}(T,U)$ *for each completed modified play π from (T,U); moreover, no pair can appear at two different positions in π (we thus have no repeat of a pair in π).*

Proof. The eq-levels of pairs in $\pi = \mu_1, \mu_2, \ldots, \mu_\ell$ are dropping in each μ_j; we have $\text{EL}(\text{END}(\mu_j)) = \text{EL}(\text{START}(\mu_{j+1}))$ but $\text{END}(\mu_j) \neq \text{START}(\mu_{j+p})$ for $p \geq 1$ by definition (which includes the normalization). $\square$

We also define a partial operation on the set of modified plays that is called the *eqlevel-concatenation* and denoted by $\odot$. For modified plays $\pi = \mu_1, \mu_2, \ldots, \mu_k$ and $\rho = \nu_1, \nu_2, \ldots, \nu_\ell$, the eqlevel-concatenation $\pi \odot \rho$ is defined if (and only if) $\text{EL}(\text{END}(\pi)) = \text{EL}(\text{START}(\rho))$; we recall that $\text{END}(\pi) = \text{END}(\mu_k)$ and $\text{START}(\rho) = \text{START}(\nu_1)$. Suppose that $\pi \odot \rho$, in the above notation, is defined. If $\text{END}(\mu_k) \neq \text{START}(\nu_1)$, then $\pi \odot \rho = \mu_1, \mu_2, \ldots, \mu_k, \nu_1, \nu_2, \ldots, \nu_\ell$; if $\text{END}(\mu_k) = \text{START}(\nu_1)$, then $\pi \odot \rho = \mu_1, \mu_2, \ldots, \mu_{k-1}, \mu_k\nu_1, \nu_2, \nu_3, \ldots, \nu_\ell$. (We implicitly assume a normalization in the end, if necessary; but this will not be needed in our concrete cases.)

We note that the operation $\odot$ is associative.

In what follows, by writing the expression $\pi \odot \rho$ for modified plays π, ρ we implicitly claim that $\pi \odot \rho$ is defined (and we refer to the resulting modified play $\pi \odot \rho$). By writing $\pi\rho$ we implicitly claim that $\text{END}(\pi) = \text{START}(\rho)$, and $\pi\rho$ refers to the modified play $\pi \odot \rho$.
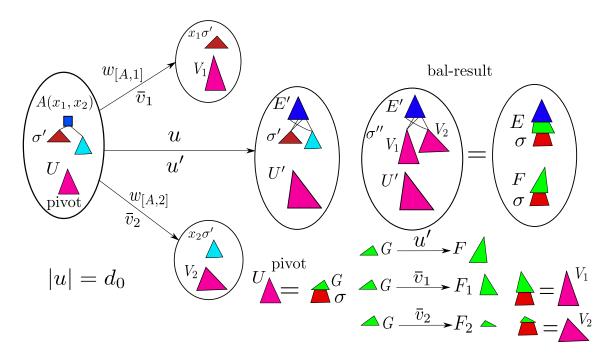
We now show a particular modification of plays, a first step towards creating (n,s,g)-sequences. In this process we will frequently replace a (sub)play of the type $\rho = T \xrightarrow[U]{u} T' \atop U\ u'\ U'$ with a modified play $\rho' = T \xrightarrow[U]{u} T' \odot T'' \atop U\ u'\ U'\ U'$ that has the same length by definition; essentially it means that we have replaced T' with T'' while guaranteeing that $\text{EL}(T',U') = \text{EL}(T'',U')$.

Balancing steps, their pivots and balanced results. Informally speaking, a play $T \xrightarrow[U]{u} T' \atop U\ u'\ U'$ enables left balancing if $T \xrightarrow{u} T'$ misses the opportunity to sink to a root-successor as quickly as possible (recall $w_{[A,i]}$ and d_0 defined around (6)). A left balancing is illustrated in Figure 6 (in both a pictorial and a textual form). We start with a simple example, and only then we give a formal definition.

Let us consider a play of the form

$$T = A(G_1,G_2) \xrightarrow[r_1']{r_1} B(C(G_2,G_1),G_1) \xrightarrow[U_1]{r_2} B'(G_1,C(G_2,G_1)) = T'$$
$$\quad U \qquad\qquad\qquad\qquad r_2' \qquad\qquad U'$$

where r_1 is $A(x_1,x_2) \xrightarrow{a_1} B(C(x_2,x_1),x_1)$, and r_2 is $B(x_1,x_2) \xrightarrow{a_2} B'(x_2,x_1)$. Let r_3 be $A(x_1,x_2) \xrightarrow{a_3} x_1$, hence we also have $A(G_1,G_2) \xrightarrow{a_3} G_1$. (Therefore the path $T \xrightarrow{r_1 r_2} T'$ clearly missed the opportunity to sink to G_1 as quickly as possible.) Since $T \xrightarrow{a_3} G_1$, there must be a transition $U \xrightarrow{a_3} V_1$, generated by a rule r_3', such that $\text{EL}(G_1,V_1) \geq \text{EL}(T,U) -$

$$A(x_1,...,x_m)\sigma' \xrightarrow{w_{[A,i]}} x_i\sigma' \text{ (or } w_{[A,i]} \text{ does not exist)}$$

$$\rho = \begin{matrix} A(x_1,...,x_m)\sigma' \\ U \end{matrix} \xrightarrow[u']{u} \begin{matrix} E'\sigma' \\ U' \end{matrix} \qquad \rho' = \begin{matrix} A(x_1,...,x_m)\sigma' \\ U \end{matrix} \xrightarrow[u']{u} \begin{matrix} E'\sigma' \\ U' \end{matrix} \odot \begin{matrix} E'\sigma'' \\ U' \end{matrix} \qquad \text{where } x_i\sigma'' = V_i$$

$$U \xrightarrow{\bar{v}_i} V_i \ (V_i = U \text{ when } w_{[A,i]} \text{ does not exist})$$

Figure 6: Balancing step $\rho \vdash_L \rho'$ ($|u| = d_0$, i ranges over $[1,m]$, $\text{EL}(x_i\sigma', V_i) > \text{EL}(E'\sigma', U')$)

1 (by Proposition 1(3)); hence $\text{EL}(G_1, V_1) > \text{EL}(T', U')$ (since $\text{EL}(T', U') = \text{EL}(T, U) - 2$ by the definition of plays). In $T' = B'(G_1, C(G_2, G_1))$ we can thus replace G_1 with V_1 without affecting $\text{EL}(T', U')$; indeed, we have $\text{EL}(T', B'(V_1, C(G_2, V_1)) \geq \text{EL}(G_1, V_1)$ (using Proposition 3(2)), and $\text{EL}(G_1, V_1) > \text{EL}(T', U')$ thus entails that $\text{EL}(B'(V_1, C(G_2, V_1)), U') = \text{EL}(T', U')$ (by Proposition 1(1)). If also G_2 can be reached from $A(G_1, G_2)$ in less than two steps, we similarly get V_2, where $U \xrightarrow{r'_4} V_2$ for some r'_4, so that $\text{EL}(B'(V_1, C(V_2, V_1)), U') = \text{EL}(T', U')$; hence

$$\begin{matrix} A(G_1, G_2) \\ U \end{matrix} \xrightarrow[r'_1 r'_2]{r_1 r_2} \begin{matrix} T' \\ U' \end{matrix} \odot \begin{matrix} B'(V_1, C(V_2, V_1)) \\ U' \end{matrix}$$

is a well-defined modified play in this case. Here U is the "pivot", and we note that U', V_1, V_2 are all reachable from U in at most two steps. Hence if we present U in a "2-top form", say $U = G\sigma$ where $G = A_0(A_1(x_1, x_2), A_2(x_3, x_4))$, then we have $U' = F\sigma$, $V_1 = F_1\sigma$, $V_2 = F_2\sigma$ where $G \xrightarrow{r'_1 r'_2} F$, $G \xrightarrow{r'_3} F_1$, $G \xrightarrow{r'_4} F_2$. Now the "bal-result" $(T'', U') = (B'(V_1, C(V_2, V_1)), U')$ can be presented as $(E\sigma, F\sigma)$ where $E = B'(F_1, C(F_2, F_1))$; we note that in $U = G\sigma$ the top G is small, hence also E, F are small, while the terms $x\sigma$ might be large. We now formalize (and generalize) the observation that has been exemplified.

We again consider a fixed general grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, and the numbers m (7) and d_0 (6). We say that a *play* $\rho = \begin{smallmatrix} T \\ U \end{smallmatrix} \xrightarrow[u']{u} \begin{smallmatrix} T' \\ U' \end{smallmatrix}$ *enables L-balancing* if $|u| = d_0$ (hence also $|u'| = d_0$)

and $T \xrightarrow{u} T'$ is root-performable, i.e., $T = A(x_1, \ldots, x_m)\sigma'$, $A(x_1, \ldots, x_m) \xrightarrow{u} E'$, and thus $T' = E'\sigma'$ (where $A \in \mathcal{N}$, $E' \in \text{TERMS}_\mathcal{N}$, $\text{VAR}(E') \subseteq \{x_1, \ldots, x_m\}$). We can thus write

$$\rho = \begin{array}{c} T \\ U \end{array} \xrightarrow[u']{u} \begin{array}{c} T' \\ U' \end{array} = \begin{array}{c} A(x_1, \ldots, x_m)\sigma' \\ U \end{array} \xrightarrow[u']{u} \begin{array}{c} E'\sigma' \\ U' \end{array}.$$

(We have not excluded that $E' = x_i$ for some $i \in [1, m]$.)

In the described case, in T' we can replace each occurrence of a root-successor of T (which is $x_i\sigma'$ for $i \in [1, m]$) with a term that is shortly reachable from U so that $\text{EL}(T', U')$ is unaffected by this replacement; we now make this claim more precise, referring again to the illustration in Figure 6.

Suppose $A(x_1, \ldots, x_m) \xrightarrow{w_{[A,i]}} x_i$ (cf. the definitions around (6)), hence $T \xrightarrow{w_{[A,i]}} x_i\sigma'$; since $\text{EL}(T, U) = d_0 + \text{EL}(T', U') \geq d_0$ and $|w_{[A,i]}| < d_0$, there must be $\bar{v}_i \in \mathcal{R}^+$ and a term V_i such that $|\bar{v}_i| = |w_{[A,i]}|$, $\text{LAB}(\bar{v}_i) = \text{LAB}(w_{[A,i]})$, $U \xrightarrow{\bar{v}_i} V_i$, and $\text{EL}(x_i\sigma', V_i) \geq \text{EL}(T, U) - |w_{[A,i]}| > \text{EL}(T, U) - d_0 = \text{EL}(T', U')$ (we use Proposition 1(3)). We can thus reason for all $i \in [1, m]$. If there is no $w_{[A,i]}$ for some $i \in [1, m]$, then $x_i\sigma'$ is not "exposable" in $T = A(x_1, \ldots, x_m)\sigma'$, hence not in $T' = E'\sigma'$ either, and $x_i\sigma'$ can be replaced by any term without changing the equivalence class of T'; in this case we put $V_i = U$, thus having $U \xrightarrow{\varepsilon} V_i$. Therefore $\text{EL}(E'\sigma', U') = \text{EL}(E'\sigma'', U')$ where $x_i\sigma'' = V_i$ for all $i \in [1, m]$ (by using Propositions 3(2) and 1(1)).

Hence for a play $\rho = \begin{array}{c} T \\ U \end{array} \xrightarrow[u']{u} \begin{array}{c} T' \\ U' \end{array}$ in the above notation we can soundly define an *L-balancing step* $\rho \vdash_L \rho'$ where ρ' is a modified play $\rho' = \rho \odot (E'\sigma'', U')$, depicted in Fig.6. For such an L-balancing step $\rho \vdash_L \rho'$, the term U is called the *pivot* and the pair $(E'\sigma'', U')$ is called the *bal-result*.

An *R-balancing step* $\rho \vdash_R \rho'$ is defined symmetrically: if in $\rho = \begin{array}{c} T \\ U \end{array} \xrightarrow[u']{u} \begin{array}{c} T' \\ U' \end{array}$ we have $|u| = |u'| = d_0$ and $U \xrightarrow{u'} U'$ is root-performable, and presented as $A(x_1, \ldots, x_m)\sigma' \xrightarrow{u'} F'\sigma'$, then we can soundly define

$$\begin{array}{c} T \\ A(x_1, \ldots, x_m)\sigma' \end{array} \xrightarrow[u']{u} \begin{array}{c} T' \\ F'\sigma' \end{array} \vdash_R \begin{array}{c} T \\ A(x_1, \ldots, x_m)\sigma' \end{array} \xrightarrow[u']{u} \begin{array}{c} T' \\ F'\sigma' \end{array} \odot \begin{array}{c} T' \\ F'\sigma'' \end{array};$$

here T is the *pivot* and $(T', F'\sigma'')$ is the *bal-result*.

Relation of the tops of the pivot and of the bal-result. We now look in more detail at the fact that the pivot of a balancing step and the respective bal-result can be written $G\sigma$ and $(E\sigma, F\sigma)$ for specifically related small "tops" G, E, F (as is also depicted in Figure 6).

We say that a finite term G is a *p-top*, for $p \in \mathbb{N}_+$, if $\text{HEIGHT}(G) \leq p$, each depth-p subterm is a variable, and $\text{VAR}(G) = \{x_1, \ldots, x_n\}$ for some $n \in \mathbb{N}$; hence $n \leq m^p$ (for m being the maximum arity of nonterminals (7)).

We note that each term W has a *p-top form* $G\sigma$, i.e., $W = G\sigma$, G is a p-top, $\text{SUPP}(\sigma) \subseteq \text{VAR}(G)$, and we have $x\sigma \in \text{VAR}$ for each x occurring in G in depth less than p. (Only a branch of W that finishes with a variable in depth less than p gives rise to such a branch in G.) E.g., a 2-top form of $A(B(x_9, C(x_3, x_6)), x_9)$ is $G\sigma$ where $G = A(B(x_1, x_2), x_3)$ and $\sigma = [x_1/x_9, x_2/C(x_3, x_6), x_3/x_9]$; another 2-top form of this term is $G'\sigma'$ where $G' = A(B(x_1, x_2), x_1)$ and $\sigma' = [x_1/x_9, x_2/C(x_3, x_6)]$. (We could strengthen the definition to get the unique p-top form to each term, but this is not necessary.)

We say that $G\sigma$ is a *p-safe form* of W if $W = G\sigma$ and $W \xrightarrow{v}$, $|v| \leq p$, implies $G \xrightarrow{v}$ (i.e., each word $v \in \mathcal{R}^*$ of length at most p that is performable from W is also performable from G). We easily observe that each p-top form $G\sigma$ of W is also a p-safe form of W.

The next proposition follows immediately from the definition of balancing steps.

Proposition 12. *Let W be the pivot and (T'', U'') the bal-result of an L-balancing step. Then for any d_0-safe form $G\sigma$ of W we have $(T'', U'') = (E\sigma, F\sigma)$ where*

- $G \xrightarrow{u'} F$ *for some* $u' \in \mathcal{R}^+$, $|u'| = d_0$;

- $E = E'\overline{\sigma}$ *where* $A(x_1, \ldots, x_m) \xrightarrow{u} E'$ *for some* $A \in \mathcal{N}$, $u \in \mathcal{R}^+$, $|u| = d_0$, *and for all* $i \in [1, m]$ *we have* $G \xrightarrow{\overline{v}_i} F_i$ *where* $F_i = x_i\overline{\sigma}$, *for some* $\overline{v}_i$, $|\overline{v}_i| < d_0$ *(hence* $T'' = E\sigma = E'\overline{\sigma}\sigma = E'\sigma''$ *where* $W \xrightarrow{\overline{v}_i} x_i\sigma''$, *for all* $i \in [1, m]$*)*.

A symmetric claim holds if W, (T'', U'') correspond to an R-balancing step.

We note a concrete consequence for future use. (Fig. 2 might be again helpful.)

Corollary 13. *Let $G\sigma$ be a d_0-safe form of W. If W is the pivot of a balancing step, then the respective bal-result can be written as $(E\sigma, F\sigma)$ where $\mathrm{VAR}(E, F) \subseteq \mathrm{VAR}(G)$ and*

$$\mathrm{SIZE}(E, F) \leq \mathrm{SIZE}(G) + (m{+}2) \cdot d_0 \cdot \mathrm{SINC}.$$

Proof. W.l.o.g. we assume an L-balancing step, and use $E = E'\overline{\sigma}$ and F guaranteed by Proposition 12, where $x_i\overline{\sigma} = F_i$ for all $i \in [1, m]$. We thus have

$$\mathrm{SIZE}(E, F) \leq \mathrm{NTSIZE}(E') + \mathrm{SIZE}(\{F, F_1, F_2, \ldots, F_m\}),$$

since for presenting E we redirect each arc in E' that leads to x_i towards the root of F_i (for $i \in [1, m]$). Since $A(x_1, \ldots, x_m) \xrightarrow{u} E'$ where $|u| = d_0$, we have $\mathrm{NTSIZE}(E') \leq d_0 \cdot \mathrm{SINC}$. Since all $F, F_1, F_2, \ldots, F_m$ are reachable from G in at most d_0 steps, we get $\mathrm{SIZE}(\{F, F_1, F_2, \ldots, F_m\}) \leq \mathrm{SIZE}(G) + (m + 1) \cdot d_0 \cdot \mathrm{SINC}$ by Proposition 5(3); moreover, all sets $\mathrm{VAR}(F)$ and $\mathrm{VAR}(F_i)$, $i \in [1, m]$, are thus subsets of $\mathrm{VAR}(G)$. The claim follows. $\square$

We derive a small bound on the number of bal-results when the pivot is fixed. We put

$$d_1 = 2 \cdot |\mathcal{N}| \cdot (\max\{d_0, |\mathcal{R}|^{d_0}\})^{m+2} \tag{16}$$

(referring to the grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$).

Proposition 14. *The number of bal-results related to a fixed pivot W is at most d_1.*

Proof. Given W, we fix its d_0-safe form $G\sigma$ (e.g., a d_0-top form). Now we suppose that $W = G\sigma$ is the pivot of an L-balancing step; let $(E\sigma, F\sigma) = (E'\overline{\sigma}\sigma, F\sigma)$ be the respective bal-result, as captured by Proposition 12. We have at most $|\mathcal{R}|^{d_0}$ options for u' determining F, and at most $|\mathcal{N}| \cdot |\mathcal{R}|^{d_0}$ options for E'. For each $i \in [1, m]$, we have at most $1 + |\mathcal{R}|^1 + |\mathcal{R}|^2 \cdots + |\mathcal{R}|^{d_0-1} \leq \max\{d_0, |\mathcal{R}|^{d_0}\}$ options for F_i. Altogether we get no more than $|\mathcal{N}| \cdot (\max\{d_0, |\mathcal{R}|^{d_0}\})^{m+2}$ options for the bal-result. The same number bounds the possible bal-results of R-balancing steps with the pivot W, hence the claim follows. $\square$

Balanced modified plays, and pivot paths. We now describe a balancing policy, yielding a sequence of balancing steps that transform a completed play to a "balanced" modified play; the idea of this policy (in a different framework) can be traced back to Sénizergues [1] (and was also used by Stirling [14]).

Let $T_0 \not\sim U_0$ and let π be a completed play π from (T_0, U_0). We show a sequence of transformation phases; after j phases we will get a completed modified play from (T_0, U_0) of the form

$$\pi_j = \mu_0 \rho_1' \mu_1 \rho_2' \cdots \mu_{j-1} \rho_j' \pi_j'$$

where π_j' is a play to be transformed in the $(j+1)$-th phase. We start with $\pi_0 = \pi_0' = \pi$. In general π_j' is not a suffix of π but the lengths of the modified plays $\pi_0, \pi_1, \pi_2, \ldots$ are the same (recall Proposition 11). In the end we get a *balanced modified play* $\pi_\ell = \mu_0 \rho_1' \mu_1 \rho_2' \cdots \mu_{\ell-1} \rho_\ell' \pi_\ell'$ (for some $\ell \geq 0$) where π_ℓ' is non-transformable; this final modified play $\pi_\ell = \mu_0 \rho_1' \mu_1 \rho_2' \cdots \mu_{\ell-1} \rho_\ell' \mu_\ell$ (where $\mu_\ell = \pi_\ell'$) can be also presented as

$$\begin{bmatrix} T_0 & \xrightarrow{v_0} & T_1 \\ U_0 & v_0' & U_1 \end{bmatrix} \begin{bmatrix} T_1 & \xrightarrow{u_1} & T_1' & \odot & T_1'' \\ U_1 & u_1' & U_1' & & U_1'' \end{bmatrix} \begin{bmatrix} T_1'' & \xrightarrow{v_1} & T_2 \\ U_1'' & v_1' & U_2 \end{bmatrix} \begin{bmatrix} T_2 & \xrightarrow{u_2} & T_2' & \odot & T_2'' \\ U_2 & u_2' & U_2' & & U_2'' \end{bmatrix} \cdots \begin{bmatrix} T_\ell & \xrightarrow{u_\ell} & T_\ell' & \odot & T_\ell'' \\ U_\ell & u_\ell' & U_\ell' & & U_\ell'' \end{bmatrix} \begin{bmatrix} T_\ell'' & \xrightarrow{v_\ell} & T_{\ell+1} \\ U_\ell'' & v_\ell' & U_{\ell+1} \end{bmatrix}$$

where μ_j is $\begin{smallmatrix} T_0 & \xrightarrow{v_0} & T_1 \\ U_0 & v_0' & U_1 \end{smallmatrix}$ for $j = 0$ and $\begin{smallmatrix} T_j'' & \xrightarrow{v_j} & T_{j+1} \\ U_j'' & v_j' & U_{j+1} \end{smallmatrix}$ for $j \in [1, \ell]$, and ρ_j' is $\begin{smallmatrix} T_j & \xrightarrow{u_j} & T_j' & \odot & T_j'' \\ U_j & u_j' & U_j' & & U_j'' \end{smallmatrix}$ (for $j \in [1, \ell]$). By ρ_j we denote $\begin{smallmatrix} T_j & \xrightarrow{u_j} & T_j' \\ U_j & u_j' & U_j' \end{smallmatrix}$, and we have either $\rho_j \vdash_L \rho_j'$ or $\rho_j \vdash_R \rho_j'$. (Hence all μ_j and ρ_j are plays, while ρ_j' is a modified play resulting from ρ_j by a balancing step.) By our conventions (and associativity of $\odot$) we can present $\pi_\ell = \mu_0 \rho_1' \mu_1 \rho_2' \cdots \mu_{\ell-1} \rho_\ell' \mu_\ell$ also as

$$\begin{bmatrix} T_0 & \xrightarrow{v_0} & T_1 & \xrightarrow{u_1} & T_1' \\ U_0 & v_0' & U_1 & u_1' & U_1' \end{bmatrix} \odot \begin{bmatrix} T_1'' & \xrightarrow{v_1} & T_2 & \xrightarrow{u_2} & T_2' \\ U_1'' & v_1' & U_2 & u_2' & U_2' \end{bmatrix} \odot \begin{bmatrix} T_2'' & \xrightarrow{v_2} & \cdots \cdots & T_\ell & \xrightarrow{u_\ell} & T_\ell' \\ U_2'' & v_2' & & U_\ell & u_\ell' & U_\ell' \end{bmatrix} \odot \begin{bmatrix} T_\ell'' & \xrightarrow{v_\ell} & T_{\ell+1} \\ U_\ell'' & v_\ell' & U_{\ell+1} \end{bmatrix}.$$
$$(17)$$

There are ℓ (occurrences of) pivots $W_1, W_2, \cdots, W_\ell$ in (17), where $W_j \in \{T_j, U_j\}$ for each $j \in [1, \ell]$; the bal-result corresponding to W_j is (T_j'', U_j''). Though the pivots W_j can be changing their sides (we can have, e.g., $W_j = U_j$ and $W_{j+1} = T_{j+1}$), they will be on one specific *pivot path* in the LTS $\mathcal{L}_\mathcal{G}^R$, denoted

$$W_0 \xrightarrow{w_0} W_1 \xrightarrow{w_1} W_2 \cdots \xrightarrow{w_{\ell-1}} W_\ell \xrightarrow{w_\ell} W_{\ell+1} \tag{18}$$

and defined below; we will have $W_0 \in \{T_0, U_0\}$ and $W_{\ell+1} \in \{T_{\ell+1}, U_{\ell+1}\}$ but $W_0, W_{\ell+1}$ are no pivots, except the case $w_0 = \varepsilon$ and $W_0 = W_1$. The pivot path will be a useful ingredient for applying our bound on (n, s, g)-sequences (Lemma 10).

Now we describe the transformation phases (as non-effective procedures), giving also a finer presentation of μ_j ($j \in [1, \ell]$) as $\mu_j = \mu_j^U$ or $\mu_j = \mu_j^U \mu_j^S$ (U for "unclear", S for "sinking") to be discussed later. The first phase, starting with $\pi_0 = \pi$, works as follows:

1. If possible, present π_0 as $\mu_0 \rho_1 \pi'$ where ρ_1 enables a balancing step (on any side) and $\mu_0 \rho_1$ is the shortest possible. If there is no such presentation of π_0, then put $\mu_0 = \pi_0$ and *halt* (here $\ell = 0$). In this case we do not need to define the path (18).

2. Replace ρ_1 with ρ_1' where $\rho_1 \vdash_L \rho_1'$ or $\rho_1 \vdash_R \rho_1'$ (choosing arbitrarily when ρ_1 allows both L-balancing and R-balancing). Finally replace π' with a completed play π_1' from the bal-result, i.e., from $\text{END}(\rho_1')$, thus getting $\pi_1 = \mu_0 \rho_1' \pi_1'$ where $\mu_0 \rho_1' = \begin{smallmatrix} T_0 & \xrightarrow{v_0} & T_1 & \xrightarrow{u_1} & T_1' & \odot & T_1'' \\ U_0 & v_0' & U_1 & u_1' & U_1' & & U_1'' \end{smallmatrix}$.

We also define the prefix $W_0 \xrightarrow{w_0} W_1$ of (18): if we have $\rho_1 \vdash_L \rho_1'$, hence $W_1 = U_1$, then this prefix is $U_0 \xrightarrow{v_0'} U_1$; if $\rho_1 \vdash_R \rho_1'$, hence $W_1 = T_1$, then the prefix is $T_0 \xrightarrow{v_0} T_1$.

For $j \geq 1$, the $(j{+}1)$-th phase starts with $\pi_j = \mu_0 \rho_1' \mu_1 \rho_2' \cdots \mu_{j-1} \rho_j' \pi_j'$ where the last balancing step was either left, $\rho_j \vdash_L \rho_j'$, or right, $\rho_j \vdash_R \rho_j'$. We describe the $(j{+}1)$-th phase for the case $\rho_j \vdash_L \rho_j'$; the other case is symmetric. We recall Figure 6 and present $\rho_j' \pi_j'$ as

$$A(x_1,\ldots,x_m)\sigma' \xrightarrow[U_j]{u_j} E'\sigma' \xrightarrow[u_j' \ \ U_j']{} \odot E'\sigma'' \xrightarrow[U_j' \ \ v']{v} .$$

We have also already defined the prefix $W_0 \xrightarrow{w_0} W_1 \xrightarrow{w_1} W_2 \cdots \xrightarrow{w_{j-1}} W_j$ of (18), and we have $W_j = U_j$ in our considered case $\rho_j \vdash_L \rho_j'$.

Informally, the $(j{+}1)$-phase aims to make a balancing step in π_j' as early as possible but balancing at the opposite side than previously is a bit constrained. In our case a future right balancing would entail that the next pivot is on the path $E'\sigma'' \xrightarrow{v}$, and we first have to wait until a term $x_i \sigma''$ is exposed (i.e., until a prefix of v exposes one of V_1, V_2 in Figure 6, where U represents the last pivot U_j). Only then a right balancing is allowed. This exposing must obviously happen soon (i.e., for a short prefix of v) if a further left balancing is not enabled for a while (since in this case $E'\sigma'' \xrightarrow{v}$ must be quickly sinking along a branch of E'). Hence if even under this constraint the earliest next balancing will be a right balancing, then the next pivot T_{j+1} will be the final term on a path $E'\sigma'' \xrightarrow{v_{j1}} x_i \sigma'' \xrightarrow{v_{j2}} T_{j+1}$ where $v_{j1}v_{j2}$ is a prefix of v and v_{j1} is short. Since $x_i \sigma''$ is reachable by a short $\bar{v}$ from the last pivot U_j (e.g., if $x_i \sigma'' = V_2$ in Figure 6, then we use $U \xrightarrow{\bar{v}_2} V_2$), we continue building the pivot path smoothly: in our case $W_j \xrightarrow{w_j} W_{j+1}$ will be defined as $U_j \xrightarrow{\bar{v}} x_i \sigma'' \xrightarrow{v_{j2}} T_{j+1}$. When a left (unconstrained) balancing is the earliest possibility, $W_j \xrightarrow{w_j} W_{j+1}$ will be defined simply as $U_j \xrightarrow{u_j' v_j'} U_{j+1}$ for the respective prefix v_j' of v'. (We note that the pivot path gets a bit shorter than the modified play (17) whenever a switch of balancing sides occurrs.) Now we describe the $(j{+}1)$-phase more formally (assuming $\rho_j \vdash_L \rho_j'$).

1. If possible, present $\pi_j' = E'\sigma'' \xrightarrow[U_j']{v} \xrightarrow{v'}$ as $\mu_j \rho_{j+1} \pi'$ with the shortest possible $\mu_j \rho_{j+1}$ where

   a) either ρ_{j+1} enables L-balancing,

   b) or ρ_{j+1} does not enable L-balancing but it enables R-balancing and the path $E'\sigma'' \xrightarrow{v_j}$ T_{j+1} in the play $\mu_j = E'\sigma'' \xrightarrow[U_j' \ \ v_j' \ \ U_{j+1}]{v_j} T_{j+1}$ can be written $E'\sigma'' \xrightarrow{v_{j1}} x_i \sigma'' \xrightarrow{v_{j2}} T_{j+1}$ where $E' \xrightarrow{v_{j1}} x_i$, for some $i \in [1, m]$. (We recall that $U_j \xrightarrow{\bar{v}} x_i \sigma''$ where $|\bar{v}| < d_0$.)

   If there is no such presentation of π_j', then put $\mu_j = \pi_j'$ and *halt* (here $\ell = j$). In this case we have $\rho_\ell' \mu_\ell = A(x_1,\ldots,x_m)\sigma' \xrightarrow[W_\ell]{u_\ell} E'\sigma' \xrightarrow[u_\ell' \ \ U_\ell']{} \odot E'\sigma'' \xrightarrow[U_\ell' \ \ v_\ell' \ \ U_{\ell+1}]{v_\ell} T_{\ell+1}$ and we define $W_\ell \xrightarrow{w_\ell} W_{\ell+1}$ as $W_\ell \xrightarrow{u_\ell' v_\ell'} U_{\ell+1}$.

   In each case we get $\mu_j = E'\sigma'' \xrightarrow[U_j' \ \ v_j' \ \ U_{j+1}]{v_j} T_{j+1}$, and if $E'\sigma'' \xrightarrow{v_j} T_{j+1}$ can be written $E'\sigma'' \xrightarrow{v_{j1}} x_i \sigma'' \xrightarrow{v_{j2}} T_{j+1}$ where $E' \xrightarrow{v_{j1}} x_i$ (which holds in the case b) by definition), then we put $\mu_j = \mu_j^{\mathrm{U}} \mu_j^{\mathrm{S}}$ where $\mu_j^{\mathrm{U}} = E'\sigma'' \xrightarrow[U_j' \ \ v_{j1}' \ \ \overline{U}_j]{v_{j1}} x_i \sigma''$ and $\mu_j^{\mathrm{S}} = x_i \sigma'' \xrightarrow[\overline{U}_j \ \ v_{j2}' \ \ U_{j+1}]{v_{j2}} T_{j+1}$; otherwise $\mu_j = \mu_j^{\mathrm{U}}$.

24

(We note that the "unclear" play μ_j^{U} is always short. The "sinking" play μ_j^{S} can be nonempty even if there is no switch in balancing sides, and μ_j^{S} can be long, but both paths in μ_j^{S} are quickly sinking [since no balancing possibility appears].)

2. Replace ρ_{j+1} with ρ'_{j+1} where $\rho_{j+1} \vdash_L \rho'_{j+1}$ in the case a), and $\rho_{j+1} \vdash_R \rho'_{j+1}$ in the case b). Finally replace π' with a completed play π'_{j+1} from the bal-result, i.e., from $\mathrm{END}(\rho'_{j+1})$, thus getting $\pi_{j+1} = \mu_0 \rho'_1 \mu_1 \rho'_2 \cdots \mu_j \rho'_{j+1} \pi'_{j+1}$.

In the case $\rho_{j+1} \vdash_L \rho'_{j+1}$ we have $\rho'_j \mu_j = \begin{smallmatrix} A(x_1,\ldots,x_m)\sigma' \\ W_j \end{smallmatrix} \xrightarrow[u'_j]{u_j} \begin{smallmatrix} E'\sigma' \\ U'_j \end{smallmatrix} \odot \begin{smallmatrix} E'\sigma'' \\ U'_j \end{smallmatrix} \xrightarrow[v'_j]{v_j} \begin{smallmatrix} T_{j+1} \\ W_{j+1} \end{smallmatrix}$ and we put $w_j = u'_j v'_j$, thus defining $W_j \xrightarrow{w_j} W_{j+1}$.

In the case $\rho_{j+1} \vdash_R \rho'_{j+1}$ we have $\rho'_j \mu_j = \begin{smallmatrix} A(x_1,\ldots,x_m)\sigma' \\ W_j \end{smallmatrix} \xrightarrow[u'_j]{u_j} \begin{smallmatrix} E'\sigma' \\ U'_j \end{smallmatrix} \odot \begin{smallmatrix} E'\sigma'' \\ U'_j \end{smallmatrix} \xrightarrow[v'_{j1}]{v_{j1}} \begin{smallmatrix} x_i\sigma'' \\ \overline{U}_j \end{smallmatrix} \xrightarrow[v'_{j2}]{v_{j2}} \begin{smallmatrix} W_{j+1} \\ U_{j+1} \end{smallmatrix}$

and we define $W_j \xrightarrow{w_j} W_{j+1}$ by putting $w_j = \overline{v} \, v_{j2}$ for a respective $\overline{v}$, $|\overline{v}| < d_0$, for which $W_j \xrightarrow{\overline{v}} x_i\sigma''$.

As already mentioned, the work of the $(j+1)$-phase in the case $\rho_j \vdash_R \rho'_j$ is symmetric; here we have R-balancing in the "unconditional" case a), and L-balancing in the case b) that now requires a prefix $\mu_j^{\mathrm{U}} = \begin{smallmatrix} T'_j \\ F'\sigma'' \end{smallmatrix} \xrightarrow[v'_{j1}]{v_{j1}} \begin{smallmatrix} \overline{T}_j \\ x_i\sigma'' \end{smallmatrix}$ (where $x_i\sigma''$ is shortly reachable from the last pivot T_j).

6 Analysis of Balanced Modified Plays

Assuming a given grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, we have shown a transformation of a completed play π starting in (T_0, U_0) (where T_0, U_0 can be large regular terms) to a balanced modified play $\pi_\ell = \mu_0 \rho'_1 \mu_1 \rho'_2 \cdots \mu_{\ell-1} \rho'_\ell \mu_\ell$ in the form (17), repeated here:

$$\begin{bmatrix} T_0 & \xrightarrow[v'_0]{v_0} & T_1 & \xrightarrow[u'_1]{u_1} & T'_1 \\ U_0 & & U_1 & & U'_1 \end{bmatrix} \odot \begin{bmatrix} T''_1 & \xrightarrow[v'_1]{v_1} & T_2 & \xrightarrow[u'_2]{u_2} & T'_2 \\ U''_1 & & U_2 & & U'_2 \end{bmatrix} \odot \begin{bmatrix} T''_2 & \xrightarrow[v'_2]{v_2} & \ldots\ldots & T_\ell & \xrightarrow[u'_\ell]{u_\ell} & T'_\ell \\ U''_2 & & & U_\ell & & U'_\ell \end{bmatrix} \odot \begin{bmatrix} T''_\ell & \xrightarrow[v'_\ell]{v_\ell} & T_{\ell+1} \\ U''_\ell & & U_{\ell+1} \end{bmatrix}. \tag{19}$$

In this section we perform a technical analysis of such π_ℓ, to verify that we indeed get specific small numbers n, s, g and c yielding (8), where $\mathcal{E} = \mathcal{E}_\mathcal{B}$ for a "sound" (n, s, g)-candidate $\mathcal{B}$ (which will turn out equal to the base $\mathcal{B}_{n,s,g}$, as discussed in Section 7). First we recall the discussion at the beginning of Section 5 and give an informal overview of the future analysis.

We recall that in the pivot path

$$W_0 \xrightarrow{w_0} W_1 \xrightarrow{w_1} W_2 \cdots \xrightarrow{w_{\ell-1}} W_\ell \xrightarrow{w_\ell} W_{\ell+1}$$

we have $W_j \in \{T_j, U_j\}$ for $j \in [0, \ell+1]$, and the pivots $W_1, W_2, \ldots, W_\ell$ have the respective related (eqlevel-decreasing) bal-results $(T''_1, U''_1), (T''_2, U''_2), \ldots, (T''_\ell, U''_\ell)$. Referring to (19), we recall that u_i (hence also u'_i) are short since $|u_i| = |u'_i| = d_0$ for all $i \in [1, \ell]$ (and d_0 from (6)). Recalling the discussion around (14), we can present (19) in a refined form as

$$\begin{bmatrix} T_0 & \xrightarrow[v'_0]{v_0} & T_1 & \xrightarrow[u'_1]{u_1} & T'_1 \\ U_0 & & U_1 & & U'_1 \end{bmatrix} \odot \begin{bmatrix} T''_1 & \xrightarrow[v'_{11}]{v_{11}} & \overline{T}_1 & \xrightarrow[v'_{12}]{v_{12}} & T_2 & \xrightarrow[u'_2]{u_2} & T'_2 \\ U''_1 & & \overline{U}_1 & & U_2 & & U'_2 \end{bmatrix} \odot \cdots \odot \begin{bmatrix} T''_\ell & \xrightarrow[v'_{\ell 1}]{v_{\ell 1}} & \overline{T}_\ell & \xrightarrow[v'_{\ell 2}]{v_{\ell 2}} & T_{\ell+1} \\ U''_\ell & & \overline{U}_\ell & & U_{\ell+1} \end{bmatrix}; \tag{20}$$

25

here $\begin{bmatrix} T_j'' \xrightarrow{v_j} T_{j+1} \xrightarrow{u_{j+1}} T_{j+1}' \\ U_j'' \xrightarrow{v_j'} U_{j+1} \xrightarrow{u_{j+1}'} U_{j+1}' \end{bmatrix}$, for $j \in [1, \ell]$, is presented as $\begin{bmatrix} T_j'' \xrightarrow{v_{j1}} \overline{T}_j \xrightarrow{v_{j2}} T_{j+1} \xrightarrow{u_{j+1}} T_{j+1}' \\ U_j'' \xrightarrow{v_{j1}'} \overline{U}_j \xrightarrow{v_{j2}'} U_{j+1} \xrightarrow{u_{j+1}'} U_{j+1}' \end{bmatrix}$

where v_{j1} is short and both paths $\overline{T}_j \xrightarrow{v_{j2}} T_{j+1}$ and $\overline{U}_j \xrightarrow{v_{j2}'} U_{j+1}$ are d_0-sinking (i.e., in each segment of length d_0 of these paths a root-successor in the term that starts the segment is exposed); we can have $v_{j2} = \varepsilon$.

The first segment $W_0 \xrightarrow{w_1} W_1$ is one of the paths $T_0 \xrightarrow{v_0} T_1$ and $U_0 \xrightarrow{v_0'} U_1$; each of these two paths is d_0-sinking (since otherwise the first balancing would be possible earlier). For the segment $W_j \xrightarrow{w_j} W_{j+1}$, $j \in [1, \ell]$, we have four options:

- $U_j \xrightarrow{u_j' v_{j1}'} \overline{U}_j \xrightarrow{v_{j2}'} U_{j+1}$, if $W_j = U_j$ and $W_{j+1} = U_{j+1}$;

- $T_j \xrightarrow{u_j v_{j1}} \overline{T}_j \xrightarrow{v_{j2}} T_{j+1}$, if $W_j = T_j$ and $W_{j+1} = T_{j+1}$;

- $U_j \xrightarrow{\bar{v}} \overline{T}_j \xrightarrow{v_{j2}} T_{j+1}$ for some $\bar{v}$, $|\bar{v}| < d_0$, if $W_j = U_j$ and $W_{j+1} = T_{j+1}$;

- $T_j \xrightarrow{\bar{v}} \overline{U}_j \xrightarrow{v_{j2}'} U_{j+1}$ for some $\bar{v}$, $|\bar{v}| < d_0$, if $W_j = T_j$ and $W_{j+1} = U_{j+1}$.

Hence each $W_j \xrightarrow{w_j} W_{j+1}$ has a short "unclear" prefix (it is unclear if it sinks or not), followed by a d_0-sinking suffix (which might be empty, or short, or long ...).

This entails that *if the path $W_j \xrightarrow{w_j} W_{j+1}$ visits a subterm of W_0*, which is surely the case for $W_0 \xrightarrow{w_0} W_1$, *then W_{j+1} is shortly reachable from a subterm of W_0*. Indeed, if $W_j \xrightarrow{w_j} W_{j+1}$ is $W_j \xrightarrow{w_j'} V \xrightarrow{w_j''} W_{j+1}$ where V is the last subterm of W_0 visited by the path $W_j \xrightarrow{w_j} W_{j+1}$, then w_j'' has a short unclear prefix (maybe empty) followed by a d_0-sinking suffix; but if this suffix was not short, then it would necessarily expose a root-successor in V, which is another subterm of W_0; this would contradict the choice of V. (Figure 2 might be again helpful to realize this fact.)

For each subterm V of W_0 we certainly have only a small number of terms W that are shortly reachable from V. Since there is only a small number of possible bal-results related to each concrete pivot W, and the bal-results do not repeat, we get that *the number of indices $j \in [0, \ell]$ for which $W_j \xrightarrow{w_j} W_{j+1}$ visits a subterm of W_0 is bounded by $d \cdot \text{SIZE}(T_0, U_0)$* for a small constant d. (We recall that $W_0 \in \{T_0, U_0\}$.)

We say that a *segment of the pivot path* of the form

$$V \xrightarrow{w} W_{j+1} \xrightarrow{w_{j+1}} W_{j+2} \cdots \xrightarrow{w_{j+z-1}} W_{j+z} \xrightarrow{w'} V'$$

is *crucial* if w is a nonempty suffix of w_j, $1 \leq z \leq \ell - j$, w' is a prefix of w_{j+z}, V is a subterm of W_0, and no subterm of W_0 is visited inside the segment; moreover, either V' is a subterm of W_0 or $V' = W_{\ell+1}$ (the end of the pivot path). (We can again look at Figure 2, and imagine that W_0 is the (maybe large regular) term in the rectangle and V is its subterm determined in the third rectangle, whose root is B. The next two steps can be viewed as a prefix of a crucial segment that could finish after many steps later when some of the root-successors of B in the rectangle is exposed and becomes the current root.)

Since each crucial segment is non-sinking (until the last step), it gives rise to an (n, s, g)-sequence (for some small n, s, g), as was depicted in Figure 5 and discussed in the informal beginning of Section 5. It is thus intuitively clear that *the length of any crucial segment of*

the pivot path, as well as the length of its corresponding segment of the modified play (19), *is bounded by* $d' \cdot \mathcal{E}_{\mathcal{B}_{n,s,g}}$ for a small constant d' (and the (n,s,g)-base $\mathcal{B}_{n,s,g}$, by Lemma 10).

Since each crucial segment is fully determined by the segment $W_j \xrightarrow{w_j} W_{j+1}$ in which it starts (and which visits a subterm of W_0), there are at most $d \cdot \mathrm{SIZE}(T_0, U_0)$ crucial segments, and their overall length is thus bounded by $d \cdot \mathrm{SIZE}(T_0, U_0) \cdot d' \cdot \mathcal{E}_{\mathcal{B}_{n,s,g}}$. Hence we are approaching the required bound

$$\mathrm{EL}(T_0, U_0) \leq c \cdot \left(\mathcal{E}_{\mathcal{B}_{n,s,g}} \cdot \mathrm{SIZE}(T_0, U_0) + (\mathrm{SIZE}(T_0, U_0))^2 \right),$$

for a small constant c. The bound $c \cdot (\mathrm{SIZE}(T_0, U_0))^2$ serves for bounding the sum of lengths of subpaths of $W_j \xrightarrow{w_j} W_{j+1}$ (and the corresponding subplays in (19)) when both sides are quickly sinking "inside" the (regular) terms T_0 and U_0, respectively. (An extreme case is when there is no balancing since both paths from T_0 and U_0, respectively, are d_0-sinking all the time.) Since the eq-level drops by one in each step of each play in (19), we cannot have a repeat of a pair there. Hence there is some small c such that $c \cdot (\mathrm{SIZE}(T_0, U_0))^2$ bounds the number of those pairs in (19) in which both members are "close to" subterms of T_0 or U_0. This bounds the sum of lengths of the respective segments of (19) that are sinking "closely to T_0, U_0" on both sides.

The claim of Theorem 7 is now almost clear; it will be completed in Section 7 where we show that the respective constant $\mathcal{E} = \mathcal{E}_{\mathcal{B}_{n,s,g}}$ is indeed computable. In the rest of this section we perform a routine (and somewhat tedious) analysis to show some concrete numbers n, s, g, c (cf. Table 1 at the end of the paper).

Refined presentations of balanced modified plays. Assuming a given grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, we fix a completed play π from some (maybe large regular terms) (T_0, U_0) and its transformation $\pi_\ell = \mu_0 \rho_1' \mu_1 \rho_2' \cdots \mu_{\ell-1} \rho_\ell' \pi_\ell'$ in the previous notation; in fact, we also use a finer form and write

$$\pi_\ell = \mu_0^{\mathrm{S}} \rho_1' \mu_1^{\mathrm{U}} \mu_1^{\mathrm{S}} \rho_2' \mu_2^{\mathrm{U}} \mu_2^{\mathrm{S}} \cdots \rho_\ell' \mu_\ell^{\mathrm{U}} \mu_\ell^{\mathrm{S}} \tag{21}$$

(where the superscript U can be read as "unclear" and S as "sinking"). We add that $\mu_0^{\mathrm{S}} = \mu_0$ and that we view ε (the empty sequence) also as the *empty play*, and we put $\mu_j^{\mathrm{S}} = \varepsilon$ in the cases where μ_j^{S} has not been defined explicitly. As expected, we stipulate $\mathrm{LENGTH}(\varepsilon) = 0$, $\mathrm{PAIRS}(\varepsilon) = \emptyset$, and $\mu\varepsilon = \varepsilon\mu = \mu$ for all (modified) plays μ.

The presentation (17) is accordingly refined (as in (20)) to

$$\begin{bmatrix} T_0 & \xrightarrow{v_0} & T_1 & \xrightarrow{u_1} & T_1' \\ U_0 & v_0' & U_1 & u_1' & U_1' \end{bmatrix} \odot \begin{bmatrix} T_1'' & \xrightarrow{v_{11}} & \overline{T}_1 & \xrightarrow{v_{12}} & T_2 & \xrightarrow{u_2} & T_2' \\ U_1'' & v_{11}' & \overline{U}_1 & v_{12}' & U_2 & u_2' & U_2' \end{bmatrix} \odot \cdots \odot \begin{bmatrix} T_\ell'' & \xrightarrow{v_{\ell 1}} & \overline{T}_\ell & \xrightarrow{v_{\ell 2}} & T_{\ell+1} \\ U_\ell'' & v_{\ell 1}' & \overline{U}_\ell & v_{\ell 2}' & U_{\ell+1} \end{bmatrix} \tag{22}$$

where, for $j \in [1, \ell]$, we have $\mu_j^{\mathrm{U}} = \begin{array}{c} T_j'' \xrightarrow{v_{j1}} \overline{T}_j \\ U_j'' \xrightarrow{v_{j1}'} \overline{U}_j \end{array}$, and either $\mu_j^{\mathrm{S}} = \begin{array}{c} \overline{T}_j \xrightarrow{v_{j2}} T_{j+1} \\ \overline{U}_j \xrightarrow{v_{j2}'} U_{j+1} \end{array}$ or $\mu_j^{\mathrm{S}} = \varepsilon$ in which case $v_{j2} = v_{j2}' = \varepsilon$, $\overline{T}_j = T_{j+1}$, $\overline{U}_j = U_{j+1}$.

To explain the use of the superscript S ("sinking") in μ_j^{S}, we introduce a few notions.

An (A, i)-sink word $v \in \mathcal{R}^+$ (satisfying $A(x_1, \ldots, x_m) \xrightarrow{v} x_i$) is also called a *sink-segment*; any path of the form $V \xrightarrow{v} V'$ is then also understood as a sink-segment (presentable as $A(x_1, \ldots, x_m)\sigma \xrightarrow{v} x_i\sigma$). We say that a *path* $V \xrightarrow{v} V'$ is d_0-*sinking*, if $v = v_1 v_2 \cdots v_{k+1}$ where $|v_j| < d_0$ for all $j \in [1, k+1]$ and v_j, $j \in [1, k]$, are sink-segments. A zero-length path $V \xrightarrow{\varepsilon} V$ is d_0-sinking, by putting $k = 0$ and $v_{k+1} = \varepsilon$.

A *play* $\mu = \begin{smallmatrix} T & \overset{v}{\longrightarrow} & T' \\ U & \underset{v'}{\longrightarrow} & U' \end{smallmatrix}$ is d_0-*sinking* if both its paths $T \overset{v}{\to} T'$ and $U \overset{v'}{\to} U'$ are d_0-sinking. In particular, a zero-length play $\mu = \frac{T}{U}$ is d_0-sinking, and we also view the empty play ε as d_0-sinking.

The above transformation (of π to π_ℓ) guarantees that all plays μ_0, μ_1^{S}, μ_2^{S}, ..., μ_ℓ^{S} are d_0-sinking (therefore we have put $\mu_0 = \mu_0^{\mathrm{S}}$). Indeed, if some μ_j^{S} ($j \in [0,\ell]$) was not d_0-sinking, then there would be a possibility to make a "legal" balancing step earlier in the respective transformation phase.

The presentations (21) and (22) also yield the corresponding refined version of the pivot path (18):

$$W_0 \xrightarrow{w_0^{\mathrm{S}}} W_1 \xrightarrow{w_1^{\mathrm{U}}} \overline{W}_1 \xrightarrow{w_1^{\mathrm{S}}} W_2 \xrightarrow{w_2^{\mathrm{U}}} \overline{W}_2 \xrightarrow{w_2^{\mathrm{S}}} \cdots W_\ell \xrightarrow{w_\ell^{\mathrm{U}}} \overline{W}_\ell \xrightarrow{w_\ell^{\mathrm{S}}} W_{\ell+1} \tag{23}$$

where each segment $\overline{W}_j \xrightarrow{w_j^{\mathrm{S}}} W_{j+1}$ (for $j \in [0,\ell]$ when putting $\overline{W}_0 = W_0$) corresponds to one of the paths in the play μ_j^{S}, and is thus d_0-sinking. More concretely, $W_0 \xrightarrow{w_0^{\mathrm{S}}} W_1$ (where $w_0^{\mathrm{S}} = w_0$) is either $T_0 \xrightarrow{v_0} T_1$ or $U_0 \xrightarrow{v_0'} U_1$, and $\overline{W}_j \xrightarrow{w_j^{\mathrm{S}}} W_{j+1}$ is either $\overline{T}_j \xrightarrow{v_{j2}} T_{j+1}$ or $\overline{U}_j \xrightarrow{v_{j2}'} U_{j+1}$. Each ("unclear") segment $W_j \xrightarrow{w_j^{\mathrm{U}}} \overline{W}_j$ is one of the following paths:

- $U_j \xrightarrow{u_j' v_{j1}'} \overline{U}_j$, if $W_j = U_j$ and $W_{j+1} = U_{j+1}$ (in which case $U_j' = U_j''$);

- $T_j \xrightarrow{u_j v_{j1}} \overline{T}_{j+1}$, if $W_j = T_j$ and $W_{j+1} = T_{j+1}$ (in which case $T_j' = T_j''$);

- $U_j \xrightarrow{v} \overline{T}_j$ for some v, $|v| < d_0$, if $W_j = U_j$ and $W_{j+1} = T_{j+1}$;

- $T_j \xrightarrow{v} \overline{U}_j$ for some v, $|v| < d_0$, if $W_j = T_j$ and $W_{j+1} = U_{j+1}$.

We now note that the length of each segment $\rho_j' \mu_j^{\mathrm{U}} = \begin{smallmatrix} T_j & \overset{u_j}{\longrightarrow} & T_j' \\ U_j & \underset{u_j'}{\longrightarrow} & U_j' \end{smallmatrix} \odot \begin{smallmatrix} T_j'' & \overset{v_{j1}}{\longrightarrow} & \overline{T}_j \\ U_j'' & \underset{v_{j1}'}{\longrightarrow} & \overline{U}_j \end{smallmatrix}$, and of the respective pivot-path segment $W_j \xrightarrow{w_j^{\mathrm{U}}} \overline{W}_j$, can be bounded by the small number

$$d_2 = d_0 + (1 + d_0 \cdot \mathrm{HINC}) \cdot (d_0 - 1). \tag{24}$$

Proposition 15. *For each $j \in [1,\ell]$ we have $|w_j^{\mathrm{U}}| \leq \mathrm{LENGTH}(\rho_j' \mu_j^{\mathrm{U}}) \leq d_2$.*

Proof. We have $|w_j^{\mathrm{U}}| \leq \mathrm{LENGTH}(\rho_j' \mu_j^{\mathrm{U}})$ by the above definitions (since $\mathrm{LENGTH}(\rho_j' \mu_j^{\mathrm{U}}) \geq d_0$, and either $|w_j^{\mathrm{U}}| = \mathrm{LENGTH}(\rho_j' \mu_j^{\mathrm{U}})$ or $|w_j^{\mathrm{U}}| < d_0$).

W.l.o.g. we suppose $\rho_j \vdash_L \rho_j'$ (illustrated in Fig.6) and present $\rho_j' \mu_j^{\mathrm{U}}$ accordingly as

$$\rho_j' \mu_j^{\mathrm{U}} = \begin{smallmatrix} A(x_1,\ldots,x_m)\sigma' & \overset{u_j}{\longrightarrow} & E'\sigma' \\ U_j & \underset{u_j'}{\longrightarrow} & U_j' \end{smallmatrix} \odot \begin{smallmatrix} E'\sigma'' & \overset{v_{j1}}{\longrightarrow} & \overline{T}_j \\ U_j' & \underset{v_{j1}'}{\longrightarrow} & \overline{U}_j \end{smallmatrix}$$

where $A(x_1,\ldots,x_m) \xrightarrow{u_j} E'$ and $|u_j| = d_0$; hence $\mathrm{HEIGHT}(E') \leq 1 + d_0 \cdot \mathrm{HINC}$. We have $\overline{T}_j = T_{j+1}$ if $\mu_j^{\mathrm{S}} = \varepsilon$, and $\overline{T}_j = x_i \sigma''$ (for some $i \in [1,m]$) if $\mu_j^{\mathrm{S}} \neq \varepsilon$.

The path $E'\sigma'' \xrightarrow{v_{j1}} \overline{T}_j$ must be d_0-sinking (otherwise there would be an earlier next balancing step). Hence $|v_{j1}| \leq \mathrm{HEIGHT}(E') \cdot (d_0 - 1)$. We thus get

$$\mathrm{LENGTH}(\rho_j' \mu_j^{\mathrm{U}}) = |u_j| + |v_{j1}| \leq d_0 + (1 + d_0 \cdot \mathrm{HINC}) \cdot (d_0 - 1) = d_2.$$

$\square$

Having bounded the parts $\rho'_j\mu^{\mathrm{U}}_j$, we will now bound the total length of the suffixes of μ^{S}_j that are "close to" T_0, U_0; then we will finally bound the number and the length of so-called "crucial segments" of π_ℓ starting with pivots that are also close to T_0, U_0 in a sense.

Close sink-parts in π_ℓ. Since $\mu_0 = \mu^{\mathrm{S}}_0 = \begin{smallmatrix} T_0 \\ U_0 \end{smallmatrix} \xrightarrow[v'_0]{v_0} \begin{smallmatrix} T_1 \\ U_1 \end{smallmatrix}$ is d_0-sinking, both paths $T_0 \xrightarrow{v_0} T_1$

and $U_0 \xrightarrow{v'_0} U_1$ are frequently visiting subterms of the terms T_0 and U_0. Using the fact that no pair repeats along π_ℓ (recall Proposition 11), we now derive a bound on the length of μ_0 and other segments that are "close" to (T_0, U_0).

For each $j \in [1, \ell]$ where $\mu^{\mathrm{S}}_j \neq \varepsilon$ we define the presentation $\mu^{\mathrm{S}}_j = \mu^{\mathrm{US}}_j\mu^{\mathrm{CS}}_j$ (the superscript US for "unclear sinking" and CS for "close sinking") as follows: If some of the paths in the play $\mu^{\mathrm{S}}_j = \begin{smallmatrix} \overline{T}_j \\ \overline{U}_j \end{smallmatrix} \xrightarrow[v'_{j2}]{v_{j2}} \begin{smallmatrix} T_{j+1} \\ U_{j+1} \end{smallmatrix}$ never visits a subterm of T_0 or U_0, then $\mu^{\mathrm{US}}_j = \mu^{\mathrm{S}}_j$ and $\mu^{\mathrm{CS}}_j = \varepsilon$.

Otherwise we write μ^{S}_j as $\begin{smallmatrix} \overline{T}_j \\ \overline{U}_j \end{smallmatrix} \xrightarrow[\overline{v}'_{j2}]{\overline{v}_{j2}} \begin{smallmatrix} \overline{\overline{T}}_j \\ \overline{\overline{U}}_j \end{smallmatrix} \xrightarrow[\overline{\overline{v}}'_{j2}]{\overline{\overline{v}}_{j2}} \begin{smallmatrix} T_{j+1} \\ U_{j+1} \end{smallmatrix}$ for the shortest prefix $\mu^{\mathrm{US}}_j = \begin{smallmatrix} \overline{T}_j \\ \overline{U}_j \end{smallmatrix} \xrightarrow[\overline{v}'_{j2}]{\overline{v}_{j2}} \begin{smallmatrix} \overline{\overline{T}}_j \\ \overline{\overline{U}}_j \end{smallmatrix}$ such

that each of the paths $\overline{T}_j \xrightarrow{\overline{v}_{j2}} \overline{\overline{T}}_j$ and $\overline{U}_j \xrightarrow{\overline{v}'_{j2}} \overline{\overline{U}}_j$ visits a subterm of T_0 or U_0; in this case $\mu^{\mathrm{CS}}_j = \begin{smallmatrix} \overline{\overline{T}}_j \\ \overline{\overline{U}}_j \end{smallmatrix} \xrightarrow[\overline{\overline{v}}'_{j2}]{\overline{\overline{v}}_{j2}} \begin{smallmatrix} T_{j+1} \\ U_{j+1} \end{smallmatrix}$. (Since μ^{S}_j is d_0-sinking, both paths $\overline{\overline{T}}_j \xrightarrow{\overline{\overline{v}}_{j2}} T_{j+1}$ and $\overline{\overline{U}}_j \xrightarrow{\overline{\overline{v}}'_{j2}} U_{j+1}$ are

frequently visiting subterms of the terms T_0 and U_0.) If $\mu^{\mathrm{S}}_j = \varepsilon$, then we put $\mu^{\mathrm{US}}_j = \mu^{\mathrm{CS}}_j = \varepsilon$; we also put $\mu_0 = \mu^{\mathrm{S}}_0 = \mu^{\mathrm{CS}}_0$ (while $\mu^{\mathrm{US}}_0 = \varepsilon$).

The balanced modified play π_ℓ (21) can be thus presented in more detail as

$$\pi_\ell = \mu^{\mathrm{CS}}_0\rho'_1\mu^{\mathrm{U}}_1\mu^{\mathrm{US}}_1\mu^{\mathrm{CS}}_1\rho'_2\mu^{\mathrm{U}}_2\mu^{\mathrm{US}}_2\mu^{\mathrm{CS}}_2\cdots\rho'_\ell\mu^{\mathrm{U}}_\ell\mu^{\mathrm{US}}_\ell\mu^{\mathrm{CS}}_\ell. \tag{25}$$

We refer to μ^{CS}_j, $j \in [0, \ell]$, as to *close sink-parts*. The next proposition bounds the total length of close sink-parts in (25), using the small number

$$d_3 = (\max\{d_0, |\mathcal{R}|^{d_0}\})^2 \tag{26}$$

(determined by $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$).

Proposition 16. $\sum_{j\in[0,\ell]} \mathrm{LENGTH}(\mu^{\mathrm{CS}}_j) \leq d_3 \cdot (\mathrm{SIZE}(T_0, U_0))^2$.

Proof. The number of subterms of T_0 and U_0 is $\mathrm{SIZE}(T_0, U_0)$, and each term can reach at most $\max\{|\mathcal{R}|^{d_0}, d_0\}$ terms within less than d_0 steps (since $|\mathcal{R}|^0 + |\mathcal{R}|^1 + \cdots + |\mathcal{R}|^{d_0-1} \leq |\mathcal{R}|^{d_0}$ when $|\mathcal{R}| \geq 2$). Hence there are at most $\left(\max\{|\mathcal{R}|^{d_0}, d_0\} \cdot \mathrm{SIZE}(T_0, U_0)\right)^2$ elements in $\bigcup_{j\in[0,\ell]} \mathrm{PAIRS}(\mu^{\mathrm{CS}}_j)$. Since there is no repeat of a pair in π_ℓ, the claim follows. $\square$

Crucial segments of π_ℓ. For $\pi_\ell = \mu^{\mathrm{CS}}_0\rho'_1\mu^{\mathrm{U}}_1\mu^{\mathrm{US}}_1\mu^{\mathrm{CS}}_1\rho'_2\mu^{\mathrm{U}}_2\mu^{\mathrm{US}}_2\mu^{\mathrm{CS}}_2\cdots\rho'_\ell\mu^{\mathrm{U}}_\ell\mu^{\mathrm{US}}_\ell\mu^{\mathrm{CS}}_\ell$ and the respective pivot path $W_0 \xrightarrow{w_0} W_1 \xrightarrow{w_1} W_2 \xrightarrow{w_2} \cdots W_\ell \xrightarrow{w_\ell} W_{\ell+1}$, assuming $\ell \geq 1$, we say that W_j, $j \in [1, \ell]$ is *close* (which is another variant of closeness to (T_0, U_0)) if the path $W_{j-1} \xrightarrow{w_{j-1}} W_j$ visits a subterm of T_0 or U_0; in this case we also write $W_{j-1} \xrightarrow{w_{j-1}} W_j$ as

$$W_{j-1} \xrightarrow{w'_{j-1}} V_{j-1} \xrightarrow{w''_{j-1}} W_j$$

where V_{j-1} is the last subterm of T_0 or U_0 in the path (not excluding the cases $V_{j-1} = W_{j-1}$ and $V_{j-1} = W_j$). We note that W_1 is close, since $W_0 \in \{T_0, U_0\}$.

Let $\{j \in [1, \ell] \mid W_j \text{ is close}\} = \{k_1, k_2, \ldots, k_p\}$ where $1 = k_1 < k_2 < k_3 \cdots < k_p \leq \ell$; for technical reasons we also put $k_{p+1} = \ell+1$. The pivot path can be thus written

$$W_0 \xrightarrow{w_0'} [V_0 \xrightarrow{w_0''} W_1 \xrightarrow{w_1} \cdots W_{k_2-1}] \xrightarrow{w_{k_2-1}'} \cdots [V_{k_p-1} \xrightarrow{w_{k_p-1}''} W_{k_p} \xrightarrow{w_{k_p}} \cdots W_\ell] \xrightarrow{w_\ell} W_{\ell+1} \tag{27}$$

where the brackets are just highlighting the corresponding segments. We use the segmentation (27) of the pivot path to induce the following segmentation of π_ℓ:

$$\mu_0^{\mathrm{CS}} [\rho_1' \cdots \mu_{k_2-1}^{\mathrm{US}}] \mu_{k_2-1}^{\mathrm{CS}} [\rho_{k_2}' \cdots \mu_{k_3-1}^{\mathrm{US}}] \mu_{k_3-1}^{\mathrm{CS}} \cdots \cdots [\rho_{k_p}' \cdots \mu_\ell^{\mathrm{US}}] \mu_\ell^{\mathrm{CS}} .$$

The highlighted segments are called the *crucial segments* (of π_ℓ). The total length of "non-crucial" segments μ_0^{CS}, $\mu_{k_2-1}^{\mathrm{CS}}$, $\mu_{k_3-1}^{\mathrm{CS}}$, $\cdots$, μ_ℓ^{CS} is bounded by Proposition 16. We note that μ_j^{CS} inside the crucial segments are empty since otherwise we had a close pivot there.

For bounding the number p of crucial segments and their lengths, it is useful to use the notions of stairs and their simple-stair decompositions.

Stairs, simple stairs, simple-stair decompositions. A word $v \in \mathcal{R}^*$ is a *stair* if $v = \varepsilon$ or $v = rv'$ where $r \in \mathcal{R}$, let r be $A(x_1, \ldots, x_m) \xrightarrow{a} E$, and $E \xrightarrow{v'} F$ for some $F \notin$ VAR. If v is a stair, then any path of the form $V \xrightarrow{v} V'$ is also called a stair (in the form $A(x_1, \ldots, x_m)\sigma \xrightarrow{v} F\sigma$). Hence no prefix of a stair is a sink-segment.

We say that $v = rv' \in \mathcal{R}^+$ ($r \in \mathcal{R}$) is a *simple stair* if $A(x_1, \ldots, x_m) \xrightarrow{r} E \xrightarrow{v'} F$ (for r being $A(x_1, \ldots, x_m) \xrightarrow{a} E$) where F is a subterm of E with a nonterminal root (hence $F \notin$ VAR) and v' is a (possibly empty) concatenation of (possibly long) sink-segments (hence $v' = u_1 u_2 \cdots u_k$ where u_i, $i \in [1, k]$, are sink-segments). If v is a simple stair, then also any path $V \xrightarrow{v} V'$ is called a simple stair.

Proposition 17. *1. Any stair $v \in \mathcal{R}^*$ has the unique simple-stair decomposition $v = v_1 v_2 \cdots v_q$ ($q \in \mathbb{N}$) where v_i, $i \in [1, q]$, are simple stairs.*

2. If $G \xrightarrow{v_1 v_2 \cdots v_q} G'$ where v_i are simple stairs, then $\mathrm{SIZE}(G') \leq \mathrm{SIZE}(G) + q \cdot \mathrm{SINC}$; moreover, if G is finite, then $\mathrm{HEIGHT}(G') \leq \mathrm{HEIGHT}(G) + q \cdot \mathrm{HINC}$.

Proof. 1. By induction on $|v|$, for stairs v. If $v = \varepsilon$, then $q = 0$. If $|v| > 0$, then we write $v = v_1 v'$ for the shortest $v_1 \in \mathcal{R}^+$ such that v' is a stair; v' has the unique simple-stair decomposition by the induction hypothesis. We can easily verify that v_1 is a simple stair, and that we cannot have $v_1 v' = v_1' v''$ where v_1' is a simple stair, v'' is a stair (decomposed into simple stairs), and $v_1' \neq v_1$.

2. We recall that $A(x_1, \ldots, x_m) \xrightarrow{r} E$ entails $\mathrm{SIZE}(E\sigma) \leq \mathrm{SIZE}(A(x_1, \ldots, x_m)\sigma) + \mathrm{SINC}$, and we have $\mathrm{SIZE}(F\sigma) \leq \mathrm{SIZE}(E\sigma)$ for any subterm F of E; moreover, if $A(x_1, \ldots, x_m)\sigma$ is finite, then $\mathrm{HEIGHT}(F\sigma) \leq \mathrm{HEIGHT}(E\sigma) \leq \mathrm{HEIGHT}(A(x_1, \ldots, x_m)\sigma) + \mathrm{HINC}$. $\square$

Bounding the number of crucial segments. To bound the number p of crucial segments, we use the small number

$$d_4 = d_1 \cdot (1 + |\mathrm{SRHS}|)^{d_2 + d_0 - 1} \tag{28}$$

where $\mathrm{SRHS} = \{F \mid F \text{ is a subterm of the rhs of a rule in } \mathcal{R} \text{ and } F \notin \mathrm{VAR}\}$.

Proposition 18. *The number p of crucial segments is at most $d_4 \cdot \text{SIZE}(T_0, U_0)$.*

Proof. First we note that we can have $W_j = W_{j'}$ for different $j, j' \in [1, \ell]$; but for each W we can have $W = W_j$ for at most d_1 indices $j \in [1, \ell]$, since there are at most d_1 possible bal-results for each pivot (Proposition 14) and the bal-results (T_j'', U_j''), $j \in [1, \ell]$, are all pairwise different (Proposition 11).

Hence if we get a bound on the cardinality of the set $\text{SP} = \{W_{k_1}, W_{k_2}, \ldots, W_{k_p}\}$ of "starting pivots" of the crucial segments (where $k_1 = 1$), then multiplying this bound by d_1 yields a bound on p.

We fix $j \in [1, p]$, and note that the stair $V_{k_j-1} \xrightarrow{w_{k_j-1}''} W_{k_j}$ is a suffix of the path $W_{k_j-1} \xrightarrow{w_{k_j-1}^{\text{U}}} \overline{W}_{k_j-1} \xrightarrow{w_{k_j-1}^{\text{S}}} W_{k_j}$, where $|w_{k_j-1}^{\text{U}}| \leq d_2$ and $\overline{W}_{k_j-1} \xrightarrow{w_{k_j-1}^{\text{S}}} W_{k_j}$ can be written $\overline{W}_{k_j-1} \xrightarrow{\overline{w}} \overline{\overline{W}} \xrightarrow{\overline{\overline{w}}} W_{k_j}$ where $\overline{w}$ is a sequence of sink-segments and $|\overline{\overline{w}}| < d_0$. The simple-stair decomposition of $V_{k_j-1} \xrightarrow{w_{k_j-1}''} W_{k_j}$ is thus a sequence of at most $d_2 + (d_0 - 1)$ simple stairs.

Hence a (generous) upper bound on $|\text{SP}|$ is $\text{SIZE}(T_0, U_0) \cdot (1 + |\text{SRHS}|)^{d_2 + d_0 - 1}$. This yields $p \leq \text{SIZE}(T_0, U_0) \cdot (1 + |\text{SRHS}|)^{d_2 + d_0 - 1} \cdot d_1 = d_4 \cdot \text{SIZE}(T_0, U_0)$ as claimed. $\qquad\square$

Bounding the lengths of crucial segments. For $j \in [1, p]$, we view the number $k_{j+1} - k_j$ as the *index length* of the crucial segment $[\rho_{k_j}' \cdots \mu_{k_{j+1}-1}^{\text{US}}]$. We first bound the index length, defining n, s, g and using the bound on (n, s, g)-sequences (Lemma 10), and then we bound the standard length.

We first note that *each highlighted segment* in (27) *is a stair*. Indeed, if the path

$$\left[V_{k_j-1} \xrightarrow{w_{k_j-1}''} W_{k_j} \xrightarrow{w_{k_j}} W_{k_j+1} \xrightarrow{w_{k_j+1}} W_{k_j+2} \cdots \xrightarrow{w_{k_{j+1}-2}} W_{k_{j+1}-1}\right]$$

(for $j \in [1, p]$) had a prefix that is a sink-segment, then one of $W_{k_j+1}, W_{k_j+2}, \ldots, W_{k_{j+1}-1}$ would be also close, since V_{k_j-1} is the last subterm of T_0 or U_0 in $W_{k_j-1} \xrightarrow{w_{k_j-1}} W_{k_j}$, and each subterm of V_{k_j-1} is also a subterm of T_0 or U_0.

Thus the index length of crucial segments is bounded due to the next lemma, for which we define the following small numbers:

$$n = m^{d_0}; \tag{29}$$

$$s = m^{d_0+1} + (m+2) \cdot d_0 \cdot \text{SINC} + (d_2 + d_0 - 1) \cdot \text{SINC}; \tag{30}$$

$$g = (d_2 + d_0 - 1) \cdot \text{SINC}. \tag{31}$$

Lemma 19. *We assume a balanced modified play $\pi_\ell = \mu_0^{\text{CS}} \rho_1' \mu_1^{\text{U}} \mu_1^{\text{US}} \mu_1^{\text{CS}} \cdots \rho_\ell' \mu_\ell^{\text{U}} \mu_\ell^{\text{US}} \mu_\ell^{\text{CS}}$ and the respective pivot path $W_0 \xrightarrow{w_0} W_1 \xrightarrow{w_1} \cdots W_\ell \xrightarrow{w_\ell} W_{\ell+1}$. Let*

$$V \xrightarrow{w} W_{j+1} \xrightarrow{w_{j+1}} W_{j+2} \cdots \xrightarrow{w_{j+k-1}} W_{j+k} \tag{32}$$

be a segment of the pivot path that is a stair, where $j \geq 0$, $k \geq 1$, $j + k \leq \ell$, and w is a suffix of w_j. Let $e = 1 + \text{EL}(\text{END}(\rho_{j+1}'))$ (where $\text{END}(\rho_{j+1}')$ is the bal-result related to the pivot W_{j+1}, hence (T_{j+1}'', U_{j+1}'') in (22)).

Then $k \leq \mathcal{E}_{\mathcal{B}}$ for each (n,s,g)-candidate $\mathcal{B}$ that is full below e; in particular, $k \leq \mathcal{E}_{\mathcal{B}_{n,s,g}}$. (Here n,s,g are the numbers defined by (29), (30), (31).)

Proof. We will show that the (eqlevel-decreasing) sequence $\mathrm{END}(\rho'_{j+1})$, $\mathrm{END}(\rho'_{j+2})$, ..., $\mathrm{END}(\rho'_{j+k})$ of the bal-results corresponding to the pivots W_{j+1}, W_{j+2}, ..., W_{j+k} can be presented as an (n,s,g)-sequence

$$(E_1\sigma, F_1\sigma), (E_2\sigma, F_2\sigma), \ldots, (E_k\sigma, F_k\sigma). \tag{33}$$

The claim then follows by Lemma 10. Hence it remains to show the presentation (33) of the respective bal-results. By the definition of stairs, we can present (32) as

$$A(x_1,\ldots,x_m)\sigma' \xrightarrow{w} G'_1\sigma' \xrightarrow{w_{j+1}} G'_2\sigma' \cdots \xrightarrow{w_{j+k-1}} G'_k\sigma'$$

where

$$A(x_1,\ldots,x_m) \xrightarrow{w} G'_1 \xrightarrow{w_{j+1}} G'_2 \cdots \xrightarrow{w_{j+k-1}} G'_k; \tag{34}$$

we thus have $W_{j+i} = G'_i\sigma'$ (for $i \in [1,k]$) where G'_i are finite terms with nonterminal roots.

Recalling the refined presentation (23), we write the path $W_{j+i} \xrightarrow{w^{\mathrm{U}}_{j+i}} \overline{W}_{j+i} \xrightarrow{w^{\mathrm{S}}_{j+i}} W_{j+i+1}$, for each $i \in [0,k-1]$, as $W_{j+i} \xrightarrow{w^{\mathrm{U}}_{j+i}} \overline{W}_{j+i} \xrightarrow{\overline{w}_i} \overline{\overline{W}}_{j+i} \xrightarrow{\overline{\overline{w}}_i} W_{j+i+1}$ where $\overline{W}_{j+i} \xrightarrow{\overline{w}_i} \overline{\overline{W}}_{j+i}$ is a sequence of sink-segments of lengths less than d_0, and $|\overline{\overline{w}}_i| < d_0$. We thus present (34) as

$$A(x_1,\ldots,x_m) \xrightarrow{w} G'_1 \xrightarrow{w^{\mathrm{U}}_{j+1}} \overline{G}_1 \xrightarrow{\overline{w}_1} \overline{\overline{G}}_1 \xrightarrow{\overline{\overline{w}}_1} G'_2 \cdots \xrightarrow{w^{\mathrm{U}}_{j+k-1}} \overline{G}_{k-1} \xrightarrow{\overline{w}_{k-1}} \overline{\overline{G}}_{k-1} \xrightarrow{\overline{\overline{w}}_{k-1}} G'_k\,.$$

We recall that $|w^{\mathrm{U}}_{j+i}| \leq d_2$ (for all $i \in [0,k-1]$). Since w is a suffix of $w^{\mathrm{U}}_j w^{\mathrm{S}}_j = w^{\mathrm{U}}_j \overline{w}_0 \overline{\overline{w}}_0$, we note that the simple-stair decomposition of the stair $A(x_1,\ldots,x_m) \xrightarrow{w} G'_1$ is a sequence of at most $d_2+(d_0-1)$ simple stairs. More generally, for each $i \in [1,k]$, the simple-stair decomposition of the stair

$$A(x_1,\ldots,x_m) \xrightarrow{w} G'_1 \xrightarrow{w^{\mathrm{U}}_{j+1}} \overline{G}_1 \xrightarrow{\overline{w}_1} \overline{\overline{G}}_1 \xrightarrow{\overline{\overline{w}}_1} G'_2 \cdots \xrightarrow{w^{\mathrm{U}}_{j+i-1}} \overline{G}_{i-1} \xrightarrow{\overline{w}_{i-1}} \overline{\overline{G}}_{i-1} \xrightarrow{\overline{\overline{w}}_{i-1}} G'_i$$

is a sequence of at most $i \cdot (d_2+(d_0-1))$ simple stairs; hence

$$\mathrm{SIZE}(G'_i) \leq \mathrm{SIZE}(A(x_1,\ldots,x_m)) + i \cdot (d_2+d_0-1) \cdot \mathrm{SINC} \tag{35}$$

(recalling Proposition 17). We recall the relation of a pivot, $W_{j+i} = G'_i\sigma'$ in our case, and its bal-result, as captured by Proposition 12 (and illustrated in Figure 6). We note that $G'_i\sigma'$ might not be a d_0-safe form of W_{j+i} (due to possible short branches of G'_i). This leads us to present $V = A(x_1,\ldots,x_m)\sigma'$ in a d_0-top form, as $A(x_1,\ldots,x_m)\overline{\overline{\sigma}}\sigma$ where $A(x_1,\ldots,x_m)\overline{\overline{\sigma}}$ is the respective d_0-top.

Putting $G_i = G'_i\overline{\overline{\sigma}}$, we get $W_{j+i} = G'_i\sigma' = G'_i\overline{\overline{\sigma}}\sigma = G_i\sigma$, for each $i \in [1,k]$. We have $\mathrm{VAR}(G_i) \subseteq \mathrm{VAR}(A(x_1,\ldots,x_m)\overline{\overline{\sigma}}) \subseteq \{x_1,\ldots,x_n\}$ (for $n = m^{d_0}$), and any word $v \in \mathcal{R}^*$ with $|v| \leq d_0$ that is performable from $W_{j+i} = G_i\sigma$ is performable from G_i as well.

Since $G_i\sigma$ is thus a d_0-safe form of W_{j+i}, the bal-result related to $W_{j+i} = G_i\sigma$ can be written as $(E_i\sigma, F_i\sigma)$ where $\mathrm{VAR}(E_i, F_i) \subseteq \mathrm{VAR}(G_i) \subseteq \{x_1,\ldots,x_n\}$, and $\mathrm{SIZE}(E_i, F_i) \leq \mathrm{SIZE}(G_i)+(m+2)\cdot d_0\cdot\mathrm{SINC}$ (by Corollary 13). By mimicking the derivation of the bound (35), we get

$$\mathrm{SIZE}(G_i) \leq \mathrm{SIZE}(A(x_1,\ldots,x_m)\overline{\sigma}) + i \cdot (d_2+(d_0-1)) \cdot \mathrm{SINC}.$$

Since $\text{SIZE}(A(x_1,\ldots,x_m)\overline{\sigma}) \le m^{d_0+1}$, and $g = (d_2+d_0-1)\cdot\text{SINC}$, we get

$$\text{SIZE}(G_i) \le m^{d_0+1} + i\cdot g, \text{ for all } i \in [1,k].$$

From $\text{SIZE}(E_i, F_i) \le \text{SIZE}(G_i) + (m+2)\cdot d_0 \cdot \text{SINC}$ we derive, for all $i \in [1,k]$, that

$$\text{SIZE}(E_i, F_i) \le m^{d_0+1} + (m+2)\cdot d_0 \cdot \text{SINC} + i\cdot g = s + (i-1)\cdot g\,.$$

Hence the sequence $\text{END}(\rho'_{j+1})$, $\text{END}(\rho'_{j+2})$, ..., $\text{END}(\rho'_{j+k})$ can be indeed presented as an (n,s,g)-sequence $(E_1\sigma, F_1\sigma), (E_2\sigma, F_2\sigma),\ldots,(E_k\sigma, F_k\sigma)$. $\qquad\square$

Corollary 20. *For each crucial segment $\rho'_{k_j}\cdots\mu^{\text{US}}_{k_{j+1}-1}$ we have $k_{j+1} - k_j \le \mathcal{E}_\mathcal{B}$ for each (n,s,g)-candidate $\mathcal{B}$ that is full below $1 + \text{EL}(\text{END}(\rho'_{k_j}))$; in particular, $k_{j+1} - k_j \le \mathcal{E}_{\mathcal{B}_{n,s,g}}$.*

We will now bound the (standard) length of a crucial segment by multiplying its index length, increased by 1, by the small number

$$d_5 = (d_2 + d_0 - 1)\cdot(1 + (d_0 - 1)\cdot\text{HINC})\,. \tag{36}$$

Proposition 21. *For each $j \in [1,p]$ we have $\text{LENGTH}(\rho'_{k_j}\cdots\mu^{\text{US}}_{k_{j+1}-1}) \le d_5 \cdot (1+k_{j+1}-k_j)$.*

Proof. We fix a crucial segment $\rho'_{k_j}\cdots\mu^{\text{US}}_{k_{j+1}-1}$. We make a convenient notational change (using $j+1$ for the previous k_j, and k for $k_{j+1}-k_j$) and present this segment as

$$\rho'_{j+1}\mu^{\text{U}}_{j+1}\mu^{\text{US}}_{j+1}\rho'_{j+2}\mu^{\text{U}}_{j+2}\mu^{\text{US}}_{j+2}\cdots\rho'_{j+k}\mu^{\text{U}}_{j+k}\mu^{\text{US}}_{j+k}$$

(for $i \in [1,k-1]$ we have $\mu^{\text{US}}_{j+i} = \mu^{\text{S}}_{j+i}$ since $\mu^{\text{CS}}_{j+i} = \varepsilon$). In a more detailed presentation, the segment is a prefix of

$$\begin{array}{l}
T_{j+1} \xrightarrow{u_{j+1}} T'_{j+1} \odot T''_{j+1} \xrightarrow{v_{j+1,1}} \overline{T}_{j+1} \xrightarrow{v_{j+1,2}} T_{j+2} \ldots T_{j+k} \xrightarrow{u_{j+k}} T'_{j+k} \odot T''_{j+k} \xrightarrow{v_{j+k,1}} \overline{T}_{j+k} \xrightarrow{v_{j+k,2}} T_{j+k+1} \\
U_{j+1} \xrightarrow{u'_{j+1}} U'_{j+1} \quad U''_{j+1} \xrightarrow{v'_{j+1,1}} \overline{U}_{j+1} \xrightarrow{v'_{j+1,2}} U_{j+2} \quad U_{j+k} \xrightarrow{u'_{j+k}} U'_{j+k} \quad U''_{j+k} \xrightarrow{v'_{j+k,1}} \overline{U}_{j+k} \xrightarrow{v'_{j+k,2}} U_{j+k+1}
\end{array} \tag{37}$$

finishing somewhere in the part $\begin{array}{c}\overline{T}_{j+k} \xrightarrow{v_{j+k,2}} T_{j+k+1}\\ \overline{U}_{j+k} \xrightarrow{v'_{j+k,2}} U_{j+k+1}\end{array}$, as determined by μ^{CS}_{j+k} (which might be empty or nonempty). We also consider the related pivot-path stair

$$V \xrightarrow{w} W_{j+1} \xrightarrow{w^{\text{U}}_{j+1}} \overline{W}_{j+1} \xrightarrow{w^{\text{S}}_{j+1}} W_{j+2}\cdots \xrightarrow{w^{\text{U}}_{j+k-1}} \overline{W}_{j+k-1} \xrightarrow{w^{\text{S}}_{j+k-1}} W_{j+k} \tag{38}$$

where $V \xrightarrow{w} W_{j+1}$ is related to the part $\rho'_j\mu^{\text{U}}_j\mu^{\text{S}}_j$ that precedes our crucial segment: the path $V \xrightarrow{w} W_{j+1}$ is the suffix $V_j \xrightarrow{w''_j} W_{j+1}$ of $W_j \xrightarrow{w^{\text{U}}_j} \overline{W}_j \xrightarrow{w^{\text{S}}_j} W_{j+1}$ for the respective last subterm V_j of T_0 or U_0. We present the stair (38) similarly as the stair (32) in the proof of Lemma 19. We get $V = A(x_1,\ldots,x_m)\sigma'$ and

$$A(x_1,\ldots,x_m) \xrightarrow{w} G'_1 \xrightarrow{w^{\text{U}}_{j+1}} \overline{G}_1 \xrightarrow{\overline{\overline{w}}_1} \overline{\overline{G}}_1 \xrightarrow{\overline{\overline{\overline{w}}}_1} G'_2 \cdots \xrightarrow{w^{\text{U}}_{j+k-1}} \overline{G}_{k-1} \xrightarrow{\overline{\overline{w}}_{k-1}} \overline{\overline{G}}_{k-1} \xrightarrow{\overline{\overline{\overline{w}}}_{k-1}} G'_k\,.$$

We will show that

$$\text{LENGTH}(\rho'_{j+1}\cdots\mu^{\text{US}}_{j+k-1}) \le d_5 \cdot k - (d_0-1)\cdot\text{HEIGHT}(G'_k) \tag{39}$$

and

$$\text{LENGTH}(\rho'_{j+k}\mu^{\text{U}}_{j+k}\mu^{\text{US}}_{j+k}) \le d_5 + (d_0-1)\cdot\text{HEIGHT}(G'_k)\,, \tag{40}$$

33

which yields $\text{LENGTH}(\rho'_{j+1} \cdots \mu^{\text{US}}_{j+k}) \leq d_5 \cdot (1+k)$ and thus finishes the proof.

We show (39): Similarly as (35), we derive $\text{HEIGHT}(G'_i) \leq 1 + i \cdot (d_2+d_0-1) \cdot \text{HINC}$, for all $i \in [1,k]$. Since $|w^{\text{U}}_{j+i}| \leq \text{LENGTH}(\rho'_{j+i} \mu^{\text{U}}_{j+i}) \leq d_2$, $|\overline{w}_i \overline{\overline{w}}_i| = \text{LENGTH}(\mu^{\text{S}}_{j+i})$, $\overline{G}_i \xrightarrow{\overline{w}_i} \overline{\overline{G}}_i$ is a sequence of sink-segments of lengths less than d_0, and $|\overline{\overline{w}}_i| < d_0$, we also derive

$$|\overline{w}_i| \leq (d_0-1) \cdot (\text{HEIGHT}(\overline{G}_i) - \text{HEIGHT}(\overline{\overline{G}}_i)), \text{ and}$$

$$\text{HEIGHT}(G'_{i+1}) \leq \text{HEIGHT}(G'_i) + (d_2 + d_0-1) \cdot \text{HINC} - (\text{HEIGHT}(\overline{G}_i) - \text{HEIGHT}(\overline{\overline{G}}_i)).$$

For $\text{SUM} = \sum_{i=1}^{k-1} \left(\text{HEIGHT}(\overline{G}_i - \text{HEIGHT}(\overline{\overline{G}}_i)) \right)$ we thus get

$$\text{HEIGHT}(G'_k) \leq 1 + k \cdot (d_2 + d_0-1) \cdot \text{HINC} - \text{SUM}, \text{ and} \tag{41}$$

$$\text{LENGTH}(\rho'_{j+1} \cdots \mu^{\text{US}}_{j+k-1}) \leq (k-1) \cdot (d_2+d_0-1) + (d_0-1) \cdot \text{SUM}. \tag{42}$$

Replacing SUM in (42) with its upper bound $1 + k \cdot (d_2+d_0-1) \cdot \text{HINC} - \text{HEIGHT}(G'_k)$ (derived from (41)), we get

$$\text{LENGTH}(\rho'_{j+1} \cdots \mu^{\text{US}}_{j+k-1}) \leq k \cdot \left(d_2+d_0-1 + (d_0-1)(d_2+d_0-1) \cdot \text{HINC} \right) - (d_0-1) \cdot \text{HEIGHT}(G'_k).$$

This yields (39).

We show (40): We recall that $\text{LENGTH}(\rho'_{j+k} \mu^{\text{U}}_{j+k}) \leq d_2$, and aim to bound μ^{US}_{j+k}, assuming $\mu^{\text{US}}_{j+k} \neq \varepsilon$. In this case $\text{START}(\mu^{\text{US}}_{j+k}) = (\overline{T}_{j+k}, \overline{U}_{j+k})$, and both paths $\overline{T}_{j+k} \xrightarrow{v_{j+k,2}}, \overline{U}_{j+k} \xrightarrow{v_{j+k,2}}$ of the play $\mu^{\text{US}}_{j+k} \mu^{\text{CS}}_{j+k}$ (recall (37)) are d_0-sinking. In the worst case the play μ^{US}_{j+k} finishes when each of these two paths visits a subterm of T_0 or U_0 (in which case $\mu^{\text{CS}}_{j+k} \neq \varepsilon$ follows). Due to the construction of $\rho'_{j+k} \mu^{\text{U}}_{j+k}$ we have that both $\overline{T}_{j+k}$ and $\overline{U}_{j+k}$ are reachable from the pivot $W_{j+k} = G'_k \sigma' \in \{T_{j+k}, U_{j+k}\}$ in at most d_2 steps (in fact, one even in less than d_0 steps).

We recall that $\text{VAR}(G'_k) \subseteq \{x_1, \ldots, x_m\}$ and that $x_q \sigma'$ is a subterm of T_0 or U_0, for each $q \in [1,m]$ (since $V = A(x_1, \ldots, x_m)\sigma'$ is a subterm of T_0 or U_0). Thus if the respective paths $G'_k \sigma' \xrightarrow{\overline{v}} \overline{T}_{j+k}$ and $G'_k \sigma' \xrightarrow{\overline{\overline{v}}} \overline{U}_{j+k}$, where $|\overline{v}| \leq d_2$ and $|\overline{\overline{v}}| \leq d_2$, "sink inside" the terms $x_q \sigma'$, they visit subterms of T_0 or U_0 at such moments. The pair $(\overline{T}_{j+k}, \overline{U}_{j+k})$ can be thus surely presented as $(\overline{E}\sigma_1, \overline{F}\sigma_2)$ where $\text{VAR}(\overline{E})$ and $\text{VAR}(\overline{F})$ are subsets of $\{x_1, \ldots, x_m\}$, the terms $x_q \sigma_1$ and $x_q \sigma_2$ are subterms of T_0 or U_0, for each $q \in [1,m]$, and both $\text{HEIGHT}(\overline{E})$ and $\text{HEIGHT}(\overline{F})$ are bounded by $\text{HEIGHT}(G'_k) + d_2 \cdot \text{HINC}$.

Therefore μ^{US}_{j+k} cannot be longer than $(d_0-1) \cdot (\text{HEIGHT}(G'_k) + d_2 \cdot \text{HINC})$. This yields $\text{LENGTH}(\rho'_{j+k} \mu^{\text{U}}_{j+k} \mu^{\text{US}}_{j+k}) \leq d_2 + (d_0-1) \cdot (\text{HEIGHT}(G'_k) + d_2 \cdot \text{HINC})$, which implies (40). $\qquad \square$

7 Completing the Proof of Theorem 7

Below we repeat the statement of Theorem 7, and show a proof based on the previous results. In fact, it remains to prove that $\mathcal{E} = \mathcal{E}_{\mathcal{B}_{n,s,g}}$ is computable. The idea is that we stepwise increase an under-approximation of $\mathcal{B}_{n,s,g}$ and of the respective $\mathcal{E}$; the pairs (E, F) (of the respective sizes) that are in this process so far deemed to be equivalent (i.e., assumed to satisfy $E \sim F$) are verified by using the assumption (43) for the current (under-approximation of) $\mathcal{E}$. If we find that $\text{EL}(E, F) \leq c \cdot \left(\mathcal{E} \cdot \text{SIZE}(E, F) + (\text{SIZE}(E, F))^2 \right)$ for some of such pairs (E, F) (which can be checked Proposition 6), then we adjust (increase) the under-approximation. This process must clearly terminate, and at the end the claim (43) holds for all $T \not\sim U$, as

m	(7)	maximum arity of nonterminals
HINC	(4)	height-increase in one step
SINC	(5)	size-increase in one step
d_0	(6)	lengths of shortest (A, i)-sink words (plus 1)
d_1	(16)	number of bal-results related to one pivot
d_2	(24)	length of "unclear" part after a pivot, followed by d_0-sinking
d_3	(26)	$d_3 \cdot (\text{SIZE}(T_0, U_0))^2$ bounds the total length of close sink-parts
$n = m^{d_0}$	(29)	number of variables in "(n, s, g)-tops" (E_i, F_i) of the bal-results
		related to pivots on a pivot-path stair
s	(30)	$\text{SIZE}(E_1, F_1)$ for the first such (n, s, g)-top
g	(31)	maximal growth-rate of (n, s, g)-tops
d_4	(28)	$d_4 \cdot \text{SIZE}(T_0, U_0)$ bounds the number of crucial segments
d_5	(36)	$d_5 \cdot (1 + \mathcal{E}_{\mathcal{B}_{n,s,g}})$ bounds the length of each crucial segment
c	(44)	$\max\{d_3, 2\, d_4\, d_5\}$, the number c in (8) in Theorem 7

Table 1: Small upper bounds determined by a given grammar $\mathcal{G}$

can be easily shown by contradicting the existence of a violating pair $T \not\sim U$ with the least eq-level.

Theorem 7. *For any grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$ there is a small number c and a computable (not necessarily small) number $\mathcal{E}$ such that for all $T, U \in \text{TERMS}_{\mathcal{N}}$ we have:*

$$\text{if } T \not\sim U \text{ then } \text{EL}(T, U) \leq c \cdot \big(\mathcal{E} \cdot \text{SIZE}(T, U) + (\text{SIZE}(T, U))^2\big). \tag{43}$$

Proof. We fix a grammar $\mathcal{G} = (\mathcal{N}, \Sigma, \mathcal{R})$, which determines the small numbers in Table 1, and two terms T_0, U_0 such that $T_0 \not\sim U_0$. Let

$$\pi_\ell = \mu_0^{\text{CS}} \rho_1' \mu_1^{\text{U}} \mu_1^{\text{US}} \mu_1^{\text{CS}} \rho_2' \mu_2^{\text{U}} \mu_2^{\text{US}} \mu_2^{\text{CS}} \cdots \rho_\ell' \mu_\ell^{\text{U}} \mu_\ell^{\text{US}} \mu_\ell^{\text{CS}}$$

be a respective balanced modified play for which we use the above developed notions and notation; we recall that $\text{LENGTH}(\pi_\ell) = \text{EL}(T_0, U_0)$. Highlighting the crucial segments, we write π_ℓ as

$$\mu_0^{\text{CS}} \big[\rho_{k_1}' \cdots \mu_{k_2-1}^{\text{US}}\big] \mu_{k_2-1}^{\text{CS}} \big[\rho_{k_2}' \cdots \mu_{k_3-1}^{\text{US}}\big] \mu_{k_3-1}^{\text{CS}} \cdots \cdots \big[\rho_{k_p}' \cdots \mu_\ell^{\text{US}}\big] \mu_\ell^{\text{CS}}.$$

We have $p = 0$ (and $\ell = 0$) if $\pi_\ell = \mu_0^{\text{CS}}$; otherwise $1 = k_1 < k_2 < k_3 \cdots < k_p < k_{p+1} = \ell + 1$. The close sink-segments $\mu_{k_j-1}^{\text{CS}}$, for $j \in [1, p+1]$, might be empty or nonempty, but all close sink-segments inside the crucial segments are empty. The total length of the close sink-segments is bounded by $d_3 \cdot (\text{SIZE}(T_0, U_0))^2$ (by Proposition 16), the number p of the crucial segments is bounded by $d_4 \cdot \text{SIZE}(T_0, U_0)$ (by Proposition 18), and the length of each crucial segment is bounded by $d_5 \cdot (1 + \mathcal{E}_{\mathcal{B}_{n,s,g}})$ (by Corollary 20 and Proposition 21).

Hence $\text{LENGTH}(\pi_\ell)$ (and thus $\text{EL}(T_0, U_0)$) is bounded by

$$d_3 \cdot (\text{SIZE}(T_0, U_0))^2 + d_4 \cdot \text{SIZE}(T_0, U_0) \cdot d_5 \cdot (1 + \mathcal{E}_{\mathcal{B}_{n,s,g}}).$$

Putting

$$c = \max \big\{ d_3 , \, 2 \cdot d_4 \cdot d_5 \big\}, \tag{44}$$

and recalling that $\mathcal{E}_\mathcal{B} \geq 1$ for any (n, s, g)-candidate $\mathcal{B}$, we get

$$\text{EL}(T_0, U_0) \leq c \cdot \big(\mathcal{E}_{\mathcal{B}_{n,s,g}} \cdot \text{SIZE}(T_0, U_0) + (\text{SIZE}(T_0, U_0))^2 \big).$$

It remains to show that $\mathcal{E}_{\mathcal{B}_{n,s,g}}$ is computable. We first recall that $\mathcal{E}_{\mathcal{B}_{n,s,g}}$ in the bound $d_5 \cdot (1 + \mathcal{E}_{\mathcal{B}_{n,s,g}})$ on the length of each crucial segment can be refined, as stated in Corollary 20. For all terms T, U we thus get the following implication:

$$\text{if } T \not\sim U, \text{ then } \text{EL}(T, U) \leq c \cdot \big(\mathcal{E}_\mathcal{B} \cdot \text{SIZE}(T, U) + (\text{SIZE}(T, U))^2 \big) \tag{45}$$

for any (n, s, g)-candidate $\mathcal{B}$ that is full below $\text{EL}(T, U)$. (In this case $\mathcal{B}$ is surely full below $1 + \text{EL}(E\sigma, F\sigma)$ for the first, and each further, bal-result $(E\sigma, F\sigma)$ in any balanced modified play from (T, U), if there is any balancing step there at all.)

For $k \in \mathbb{N}$ we define the (reflexive and symmetric) relation $\approx_k$ on $\text{TERMS}_\mathcal{N}$ as follows:

$$T \approx_k U \Leftrightarrow_{df} \text{EL}(T, U) > c \cdot \big(k \cdot \text{SIZE}(T, U) + (\text{SIZE}(T, U))^2 \big);$$

hence $\sim \, \subseteq \, \approx_k$ for all $k \in \mathbb{N}$. We say that an (n, s, g)-candidate $\mathcal{B}$ is *k-sound* (for $k \in \mathbb{N}$) if $(\text{PAIRS}_{n,s} \smallsetminus \mathcal{B}) \subseteq \, \approx_k$ and, moreover, in the case $n > 0$ the $(n-1, s', g)$-candidate $\mathcal{B}'$ is k-sound (we use the notation (10)). An (n, s, g)-*candidate* $\mathcal{B}$ is *sound* if it is $\mathcal{E}_\mathcal{B}$-sound. We note that the full candidate $\mathcal{B}_{n,s,g}$ is sound (since all relevant pairs outside $\mathcal{B}_{n,s,g}$ are in $\sim$, and thus in $\approx_k$ for all k).

There is an obvious algorithm that constructs a sound (n, s, g)-candidate $\mathcal{B}$, for the above defined small n, s, g, and c. (Just a systematic brute-force search would do.)

We will now observe that for each sound (n, s, g)-candidate $\mathcal{B}$ we have $\approx_{\mathcal{E}_\mathcal{B}} \, = \, \sim$ (on the set $\text{TERMS}_\mathcal{N}$), and thus $\mathcal{B} = \mathcal{B}_{n,s,g}$; by this the proof will be finished. For the sake of contradiction we suppose a sound (n, s, g)-candidate $\mathcal{B}$ and some $(T, U) \in \, \approx_{\mathcal{E}_\mathcal{B}} \cap \not\sim$ where $\text{EL}(T, U)$ is the least possible. Then $\mathcal{B}$ is full below $\text{EL}(T, U)$ (for any (T', U') with $\text{EL}(T', U') < \text{EL}(T, U)$ we have $T' \not\approx_{\mathcal{E}_\mathcal{B}} U'$, hence all relevant (T', U') with $\text{EL}(T', U') < \text{EL}(T, U)$ must be in $\mathcal{B}$ since $\mathcal{B}$ is sound). But then (45), applied to our $T, U, \mathcal{B}$, contradicts the assumption $T \approx_{\mathcal{E}_\mathcal{B}} U$. $\qquad\square$

References

[1] G. Sénizergues, L(A)=L(B)? Decidability results from complete formal systems, Theor. Comput. Sci. 251 (1–2) (2001) 1–166.

[2] R. Milner, Communication and concurrency, Prentice-Hall, Inc., 1989.

[3] J. Baeten, J. Bergstra, J. Klop, Decidability of bisimulation equivalence for processes generating context-free languages, J.ACM 40 (3) (1993) 653–682.

[4] J. Srba, Roadmap of infinite results, in: Current Trends In Theoretical Computer Science, The Challenge of the New Century, Vol. 2, World Scientific Publishing Co., 2004, pp. 337–350, (updated version at http://users-cs.au.dk/srba/roadmap/).

[5] G. Sénizergues, The bisimulation problem for equational graphs of finite out-degree, SIAM J.Comput. 34 (5) (2005) 1025–1106, (preliminary version at FOCS'98).

[6] S. Schmitz, Complexity hierarchies beyond elementary, TOCT 8 (1) (2016) 3.

[7] C. Stirling, Deciding DPDA equivalence is primitive recursive, in: Proc. ICALP'02, Vol. 2380 of LNCS, Springer, 2002, pp. 821–832.

[8] P. Jančar, Equivalences of pushdown systems are hard, in: Proc. FoSSaCS'14, Vol. 8412 of LNCS, Springer, 2014, pp. 1–28.

[9] M. Benedikt, S. Göller, S. Kiefer, A. S. Murawski, Bisimilarity of pushdown automata is nonelementary, in: Proc. LICS 2013, IEEE Computer Society, 2013, pp. 488–498.

[10] D. Caucal, Bisimulation of context-free grammars and of pushdown automata, in: A. Ponse, M. de Rijke, Y. Venema (Eds.), CSLI volume 53 "Modal logic and process algebra", Stanford, 1995, pp. 85–106.

[11] B. Courcelle, Recursive applicative program schemes, in: Handbook of Theoretical Computer Science, vol. B, Elsevier, MIT Press, 1990, pp. 459–492.

[12] P. Jančar, Decidability of DPDA language equivalence via first-order grammars, in: Proc. LICS 2012, IEEE Computer Society, 2012, pp. 415–424.

[13] P. Jančar, Bisimulation equivalence of first-order grammars, in: Proc. ICALP'14 (II), Vol. 8573 of LNCS, Springer, 2014, pp. 232–243.

[14] C. Stirling, Decidability of DPDA equivalence, Theor. Comput. Sci. 255 (1–2) (2001) 1–31.

[15] S. Kiefer, BPA bisimilarity is EXPTIME-hard, Inf. Proc. Letters 113 (4) (2013) 101–106.

[16] O. Burkart, D. Caucal, B. Steffen, An elementary bisimulation decision procedure for arbitrary context-free processes, in: Proc. of MFCS'95, Vol. 969 of LNCS, Springer, 1995, pp. 423–433.

[17] P. Jančar, Bisimilarity on basic process algebra is in 2-ExpTime (an explicit proof), Logical Methods in Computer Science 9 (1:10) (2013) 1–19.

[18] Y. Hirshfeld, M. Jerrum, F. Moller, A polynomial algorithm for deciding bisimilarity of normed context-free processes, Theor. Comput. Sci. 158 (1996) 143–159. `doi:10.1016/0304-3975(95)00064-X`.

[19] W. Czerwiński, S. Lasota, Fast equivalence-checking for normed context-free processes, in: Proc. FSTTCS'10, Vol. 8 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2010.

[20] P. Jančar, J. Srba, Undecidability of bisimilarity by Defender's forcing, J. ACM 55 (2008) (1).

[21] Q. Yin, Y. Fu, C. He, M. Huang, X. Tao, Branching bisimilarity checking for PRS, in: Proc. ICALP'14 (II), Vol. 8573 of LNCS, Springer, 2014, pp. 363–374.

[22] C. H. Broadbent, S. Göller, On bisimilarity of higher-order pushdown automata: Undecidability at order two, in: FSTTCS 2012, Vol. 18 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2012, pp. 160–172.

[23] L. Ong, Higher-order model checking: An overview, in: Proc. LICS 2015, IEEE Computer Society, 2015, pp. 1–15.

[24] I. Walukiewicz, Automata theory and higher-order model-checking, ACM SIGLOG News 3 (4) (2016) 13–31.

[25] C. Stirling, Second-order simple grammars, in: Proc. CONCUR'06, Vol. 4137 of LNCS, Springer, 2006, pp. 509–523.

[26] P. Jančar, S. Schmitz, Bisimulation equivalence of first-order grammars is Ackermann-complete, in: 34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019, Vancouver, BC, Canada, June 24-27, 2019, IEEE, 2019, pp. 1–12. `doi:10.1109/LICS.2019.8785848`.

[27] P. Jančar, Deciding semantic finiteness of pushdown processes and first-order grammars w.r.t. bisimulation equivalence, J. Comput. Syst. Sci. 109 (2020) 22–44, (a preliminary version at MFCS'16). `doi:10.1016/j.jcss.2019.10.002`.