

# Approximation of Weighted Automata with Storage

Tobias Denkinger

Faculty of Computer Science, Technische Universität Dresden,  
Nöthnitzer Str. 46, 01062 Dresden, Germany

tobias.denkinger@tu-dresden.de

We use a non-deterministic variant of storage types to develop a framework for the approximation of automata with storage. This framework is used to provide automata-theoretic views on the approximation of multiple context-free languages and on coarse-to-fine parsing.

## 1 Introduction

Formal grammars (e.g. context-free grammars) are used to model natural languages. Language models are often incorporated into systems that have to guarantee a certain response time, e.g. translation systems or speech recognition systems. The desire for low response times and the high parsing complexity of the used formal grammars are at odds. Thus, in real-world applications, the language model is often replaced by another language model that is easier to parse but still captures the desired natural language reasonably well. This new language model is called an *approximation* of the original language model. Nederhof [25] gives an overview for the approximation of context-free grammars.

In order to approximate a context-free grammar it is common (but not exclusive [26, 4]) to first construct an equivalent pushdown automaton and then approximate this automaton [21, 29, 23, 2, 28, 13, 20], e.g. by restricting the height of the pushdown. Automata with storage [31, 14, 10, 11] generalise pushdown automata. By attaching weights to the transitions of an automaton with storage, we can model, e.g. the *multiplicity* with which a word belongs to a language or the *cost* of recognising a word [30, 9]. The resulting devices are called *weighted* automata with storage and were studied in recent literature [15, 34]. Multiple context-free languages (MCFLs) [32, 33] are currently studied as language models because they can express the non-projective constituents and discontinuous dependencies that occur in natural languages [24, 22]. Their approximation was recently investigated from a grammar-centric viewpoint [3, 6]. MCFLs can be captured by automata with specific storage [5, 7], which allows an automata-theoretic view on their approximation.

We develop a framework to study the approximation of weighted automata with arbitrary storage. To deal with non-determinism that arises due to approximation, we use automata with *data storage* [14] which allow instructions to be non-deterministic;<sup>1</sup> and we investigate their relation to automata with storage (Sec. 3). Weighted automata with data storage differ from Engelfriet’s automata with storage [10, 11] in two aspects: As instructions we allow binary relations instead of partial functions and each transition is associated with a weight from a semiring. Using a powerset construction, we show that (weighted) automata with data storage have the same expressive power as (weighted) automata with storage (Props. 9 and 31). Our formalisation of strategies for approximating data storage (called *approximation strategies*) is inspired by the storage simulation of Hoare [17, 12]. We use partial functions as approximation

---

<sup>1</sup>We add predicates to Goldstine’s original definition of data storage. This does not increase their expressiveness (Lem. 8).

strategies (Sec. 4). Properties of the approximation strategy imply properties of the while approximation process: If an approximation strategy is a total function, then we have a superset approximation (Thms. 21 and 34(i)). If an approximation strategy is injective, then we have a subset approximation (Thms. 26 and 34(ii)). In contrast to Engelfriet and Vogler [12], we do not utilise flowcharts in our constructions. We demonstrate the benefit of our framework by providing an automata-based view on the approximation of MCFLs (Sec. 5) and by describing an algorithm for coarse-to-fine parsing of weighted automata with data storage (Sec. 6).

## 2 Preliminaries

The set  $\{0, 1, 2, \dots\}$  of *natural numbers* is denoted by  $\mathbb{N}$ ,  $\mathbb{N} \setminus \{0\}$  is denoted by  $\mathbb{N}_+$ , and  $\{1, \dots, k\}$  is denoted by  $[k]$  for every  $k \in \mathbb{N}$  (note that  $[0] = \emptyset$ ). Let  $A$  be a set. The *power set of  $A$*  is denoted by  $\mathcal{P}(A)$ .

Let  $A$ ,  $B$ , and  $C$  be sets and let  $r \subseteq A \times B$  and  $s \subseteq B \times C$  be binary relations. We denote  $\{(b, a) \in B \times A \mid (a, b) \in r\}$  by  $r^{-1}$ ,  $\{b \in B \mid (a, b) \in r\}$  by  $r(a)$  for every  $a \in A$ , and  $\bigcup_{a \in A'} r(a)$  by  $r(A')$  for every  $A' \subseteq A$ . The *sequential composition of  $r$  and  $s$*  is the binary relation  $r; s = \{(a, c) \in A \times C \mid \exists b \in B: ((a, b) \in r) \wedge ((b, c) \in s)\}$ . We call  $r$  an *endorelation (on  $A$ )* if  $A = B$ . A *semiring* is an algebraic structure  $(K, +, \cdot, 0, 1)$  where  $(K, +, 0)$  is a commutative monoid,  $(K, \cdot, 1)$  is a monoid,  $0$  is absorptive with respect to  $\cdot$ , and  $\cdot$  distributes over  $+$ . We say that  $K$  is *complete* if it has a sum operation  $\sum_I: K^I \rightarrow K$  that extends  $+$  for each countable set  $I$  [8, Sec. 2]. Let  $\leq$  be a partial order on  $K$ . We say that  $K$  is *positively  $\leq$ -ordered* if  $+$  preserves  $\leq$  (i.e. for each  $a, b, c \in K$  with  $a \leq b$  holds  $a + c \leq b + c$ ),  $\cdot$  preserves  $\leq$  (i.e. for each  $a, b, c \in K$  with  $a \leq b$  holds  $a \cdot c \leq b \cdot c$  and  $c \cdot a \leq c \cdot b$ ), and  $0 \leq a$  for each  $a \in K$  (cf. Droste and Kuich [8, Sec. 2]).

The *set of partial functions from  $A$  to  $B$*  is denoted by  $A \dashrightarrow B$ . The *set of (total) functions from  $A$  to  $B$*  is denoted by  $A \rightarrow B$ . Let  $f: A \dashrightarrow B$  be a partial function. The *domain of  $f$*  and the *image of  $f$*  are defined by  $\text{dom}(f) = \{a \in A \mid \exists b \in B: f(a) = b\}$  and  $\text{img}(f) = \{b \in B \mid \exists a \in A: f(a) = b\}$ , respectively. Abusing the notation, we may sometimes write  $f(a) = \text{undefined}$  to denote that  $a \notin \text{dom}(f)$ . Note that every total function is a partial function and that each partial function is a binary relation.

## 3 Automata with data storage

In addition to the finite state control, automata with storage are allowed to check and manipulate a storage configuration that comes from a possibly infinite set. We propose a syntactic extension of automata with storage where the set of unary functions (the *instructions*) is replaced by a set of binary relations on the storage configurations.

### 3.1 Data storage

**Definition 1.** A *data storage* is a tuple  $S = (C, P, R, c_i)$  where  $C$  is a set (of *storage configurations*),  $P \subseteq \mathcal{P}(C)$  (*predicates*),  $R \subseteq \mathcal{P}(C \times C)$  (*instructions*),  $c_i \in C$  (*initial storage configuration*), and the set  $r(c)$  is finite for every  $r \in R$  and  $c \in C$ .  $\square$

Our definition of data storage differs from the original definition [14, Def. 3.1] in that we have predicates. The “data storage types” introduced by Herrmann and Vogler [16, Sec. 3] are similar to our data storages. For instructions they use partial functions that may depend on the input of the automaton in addition to the current storage configuration instead of binary relations on storage configurations.

Consider a data storage  $S = (C, P, R, c_i)$ . If every element of  $R$  is a partial function, we call  $S$  *deterministic*. The definition of “deterministic data storage” in this paper coincides with the definition of “storage type” in previous literature [15, 34].

**Example 2.** The deterministic data storage Count models simple counting (Engelfriet [10, 11, Def. 3.4]):  $\text{Count} = (\mathbb{N}, \{\mathbb{N}, \mathbb{N}_+, \{0\}\}, \{\text{inc}, \text{dec}\}, 0)$  where  $\text{inc} = \{(n, n+1) \mid n \in \mathbb{N}\}$  and  $\text{dec} = \text{inc}^{-1}$ .  $\square$

**Example 3.** The following deterministic data storage models pushdown storage:<sup>2</sup>  $\text{PD}_\Gamma = (\Gamma^*, P_{\text{pd}}, R_{\text{pd}}, \varepsilon)$  where  $\Gamma$  is a nonempty finite set (*pushdown symbols*);  $P_{\text{pd}} = \{\Gamma^*, \text{bottom}\} \cup \{\text{top}_\gamma \mid \gamma \in \Gamma\}$  with  $\text{bottom} = \{\varepsilon\}$  and  $\text{top}_\gamma = \{\gamma w \mid w \in \Gamma^*\}$  for every  $\gamma \in \Gamma$ ; and  $R_{\text{pd}} = \{\text{stay}, \text{pop}\} \cup \{\text{push}_\gamma \mid \gamma \in \Gamma\} \cup \{\text{stay}_\gamma \mid \gamma \in \Gamma\}$  with  $\text{stay} = \{(w, w) \mid w \in \Gamma^*\}$ ,  $\text{pop} = \{(\gamma w, w) \mid w \in \Gamma^*, \gamma \in \Gamma\}$ ,  $\text{push}_\gamma = \{(w, \gamma w) \mid w \in \Gamma^*\}$ , and  $\text{stay}_\gamma = \{(\gamma' w, \gamma w) \mid w \in \Gamma^*, \gamma' \in \Gamma\}$  for every  $\gamma \in \Gamma$ .  $\square$

We call a data storage  $S = (C, P, R, c_i)$  *boundedly non-deterministic* (*short: boundedly nd*) if there is a natural number  $k$  such that  $|r(c)| \leq k$  holds for every  $r \in R$  and  $c \in C$ . The following two examples illustrate that each deterministic data storage is also boundedly nd, but not vice versa.

**Example 4.**  $\text{PD}'_\Gamma$  extends  $\text{PD}_\Gamma$  (cf. Ex. 3) by adding an instruction  $\text{pop}^*$  that allows us to remove arbitrarily many symbols from the top of the pushdown:  $\text{PD}'_\Gamma = (\Gamma^*, P_{\text{pd}}, R_{\text{pd}} \cup \{\text{pop}^*\}, \varepsilon)$  where  $\text{pop}^* = \{(uw, w) \mid u, w \in \Gamma^*\}$ .

The tuple  $\text{PD}'_\Gamma$  is a data storage because  $|\text{stay}(w)| = 1$ ,  $|\text{pop}(w)| \leq 1$ ,  $|\text{push}_\gamma(w)| = 1$ ,  $|\text{stay}_\gamma(w)| \leq 1$ , and  $|\text{pop}^*(w)| = |w| + 1$  for each  $w \in \Gamma^*$  and  $\gamma \in \Gamma$  are all finite. But  $\text{PD}'_\Gamma$  is not boundedly nd. Assume that it were. Then there would be a number  $k \in \mathbb{N}$  such that  $|r(w)| \leq k$  for every  $r \in R_{\text{pd}}$  and  $w \in \Gamma^*$ . But if we take some  $w' \in \Gamma^*$  of length  $k$ , then  $|\text{pop}^*(w')| = k + 1 > k$  which contradicts our assumption.  $\square$

**Example 5.** The data storage  $\text{PD}''_\Gamma$  extends  $\text{PD}_\Gamma$  (cf. Ex. 3) by adding an instruction  $\text{push}_\Gamma$  that allows us to add an arbitrary symbol from  $\Gamma$  the top of the pushdown:  $\text{PD}''_\Gamma = (\Gamma^*, P_{\text{pd}}, R_{\text{pd}} \cup \{\text{push}_\Gamma\}, \varepsilon)$  where  $\text{push}_\Gamma = \{(w, w\gamma) \mid w \in \Gamma^*, \gamma \in \Gamma\}$ .

The data storage  $\text{PD}''_\Gamma$  is boundedly nd because if we take bound  $k = |\Gamma|$ , then  $|\text{stay}(w)| = 1 \leq k$ ,  $|\text{pop}(w)| \leq 1 \leq k$ ,  $|\text{push}_\gamma(w)| = 1 \leq k$ ,  $|\text{stay}_\gamma(w)| \leq 1 \leq k$ , and  $|\text{push}_\Gamma(w)| = |\Gamma| \leq k$ . In particular, if  $|\Gamma| > 1$ , then  $\text{PD}''_\Gamma$  is not deterministic because  $|\text{push}_\Gamma(w)| = |\Gamma| > 1$ .  $\square$

## 3.2 Automata with data storage

For the rest of this paper let  $\Sigma$  be an arbitrary non-empty finite set.

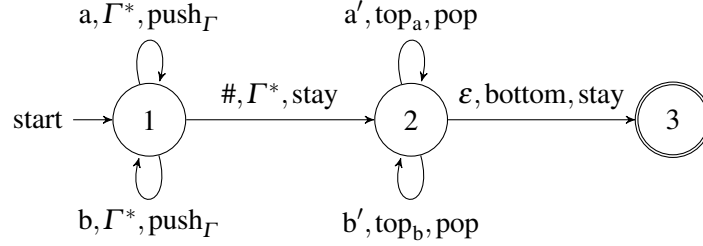
**Definition 6.** Let  $S = (C, P, R, c_i)$  be a data storage. An  $(S, \Sigma)$ -automaton is a tuple  $\mathcal{M} = (Q, T, Q_i, Q_f)$  where  $Q$  is a finite set (of *states*),  $T$  is a finite subset of  $Q \times (\Sigma \cup \{\varepsilon\}) \times P \times R \times Q$  (*transitions*),  $Q_i \subseteq Q$  (*initial states*), and  $Q_f \subseteq Q$  (*final states*).  $\square$

Let  $\mathcal{M} = (Q, T, Q_i, Q_f)$  be an  $(S, \Sigma)$ -automaton and  $S = (C, P, R, c_i)$ . An  $\mathcal{M}$ -configuration is an element of  $Q \times C \times \Sigma^*$ . For every  $\tau = (q, v, p, r, q') \in T$ , the *transition relation of  $\tau$*  is the endorelation  $\vdash_\tau$  on the set of  $\mathcal{M}$ -configurations that contains  $(q, c, vw) \vdash_\tau (q', c', w)$  for every  $w \in \Sigma^*$  and  $(c, c') \in r$  with  $c \in p$ . The *run relation of  $\mathcal{M}$*  is  $\vdash_{\mathcal{M}} = \bigcup_{\tau \in T} \vdash_\tau$ . The transition relations are extended to sequences of transitions by setting  $\vdash_{\tau_1 \dots \tau_k} = \vdash_{\tau_1}; \dots; \vdash_{\tau_k}$  for every  $k \in \mathbb{N}$  and  $\tau_1, \dots, \tau_k \in T$ . In particular, for the case  $k = 0$  we use the identity on  $Q \times C \times \Sigma^*$ :  $\vdash_\varepsilon = \{(d, d) \mid d \in Q \times C \times \Sigma^*\}$ . The *set of runs of  $\mathcal{M}$*  is the set

$$\mathbf{R}_{\mathcal{M}} = \{\theta \in T^* \mid \exists q, q' \in Q, c, c' \in C, w, w' \in \Sigma^*: (q, c, w) \vdash_\theta (q', c', w')\}. \quad (1)$$

Let  $w \in \Sigma^*$ . The *set of runs of  $\mathcal{M}$  on  $w$*  is  $\mathbf{R}_{\mathcal{M}}(w) = \{\theta \in T^* \mid \exists q \in Q_i, q' \in Q_f, c' \in C: (q, c_i, w) \vdash_\theta (q', c', \varepsilon)\}$ . The *language accepted by  $\mathcal{M}$*  is the set  $L(\mathcal{M}) = \{w \in \Sigma^* \mid \mathbf{R}_{\mathcal{M}}(w) \neq \emptyset\}$ . Let  $S$  be a data storage and  $L \subseteq \Sigma^*$ . We call  $L$   $(S, \Sigma)$ -recognisable if there is an  $(S, \Sigma)$ -automaton  $\mathcal{M}$  with  $L = L(\mathcal{M})$ .

<sup>2</sup>We allow (in comparison to Engelfriet [10, 11, Def. 3.2]) the execution of (some) instructions on the empty pushdown.

Figure 1: Graph of the  $(PD_{\Gamma}^{\prime}, \Sigma)$ -automaton  $\mathcal{M}$  from Ex. 7

**Example 7.** Recall the data storage  $PD_{\Gamma}^{\prime}$  from Ex. 5. Let  $\Sigma = \{a, b, \#, a', b'\}$  and  $\Gamma = \{a, b\}$ , and consider the  $(PD_{\Gamma}^{\prime}, \Sigma)$ -automaton  $\mathcal{M} = ([3], T, \{1\}, \{3\})$  where

$$T: \begin{array}{lll} (1, a, \Gamma^*, \text{push}_{\Gamma}, 1) & (1, b, \Gamma^*, \text{push}_{\Gamma}, 1) & (1, \#, \Gamma^*, \text{stay}, 2) \\ (2, a', \text{top}_a, \text{pop}, 2) & (2, b', \text{top}_b, \text{pop}, 2) & (2, \epsilon, \text{bottom}, \text{stay}, 3). \end{array}$$

The graph of  $\mathcal{M}$  is shown in Fig. 1. The label of each edge in the graph contains the input that is read by the corresponding transition, the predicate that is checked, and the instruction that is executed. The language recognised by  $\mathcal{M}$  is  $L(\mathcal{M}) = \{u\#v \mid u \in \{a, b\}^*, v \in \{a', b'\}^*, |u| = |v|\}$ . The automaton  $\mathcal{M}$  recognises a given word  $u\#v$  (with  $u \in \{a, b\}^*$  and  $v \in \{a', b'\}^*$ ) as follows: In state 1, it reads the prefix  $u$  and constructs any element of  $\Gamma^*$  of length  $|u|$  on the pushdown non-deterministically. It then reads  $\#$  and goes to state 2. In state 2, it reads  $a'$  for each  $a$  on the pushdown and it reads  $b'$  for each  $b$  on the pushdown until the pushdown is empty. Since the pushdown can contain any sequence over  $\{a, b\}$  of length  $|u|$ ,  $\mathcal{M}$  can read any sequence of  $\{a', b'\}$  of length  $|u|$ , ensuring that  $|u| = |v|$ .  $\square$

We call a data storage  $S = (C, P, R, c_i)$  *predicate-free* if  $P = \{C\}$ .<sup>3</sup> The following lemma shows that predicate-free-ness is a normal form among data storages.

**Lemma 8.** For every data storage  $S$  there is a predicate-free data storage  $S'$  such that the classes of  $(S, \Sigma)$ -recognisable languages and the class of  $(S', \Sigma)$ -recognisable languages are the same.

*Proof idea.* Encode the predicates of  $S$  in the instructions of  $S'$ .  $\blacksquare$

**Proposition 9.** For every data storage  $S$  there is a deterministic data storage  $\det(S)$  such that the class of  $(S, \Sigma)$ -recognisable languages is equal to the class of  $(\det(S), \Sigma)$ -recognisable languages.

*Proof.* Due to Lem. 8 we can assume that  $S$  is predicate-free. Thus, let  $S = (C, \{C\}, R, c_i)$ . Using a power set construction, we obtain the deterministic data storage  $\det(S) = (\mathcal{P}(C), \{\mathcal{P}(C)\}, \det(R), \{c_i\})$  where  $\det(R) = \{\det(r) \mid r \in R\}$  with  $\det(r) = \{(d, r(d)) \mid d \subseteq C, r(d) \neq \emptyset\}$  for every  $r \in R$ .

Let  $\mathcal{M} = (Q, T, Q_i, Q_f)$  be an  $(S, \Sigma)$ -automaton and  $\mathcal{M}' = (Q, T', Q_i, Q_f)$  be a  $(\det(S), \Sigma)$ -automaton. We say that  $\mathcal{M}$  and  $\mathcal{M}'$  are *related* if  $T' = \det(T) = \{\det(\tau) \mid \tau \in T\}$  with  $\det(\tau) = (q, v, \mathcal{P}(C), \det(r), q')$  for each  $\tau = (q, v, C, r, q') \in T$ . Clearly, for every  $(S, \Sigma)$ -automaton there is an  $(\det(S), \Sigma)$ -automaton such that both are related, and vice versa.

Now let  $\mathcal{M} = (Q, T, Q_i, Q_f)$  be an  $(S, \Sigma)$ -automaton and  $\mathcal{M}' = (Q, \det(T), Q_i, Q_f)$  be a  $(\det(S), \Sigma)$ -automaton. Note that  $\mathcal{M}$  and  $\mathcal{M}'$  are related. We extend  $\det: T \rightarrow \det(T)$  to a function  $\det: T^* \rightarrow$

<sup>3</sup>Even though  $S$  has a predicate  $C$ , we still call it predicate-free since  $C$  is trivial, i.e.  $C$  accepts any storage configuration.

$(\det(T))^*$  by point-wise application. We can show for every  $\theta \in T^*$  by induction on the length of  $\theta$  that

$$\forall q, q' \in Q, c, c' \in C, w, w' \in \Sigma^*: (q, c, w) \vdash_{\theta} (q', c', w') \iff \forall d \ni c: \exists d' \ni c': (q, d, w) \vdash_{\det(\theta)} (q', d', w') \quad (2)$$

holds. We obtain  $L(\mathcal{M}) = L(\mathcal{M}')$  from (2) and since  $\{c_i\}$  is the initial storage configuration of  $\mathcal{M}'$ . ■

For practical reasons it might be preferable to avoid the construction of power sets. The proof of the following Proposition shows a construction for boundedly nd data storages.

**Proposition 10.** Let  $S = (C, P, R, c_i)$  be a boundedly nd data storage. There is a deterministic data storage  $S'$  with the same set of storage configurations such that the class of  $(S, \Sigma)$ -recognisable languages is contained in the class of  $(S', \Sigma)$ -recognisable languages.

*Proof.* We construct the deterministic data storage  $S' = (C, P, R', c_i)$  where  $R'$  is constructed as follows: Let  $r \in R$  and  $r(c)_1, \dots, r(c)_{m_{r,c}}$  be a fixed enumeration of the elements of  $r(c)$  for every  $c \in C$ . Furthermore, let  $k = \max\{|r(c)| \mid r \in R, c \in C\}$ . Since  $S$  is boundedly nd, the number  $k$  is well defined. We define for each  $i \in [k]$  an instruction  $r'_i$  by  $r'_i(c) = r(c)_i$  if  $i \leq m_{r,c}$  and  $r'_i(c) = \text{undefined}$  otherwise. Let  $R'$  contain the instruction  $r'_i$  for every  $r \in R$  and  $i \in [k]$ . Now let  $\mathcal{M} = (Q, T, Q_i, Q_f)$  be an  $(S, \Sigma)$ -automaton. We construct the  $(S', \Sigma)$ -automaton  $\mathcal{M}' = (Q, T', Q_i, Q_f)$  where  $T'$  contains for every transition  $t = (q, v, p, r, q') \in T$  and  $i \in [k]$  the transition  $t'_i = (q, v, p, r'_i, q')$ . Then  $\vdash_{\mathcal{M}} = \bigcup_{t \in T} \vdash_t = \bigcup_{t=(q,v,p,r,q') \in T} \bigcup_{i \in [k]} \vdash_{t'_i} = \bigcup_{t' \in T'} \vdash_{t'} = \vdash_{\mathcal{M}'}$  and thus  $L(\mathcal{M}) = L(\mathcal{M}')$ . ■

The above construction fails for data storages that are not boundedly nd. Consider the data storage  $\text{PD}'_{\Gamma}$  from Ex. 4. Then there exists no bound  $k_{\text{pop}^*} \in \mathbb{N}$  as would be required in the proof.

The containment shown in Prop. 10 is strict as the following example reveals.

**Example 11** (due to Nederhof [27]). Recall the data storage  $\text{PD}''_{\Gamma}$  from Ex. 5. Consider the similar data storage  $\text{PD}^{\dagger}_{\Gamma} = (\Gamma^*, \{\Gamma^*, \text{bottom}\}, \{\text{stay}, \text{push}_{\Gamma}\} \cup \{\text{pop}_{\gamma} \mid \gamma \in \Gamma\}, \varepsilon)$  where  $\text{pop}_{\gamma} = \{(\gamma w, w) \mid \gamma \in \Gamma, w \in \Gamma^*\}$  for each  $\gamma \in \Gamma$ . We can again think of  $\Gamma^*$  as a pushdown. Now, starting from  $\text{PD}^{\dagger}_{\Gamma}$ , we construct the deterministic data storage  $(\text{PD}^{\dagger}_{\Gamma})'$  by the construction given in Prop. 10. We thereby obtain  $(\text{PD}^{\dagger}_{\Gamma})' = (\Gamma^*, \{\Gamma^*, \text{bottom}\}, \{\text{stay}\} \cup \{\text{push}_{\gamma} \mid \gamma \in \Gamma\} \cup \{\text{pop}_{\gamma} \mid \gamma \in \Gamma\}, \varepsilon)$ . The only difference between  $\text{PD}^{\dagger}_{\Gamma}$  and  $(\text{PD}^{\dagger}_{\Gamma})'$  is that the instruction  $\text{push}_{\Gamma}$  is replaced by the  $|\Gamma|$  instructions in the set  $\{\text{push}_{\gamma} \mid \gamma \in \Gamma\}$ .

Now consider the sets  $\Sigma = \{a, b\}$  and  $\Gamma = \Sigma$ , and the language  $L = \{ww^R \mid w \in \Sigma^*\} \subseteq \Sigma^*$  where  $w^R$  denotes the reverse of  $w$  for each  $w \in \Sigma^*$ . The following  $((\text{PD}^{\dagger}_{\Gamma})', \Sigma)$ -automaton  $\mathcal{M}'$  recognises  $L$  and thus demonstrates that  $L$  is  $((\text{PD}^{\dagger}_{\Gamma})', \Sigma)$ -recognisable:  $\mathcal{M}' = ([3], T', \{1\}, \{3\})$  with

$$T': \begin{array}{lll} (1, a, \Gamma^*, \text{push}_a, 1) & (1, b, \Gamma^*, \text{push}_b, 1) & (1, \varepsilon, \Gamma^*, \text{stay}, 2) \\ (2, a, \Gamma^*, \text{pop}_a, 2) & (2, b, \Gamma^*, \text{pop}_b, 2) & (2, \varepsilon, \text{bottom}, \text{stay}, 3). \end{array}$$

In state 1,  $\mathcal{M}'$  stores the input in reverse on the pushdown until it decides non-deterministically go to state 2. In state 2,  $\mathcal{M}'$  accepts the sequence of symbols that is stored on the pushdown. We can only enter the final state 3 if the pushdown is empty, thus  $\mathcal{M}'$  recognises  $L$ .

On the other hand, there is no  $(\text{PD}^{\dagger}_{\Gamma}, \Sigma)$ -automaton  $\mathcal{M}$  that recognises  $L$ . Assume that some  $(\text{PD}^{\dagger}_{\Gamma}, \Sigma)$ -automaton  $\mathcal{M}$  recognises  $L$ . Then  $\mathcal{M}$  would have to encode the first half of the input in the pushdown since this unbounded information can not be stored in the states. The only instruction that adds information to the pushdown is  $\text{push}_{\Gamma}$ . Thus, in the first half of the input, whenever we read the symbol  $a$ , we have to execute  $\text{push}_{\Gamma}$ ; and whenever we read the symbol  $b$ , we also have to execute  $\text{push}_{\Gamma}$ . This offers no means of distinguishing the two situations (reading symbol  $a$  and reading symbol  $b$ ) and hence no means of encoding the first half of the input in the pushdown. □

**Proposition 12.** Let  $S = (C, P, R, c_i)$  be a data storage and  $L$  be an  $(S, \Sigma)$ -recognisable language. If  $C$  is finite, then  $L$  is recognisable (by a finite state automaton).

*Proof.* We will use a product construction. In particular, the states of the constructed finite state automaton are elements of  $Q \times C$ . For this we employ non-deterministic finite-state automata with extended transition function (short: fsa) from Hopcroft and Ullman [18, Sec. 2.3] in a notation similar to that of automata with storage. (We simply leave out the storage-related parts of the transitions.)

Let  $\mathcal{M} = (Q, T, Q_i, Q_f)$ . We construct the fsa  $\mathcal{M}' = (Q \times C, \Sigma, T', Q_i \times \{c_i\}, Q_f \times C)$  where  $T' = \{((q, c), v, (q', c')) \mid (q, v, p, r, q') \in T, (c, c') \in r, c \in p\}$ . We can show

$$\forall q, q' \in Q, c, c' \in C, w, w' \in \Sigma^*: (q, c, w) \vdash_{\mathcal{M}}^* (q', c', w') \iff ((q, c), w) \vdash_{\mathcal{M}'}^* ((q', c'), w'). \quad (3)$$

by straight-forward induction on the length of runs. Using (1) and (3), we then derive  $L(\mathcal{M}) = L(\mathcal{M}')$ .  $\blacksquare$

## 4 Approximation of automata with data storage

An approximation strategy maps a data storage to another data storage. It is specified in terms of storage configurations and naturally extended to predicates and instructions.

**Definition 13.** Let  $S = (C, P, R, c_i)$  be a data storage. An *approximation strategy* is a partial function  $A: C \dashrightarrow C'$  for some set  $C'$ . We call  $A$  *S-proper* if  $(A^{-1}; r; A)(c')$  is finite for every  $r \in R$  and  $c' \in C'$ .  $\square$

**Definition 14.** Let  $S = (C, P, R, c_i)$  be a data storage and  $A: C \dashrightarrow C'$  be an *S-proper* approximation strategy. The *approximation of  $S$  with respect to  $A$*  is the data storage  $A(S) = (C', A(P), A(R), A(c_i))$  where  $A(P) = \{A(p) \mid p \in P\}$  with  $A(p) = \{A(c) \mid c \in p\}$  for every  $p \in P$ , and  $A(R) = \{A(r) \mid r \in R\}$  with  $A(r) = A^{-1}; r; A$  for every  $r \in R$ .  $\square$

**Example 15.** Consider the approximation strategy  $A_o: \mathbb{N} \rightarrow \{\text{odd}\} \cup \{2n \mid n \in \mathbb{N}\}$  that assigns to every odd number the value odd and to every even number the number itself. Then  $A_o$  is *not* Count-proper since  $(A_o^{-1}; \text{inc}; A_o)(\text{odd}) = (A_o^{-1}; \text{dec}; A_o)(\text{odd}) = \{2n \mid n \in \mathbb{N}\}$  is not finite.

On the other hand, consider the approximation strategy  $A_{eo}: \mathbb{N} \rightarrow \{\text{even}, \text{odd}\}$  that returns odd for every odd number and even otherwise. Then  $A_{eo}$  is Count-proper since  $(A_{eo}^{-1}; \text{inc}; A_{eo})(\text{even}) = \{\text{odd}\} = (A_{eo}^{-1}; \text{dec}; A_{eo})(\text{even})$  and  $(A_{eo}^{-1}; \text{inc}; A_{eo})(\text{odd}) = \{\text{even}\} = (A_{eo}^{-1}; \text{dec}; A_{eo})(\text{odd})$  are finite.  $\square$

**Definition 16.** Let  $\mathcal{M} = (Q, T, Q_i, Q_f)$  be an  $(S, \Sigma)$ -automaton and  $A$  an *S-proper* approximation strategy. The *approximation of  $\mathcal{M}$  with respect to  $A$*  is the  $(A(S), \Sigma)$ -automaton  $A(\mathcal{M}) = (Q, A(T), Q_i, Q_f)$  where  $A(T) = \{A(\tau) \mid \tau \in T\}$  and  $A(\tau) = (q, v, A(p), A(r), q')$  for each  $\tau = (q, v, p, r, q') \in T$ .  $\square$

**Example 17.** Let  $\Sigma = \{a, b\}$ . Consider the  $(\text{Count}, \Sigma)$ -automaton  $\mathcal{M} = ([3], T, \{1\}, \{3\})$  and its approximation  $A_{eo}(\mathcal{M}) = ([3], A_{eo}(T), \{1\}, \{3\})$  with

$$\begin{array}{ll} T: \tau_1 = (1, a, \mathbb{N}, \text{inc}, 1) & A_{eo}(T): \tau'_1 = (1, a, A_{eo}(\mathbb{N}), A_{eo}(\text{inc}), 1) \\ \tau_2 = (1, b, \mathbb{N}, \text{dec}, 2) & \tau'_2 = (1, b, A_{eo}(\mathbb{N}), A_{eo}(\text{dec}), 2) \\ \tau_3 = (2, b, \mathbb{N}, \text{dec}, 2) & \tau'_3 = (2, b, A_{eo}(\mathbb{N}), A_{eo}(\text{dec}), 2) \\ \tau_4 = (2, \varepsilon, \{0\}, \text{inc}, 3) & \tau'_4 = (2, \varepsilon, A_{eo}(\{0\}), A_{eo}(\text{inc}), 3) \end{array}$$

where  $A_{eo}(\mathbb{N}) = A_{eo}(\mathbb{N}_+) = \{\text{even}, \text{odd}\}$  and  $A_{eo}(\{0\}) = \{\text{even}\}$  are the predicates of  $A_{eo}(\text{Count})$ , and  $A_{eo}(\text{inc}) = A_{eo}(\text{dec}) = \{(\text{even}, \text{odd}), (\text{odd}, \text{even})\}$  is the instruction of  $A_{eo}(\text{Count})$ . The word  $\text{aabb} \in \{a, b\}^*$  is recognised by both automata:

$$\begin{array}{cccccc} (1, 0, \text{aabb}) \vdash_{\tau_1} (1, 1, \text{abb}) & \vdash_{\tau_1} (1, 2, \text{bb}) & \vdash_{\tau_2} (2, 1, \text{b}) & \vdash_{\tau_3} (2, 0, \varepsilon) & \vdash_{\tau_4} (3, 1, \varepsilon) \\ (1, \text{even}, \text{aabb}) \vdash_{\tau'_1} (1, \text{odd}, \text{abb}) & \vdash_{\tau'_1} (1, \text{even}, \text{bb}) & \vdash_{\tau'_2} (2, \text{odd}, \text{b}) & \vdash_{\tau'_3} (2, \text{even}, \varepsilon) & \vdash_{\tau'_4} (3, \text{odd}, \varepsilon). \end{array}$$

On the other hand, the word  $\text{bb}$  can be recognised by  $A_{\text{eo}}(\mathcal{M})$  but not by  $\mathcal{M}$ :

$$(1, \text{even}, \text{bb}) \vdash_{\tau'_2} (2, \text{odd}, \text{b}) \vdash_{\tau'_3} (2, \text{even}, \varepsilon) \vdash_{\tau'_4} (3, \text{odd}, \varepsilon). \quad \square$$

**Observation 18.** Let  $S = (C, P, R, c_i)$ ,  $\mathcal{M}$  be an  $(S, \Sigma)$ -automaton, and  $A_1: C \dashrightarrow \bar{C}$  and  $A_2: \bar{C} \dashrightarrow C'$  be approximation strategies. If  $A_1$  is  $S$ -proper and  $A_2$  is  $A_1(S)$ -proper, then  $A_2(A_1(\mathcal{M})) = (A_1; A_2)(\mathcal{M})$ . ■

We call an approximation strategy *total* if it is a total function and we call it *injective* if it is an injective partial function. The distinction between *total* and *injective* approximation strategies allows us to define two preorders on approximation strategies (Def. 19) and provides us with simple criteria to ensure that an approximation strategy leads to a superset (Thm. 21) or a subset approximation (Thm. 26).

**Definition 19.** Let  $A_1: C \dashrightarrow C_1$  and  $A_2: C \dashrightarrow C_2$  be approximation strategies. We call  $A_1$  *finer than*  $A_2$ , denoted by  $A_1 \preceq A_2$ , if there is a total approximation strategy  $A: C_1 \rightarrow C_2$  with  $A_1; A = A_2$ . We call  $A_1$  *less partial than*  $A_2$ , denoted by  $A_1 \sqsubseteq A_2$ , if there is an injective approximation strategy  $A: C_1 \dashrightarrow C_2$  with  $A_1; A = A_2$ . ■

## 4.1 Superset approximations

In this section we will show that total approximation strategies (i.e. total functions) lead to superset approximations.

**Lemma 20.** Let  $\mathcal{M} = (Q, T, Q_i, Q_f)$  be an  $(S, \Sigma)$ -automaton,  $S = (C, P, R, c_i)$ , and  $A$  be an  $S$ -proper total approximation strategy. We extend  $A: T \rightarrow A(T)$  to sequences of transitions by point-wise application. Then for each  $\theta \in T^*$ ,  $q, q' \in Q$ ,  $c, c' \in C$ ,  $w, w' \in \Sigma^*$ :  $(q, c, w) \vdash_{\theta} (q', c', w') \implies (q, A(c), w) \vdash_{A(\theta)} (q', A(c'), w')$ .

*Proof idea.* The claim can be shown by straightforward induction on the length of  $\theta$ . ■

**Theorem 21.** Let  $\mathcal{M}$  be an  $(S, \Sigma)$ -automaton and  $A$  be an  $S$ -proper total approximation strategy. Then  $L(A(\mathcal{M})) \supseteq L(\mathcal{M})$ .

*Proof.* The claim follows immediately from Lem. 20 and the definition of  $A(\mathcal{M})$ . ■

**Example 22.** Recall  $\mathcal{M}$  and  $A_{\text{eo}}(\mathcal{M})$  from Ex. 17. Their recognised languages are  $L(\mathcal{M}) = \{a^n b^n \mid n \in \mathbb{N}_+\}$  and  $L(A_{\text{eo}}(\mathcal{M})) = \{a^m b^n \mid m \in \mathbb{N}, n \in \mathbb{N}_+, m \equiv n \pmod{2}\}$ . Thus  $L(A_{\text{eo}}(\mathcal{M}))$  is a superset of  $L(\mathcal{M})$ . ■

**Corollary 23.** Let  $\mathcal{M}$  be an  $(S, \Sigma)$ -automaton, and  $A_1$  and  $A_2$  be  $S$ -proper approximation strategies. If  $A_1$  is finer than  $A_2$ , then  $L(A_1(\mathcal{M})) \subseteq L(A_2(\mathcal{M}))$ .

*Proof.* Since  $A_1$  is finer than  $A_2$ , there is a total approximation strategy  $A$  such that  $A_1; A = A_2$ . It follows from the fact that  $A_2$  is  $S$ -proper and from  $A_1; A = A_2$  that  $A$  must be  $A_1(S)$ -proper. Hence we obtain  $L(A_1(\mathcal{M})) \stackrel{\text{Thm. 21}}{\subseteq} L(A(A_1(\mathcal{M}))) \stackrel{\text{Obs. 18}}{=} L((A_1; A)(\mathcal{M})) = L(A_2(\mathcal{M}))$ . ■

The following example shows four approximation strategies that occur in the literature. The first three approximation strategies approximate a context-free language by a recognisable language (taken from Nederhof [26, Sec. 7]). The fourth approximation strategy approximates a context-free language by another context-free language. It is easy to see that the shown approximation strategies are total and thus lead to superset approximations.

**Example 24.** Let  $\Gamma$  be a finite set and  $k \in \mathbb{N}_+$ .

- (i) Evans [13] proposed to map each pushdown to its top-most element. The same result is achieved by dropping condition 7 and 8 from Baker [1]. This idea is expressed by the total approximation strategy  $A_{\text{top}}: \Gamma^* \rightarrow \Gamma \cup \{\text{@}\}$  with  $A_{\text{top}}(\varepsilon) = \text{@}$  and  $A_{\text{top}}(\gamma w) = \gamma$  for every  $w \in \Gamma^*$  and  $\gamma \in \Gamma$ , where @ is a new symbol that is not in  $\Gamma$ .
- (ii) Bermudez and Schimpf [2] proposed to map each pushdown to its top-most  $k$  elements. The total approximation strategy  $A_{\text{top},k}: \Gamma^* \rightarrow \{w \in \Gamma^* \mid |w| \leq k\}$  implements this idea where  $A_{\text{top},k}(w) = w$  if  $|w| \leq k$  and  $A_{\text{top},k}(w) = u$  if  $w$  is of the form  $uv$  for some  $u \in \Gamma^k$  and  $v \in \Gamma^+$ .
- (iii) Pereira and Wright [28] proposed to map each pushdown to one where no pushdown symbol occurs more than once. To achieve this, they replace each substrings of the form  $\gamma w' \gamma$  (for some  $\gamma \in \Gamma$  and  $w' \in \Gamma^*$ ) in the given pushdown by  $\gamma$ . Consider  $A_{\text{uniq}}: \Gamma^* \rightarrow \text{Seq}_{\text{nr}}(\Gamma)$  with  $A_{\text{uniq}}(w) = A_{\text{uniq}}(u\gamma v)$  if  $w$  is of form  $u\gamma w' \gamma v$  for some  $\gamma \in \Gamma$  and  $A_{\text{uniq}}(w) = w$  otherwise, where  $\text{Seq}_{\text{nr}}(\Gamma)$  denotes the set of all sequences over  $\Gamma$  without repetition.
- (iv) In their coarse-to-fine parsing approach for context-free grammars (short: CFG), Charniak et al. [4] propose, given an equivalence relation  $\equiv$  on the set of non-terminals  $N$  of some CFG  $G$ , to construct a new CFG  $G'$  whose non-terminals are the equivalence classes of  $\equiv$ .<sup>4</sup> Let  $\Sigma$  be the terminal alphabet of  $G$ . Say that  $g: N \rightarrow N/\equiv$  is the function that assigns for a nonterminal of  $G$  its corresponding equivalence class; and let  $g': (N \cup \Sigma)^* \rightarrow ((N/\equiv) \cup \Sigma)^*$  be an extension of  $g \cup \{(\sigma, \sigma) \mid \sigma \in \Sigma\}$ . Then  $g'$  is  $\text{PD}_{N \cup \Sigma}$ -proper and  $L(g'(\mathcal{M})) = L(G')$  where  $\mathcal{M}$  is the  $(\text{PD}_{N \cup \Sigma}, \Sigma)$ -automaton obtained from  $G$  by the usual construction [18, Thm. 5.3].  $\square$

## 4.2 Subset approximations

In this section we will show that injective approximation strategies lead to a subset approximation, this is proved by a variation of the proof of Thm. 21.

**Lemma 25.** Let  $\mathcal{M} = (Q, T, Q_i, Q_f)$  be an  $(S, \Sigma)$ -automaton,  $S = (C, P, R, c_i)$ , and  $A$  be an  $S$ -proper injective approximation strategy. Then for each  $\theta \in T^*$ ,  $q, q' \in Q$ ,  $c, c' \in \text{img}(A)$ ,  $w, w' \in \Sigma^*$ :  $(q, c, w) \vdash_{A(\theta)} (q', c', w') \implies (q, A^{-1}(c), w) \vdash_{\theta} (q', A^{-1}(c'), w')$ .

*Proof idea.* The claim can be shown by straightforward induction on the length of  $\theta$ .  $\blacksquare$

**Theorem 26.** Let  $\mathcal{M}$  be an  $(S, \Sigma)$ -automaton and  $A$  be an  $S$ -proper injective approximation strategy. Then  $L(A(\mathcal{M})) \subseteq L(\mathcal{M})$ .

*Proof.* Then the claim follows immediately from Lem. 25 and the definition of  $A(\mathcal{M})$ .  $\blacksquare$

**Corollary 27.** Let  $\mathcal{M}$  be an  $(S, \Sigma)$ -automaton and  $A_1$  and  $A_2$  be  $S$ -proper approximation strategies. If  $A_1$  is less partial than  $A_2$ , then  $L(A_1(\mathcal{M})) \supseteq L(A_2(\mathcal{M}))$ .

*Proof.* Since  $A_1$  is less partial than  $A_2$ , we know that there is an injective approximation strategy  $A$  such that  $A_1; A = A_2$ . As in the proof of Cor. 23 we know that  $A$  is  $A_1(S)$ -proper. Hence we obtain

$$L(A_1(\mathcal{M})) \stackrel{\text{Thm. 26}}{\supseteq} L(A(A_1(\mathcal{M}))) \stackrel{\text{Obs. 18}}{=} L((A_1; A)(\mathcal{M})) = L(A_2(\mathcal{M})). \quad \blacksquare$$

The following example approximates a context-free language with a recognisable language (taken from Nederhof [26, Sec. 7]). It is easy to see that the shown approximation strategy is injective and thus leads to subset approximations.

<sup>4</sup>Charniak et al. [4] actually considered probabilistic CFGs, but for the sake of simplicity we leave out the probabilities here.



**Example 28.** Let  $\Gamma$  be a finite set and  $k \in \mathbb{N}_+$ . Krauwer and des Tombe [21], Pulman [29], and Langendoen and Langsam [23] proposed to disallow pushdowns of height greater than  $k$ . This can be achieved by the partial identity  $A_{\text{bd},k}: \Gamma^+ \dashrightarrow \{w \in \Gamma \mid |w| \leq k\}$  where  $A_{\text{bd},k}(w) = w$  if  $|w| \leq k$  and  $A_{\text{bd},k}(w) = \text{undefined}$  if  $|w| > k$ .  $\square$

### 4.3 Potentially incomparable approximations

The following example shows that our framework is also capable of expressing approximation strategies that lead neither to superset nor to subset approximations.

**Example 29.** Let  $\Gamma$  be a (not necessarily finite) set,  $\Delta$  be a finite set,  $k \in \mathbb{N}_+$ , and  $g: \Gamma \rightarrow \Delta$  be a total function. For pushdown automata with an infinite pushdown alphabet, Johnson [20, end of Section 1.4] proposed to first approximate the infinite pushdown alphabet with a finite set and then restrict the pushdown height to  $k$ . This can be easily expressed as the composition of two approximations:

$$A_{\text{incomp},k}: \Gamma^+ \dashrightarrow \{w \mid w \in \Delta, |w| \leq k\} \quad A_{\text{incomp},k} = \hat{g}; A_{\text{bound},k}$$

where  $\hat{g}: \Gamma^+ \rightarrow \Delta^+$  is the point-wise application of  $g$ . Let  $|\Delta| < |\Gamma|$ . Then  $\hat{g}$  is total but not injective,  $A_{\text{bound},k}$  is injective but not total, and  $A_{\text{incomp},k}$  is neither total nor injective. Hence Thms. 21 and 26 provide no further insights about the approximation strategy  $A_{\text{incomp},k}$ . This concurs with the observation of Johnson [20, end of Section 1.4] that  $A_{\text{incomp},k}$  is not guaranteed to induce either subset or superset approximations.  $\square$

### 4.4 Approximation of weighted automata with storage

**Definition 30.** Let  $S$  be a data storage and  $K$  be a complete semiring. An  $(S, \Sigma, K)$ -automaton is a tuple  $\mathcal{M} = (Q, T, Q_i, Q_f, \delta)$  where  $(Q, T, Q_i, Q_f)$  is an  $(S, \Sigma)$ -automaton and  $\delta: T \rightarrow K$  (transition weights). We sometimes denote  $(Q, T, Q_i, Q_f)$  by  $\mathcal{M}_{\text{uw}}$  (“uw” stands for unweighted).  $\square$

Consider the  $(S, \Sigma, K)$ -automaton  $\mathcal{M} = (Q, T, Q_i, Q_f, \delta)$ . The  $\mathcal{M}$ -configurations, the run relation of  $\mathcal{M}$ , and the set of runs of  $\mathcal{M}$  on  $w$  for every  $w \in \Sigma^*$  are the same as for  $\mathcal{M}_{\text{uw}}$ . The weight of  $\theta$  in  $\mathcal{M}$  is the value  $\text{wt}_{\mathcal{M}}(\theta) = \delta(\tau_1) \cdot \dots \cdot \delta(\tau_k)$  for every  $\theta = \tau_1 \dots \tau_k$  with  $\tau_1, \dots, \tau_k \in T$ . In particular, we let  $\text{wt}_{\mathcal{M}}(\varepsilon) = 1$ . The weighted language induced by  $\mathcal{M}$  is the function  $\llbracket \mathcal{M} \rrbracket: \Sigma^* \rightarrow K$  where

$$\llbracket \mathcal{M} \rrbracket(w) = \sum_{\theta \in R_{\mathcal{M}}(w)} \text{wt}_{\mathcal{M}}(\theta) \quad (4)$$

For every  $w \in \Sigma^*$ . Let  $S$  be a data storage,  $K$  be a complete semiring, and  $r: \Sigma^* \rightarrow K$ . We call  $r$   $(S, \Sigma, K)$ -recognisable if there is an  $(S, \Sigma, K)$ -automaton  $\mathcal{M}$  with  $r = \llbracket \mathcal{M} \rrbracket$ .

We extend Prop. 9 to the weighted case, using the functions  $\det$  as defined in Prop. 9.

**Proposition 31.** The classes of  $(S, \Sigma, K)$ -recognisable and of  $(\det(S), \Sigma, K)$ -recognisable languages are the same for every data storage  $S$  and semiring  $K$ .

*Proof.* Let  $\mathcal{M} = (Q, T, Q_i, Q_f, \delta)$  be an  $(S, \Sigma, K)$ -automaton and  $\mathcal{M}' = (Q', T', Q'_i, Q'_f, \delta')$  a  $(\det(S), \Sigma, K)$ -automaton. We call  $\mathcal{M}$  and  $\mathcal{M}'$  related if  $\mathcal{M}_{\text{uw}}$  and  $\mathcal{M}'_{\text{uw}}$  are related, and  $\delta'(\det(\tau)) = \delta(\tau)$  for every  $\tau \in T$ . Note that  $\det: T \rightarrow \det(T)$  is a bijection. Clearly, for every  $(S, \Sigma, K)$ -automaton  $\mathcal{M}$  there is an  $(\det(S), \Sigma, K)$ -automaton  $\mathcal{M}'$  such that  $\mathcal{M}$  and  $\mathcal{M}'$  are related and vice versa. It remains to be shown that  $\llbracket \mathcal{M} \rrbracket = \llbracket \mathcal{M}' \rrbracket$ . For every  $w \in \Sigma^*$ , we derive

$$\llbracket \mathcal{M} \rrbracket(w) \stackrel{(4)}{=} \sum_{\theta \in R_{\mathcal{M}}(w)} \text{wt}_{\mathcal{M}}(\theta) = \sum_{\theta \in R_{\mathcal{M}}(w)} \text{wt}_{\mathcal{M}'}(\det(\theta)) \stackrel{(2)}{=} \sum_{\theta' \in R_{\mathcal{M}'}(w)} \text{wt}_{\mathcal{M}'}(\theta') \stackrel{(4)}{=} \llbracket \mathcal{M}' \rrbracket(w). \quad \blacksquare$$

**Definition 32.** Let  $\mathcal{M} = (Q, T, Q_i, Q_f, \delta)$  be an  $(S, \Sigma, K)$ -automaton and  $A$  be an  $S$ -proper approximation strategy. The *approximation of  $\mathcal{M}$  with respect to  $A$*  is the  $(A(S), \Sigma, K)$ -automaton  $A(\mathcal{M}) = (Q, A(T), Q_i, Q_f, A(\delta))$  where  $A(S)$  and  $A(T)$  are defined as in Def. 16, and  $A(\delta)(\tau') = \sum_{\tau \in T: A(\tau) = \tau'} \delta(\tau)$  for every  $\tau' \in A(T)$ .  $\square$

**Lemma 33.** Let  $\mathcal{M}$  be an  $(S, \Sigma, K)$ -automaton,  $A$  be an  $S$ -proper approximation strategy,  $\leq$  be a partial order on  $K$ , and  $K$  be positively  $\leq$ -ordered.

- (i)  $\text{wt}_{A(\mathcal{M})}(\theta') \geq \sum_{\theta \in R_{\mathcal{M}}: A(\theta) = \theta'} \text{wt}_{\mathcal{M}}(\theta)$  for every  $\theta' \in R_{A(\mathcal{M})}$ .
- (ii) If  $A$  is injective, then  $\text{wt}_{A(\mathcal{M})}(\theta') = \sum_{\theta \in R_{\mathcal{M}}: A(\theta) = \theta'} \text{wt}_{\mathcal{M}}(\theta)$  for every  $\theta' \in R_{A(\mathcal{M})}$ .

*Proof.* **ad (i):** We proof the claim by induction on the length of  $\theta'$ . For  $\theta' = \varepsilon$ , we derive

$$\text{wt}_{A(\mathcal{M})}(\varepsilon) = 1 \geq 1 = \text{wt}_{\mathcal{M}}(\varepsilon) = \sum_{\theta \in R_{\mathcal{M}}: A(\theta) = \varepsilon} \text{wt}_{\mathcal{M}}(\theta).$$

For  $\theta' \tau' \in R_{A(\mathcal{M})}$  with  $\tau' \in A(T)$ , we derive

$$\begin{aligned} \text{wt}_{A(\mathcal{M})}(\theta' \tau') &= \text{wt}_{A(\mathcal{M})}(\theta') \cdot A(\delta)(\tau') \\ &\geq \left( \sum_{\theta \in R_{\mathcal{M}}: A(\theta) = \theta'} \text{wt}_{\mathcal{M}}(\theta) \right) \cdot A(\delta)(\tau') && \text{(by IH and since } \cdot \text{ preserves } \leq) \\ &= \left( \sum_{\theta \in R_{\mathcal{M}}: A(\theta) = \theta'} \text{wt}_{\mathcal{M}}(\theta) \right) \cdot \left( \sum_{\tau \in T: A(\tau) = \tau'} \delta(\tau) \right) && \text{(by Def. 32)} \\ &= \sum_{\theta \in R_{\mathcal{M}}, \tau \in T: (A(\theta) = \theta') \wedge (A(\tau) = \tau')} \text{wt}_{\mathcal{M}}(\theta) \cdot \delta(\tau) && \text{(by distributivity of } K) \\ &\geq \sum_{\theta \in R_{\mathcal{M}}, \tau \in T: \theta \tau \in R_{\mathcal{M}} \wedge (A(\theta \tau) = \theta' \tau')} \text{wt}_{\mathcal{M}}(\theta) \cdot \delta(\tau) && \text{(by } (*) \text{ and since } + \text{ preserves } \leq) \\ &= \sum_{\bar{\theta} \in R_{\mathcal{M}}: (A(\bar{\theta}) = \theta' \tau')} \text{wt}_{\mathcal{M}}(\bar{\theta}) && \text{(by Def. 32)} \end{aligned}$$

For  $(*)$ , we note that the index set of the left sum subsumes that of the right sum and hence  $\geq$  is justified.

**ad (ii):** The proof follows the same structure as the proof of (i). But we make the following modifications: In the induction base, we can write “=” instead of “ $\geq$ ” since  $1 = 1$ . For the induction step, we assume that (ii) holds for every  $\theta'$  of length  $n$ . Then the “ $\geq$ ” in the second line of the induction step can be replaced by “=” . In order to turn the “ $\geq$ ” in the fifth line of the induction step into “=”, we propose that the index sets of the left and the right sum are the same. This holds since  $A$  is injective,  $\theta' \tau'$  is in  $R_{A(\mathcal{M})}$ , and hence (by Lem. 25) each  $\theta \tau$  with  $A(\theta \tau) = \theta' \tau'$  is in  $R_{\mathcal{M}}$ .  $\blacksquare$

**Theorem 34.** Let  $\mathcal{M}$  be an  $(S, \Sigma, K)$ -automaton,  $A$  be an  $S$ -proper approximation strategy, and  $\leq$  be a partial order on  $K$ , and  $K$  be positively  $\leq$ -ordered.

- (i) If  $A$  is total, then  $\llbracket A(\mathcal{M}) \rrbracket(w) \geq \llbracket \mathcal{M} \rrbracket(w)$  for every  $w \in \Sigma^*$ .
- (ii) If  $A$  is injective, then  $\llbracket A(\mathcal{M}) \rrbracket(w) \leq \llbracket \mathcal{M} \rrbracket(w)$  for every  $w \in \Sigma^*$ .

*Proof.* **ad (i):** For every  $w \in \Sigma^*$ , we derive

$$\begin{aligned} \llbracket A(\mathcal{M}) \rrbracket(w) &\stackrel{(4)}{=} \sum_{\theta' \in R_{A(\mathcal{M})}(w)} \text{wt}_{A(\mathcal{M})}(\theta') \stackrel{(*)}{\geq} \sum_{\theta' \in R_{A(\mathcal{M})}(w)} \sum_{\theta \in R_{\mathcal{M}}: A(\theta) = \theta'} \text{wt}_{\mathcal{M}}(\theta) \\ &\stackrel{\text{Def. 16}}{=} \sum_{\theta' \in R_{A(\mathcal{M})}(w)} \sum_{\theta \in R_{\mathcal{M}}(w): A(\theta) = \theta'} \text{wt}_{\mathcal{M}}(\theta) \stackrel{(\dagger)}{=} \sum_{\theta \in R_{\mathcal{M}}(w)} \text{wt}_{\mathcal{M}}(\theta) \stackrel{(4)}{=} \llbracket \mathcal{M} \rrbracket(w) \end{aligned}$$

where  $(*)$  follows from Lem. 33 (i) and the fact that  $+$  preserves  $\leq$ . For  $(\dagger)$ , we argue that for each  $\theta \in R_{\mathcal{M}}(w)$  there is exactly one  $\theta' \in R_{A(\mathcal{M})}(w)$  with  $A(\theta) = \theta'$  since  $A$  is total. Hence the left side and

the right side of the equation have exactly the same addends. Then, since  $+$  is commutative, the “=” is justified.

**ad (ii):** For every  $w \in \Sigma^*$ , we derive

$$\begin{aligned} \llbracket A(\mathcal{M}) \rrbracket(w) &\stackrel{(4)}{=} \sum_{\theta' \in R_{A(\mathcal{M})}(w)} \text{wt}_{A(\mathcal{M})}(\theta') \stackrel{\text{Lem. 33 (ii)}}{=} \sum_{\theta' \in R_{A(\mathcal{M})}(w)} \sum_{\theta \in R_{\mathcal{M}}: A(\theta)=\theta'} \text{wt}_{\mathcal{M}}(\theta) \\ &\stackrel{\text{Def. 16}}{=} \sum_{\theta' \in R_{A(\mathcal{M})}(w)} \sum_{\theta \in R_{\mathcal{M}}: A(\theta)=\theta'} \text{wt}_{\mathcal{M}}(\theta) \stackrel{(\ddagger)}{\leq} \sum_{\theta \in R_{\mathcal{M}}(w)} \text{wt}_{\mathcal{M}}(\theta) \stackrel{(4)}{=} \llbracket \mathcal{M} \rrbracket(w). \end{aligned}$$

For  $(\ddagger)$ , we argue that for each  $\theta \in R_{\mathcal{M}}(w)$  there is at most one  $\theta' \in R_{A(\mathcal{M})}(w)$  with  $A(\theta) = \theta'$  since  $A$  is a partial function. Hence all the addends on the left side of the inequality also occur on the right side. But there may be an addend  $\text{wt}_{\mathcal{M}}(\theta)$  on the right side which does not occur on the left side because  $A(\theta) = \text{undefined}$ . Since  $+$  preserves  $\leq$ , the “ $\leq$ ” is justified.  $\blacksquare$

## 5 Approximation of multiple context-free languages

Due to the equivalence of pushdown automata and context-free grammars [18, Thms. 5.3 and 5.4], the approximation strategies in Exs. 24 and 28 can be used for the approximation of context-free languages. The framework presented in this paper together with the automata characterisation of multiple context-free languages [7, Thm. 18] allows an automata-theoretic view on the approximation of multiple context-free languages. The automata characterisation uses an excursion-restricted form of automata with tree-stack storage [7]. A tree-stack is a tree with a designated position inside of it (the *stack pointer*). The automaton can read the label under the stack pointer, can determine whether the stack pointer is at the bottom (i.e. the root), and can modify the tree stack by moving the stack pointer or by adding a node. The excursion-restriction bounds how often the stack pointer may enter a position from its parent node.

**Definition 35.** Let  $\Gamma$  be a finite set. The *tree-stack storage over  $\Gamma$*  is the deterministic data storage  $\text{TSS}_{\Gamma} = (\text{TS}_{\Gamma}, P_{\text{ts}}, R_{\text{ts}}, c_{\text{i,ts}})$  where

- $\text{TS}_{\Gamma}$  is the set of tuples  $\langle \xi, \rho \rangle$  where  $\xi: \mathbb{N}_+^* \dashrightarrow \Gamma \cup \{\@\}$ ,  $\text{dom}(\xi)$  is finite and prefix-closed,<sup>5</sup>  $\rho \in \text{dom}(\xi)$ , and  $\xi(\rho') = \@$  iff  $\rho' = \varepsilon$  (We call  $\xi$  the *stack* and  $\rho$  the *stack pointer* of  $\langle \xi, \rho \rangle$ .);
- $c_{\text{i,ts}} = \langle \{(\varepsilon, @)\}, \varepsilon \rangle$ ;
- $P_{\text{ts}} = \{\text{TS}_{\Gamma}, \text{bottom}\} \cup \{\text{top}_{\gamma} \mid \gamma \in \Gamma\}$  with  $\text{bottom} = \{\langle \xi, \rho \rangle \in \text{TS}_{\Gamma} \mid \rho = \varepsilon\}$  and  $\text{top}_{\gamma} = \{\langle \xi, \rho \rangle \in \text{TS}_{\Gamma} \mid \xi(\rho) = \gamma\}$  for every  $\gamma \in \Gamma$ ; and
- $R_{\text{ts}} = \{\text{down}\} \cup \{\text{up}_n, \text{push}_{n,\gamma} \mid n \in \mathbb{N}, \gamma \in \Gamma\}$  where for each  $n \in \mathbb{N}_+$  and  $\gamma \in \Gamma$ :
  - $\text{up}_n = \{(\langle \xi, \rho \rangle, \langle \xi, \rho n \rangle) \mid \langle \xi, \rho \rangle \in \text{TS}_{\Gamma}, \rho n \in \text{dom}(\xi)\}$ ,
  - $\text{down} = \bigcup_{n \in \mathbb{N}_+} \text{up}_n^{-1}$ , and
  - $\text{push}_{n,\gamma} = \{(\langle \xi, \rho \rangle, \langle \xi \cup \{(\rho n, \gamma)\}, \rho n \rangle) \mid \langle \xi, \rho \rangle \in \text{TS}_{\Gamma}, \rho n \notin \text{dom}(\xi)\}$ .  $\square$

**Example 36.** Consider  $\Sigma = \{a, b, c\}$ ,  $\Gamma = \{*, \#\}$ , the  $(\text{TSS}_{\Gamma}, \Sigma)$ -automaton  $\mathcal{M} = ([4], T, \{1\}, \{4\})$ , and

$$\begin{aligned} T: \tau_1 &= (1, a, \text{TS}_{\Gamma}, \text{push}_{1,*}, 1) & \tau_4 &= (2, b, \text{top}_* \quad, \text{down}, 2) & \tau_7 &= (3, \varepsilon, \text{top}_{\#}, \text{down}, 4.) \\ \tau_2 &= (1, \varepsilon, \text{TS}_{\Gamma}, \text{push}_{1,\#}, 2) & \tau_5 &= (2, \varepsilon, \text{bottom}, \text{up}_1 \quad, 3) \\ \tau_3 &= (2, \varepsilon, \text{top}_{\#}, \text{down} \quad, 2) & \tau_6 &= (3, c, \text{top}_* \quad, \text{up}_1 \quad, 3) \end{aligned}$$

<sup>5</sup>A set  $D \subseteq \mathbb{N}_+^*$  is *prefix closed* if for each  $w \in D$ , every prefix of  $w$  is also in  $D$ .

$(1, a, \Gamma^*$	, push <sub>*</sub>	, 1)	$(1, a, \Gamma_{@}$	, $\{(\gamma, *) \mid \gamma \in \Gamma_{@}\}$	, 1)
$(1, \varepsilon, \Gamma^*$	, push <sub>#</sub>	, 2)	$(1, \varepsilon, \Gamma_{@}$	, $\{(\gamma, \#) \mid \gamma \in \Gamma_{@}\}$	, 2)
$(2, \varepsilon, \text{top}_{\#}$	, pop	, 2)	$(2, \varepsilon, \{\#\}$	, $\{(\gamma, \gamma') \mid \gamma, \gamma' \in \Gamma_{@}\}$	, 2)
$(2, b, \text{top}_{*}$	, pop	, 2)	$(2, b, \{*\}$	, $\{(\gamma, \gamma') \mid \gamma, \gamma' \in \Gamma_{@}\}$	, 2)
$(2, \varepsilon, \text{bottom}, \text{push}_{*} \cup \text{push}_{\#}$	, 3)		$(2, \varepsilon, \{@\}$	, $\{(\gamma, \gamma') \mid \gamma \in \Gamma_{@}, \gamma' \in \Gamma\}$	, 3)
$(3, c, \text{top}_{*}$	, push <sub>*</sub> $\cup$ push <sub>#</sub>	, 3)	$(3, c, \{*\}$	, $\{(\gamma, \gamma') \mid \gamma \in \Gamma_{@}, \gamma' \in \Gamma\}$	, 3)
$(3, \varepsilon, \text{top}_{\#}$	, pop	, 4)	$(3, \varepsilon, \{\#\}$	, $\{(\gamma, \gamma') \mid \gamma, \gamma' \in \Gamma_{@}\}$	, 4)

Figure 2: Transitions of  $A_{\text{cf},\Gamma}(\mathcal{M})$  (left) and  $(A_{\text{cf},\Gamma}; A_{\text{top}})(\mathcal{M})$  (right)

The runs of  $\mathcal{M}$  all have a specific form:  $\mathcal{M}$  executes  $\tau_1$  arbitrarily often (say  $n$  times) until it executes  $\tau_2$ , leading to the storage configuration  $\zeta = \langle \{(\varepsilon, @), (1, *), \dots, (1^n, *), (1^{n+1}, \#)\}, 1^{n+1} \rangle$  where  $1^k$  means that 1 is repeated  $k$  times. The stack of  $\zeta$  is a monadic tree where the leaf is labelled with #, the root is labelled with @, and the remaining  $n$  nodes are labelled with \*. The stack pointer of  $\zeta$  points to the leaf. From this configuration  $\mathcal{M}$  executes  $\tau_3$  once and  $\tau_4$   $n$  times (i.e. for each \* on the stack), moving the stack pointer to the root. Then  $\mathcal{M}$  executes  $\tau_5$  once and  $\tau_6$   $n$  times, leading to the final state. Hence the language of  $\mathcal{M}$  is  $L(\mathcal{M}) = \{a^n b^n c^n \mid n \in \mathbb{N}\}$ , which is not context-free.  $\square$

**Example 37.** The following two approximation strategies for multiple context-free languages are taken from the literature. Let  $\Gamma$  be a finite set.

- (i) Van Cranenburgh [6, Sec. 4] observed that the idea of Ex. 24 (iv) also applies to multiple context-free grammars (short: MCFG). The idea can be applied to tree-stack automata similarly to the way it was applied to pushdown automata in Ex. 24 (iv). The resulting data storage is still a tree-stack storage. This approximation strategy is total and thus leads to a superset approximation.
- (ii) Burden and Ljunglöf [3, Sec. 4] and van Cranenburgh [6, Sec. 4] proposed to split each production of a given MCFG into multiple productions, each of fan-out 1. Since the resulting grammar is of fan-out 1, it produces a context-free language and can be recognised by a pushdown automaton. The corresponding approximation strategy in our framework is  $A_{\text{cf},\Gamma}: \text{TS}_{\Gamma} \rightarrow \Gamma^*$  with  $A_{\text{cf},\Gamma}((\xi, n_1 \cdots n_k)) = \xi(n_1 \cdots n_k) \cdots \xi(n_1 n_2) \xi(n_1)$  for every  $(\xi, n_1 \cdots n_k) \in \text{TS}_{\Gamma}$  with  $n_1, \dots, n_k \in \mathbb{N}_+$ . The resulting data storage is a pushdown storage.  $A_{\text{cf},\Gamma}$  is total and thus leads to a superset approximation.  $\square$

**Example 38.** Let us consider the  $(\text{TSS}_{\Gamma}, \Sigma)$ -automaton  $\mathcal{M}$  from Ex. 36. Figure 2 shows the transitions of the  $(A_{\text{cf},\Gamma}(\text{TSS}_{\Gamma}), \Sigma)$ -automaton  $A_{\text{cf},\Gamma}(\mathcal{M})$  (cf. Ex. 37) and the  $((A_{\text{cf},\Gamma}; A_{\text{top}})(\text{TSS}_{\Gamma}), \Sigma)$ -automaton  $(A_{\text{cf},\Gamma}; A_{\text{top}})(\mathcal{M})$  (cf. also Ex. 24). The languages recognised by the two automata are  $L(A_{\text{cf},\Gamma}(\mathcal{M})) = \{a^n b^n c^m \mid n, m \in \mathbb{N}\}$  and  $L((A_{\text{cf},\Gamma}; A_{\text{top}})(\mathcal{M})) = \{a^n b^m c^k \mid n, m, k \in \mathbb{N}\}$ . Clearly,  $L(A_{\text{cf},\Gamma}(\mathcal{M}))$  is a context-free language. Since  $(A_{\text{cf},\Gamma}; A_{\text{top}})(\mathcal{M})$  has finitely many storage configurations, its language is recognisable by a finite state automaton (Prop. 12).  $\square$

## 6 Coarse-to-fine $n$ -best parsing for weighted automata with storage

Parsing is a process that takes a finite representation  $\mathcal{R}$  of a language  $L(\mathcal{R}) \subseteq \Sigma^*$  and a word  $w \in \Sigma^*$ , and outputs analyses of  $w$  in  $\mathcal{R}$ . If  $\mathcal{R}$  is a grammar, then the analyses of  $w$  are the *parse trees* in  $\mathcal{R}$  for  $w$ . If  $\mathcal{R}$  is an automaton (with storage), then the analyses of  $w$  are the runs of  $\mathcal{R}$  on  $w$ . Since this paper is concerned with weighted automata with storage, let  $\mathcal{R}$  be an  $(S, \Sigma, K)$ -automaton. Also, let  $K$  be partially ordered by a relation  $\leq$ . We will call a run  $\theta$  „better than“ a run  $\theta'$  if  $\text{wt}_{\mathcal{R}}(\theta) \geq \text{wt}_{\mathcal{R}}(\theta')$ .

Using  $\text{wt}_{\mathcal{R}}$ , we can assign weights to the runs of  $\mathcal{R}$  on  $w$  and enumerate those runs in descending order (with respect to  $\leq$ ) of their weights.<sup>6</sup> If we output the first  $n$  from the descending list of runs, we call the parsing *n-best parsing* [19].

*Coarse-to-fine parsing* [4] employs a simpler (i.e. easier to parse) automaton  $\mathcal{R}'$  to parse  $w$  and uses the runs of  $\mathcal{R}'$  on  $w$  to narrow the search space for the runs of  $\mathcal{R}$  on  $w$ . To ensure that there are runs of  $\mathcal{R}'$  on  $w$  whenever there are runs of  $\mathcal{R}$  on  $w$ , we require that  $L(\mathcal{R}') \supseteq L(\mathcal{R})$ . The automaton  $\mathcal{R}'$  is obtained by superset approximation. In particular, we require  $\mathcal{R}' = A(\mathcal{R})$  for some total approximation strategy  $A$ .

---

**Algorithm 3** Coarse-to-fine  $n$ -best parsing for weighted automata with storage

---

**Input:**  $(S, \Sigma, K)$ -automaton  $\mathcal{M}$ ,  $S$ -proper total approximation strategy  $A$ ,  $n \in \mathbb{N}$ , word  $w \in \Sigma^*$

**Output:** some set of  $n$  greatest (with respect to the image under  $\text{wt}_{\mathcal{M}}$  and  $\leq$ ) runs of  $\mathcal{M}$  on  $w$

---

- 1:  $X \leftarrow \emptyset$   $\triangleright X$  is the set of runs of  $\mathcal{M}$  on  $w$  that were already found
  - 2:  $Y \leftarrow R_{A(\mathcal{M})}(w)$   $\triangleright Y$  is the set of runs of  $A(\mathcal{M})$  on  $w$  that were not yet considered
  - 3: **while**  $|X| < n$  **or**  $\min_{\theta \in X} \text{wt}_{\mathcal{M}}(\theta) < \max_{\theta' \in Y} \text{wt}_{A(\mathcal{M})}(\theta')$  **do**
  - 4:      $\theta' \leftarrow$  smallest element of  $Y$  with respect to the image under  $\text{wt}_{A(\mathcal{M})}$
  - 5:      $Y \leftarrow Y \setminus \{\theta'\}$
  - 6:     **for each**  $\theta \in A^{-1}(\theta')$  that is a sequence of transitions in  $\mathcal{M}$  **do**
  - 7:         **if**  $\theta \in R_{\mathcal{M}}$  **then**  $X \leftarrow X \cup \{\theta\}$   $\triangleright$  it is sufficient to only check the storage behaviour for  $\theta$
  - 8: **return** a set of  $n$  greatest elements of  $X$  with respect to the image under  $\text{wt}_{\mathcal{M}}$
- 

Algorithm 3 describes coarse-to-fine  $n$ -best parsing for weighted automata with storage. The inputs are an  $(S, \Sigma, K)$ -automaton  $\mathcal{M}$ , an  $S$ -proper approximation strategy  $A$  which will be used to construct an approximation of  $\mathcal{M}$ , a natural number  $n$  which specifies how many runs should be computed, and a word  $w \in \Sigma^*$  which we want to parse. The output is a set of  $n$ -best runs of  $\mathcal{M}$  on  $w$ . The algorithm starts with a set  $X$  that is empty (line 1) and a set  $Y$  that contains all the runs of  $A(\mathcal{M})$  on  $w$  (line 2). Then, as long as  $X$  has less than  $n$  elements or an element of  $Y$  is greater than the smallest element in  $X$  with respect to their weights (line 3), we take the greatest element  $\theta'$  of  $Y$  (line 4), remove  $\theta'$  from  $Y$  (line 5), calculate the corresponding sequences  $\theta$  of transitions from  $\mathcal{M}$  (line 6), and add  $\theta$  to  $X$  if  $\theta$  is a run of  $\mathcal{M}$  (line 7).

We can restrict the automaton  $A(\mathcal{M})$  to the input  $w$  with the usual product construction. The set of runs of the resulting product automaton (let us call it  $\mathcal{M}_{A,w}$ ) can be mapped onto  $R_{A(\mathcal{M})}(w)$  by some projection  $\varphi$ . Hence  $\mathcal{M}_{A,w}$  (finitely) represents  $R_{A(\mathcal{M})}(w)$ . The automaton  $\mathcal{M}_{A,w}$  can be construed as a (not necessarily finite) graph  $G_{A,w}$  with the  $\mathcal{M}_{A,w}$ -configurations as nodes. The edges shall be labelled with the images of the corresponding transitions of  $\mathcal{M}_{A,w}$  under  $\varphi$ . Then the paths (i.e. sequences of edge labels) in  $G_{A,w}$  from the initial  $\mathcal{M}_{A,w}$ -configuration to all the final  $\mathcal{M}_{A,w}$ -configurations are exactly the elements of  $R_{A(\mathcal{M})}(w)$ . Those paths can be enumerated in descending order of their weights using a variant of Dijkstra's algorithm. This provides us with a method to compute  $\max_{\theta' \in Y} \text{wt}_{A(\mathcal{M})}(\theta')$  on line 3 and  $\theta'$  on line 4 of Alg. 3.

**Example 39.** Let  $\Gamma = \{a, b, c\}$ ,  $\Sigma = \Gamma \cup \{\#\}$ ,  $K$  be the Viterbi semiring  $(\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$  with linear order  $\leq$ , and  $A_{\#}: \Gamma^* \rightarrow \mathbb{N}, u \mapsto |u|$  be a total approximation strategy. Note that  $A_{\#}(\text{PD}_{\Gamma}) = \text{Count}$ . Now consider the  $(\text{PD}_{\Gamma}, \Sigma, K)$ -automaton  $\mathcal{M} = ([3], T, \{1\}, \{3\}, \delta)$  and the  $(\text{Count}, \Sigma, K)$ -automaton

---

<sup>6</sup>The resulting list of runs is not unique since different runs may get the same weight and since we only have a partial order.

$A_{\#}(\mathcal{M}) = ([3], T', \{1\}, \{3\}, \delta')$  where  $T = \{\tau_1, \dots, \tau_8\}$  and  $T' = \{\tau'_1, \tau'_2, \tau'_4, \tau'_5, \tau'_6, \tau'_7, \tau'_8\}$  with

$$\begin{array}{llll} \tau_1 = (1, a, \Gamma^*, \text{push}_a, 1) & \tau_2 = (1, \varepsilon, \Gamma^*, \text{push}_b, 1) & \tau_3 = (1, \varepsilon, \Gamma^*, \text{push}_c, 1) & \tau_4 = (1, \#, \Gamma^*, \text{stay}, 2) \\ \tau_5 = (2, a, \text{top}_a, \text{pop}, 2) & \tau_6 = (2, b, \text{top}_b, \text{pop}_b, 2) & \tau_7 = (2, c, \text{top}_c, \text{pop}, 2) & \tau_8 = (2, \varepsilon, \text{bottom}, \text{stay}, 3) \\ \tau'_1 = (1, a, \mathbb{N}, \text{inc}, 1) & \tau'_2 = (1, \varepsilon, \mathbb{N}, \text{inc}, 1) & & \tau'_4 = (1, \#, \mathbb{N}, \text{id}, 2) \\ \tau'_5 = (2, a, \mathbb{N}_+, \text{dec}, 2) & \tau'_6 = (2, b, \mathbb{N}_+, \text{dec}, 2) & \tau'_7 = (2, c, \mathbb{N}_+, \text{dec}, 2) & \tau'_8 = (2, \varepsilon, \{0\}, \text{id}, 3), \end{array}$$

$\delta(\tau) = 1$  for each  $\tau \in T$ , and  $\delta'(\tau') = 1$  for each transition  $\tau' \in T'$ .<sup>7</sup>

We use Alg. 3 to obtain the 1-best run of  $w = a\#ba$ : On line 4, we get  $\theta' = \tau'_1 \tau'_2 \tau'_4 \tau'_5 \tau'_8$  (the only run of  $A_{\#}(\mathcal{M})$  on  $w$ ). Then there are only two possible values for  $\theta$  on line 7, namely  $\theta_1 = \tau_1 \tau_2 \tau_4 \tau_7 \tau_5 \tau_8$  and  $\theta_2 = \tau_1 \tau_3 \tau_4 \tau_7 \tau_5 \tau_8$  of which only  $\theta_2$  is a run of  $\mathcal{M}$ , hence the algorithm returns  $\{\theta_2\}$ .  $\square$

**Outlook.** The author intends to extend Alg. 3 to use multiple levels of approximation (i.e. multiple approximation strategies that can be applied in sequence) and to investigate the viability of this extension for parsing multiple context-free languages in the context of natural languages.

## Acknowledgements

The author thanks Mark-Jan Nederhof for fruitful discussions and the anonymous reviewers of a previous version of this paper for their helpful comments. In particular, Ex. 4 is due to a reviewer's comment and Ex. 11 is due to Mark-Jan Nederhof.

## References

- [1] T.P. Baker (1981): *Extending lookahead for LR parsers*. *JCSS* 22(2), pp. 243–259, doi:10.1016/0022-0000(81)90030-1.
- [2] M.E. Bermudez & K.M. Schimpf (1990): *Practical arbitrary lookahead LR parsing*. *JCSS* 41(2), pp. 230–250, doi:10.1016/0022-0000(90)90037-1.
- [3] H. Burden & P. Ljunglöf (2005): *Parsing Linear Context-free Rewriting Systems*. In: *Proc. of IWPT*, pp. 11–17.
- [4] E. Charniak, M. Pozar, T. Vu, M. Johnson, M. Elsner, J. Austerweil, D. Ellis, I. Haxton, C. Hill, R. Shrivaths & J. Moore (2006): *Multilevel coarse-to-fine PCFG parsing*. In: *Proc. of NAACL HLT*, pp. 168–175, doi:10.3115/1220835.1220857.
- [5] É. Villemonte de la Clergerie (2002): *Parsing Mildly Context-Sensitive Languages with Thread Automata*. In: *Proc. of COLING*, pp. 1–7, doi:10.3115/1072228.1072256.
- [6] A. van Cranenburgh (2012): *Efficient Parsing with Linear Context-free Rewriting Systems*. In: *Proc. of EACL*, pp. 460–470.
- [7] T. Denking (2016): *An Automata Characterisation for Multiple Context-Free Languages*. In: *Proc. of DLT*, pp. 138–150, doi:10.1007/978-3-662-53132-7\_12.
- [8] M. Droste & W. Kuich (2009): *Semirings and Formal Power Series*. In: *Handbook of Weighted Automata*, Springer, pp. 3–28, doi:10.1007/978-3-642-01492-5\_1.
- [9] S. Eilenberg (1974): *Automata, languages, and machines*. Academic Press.
- [10] J. Engelfriet (1986): *Context-free grammars with storage*. Technical Report I86-11, Leiden University.

<sup>7</sup> $L(\mathcal{M}) = \{a^k \# w \mid k \in \mathbb{N}, w \in \{a, b, c\}^*, a \text{ occurs } k \text{ times in } w\}$  and  $L(A_{\#}(\mathcal{M})) = \{a^k \# w \mid k \in \mathbb{N}, w \in \{a, b, c\}^*, |w| \geq k\}$ .

- [11] J. Engelfriet (2014): *Context-free grammars with storage*. CoRR.
- [12] J. Engelfriet & H. Vogler (1986): *Pushdown machines for the macro tree transducer*. TCS 42(3), pp. 251–368, doi:10.1016/0304-3975(86)90052-6.
- [13] E.G. Evans (1997): *Approximating context-free grammars with a finite-state calculus*. In: Proc. of EACL, pp. 452–459, doi:10.3115/979617.979675.
- [14] J. Goldstine (1979): *A rational theory of AFLs*. In: Automata, Languages and Programming, Springer, pp. 271–281, doi:10.1007/3-540-09510-1\_21.
- [15] L. Herrmann & H. Vogler (2015): *A Chomsky-Schützenberger Theorem for Weighted Automata with Storage*. In: Proc. of CAI, pp. 90–102, doi:10.1007/978-3-319-23021-4\_11.
- [16] L. Herrmann & H. Vogler (2016): *Weighted Symbolic Automata with Data Storage*. In: Developments in Language Theory, Springer, pp. 203–215, doi:10.1007/978-3-662-53132-7\_17.
- [17] C.A.R. Hoare (1972): *Proof of correctness of data representations*. Acta Informatica 1(4), doi:10.1007/bf00289507.
- [18] J.E. Hopcroft & J.D. Ullman (1979): *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley.
- [19] L. Huang & D. Chiang (2005): *Better k-best Parsing*. In: Proc. of IWPT, pp. 53–64.
- [20] M. Johnson (1998): *Finite-state approximation of constraint-based grammars using left-corner grammar transforms*. In: Proc. of COLING, pp. 619–623, doi:10.3115/980451.980948.
- [21] S. Krauwer & L. des Tombe (1981): *Transducers and Grammars as Theories of Language*. Theoretical Linguistics 8(1–3), pp. 173–202, doi:10.1515/thli.1981.8.1-3.173.
- [22] M. Kuhlmann & G. Satta (2009): *Treebank grammar techniques for non-projective dependency parsing*. In: Proc. of EACL, pp. 478–486, doi:10.3115/1609067.1609120.
- [23] D.T. Langendoen & Y. Langsam (1987): *On the design of finite transducers for parsing phrase-structure languages*. Mathematics of Language, pp. 191–235, doi:10.1075/z.35.11lan.
- [24] W. Maier (2010): *Direct Parsing of Discontinuous Constituents in German*. In: Proc. of NAACL HLT, pp. 58–66.
- [25] M.-J. Nederhof (2000): *Practical experiments with regular approximation of context-free languages*. Computational Linguistics 26(1), pp. 17–44, doi:10.1162/089120100561610.
- [26] M.-J. Nederhof (2000): *Regular approximation of CFLs: a grammatical view*. In: Advances in Probabilistic and other Parsing Technologies, Springer, pp. 221–241, doi:10.1007/978-94-015-9470-7\_12.
- [27] M.-J. Nederhof (2017): *personal communication*.
- [28] F.C.N. Pereira & R.N. Wright (1991): *Finite-state approximation of phrase structure grammars*. In: Proc. of ACL, pp. 246–255, doi:10.3115/981344.981376.
- [29] S.G. Pulman (1986): *Grammars, parsers, and memory limitations*. Language and Cognitive Processes 1(3), pp. 197–225, doi:10.1080/01690968608407061.
- [30] M.-P. Schützenberger (1962): *Certain elementary families of automata*. In: Proc. Symp. on Mathematical Theory of Automata, pp. 139–153.
- [31] D. Scott (1967): *Some definitional suggestions for automata theory*. JCSS 1(2), pp. 187–212, doi:10.1016/s0022-0000(67)80014-x.
- [32] H. Seki, T. Matsumura, M. Fujii & T. Kasami (1991): *On multiple context-free grammars*. TCS 88(2), pp. 191–229, doi:10.1016/0304-3975(91)90374-B.
- [33] K. Vijay-Shanker, D.J. Weir & A.K. Joshi (1987): *Characterizing Structural Descriptions Produced by Various Grammatical Formalisms*. In: Proc. of ACL, pp. 104–111, doi:10.3115/981175.981190.
- [34] H. Vogler, M. Droste & L. Herrmann (2016): *A Weighted MSO Logic with Storage Behaviour and Its Büchi-Elgot-Trakhtenbrot Theorem*. In: Proc. of LATA, pp. 127–139, doi:10.1007/978-3-319-30000-9\_10.