On the Property of Preserving Regularity for String-Rewriting Systems

Friedrich Otto

Fachbereich Mathematik/Informatik, Universität Kassel, D-34109 Kassel
Internet: `otto@theory.informatik.uni-kassel.de`

Abstract. Some undecidability results concerning the property of preserving regularity are presented that strengthen corresponding results of Gilleron and Tison (1995). In particular, it is shown that it is undecidable in general whether a finite, length-reducing, and confluent string-rewriting system yields a regular set of normal forms for each regular language.

1 Introduction

If R is a left-linear term-rewriting system on a signature F, then the set of normal forms $\mathrm{IRR}(R)$ is a regular tree language. The system R is called *F-regularity preserving*, if, for each regular tree language $S \subseteq T(F)$, the set of descendants $\Delta_R^*(S)$ is again a regular tree language. Thus, if R is a left-linear and convergent term-rewriting system that is F-regularity preserving, then, for each regular tree language $S \subseteq T(F)$, the set of normal forms $\mathrm{NF}_R(S) := \Delta_R^*(S) \cap \mathrm{IRR}(R)$ of terms in S is a regular tree language. Hence, for systems of this form, various decision problems can be solved efficiently (see, e.g., [GS84, GT95]).

It is known that F-regularity is preserved by term-rewriting systems that contain only ground rules [Bra69], by term-rewriting systems that are linear and monadic [Sal88], that are linear and semi-monadic [CDGV94], or that are linear and generalized semi-monadic [GV97].

However, the property of preserving F-regularity is undecidable in general [GT95]. In addition, also the property that $\mathrm{NF}_R(S)$ is regular for each regular tree language $S \subseteq T(F)$ is undecidable in general [Gil91]. Actually, the latter undecidability result remains valid even for the class of finite term-rewriting systems that are convergent [Gil91, GT95].

It has been observed by Gyenizse and Vágvölgyi that the property of preserving F-regularity does not only depend on the term-rewriting system R considered, but also on the actual signature F [GV97]. Accordingly, they call a system R on a signature F *regularity preserving*, if it is F_1-regularity preserving for each signature F_1 containing F, and they ask whether this property is undecidable in general. Further, Gyenizse and Vágvölgyi conjecture that for string-rewriting systems the properties of preserving F-regularity and of preserving regularity are equivalent. Here we prove this conjecture.

If R is a string-rewriting system on Σ, then with R we can associate a term-rewriting system S_R over the signature F_Σ, which contains unary function symbols and a single constant only. We shall see that R preserves regularity if and only if the term-rewriting system S_R does. Since the property of preserving regularity is undecidable for finite string-rewriting systems, this answers the question above in the negative.

For string-rewriting systems the property of being a generalized semi-monadic system is just the property of being a left-basic string-rewriting system (see, e.g., [Sén90]), and it is known that a finite, length-reducing, confluent, and left-basic string-rewriting system R yields a regular set $\mathrm{NF}_R(S)$ for each regular language S [Sak79]. Thus, for string-rewriting systems the above-mentioned result that a linear, generalized semi-monadic system preserves F-regularity can be seen as a generalization of Sakarovitch's result. On the other hand, regularity is not preserved by finite, length-reducing, and confluent string-rewriting systems in general [BO93].

Here we give a simple proof for the fact that the property of preserving regularity is undecidable for finite string-rewriting systems in general. Further, we construct a finite, length-reducing, and confluent string-rewriting system R such that it is undecidable whether, for a regular language $S \subseteq \Sigma^*$, the set $\Delta_R^*(S)$ is a regular language. The same result is also established for the set of normal forms $\mathrm{NF}_R(S)$. In addition, a finite, length-reducing, and confluent string-rewriting system R and an infinite family of regular languages $(S_i)_{i \in \mathrm{N}}$ are constructed such that, for each $i \in \mathrm{N}$, the set of normal forms $\mathrm{NF}_R(S_i)$ is a singleton, but it is undecidable in general whether $\Delta_R^*(S_i)$ is a regular language.

Finally, using the undecidability of the *strong boundedness property* of single-tape Turing machines we prove that it is undecidable in general whether a finite, length-reducing, and confluent string-rewriting system R yields a regular set of normal forms $\mathrm{NF}_R(S)$ for each regular language S.

This paper is organized as follows. In Section 2 we restate the basis definitions used in short in order to establish notation. In Section 3 we prove the above-mentioned conjecture of Gyenizse and Vágvölgyi, and in Section 4 we present the first of the undecidability results mentioned above. Then in Section 5 we consider the strong boundedness problem for single-tape Turing machines, and finally in Section 6 we reduce this problem to the problem of deciding whether a finite, length-reducing, and confluent string-rewriting system yields a regular set of normal forms for each regular language. The paper closes with a short discussion of some open problems.

2 String-rewriting systems and monoid presentations

After establishing notation we describe the problems considered in this paper in detail. For more information and a detailed discussion of the notions introduced the reader is refered to the literature, e.g., [BO93].

Let Σ be a finite alphabet. Then Σ^* denotes the set of all strings over Σ including the empty string λ. A *string-rewriting system* (srs) R on Σ is a subset of $\Sigma^* \times \Sigma^*$, the elements of which are called (*rewrite*) *rules*. The *domain* of R

is the set $\mathrm{dom}(R) := \{\ell \mid \exists r \in \Sigma^* : (\ell \to r) \in R\}$. The srs R is called *length-reducing* if $|\ell| > |r|$ holds for each rule $(\ell \to r)$ of R, and it is called *monadic* if it is length-reducing, and $r \in \Sigma \cup \{\lambda\}$ for each rule $(\ell \to r)$ of R.

The *single-step reduction relation* induced by R is the binary relation $\to_R := \{(x\ell y, xry) \mid x, y \in \Sigma^*, (\ell \to r) \in R\}$. Its reflexive and transitive closure $\to_R^*$ is the *reduction relation* induced by R, and its reflexive, symmetric, and transitive closure $\leftrightarrow_R^*$ is the *Thue congruence* generated by R. For $u \in \Sigma^*$, $\Delta_R^*(u) := \{v \in \Sigma^* \mid u \to_R^* v\}$ is the *set of descendants* of u $(\mathrm{mod}\, R)$, and $[u]_R := \{v \in \Sigma^* \mid u \leftrightarrow_R^* v\}$ is the *congruence class* of u $(\mathrm{mod}\, R)$. For $L \subseteq \Sigma^*$, $\Delta_R^*(L) = \bigcup_{u \in L} \Delta_R^*(u)$ and $[L]_R := \bigcup_{u \in L} [u]_R$.

Let Σ be a finite alphabet, and let R be a srs on Σ. We say that R *preserves Σ-regularity* if, for each regular language $L \subseteq \Sigma^*$, $\Delta_R^*(L)$ is again a regular language. We say that R *preserves regularity* if R preserves Γ-regularity for each finite alphabet Γ containing all the letters that have occurrences in the rules of R. In the next section we will investigate the relationship between the property of preserving regularity and the property of preserving Σ-regularity, where Σ is the smallest alphabet that contains all the letters with occurrences in R. After that we will address the following decision problems:

Problem 1: Preserving regularity.

Instance: A finite srs R.

Question: Does R preserve regularity?

Problem 2: Regular set of descendants for a given language.

Instance: A finite srs R on Σ, and a regular language $L \subseteq \Sigma^*$.

Question: Is $\Delta_R^*(L)$ a regular language?

A string $u \in \Sigma^*$ is called *reducible* $(\mathrm{mod}\, R)$, if there exists a string $v \in \Sigma^*$ such that $u \to_R v$; otherwise, u is called *irreducible*. By $\mathrm{IRR}(R)$ we denote the set of all irreducible strings, and by $\mathrm{RED}(R)$ we denote the set of strings that are reducible $(\mathrm{mod}\, R)$.

If $u \in \Sigma^*$ and $v \in \mathrm{IRR}(R)$ such that $u \to_R^* v$, then v is called a *normal form* of u. Accordingly, $\mathrm{NF}_R(u) := \Delta_R^*(u) \cap \mathrm{IRR}(R)$ is the set of all normal forms of u, and for $L \subseteq \Sigma^*$, $\mathrm{NF}_R(L) := \Delta_R^*(L) \cap \mathrm{IRR}(R)$. If a finite srs R preserves regularity, then the set $\mathrm{NF}_R(L)$ is a regular language for each regular language $L \subseteq \Sigma^*$. However, $\mathrm{NF}_R(L)$ can be a regular language, even if $\Delta_R^*(L)$ is not. Thus, also the following decision problems are of interest:

Problem 3: Regular sets of normal forms.

Instance: A finite srs R on Σ.

Question: Is $\mathrm{NF}_R(L)$ regular for each regular language $L \subseteq \Sigma^*$?

Problem 4: Regular set of normal forms for a given language.

Instance: A finite srs R on Σ, and a regular language $L \subseteq \Sigma^*$.

Question: Is $\mathrm{NF}_R(L)$ a regular language?

The monoid $M_R := \Sigma^* / \leftrightarrow_R^*$ is uniquely determined by Σ and R (up to isomorphism). Hence, the pair $(\Sigma; R)$ is called a *monoid-presentation* of M_R.

We are in particular interested in the decision problems above for those finite srss that are length-reducing and confluent. Here a srs R on Σ is called *confluent*, if, for all $u, v, w \in \Sigma^*$, $u \to_R^* v$ and $u \to_R^* w$ imply that $v \to_R^* z$ and $w \to_R^* z$ hold for some $z \in \Sigma^*$.

If R is length-reducing and confluent, then each congruence class $[u]_R$ contains a unique irreducible string u_0, and u_0 can be obtained from u by a finite sequence of reduction steps. In fact, the word problem is decidable in linear time for each finite, length-reducing, and confluent srs [BO93].

3 Independence of the alphabet considered

A signature F consists of a (finite) set of function symbols, each equipped with a fixed arity. If $F := \{f, g, a\}$, where f and g are unary symbols and a is a symbol of arity 0 (a constant), then for $R := \{f(g(x)) \to f(f(g(g(x)))), f(a) \to a, g(a) \to a, a \to f(a), a \to g(a)\}$ it is easily seen that $\Delta_R^*(t) = T(F)$ holds for all ground terms $t \in T(F)$. However, if $F_1 := F \cup \{h\}$, where h is another unary function symbol, then $\Delta_R^*(\{f(g(h(a)))\}) = \{f^n(g^n(h(t))) \mid t \in T(F)\}$, which is not regular. Thus, R does preserve F-regularity, but not F_1-regularity [GV97].

Based on the following technical lemma we obtain a simple proof that this phenomenon does not occur for string-rewriting systems.

Lemma 1. *Let R be a srs on some finite alphabet Σ, let a be an additional letter not in Σ, and let $\Gamma := \Sigma \cup \{a\}$. If R preserves Σ-regularity, then it also preserves Γ-regularity.*

Proof. Let $L \subseteq \Gamma^*$ be a regular language. We must verify that, under the hypothesis that R preserves Σ-regularity, $\Delta_R^*(L) := \{v \in \Gamma^* \mid \exists u \in L : u \to_R^* v\}$ is a regular language.

Assume that $L \not\subseteq \Sigma^*$. There exists a deterministic finite state-acceptor (fsa) $A = (Q, \Gamma, q_0, F, \delta)$ such that $L(A) = L$. Let (1.) $q_1 \xrightarrow{a} p_1$, (2.) $q_2 \xrightarrow{a} p_2, \ldots,$ (m.) $q_m \xrightarrow{a} p_m$ be the set of all a-transitions of A, which are numbered in an arbitrary, but fixed way. Observe that the states $q_1, \ldots, q_m$ are pairwise distinct, but that possibly some of the states $p_1, \ldots, p_m$ coincide. Based on A we define some auxiliary languages.

$$\begin{array}{lll}
L_0 & := \{w \in \Sigma^* \mid \delta(q_0, w) \in F\} = L \cap \Sigma^*, & \\
P_i & := \{w \in \Sigma^* \mid \delta(q_0, w) = q_i\}, \;\; i = 1, \ldots, m, & \\
S_j & := \{w \in \Sigma^* \mid \delta(p_j, w) \in F\}, \;\; j = 1, \ldots, m, & \\
K_{i,j,k} & := \{w \in \Gamma^* \mid \delta(p_i, w) = q_j \text{ and, for all } u, v \in \Gamma^*, \text{ if } w = uav, \text{ then} & \\
 & \qquad\qquad\qquad \delta(p_i, u) = q_\ell \text{ for some } \ell \leq k\}, \;\; i, j = 1, \ldots, m, & \\
 & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad k = 0, 1, \ldots, m. &
\end{array}$$

Thus, $w \in K_{i,j,k}$ if and only if $\delta(p_i, w) = q_j$, and the only a-transitions that are used by following the computation of $\delta(p_i, w)$ are those with number at most k.

Obviously, the languages L_0, P_i $(i = 1, \ldots, m)$, and S_j $(j = 1, \ldots, m)$ are regular. In addition, $K_{i,j,k}$ is accepted by the automaton $(Q, \Gamma, p_i, \{q_j\}, \delta_k)$, where δ_k is obtained from δ by removing the a-transitions with number larger than k. Hence, these languages are regular, too.

Since $L = L_0 \cup \bigcup_{i=1}^{m} (P_i \cdot \{a\} \cdot S_i) \cup \bigcup_{i,j=1}^{m} (P_i \cdot \{a\} \cdot K_{i,j,m} \cdot \{a\} \cdot S_j)$, and since the operations of union and of computing descendants mod R commute, we obtain

$$\Delta_R^*(L) = \Delta_R^*(L_0) \cup \bigcup_{i=1}^{m} \Delta_R^*(P_i \cdot \{a\} \cdot S_i) \cup \bigcup_{i,j=1}^{m} \Delta_R^*(P_i \cdot \{a\} \cdot K_{i,j,m} \cdot \{a\} \cdot S_j).$$

The next two claims state that also the operations of concatenation and of computing descendants mod R commute in certain instances. They follow from the fact that no rule of R contains an occurrence of the symbol a.

Claim 1: $\Delta_R^*(P_i \cdot \{a\} \cdot S_i) = \Delta_R^*(P_i) \cdot \{a\} \cdot \Delta_R^*(S_i)$.

Claim 2: $\Delta_R^*(P_i \cdot \{a\} \cdot K_{i,j,m} \cdot \{a\} \cdot S_j) = \Delta_R^*(P_i) \cdot \{a\} \cdot \Delta_R^*(K_{i,j,m}) \cdot \{a\} \cdot \Delta_R^*(S_j)$.

The languages $\Delta_R^*(L_0)$, $\Delta_R^*(P_i)$, and $\Delta_R^*(S_i)$ $(i = 1, \ldots, m)$ are regular by our assumption. It remains to prove the following claim.

Claim 3: $\Delta_R^*(K_{i,j,k})$ is a regular set for all $i, j = 1, \ldots, m$, and $k = 0, 1, \ldots, m$.

Proof. This claim can easily be proved by induction on k, since $K_{i,j,0} \subseteq \Sigma^*$, $K_{i,j,1} = K_{i,j,0} \cup K_{i,1,0} \cdot (\{a\} \cdot K_{1,1,0})^* \cdot \{a\} \cdot K_{1,j,0}$, and $K_{i,j,k+1} = K_{i,j,k} \cup K_{i,k+1,k} \cdot (\{a\} \cdot K_{k+1,k+1,k})^* \cdot \{a\} \cdot K_{k+1,j,k}$ for all $i, j = 1, \ldots, m$ and $k = 0, 1, \ldots, m-1$. □

It follows that the set $\Delta_R^*(L)$ is indeed regular. □

From this lemma we obtain Gynizse's and Vágvölgyi's conjecture.

Theorem 2. *Let R be a srs, and let Σ be the alphabet consisting of all the letters that have occurrences in R. Assume that Σ is finite. Then R preserves Σ-regularity if and only if R preserves regularity.*

Observe that the proof of Lemma 1 is constructive in that, if R preserves Σ-regularity in an effective way, then it preserves Γ-regularity in an effective way, too. That is, if a regular language $L \subseteq \Gamma^*$ is given through some fsa, then a fsa for the language $\Delta_R^*(L)$ can be constructed effectively.

Let R be a srs on Σ. Consider the signature $F_\Sigma := \{a(.) \mid a \in \Sigma\} \cup \{¢\}$, where each letter $a \in \Sigma$ is interpreted as a unary function symbol $a(.)$ and $¢$ is a constant, and let $S_R := \{\ell(x) \to r(x) \mid (\ell \to r) \in R\}$ be the corresponding term-rewriting system (trs). Then the trs S_R preserves F_Σ-regularity if and only if the srs R preserves (Σ-)regularity.

Now S_R preserves regularity if it preserves F-regularity for each signature F containing F_Σ. Observe that F may contain additional constants and function symbols of arity larger than one in contrast to the case of strings considered above. Nevertheless, using the same basic idea the following can be shown.

Theorem 3. *The trs* S_R *preserves regularity if and only if it preserves* F_Σ-*regularity.*

Together with Theorem 6 below this gives the following negative answer to a question of [GV97].

Corollary 4. *It is undecidable in general whether a finite term-rewriting system preserves regularity.*

4 The property of preserving regularity is undecidable

Here we establish the first of the announced undecidability results.

Lemma 5. *If* R *is a finite srs such that the monoid* M_R *is a group, then the following two statements are equivalent:*

(a) the srs $R \cup R^{-1}$ *preserves regularity,*
(b) the group M_R *is finite.*

Here R^{-1} *denotes the srs* $R^{-1} := \{r \rightarrow \ell \mid (\ell \rightarrow r) \in R\}$.

Proof. (b) ⇒ (a): Let $L \subseteq \Sigma^*$. Since M_R is a finite group, there are finitely many strings $w_1, \ldots, w_n \in L$ such that $\Delta^*_{R \cup R^{-1}}(L) = [L]_R = \bigcup_{i=1,\ldots,n} [w_i]_R$, and $[w]_R$ is a regular language for each $w \in \Sigma^*$. Thus, $\Delta^*_{R \cup R^{-1}}(L) = \bigcup_{i=1,\ldots,n} [w_i]_R$ is a regular language.

(a) ⇒ (b): The singleton set $\{\lambda\} \subseteq \Sigma^*$ is a regular language, and $\Delta^*_{R \cup R^{-1}}(\lambda) = [\lambda]_R$. Hence, $[\lambda]_R$ is a regular language, implying that M_R is a regular group. However, a finitely presented group is regular if and only if it is finite [Ani71]. Hence, M_R is a finite group. □

The property of being finite is a Markov property of finitely presented groups, and hence, it is undecidable in general (see, e.g., [LS77]). Since Lemma 5 reduces this undecidable problem to the problem of deciding whether or not a finite srs preserves regularity, this gives the following result.

Theorem 6. *For finite srss the problem of preserving regularity is undecidable in general.*

The proof of Lemma 5 shows that the group M_R is finite if and only if the language $\Delta^*_{R \cup R^{-1}}(\lambda)$ is regular. Thus, we even obtain the following stronger undecidability result.

Corollary 7. *The following problem is undecidable in general:*

Instance: *A finite srs* R *on* Σ, *and a regular language* $L \subseteq \Sigma^*$.
Question: *Is the language* $\Delta^*_R(L)$ *regular?*

In fact, Corollary 7 remains valid even if the language L is fixed to the set $L := \{\lambda\}$. Theorem 6 and Corollary 7 improve upon Theorem 7 and Theorem 8 of [GT95], since here we are only dealing with strings, that is, with signatures containing only unary function symbols and possibly a single constant.

Our next undecidability result improves upon Corollary 7. It states that the problem considered in Corollary 7 remains undecidable even if R is restricted to be a finite, length-reducing, and confluent srs. Actually, a fixed system R can be chosen here, if it is constructed accordingly. Below such a construction is given.

Let $M = (Q, \Sigma, \delta, q_0, q_a)$ be a deterministic single-tape Turing machine (TM) that accepts a language $L \subseteq \Sigma^*$, where $\Sigma_b := \Sigma \cup \{b\}$ is the tape alphabet, b denotes the blank symbol, Q is the set of states, $\delta : \Sigma_b \times (Q \smallsetminus \{q_a\}) \rightarrow Q \times (\Sigma \cup \{\ell, r\})$ is the transition function, $q_0 \in Q$ is the initial state, $q_a \in Q$ is the final state, and $\vdash_M^*$ denotes the reflexive and transitive closure of the single-step computation relation $\vdash_M$ of M. Observe that we assume without loss of generality that the TM M cannot print the blank symbol b, that is, each tape square that has been visited by the head of M contains a non-blank symbol afterwards. Thus, for all $w \in \Sigma^*$, $w \in L$ if and only if $q_0 w \vdash_M^* u q_a v$ for some $u, v \in \Sigma^*$. Without loss of generality we may also assume that $\Sigma_b \cap Q = \emptyset$.

From M we construct another single-tape TM $\overline{M} = (\overline{Q}, \Gamma, \bar{\delta}, q_0, \underline{q_a})$. Let $\Gamma := \Sigma \cup \{1, 2, \uparrow\}$, where 1, 2, and $\uparrow$ are three new symbols, and define $\overline{Q}$ and $\bar{\delta}$ in such a way that $\overline{M}$ simulates the TM M as follows:

Whenever $q_0 w \vdash_M^n \underline{u_1} q_1 v_1 \vdash_M u_2 q_2 v_2$, where $w \in \Sigma^*$, $u_1, v_1, u_2, v_2 \in \Sigma^*$, $q_1, q_2 \in Q$, and $n \geq 0$, then $\overline{M}$ performs the following computation:

$$\begin{aligned} q_0 w \vdash_{\overline{M}}^* 1^n 2^n u_1 q_1 v_1 \;\; &\vdash_{\overline{M}}^* \;\; \tilde{q}_1 1^n 2^n u_1 \uparrow v_1 \;\; \vdash_{\overline{M}}^* \;\; 1^{n+1} 2^{n+1} \hat{q}_1 u_1 \uparrow v_1 \\ \vdash_{\overline{M}}^* 1^{n+1} 2^{n+1} u_1 \bar{q}_1 v_1 \;\; &\vdash_{\overline{M}} \;\; 1^{n+1} 2^{n+1} u_2 q_2 v_2 . \end{aligned}$$

Thus, if $q_0 w \vdash_M^n uqv$ for some $\underline{w} \in \Sigma^*$, $u, v \in \Sigma^*$, $q \in Q$, and $n \in \mathbf{N}$, then $q_0 w \vdash_{\overline{M}}^* 1^n 2^n uqv$. In particular, $\overline{M}$ also accepts the language L.

For $w \in \Sigma^*$, let $\Delta_{\overline{M}}(w) := \{uqv \mid q_0 w \vdash_{\overline{M}}^* uqv \in \Gamma^* \cdot \overline{Q} \cdot \Gamma^*\}$.

Lemma 8. *For $w \in \Sigma^*$, the following two statements are equivalent:*

(a) $\Delta_{\overline{M}}(w)$ is a regular language;
(b) $w \in L$.

Proof. (b) $\Rightarrow$ (a): If $w \in L$, then $\overline{M}$ accepts on input w, that is, $\overline{M}$ halts on input w after performing a finite number of steps. Thus, the set $\Delta_{\overline{M}}(w)$ is finite.

(a) $\Rightarrow$ (b): If $w \notin L$, then $\overline{M}$ does not halt on input w, and the same is true for M. Thus, there is an infinite computation of the form $q_0 w \vdash_M u_1 q_1 v_1 \vdash_M u_2 q_2 v_2 \vdash_M \ldots$. Hence, $\overline{M}$ performs an infinite computation of the following form: $q_0 w \vdash_{\overline{M}}^* 12 u_1 q_1 v_1 \vdash_{\overline{M}}^* 1^2 2^2 u_2 q_2 v_2 \vdash_{\overline{M}}^* \ldots$. Consider the set $\Delta'(w) := \Delta_{\overline{M}}(w) \cap 1^+ \cdot 2^+ \cdot ((\Gamma \smallsetminus \{1, 2\}) \cup \overline{Q})^+$. Since $\Delta'(w) = \{1^i 2^i u \mid i > 0, u \in C(u_i q_i v_i)\}$ for some finite sets $C(u_i q_i v_i) \subset ((\Gamma \smallsetminus \{1, 2\}) \cup \overline{Q})^+$, $i > 0$, this set does not satisfy the pumping lemma for regular languages, and hence, it is not regular. Accordingly, the set $\Delta_{\overline{M}}(w)$ is not regular, either. □

From the TM $\overline{M}$ we now construct a finite, length-reducing, and confluent srs $R(\overline{M})$ that simulates the computations of $\overline{M}$. Let \$, ¢, and d be three additional symbols, and let $\Gamma_0 := \Gamma_b \cup \overline{Q} \cup \{\$, \text{¢}, d\}$. The symbols \$ and ¢ will serve as left and right end markers, respectively, of configurations of $\overline{M}$, while the symbol d is being used to ensure that the rules of $R(\overline{M})$ are length-reducing. The system $R(\overline{M})$ consists of the following three groups of rules:

(1.) Rules to simulate the stepwise behaviour of $\overline{M}$:

$$\begin{array}{rll}
q_i a_k dd \to q_j a_\ell & \text{if } \bar{\delta}(q_i, a_k) = (q_j, a_\ell) \\
q_i \text{¢} dd \to q_j a_\ell \text{¢} & \text{if } \bar{\delta}(q_i, b) = (q_j, a_\ell) \\
q_i a_k dd \to a_k q_j & \text{if } \bar{\delta}(q_i, a_k) = (q_j, r) \\
q_i \text{¢} dd \to b q_j \text{¢} & \text{if } \bar{\delta}(q_i, b) = (q_j, r)
\end{array}$$

$$\left.\begin{array}{rll}
a_\ell q_i a_k dd \to q_j a_\ell a_k & \text{if } \bar{\delta}(q_i, a_k) = (q_j, \ell) \\
a_\ell q_i \text{¢} dd \to q_j a_\ell \text{¢} & \text{if } \bar{\delta}(q_i, b) = (q_j, \ell)
\end{array}\right\} a_\ell \in \Gamma_b$$

$$\begin{array}{rll}
\$ q_i a_k dd \to \$ q_j b a_k & \text{if } \bar{\delta}(q_i, a_k) = (q_j, \ell) \\
\$ q_i \text{¢} dd \to \$ q_j b \text{¢} & \text{if } \bar{\delta}(q_i, b) = (q_j, \ell)
\end{array}$$

(2.) Rules to shift occurrences of the symbol d to the left:

$$\left.\begin{array}{r}
a_i a_j dd \to a_i d a_j \\
a_i d a_j dd \to a_i dd a_j
\end{array}\right\} a_i \in \Gamma_b, a_j \in \Gamma_b \cup \{\text{¢}\}$$

(3.) Rules to erase halting configurations:

$$\begin{array}{r}
q_a a_i dd \to q_a \\
a_i q_a \text{¢} d \to q_a \text{¢} \\
\$ q_a \text{¢} d \to \$ q_a \text{¢}
\end{array}$$

It is rather straightforward to verify that the system $R(\overline{M})$ has the following properties.

Proposition 9. (cf. [Ott91], Proposition 3.1)

(a) The srs $R(\overline{M})$ is finite, length-reducing, and confluent.
(b) For $w \in \Sigma^*$, the following two statements are equivalent:
  (1.) $w \in L$; and
  (2.) $\exists m \in \mathbb{N}\ \forall n \geq m : \$ q_0 w \text{¢} d^n \to^*_{R(\overline{M})} \$ q_a \text{¢} d^{n-m} \to^*_{R(\overline{M})} \$ q_a \text{¢}$.

From Lemma 8 and Proposition 9(b) we obtain the following.

Lemma 10. *For $w \in \Sigma^*$, the following two statements are equivalent:*

(a) $\Delta^*_{R(\overline{M})}(\$ q_0 w \text{¢} \cdot d^*)$ is a regular language;
(b) $w \in L$.

For a nonrecursive language L, this yields the following undecidability result.

Theorem 11. *There exists a finite, length-reducing, and confluent srs R such that the following problem is undecidable:*

Instance: *A regular set $S \subseteq \Sigma^*$.*

Question: *Is the set of descendants $\Delta_R^*(S)$ regular?*

If $w \in L$, then $\Delta^*_{R(\overline{M})}(\$q_0 w\text{¢} \cdot d^n) \cap \mathrm{IRR}(R(\overline{M})) = \{\$q_a\text{¢}\}$ for all $n \geq m$. Thus, $\mathrm{NF}_{R(\overline{M})}(\$q_0 w\text{¢} \cdot d^*) := \Delta^*_{R(\overline{M})}(\$q_0 w\text{¢} \cdot d^*) \cap \mathrm{IRR}(R(\overline{M}))$ is a finite set, and hence, it is regular. On the other hand, if $w \notin L$, then $\mathrm{NF}_{R(\overline{M})}(\$q_0 w\text{¢} \cdot d^*)$ is not regular, that is, the set of normal forms in the language $\Delta^*_{R(\overline{M})}(\$q_0 w\text{¢} \cdot d^*)$ is regular if and only if $w \in L$. Thus, we obtain the following corollary, which improves upon Theorem 10 of [GT95].

Corollary 12. *There exists a finite, length-reducing, and confluent srs R such that the following problem is undecidable:*

Instance: *A regular set $S \subseteq \Sigma^*$.*

Question: *Is the set $\mathrm{NF}_R(S)$ regular?*

For the srs $R(\overline{M})$ and the languages of the form $\$q_0 w\text{¢} \cdot d^*$ $(w \in \Sigma^*)$, we have that $\mathrm{NF}_{R(\overline{M})}(\$q_0 w\text{¢} \cdot d^*)$ is regular if and only if $\Delta^*_{R(\overline{M})}(\$q_0 w\text{¢} \cdot d^*)$ is regular. However, we can avoid this equivalence.

Let $\Gamma_1 := \Gamma_0 \cup \{\#, z\}$, where $\#$ and z are two new symbols, and let $R_0(\overline{M}) := R(\overline{M}) \cup R_0$, where $R_0 := \{\text{¢}\# \rightarrow z, \text{¢}d\# \rightarrow z, za \rightarrow z, az \rightarrow z \mid a \in \Gamma_1\}$. Then $R_0(\overline{M})$ is a finite length-reducing srs, and it can easily be verified that $R_0(\overline{M})$ is confluent.

For $w \in \Sigma^*$, consider the language $S(w) := \$q_0 w\text{¢} \cdot d^* \cdot \#$. If $w \in L$, and if n is sufficiently large, then $\$q_0 w\text{¢} \cdot d^n \cdot \# \rightarrow^*_{R(\overline{M})} \$q_a\text{¢}\# \rightarrow^*_{R(\overline{M})} z$. If n is small, then $\$q_0 w\text{¢} \cdot d^n \cdot \# \rightarrow^*_{R(\overline{M})} \$uqv\text{¢}d^\varepsilon\# \rightarrow^*_{R(\overline{M})} z$, where $\varepsilon \in \{0, 1\}$. In this situation, $\Delta^*_{R_0(\overline{M})}(S(w))$ is a regular language, and $\mathrm{NF}_{R_0(\overline{M})}(S(w)) = \{z\}$. If $w \notin L$, then, for all $n \in \mathbb{N}$, $\$q_0 w\text{¢} \cdot d^n \cdot \# \rightarrow^*_{R(\overline{M})} \$uqv\text{¢}d^\varepsilon\# \rightarrow^*_{R(\overline{M})} z$, where $\varepsilon \in \{0, 1\}$. The rules of R_0 cannot be used before all the d's (but one) to the right of the ¢-symbol have been used up. Thus, it follows that $\Delta^*_{R_0(\overline{M})}(S(w))$ is not regular. However, $\mathrm{NF}_{R_0(\overline{M})}(S(w)) = \{z\}$ also holds in this situation. Thus, the following undecidability result follows.

Theorem 13. *There exists a finite, length-reducing, and confluent srs R such that the following problem is undecidable:*

Instance: *A regular set $S \subseteq \Sigma^*$ such that $\mathrm{NF}_R(S)$ is a singleton.*

Question: *Is the set of descendants $\Delta_R^*(S)$ regular?*

So far we have seen that Problem 2 and Problem 4 are undecidable, even for a fixed finite srs that is length-reducing and confluent (Theorem 11 and Corollary 12). In the remaining part of this paper we consider Problem 3. For that, however, we will need the strong boundedness problem for single-tape Turing machines.

5 The strong boundedness problem for Turing machines

First it should be stressed that we are only dealing with TMs that are deterministic. A possibly infinite configuration C of a single-tape TM M is called *immortal* if M does never halt when starting with C. In [Hoo66] Hooper shows that it is undecidable whether a TM has an immortal configuration. Actually, Hooper only considers *2-symbol TMs*, that is, TMs that only have a single tape symbol in addition to the blank symbol.

We call a TM M *strongly bounded* if there exists an integer k such that, for each finite configuration C, M halts after at most k steps when starting in configuration C. Here a configuration is called finite if almost all tape squares contain the blank symbol b. We are interested in the *strong boundedness problem* for TMs, which is the following decision problem:

Instance: A single-tape TM M.

Question: Is M strongly bounded?

In [Hoo66] Hooper describes a construction of a 2-symbol TM $\overline{M}$ from a two-counter Minsky machine $\hat{M}$ such that $\overline{M}$ has an immortal finite configuration if and only if it has an immortal configuration if and only if $\hat{M}$ does not halt from its initial configuration with empty counters. Since the halting problem is undecidable even for this restricted class of Minsky machines, it follows that the immortality problem is undecidable for 2-symbol TMs, even when it is restricted to finite configurations.

Now assume that the TM $\overline{M}$ has finite computations of arbitrary length. Then it must also have an infinite computation, though one that possibly starts with an infinite configuration (cf. the proof of Corollary 6 of [KTU93]). But then $\overline{M}$ also has an infinite computation that starts with a finite configuration. Thus, if $\overline{M}$ has no immortal finite configuration, then it is strongly bounded. Since the converse is obvious, we obtain the following undecidability result.

Proposition 14. *The strong boundedness problem is undecidable for 2-symbol single-tape TMs.*

6 The reduction

We will prove that Problem 3 is undecidable for finite srss that are length-reducing and confluent by a reduction from the strong boundedness problem for TMs. For this, we use a simulation of TMs through finite, length-reducing, and confluent srss that is based on the simple simulation given in Section 4.

Let $M = (Q, \Sigma, q_0, q_a, \delta)$ be a deterministic single-tape TM, where we assume that Σ consists of a symbol a and the blank symbol b only. From M we now construct a finite srs R for simulating M.

Let $\overline{\Sigma} := \{\bar{a}, \bar{b}\}$, let $\overline{Q}$ be another new alphabet that is in 1-to-1 correspondence to Q, and let $\Gamma := Q \cup \overline{Q} \cup \Sigma \cup \overline{\Sigma} \cup \{1, 2, \$, ¢, d, \bar{d}, \hat{d}, d_0, \hat{c}, \bar{c}, 0\}$, where $1, 2, \$, ¢, d, \bar{d}, \hat{d}, d_0, \hat{c}, \bar{c}, 0$ are 11 additional new symbols.

The srs R will consist of two main parts, that is, $R := R_1 \cup R_2$, where R_1 is a system that simulates the computations of the TM M step by step, and R_2 is a system that destroys unwanted strings. We first define the system R_1. It consists of the following 5 groups of rules.

(1.) Rules to simulate the TM M:

$$\begin{array}{lll}
q_i a_k d d a_r & \to \bar{q}_j a_\ell a_r & a_r \in \Sigma \cup \{\cent\}, \text{if } \delta(q_i, a_k) = (q_j, a_\ell) \\
q_i \cent d_0 d_0 & \to \bar{q}_j a_\ell \cent & \text{if } \delta(q_i, b) = (q_j, a_\ell) \\
q_i a_k d d a_r & \to \bar{a}_k \bar{q}_j a_r & a_r \in \Sigma \cup \{\cent\}, \text{if } \delta(q_i, a_k) = (q_j, r) \\
q_i \cent d_0 d_0 & \to \bar{b} \bar{q}_j \cent & \text{if } \delta(q_i, b) = (q_j, r) \\
\bar{a}_\ell q_i a_k d d a_r & \to \bar{q}_j a_\ell a_k a_r & a_r \in \Sigma \cup \{\cent\}, \text{if } \delta(q_i, a_k) = (q_j, \ell) \\
\bar{a}_\ell q_i \cent d_0 d_0 & \to \bar{q}_j a_\ell \cent & \text{if } \delta(q_i, b) = (q_j, \ell) \\
\$ q_i a_k d d a_r & \to \$ \bar{q}_j b a_k a_r & a_r \in \Sigma \cup \{\cent\}, \text{if } \delta(q_i, a_k) = (q_j, \ell) \\
\$ q_i \cent d_0 d_0 & \to \$ \bar{q}_j b \cent & \text{if } \delta(q_i, b) = (q_j, \ell).
\end{array}
\quad \left.\begin{array}{l} \\ \\ \end{array}\right\} \bar{a}_\ell \in \overline{\Sigma}$$

(2.) Rules to shift d to the left:

$$\left.\begin{array}{ll}
a_i a_j d d a_r & \to a_i d a_j a_r \\
a_i d a_j d d a_r & \to a_i d d a_j a_r
\end{array}\right\} a_i \in \Sigma \cup \overline{Q}, a_j \in \Sigma, \text{ and } a_r \in \Sigma \cup \{\cent\}$$

$$\left.\begin{array}{ll}
a_i \cent d_0 d_0 & \to a_i d \cent \\
a_i d \cent d_0 d_0 & \to a_i d d \cent
\end{array}\right\} a_i \in \Sigma \cup \overline{Q}$$

$$\left.\begin{array}{ll}
\bar{a}_i \bar{q}_j d d a_r & \to \bar{a}_i \bar{d} \bar{q}_j a_r \\
\bar{a}_i \bar{d} \bar{q}_j d d a_r & \to \bar{a}_i \bar{d} \bar{d} \bar{q}_j a_r
\end{array}\right\} \bar{a}_i \in \overline{\Sigma} \cup \{\$\}, \bar{q}_j \in \overline{Q}, \text{ and } a_r \in \Sigma \cup \{\cent\}$$

(3.) Rules to shift $\bar{d}$ and $\hat{d}$ to the left:

$$\left.\begin{array}{ll}
\bar{a}_i \bar{a}_j \bar{d} \bar{d} \bar{a}_r & \to \bar{a}_i \bar{d} \bar{a}_j \bar{a}_r \\
\bar{a}_i \bar{d} \bar{a}_j \bar{d} \bar{d} \bar{a}_r & \to \bar{a}_i \bar{d} \bar{d} \bar{a}_j \bar{a}_r
\end{array}\right\} \bar{a}_i \in \overline{\Sigma} \cup \{\$\}, \bar{a}_j \in \overline{\Sigma}, \text{ and } \bar{a}_r \in \overline{\Sigma} \cup \overline{Q} \cup Q$$

$$\left.\begin{array}{ll}
2\$ \bar{d} \bar{d} \bar{a}_r & \to 2 \hat{d} \$ \bar{a}_r \\
2 \hat{d} \$ \bar{d} \bar{d} \bar{a}_r & \to 2 \hat{d} \hat{d} \$ \bar{a}_r
\end{array}\right\} \bar{a}_r \in \overline{\Sigma} \cup \overline{Q} \cup Q$$

$$\left.\begin{array}{ll}
22 \hat{d} \hat{d} a & \to 2 \hat{d} 2 a \\
2 \hat{d} 2 \hat{d} \hat{d} a & \to 2 \hat{d} \hat{d} 2 a
\end{array}\right\} a \in \{2, \$\}$$

(4.) Rules to increase the number of 1's and 2's:

$$\left.\begin{array}{ll}
12 \hat{d} \hat{d} a & \to 1 \hat{d} 2 a \\
1 \hat{d} 2 \hat{d} \hat{d} a & \to 11 \hat{c} 2 a \\
1 \hat{c} 2 \hat{d} \hat{d} a & \to 122 \hat{c} a
\end{array}\right\} a \in \{2, \$\}$$

(5.) Rules to shift $\hat{c}$ and $\bar{c}$ to the right:

$$\begin{array}{lll}
2 \hat{c} 2 \hat{d} \hat{d} a & \to 22 \hat{c} a & a \in \{2, \$\} \\
2 \hat{c} \$ \bar{d} \bar{d} \bar{a} & \to 2 \$ \bar{c} \bar{a} & \multirow{2}{*}{$\Big\}\ \bar{a} \in \overline{\Sigma} \cup \overline{Q} \cup Q, \bar{a}_i \in \overline{\Sigma} \cup \{\$\}, \text{ and } \bar{a}_j \in \overline{\Sigma}$} \\
\bar{a}_i \bar{c} \bar{a}_j \bar{d} \bar{d} \bar{a} & \to \bar{a}_i \bar{a}_j \bar{c} \bar{a} & \\
\bar{a}_i \bar{c} \bar{q}_j d d a & \to \bar{a}_i q_j a & \bar{a}_i \in \overline{\Sigma} \cup \{\$\}, \bar{q}_j \in \overline{Q}, \text{ and } a \in \Sigma \cup \{\cent\}.
\end{array}$$

Obviously, R_1 is a finite and length-reducing system. Since the TM M is deterministic, it is easily verified that R_1 is in addition confluent.

The rules of group (1.) of R_1 simulate the stepwise computation of M. The auxiliary symbols $d_0, d, \bar{d}$, and $\hat{d}$ ensure that R_1 is length-reducing, and the rules of (2.) and (3.) shift occurrences of these auxiliary symbols to the left. After simulating a step of M, the prefix $1^{\ell}2^{k}$ of the encoding of the actual configuration of M is incremented to $1^{\ell+1}2^{k+1}$ through the rules of group (4.). Finally, the auxiliary symbols $\hat{c}$ and $\bar{c}$, and the copies $\bar{Q}$ of the actual state symbols Q are used to ensure that the next step of M can be simulated only after the prefix $1^{\ell}2^{k}$ has been incremented (see (5.)). Thus, the following technical result can be established for R_1, where $^{-} : \Sigma^* \to \overline{\Sigma}^*$ denotes the canonical bijection.

Lemma 15. *Let uqv be a configuration of the TM M such that $uqv \vdash_M u_1q_1v_1$, and let $k \geq 1$. Then there exists an integer $p > 0$ such that $1^{\ell}2^{k}\$\bar{u}qv\text{¢}d_0^p \to_{R_1}^* 1^{\ell+1}2^{k+1}\$\bar{u}_1q_1v_1\text{¢}$ holds for all $\ell \geq 1$.*

Let $u_0q_0v_0$ is an immortal finite configuration of M, that is, M has an infinite computation of the form $u_0q_0v_0 \vdash_M u_1q_1v_1 \vdash_M u_2q_2v_2 \vdash_M \ldots$. Consider the regular language $S := \{12\$\bar{u}_0q_0v_0\text{¢} \cdot d_0^i \mid i \geq 0\}$. From Lemma 15 we see that, for each $k \geq 1$, there exists an integer $p_k \in \mathbf{N}_+$ such that $12\$\bar{u}_0q_0v_0\text{¢} \cdot d_0^{p_k} \to_{R_1}^* 1^{k+1}2^{k+1}\$\bar{u}_kq_kv_k\text{¢}$. Hence, it follows that $\Delta_{R_1}^*(S) \cap 1^+ \cdot 2^+ \cdot \$ \cdot \overline{\Sigma}^* \cdot Q \cdot \Sigma^* \cdot \text{¢} = \{1^{k+1}2^{k+1}\$\bar{u}_kq_kv_k\text{¢} \mid k \geq 0\}$. Thus, the language $\Delta_{R_1}^*(S)$ is not regular. This proves the following result.

Lemma 16. *If the TM M has an immortal finite configuration, then R_1 does not preserve regularity.*

Observe that the strings in the set $\Delta_{R_1}^*(S) \cap 1^+ \cdot 2^+ \cdot \$ \cdot \overline{\Sigma}^* \cdot Q \cdot \Sigma^* \cdot \text{¢}$ are all irreducible mod R_1. Thus, $\mathrm{NF}_{R_1}(S) \cap 1^+ \cdot 2^+ \cdot \$ \cdot \overline{\Sigma}^* \cdot Q \cdot \Sigma^* \cdot \text{¢} = \Delta_{R_1}^*(S) \cap 1^+ \cdot 2^+ \cdot \$ \cdot \overline{\Sigma}^* \cdot Q \cdot \Sigma^* \cdot \text{¢}$, and hence, R_1 does not even give regular sets of normal forms for regular languages, if M has an immortal finite configuration.

We would like to also prove the converse of this statement. However, for this we must consider all regular languages over Γ, not just the languages containing only encodings of configurations of M. To get around this problem, we introduce a finite monadic srs R_2, which constitutes the second part of the system R. It consists of the following 18 groups of monadic rules:

(1.) $s1 \to 0$ for all $s \in \Gamma \setminus \{1\}$;
(2.) $s2 \to 0$ for all $s \in \Gamma \setminus \{1, 2, \hat{c}, \hat{d}\}$;
(3.) $s\hat{d} \to 0$ for all $s \in \Gamma \setminus \{1, 2, \hat{d}\}$;
(4.) $1\hat{d}\hat{d} \to 0$
$2\hat{d}\hat{d}\hat{d} \to 0$
(5.) $s\hat{c} \to 0$ for all $s \in \Gamma \setminus \{1, 2\}$;
(6.) $s\$ \to 0$ for all $s \in \Gamma \setminus \{2, \hat{c}, \hat{d}\}$;
(7.) $s\bar{a} \to 0$ for all $\bar{a} \in \overline{\Sigma}$ and $s \in \Gamma \setminus (\overline{\Sigma} \cup \{\$, \bar{c}, \bar{d}\})$;
(8.) $s\bar{q} \to 0$ for all $\bar{q} \in \overline{Q}$ and $s \in \Gamma \setminus (\overline{\Sigma} \cup \{\$, \bar{c}, \bar{d}\})$;
(9.) $s\bar{d} \to 0$ for all $s \in \Gamma \setminus (\overline{\Sigma} \cup \{\$, \bar{d}\}$;

(10.) $s\bar{d}\bar{d}\bar{d} \to 0$ for all $s \in \overline{\Sigma} \cup \{\$\}$;
(11.) $s\bar{c} \to 0$ for all $s \in \Gamma \smallsetminus (\overline{\Sigma} \cup \{\$\})$;
(12.) $sq \to 0$ for all $q \in Q$ and $s \in \Gamma \smallsetminus (\overline{\Sigma} \cup \{\$, \bar{c}, \bar{d}\})$;
(13.) $sa \to 0$ for all $a \in \Sigma$ and $s \in \Gamma \smallsetminus (Q \cup \overline{Q} \cup \Sigma \cup \{d\})$;
(14.) $sd \to 0$ for all $s \in \Gamma \smallsetminus (\overline{Q} \cup \Sigma \cup \{d\})$;
(15.) $sddd \to 0$ for all $s \in \overline{Q} \cup \Sigma$;
(16.) $s\text{¢} \to 0$ for all $s \in \Gamma \smallsetminus (Q \cup \overline{Q} \cup \Sigma \cup \{d\})$;
(17.) $sd_0 \to 0$ for all $s \in \Gamma \smallsetminus \{\text{¢}, d_0\}$;
(18.) $$\left.\begin{array}{ll} s0 & \to 0 \\ 0s & \to 0 \end{array}\right\} \text{for all } s \in \Gamma.$$

Lemma 17. *(a) The set* $\mathrm{IRR}(R_2)$ *consists of all the factors of the strings in the following language* $\mathrm{CONF} := 1^* \cdot (\{\hat{c}, \hat{d}\} \cup (\{\lambda, \hat{c}, \hat{d}\} \cdot (2 \cdot \{\lambda, \hat{c}, \hat{d}, \hat{d}\hat{d}\})^+)) \cdot \$ \cdot (\{\lambda, \bar{c}, \bar{d}, \bar{d}\bar{d}\} \cdot \overline{\Sigma})^* \cdot \{\lambda, \bar{c}, \bar{d}, \bar{d}\bar{d}\} \cdot ((\overline{Q} \cdot \{\lambda, d, dd\}) \cup Q) \cdot (\Sigma \cdot \{\lambda, d, dd\})^* \cdot \text{¢} \cdot d_0^* \cup \{0\}$.
(b) $\Delta^*_{R_1}(w) \subseteq \mathrm{IRR}(R_2)$ *for all* $w \in \mathrm{IRR}(R_2)$.
(c) The combined system $R := R_1 \cup R_2$ *is finite, length-reducing, and confluent.*

From Lemma 16 and Lemma 17(b) we obtain the following.

Corollary 18. *If the TM* M *has an immortal finite configuration, then* R *does not preserve regularity. In fact, in this situation there exists a regular language* $S \subseteq \Gamma^*$ *such that not even the set* $\mathrm{NF}_R(S)$ *is regular.*

It remains to prove the converse of this corollary. Actually, we will prove the following statement: if the TM M is strongly bounded, then $\mathrm{NF}_R(S)$ is regular for each regular language $S \subseteq \Gamma^*$.

First, observe that it suffices to look at rather restricted regular languages $S \subseteq \Gamma^*$. Let $S \subseteq \Gamma^*$ be a regular language. Then $S = S_1 \cup S_2$, where $S_1 := S \cap \mathrm{IRR}(R_2)$ and $S_2 := S \cap \mathrm{RED}(R_2)$ both are regular. Now $\mathrm{NF}_R(S) = \mathrm{NF}_{R_1}(S_1) \cup \mathrm{NF}_R(S_2)$, and $\mathrm{NF}_R(S_2) = \begin{cases} \emptyset & \text{if } S_2 = \emptyset, \\ \{0\} & \text{if } S_2 \neq \emptyset. \end{cases}$ Thus, $\mathrm{NF}_R(S)$ is regular if and only if $\mathrm{NF}_{R_1}(S_1)$ is regular. Hence, we can restrict our attention to regular sets S that are contained in $\mathrm{IRR}(R_2)$.

Let $S \subseteq \mathrm{IRR}(R_2)$ be a regular language. Again we can partition S into two regular subsets $S_1 := S \cap \Gamma^* \cdot d_0^+$ and $S_2 := S \cap (\Gamma \smallsetminus \{d_0\})^*$. The strings of S_2 do not contain any ocurrences of the symbol d_0. Hence, a generalized sequential machine (gsm) G_2 can be constructed such that $G_2(S_2) = \mathrm{NF}_{R_1}(S_2)$ [Ott96]. This gives the following result.

Lemma 19. *If* $S_2 \subseteq \mathrm{IRR}(R_2) \cap (\Gamma \smallsetminus \{d_0\})^*$ *is a regular language, then so is the language* $\mathrm{NF}_{R_1}(S_2)$.

Further, we partition the set S_1 as $S_1 = S_3 \cup S_4$, where $S_3 := S_1 \cap (\Gamma \smallsetminus \{d_0\})^* \cdot \{d_0, d_0^2, \ldots, d_0^{\nu-1}\}$ and $S_4 := S_1 \cap \Gamma^* \cdot d_0^{\nu}$. Here ν denotes the constant that the pumping lemma for regular languages yields for S_1. Each string in S_3 contains at most $\nu - 1$ occurrences of the symbol d_0. This allows us to prove the following in analogy to the previous lemma.

Lemma 20. *If S_1 is a regular language, then so is the language* $\mathrm{NF}_{R_1}(S_3)$.

It remains to deal with the regular language $S_4 = S_1 \cap \Gamma^* \cdot d_0^\nu$. Let S_5 be the regular language $S_5 := \{w \in (\Gamma \setminus \{d_0\})^* \mid \exists m \geq \nu : wd_0^m \in S_4\}$. For $m \in \{0, 1, \ldots, \nu - 1\}$ and $\alpha \in \{1, \ldots, \nu\}$, define the languages $S_{5,m,\alpha}$ and $S_{4,m,\alpha}$ as $S_{5,m,\alpha} := \{w \in S_5 \mid w \cdot d_0^m \cdot (d_0^\alpha)^* \subseteq S_4\}$ and $S_{4,m,\alpha} := S_{5,m,\alpha} \cdot d_0^m \cdot (d_0^\alpha)^*$. It follows from the pumping lemma that $S_4 = \bigcup_{m=0}^{\nu-1} \bigcup_{\alpha=1}^{\nu} S_{4,m,\alpha}$. Obviously, each of the languages $S_{5,m,\alpha}$ and $S_{4,m,\alpha}$ is regular.

Since $\mathrm{NF}_{R_1}(S_4) = \bigcup_{m=0}^{\nu-1} \bigcup_{\alpha=1}^{\nu} \mathrm{NF}_{R_1}(S_{4,m,\alpha})$, it suffices to look at one of the languages $S_{4,m,\alpha}$. Exploiting the fact that the TM M is strongly bounded, a gsm G can now be constructed such that $G(S_{4,m,\alpha}) = \mathrm{NF}_{R_1}(S_{4,m,\alpha})$ [Ott96]. Hence, with S_4 also the language $\mathrm{NF}_{R_1}(S_{4,m,\alpha})$ is regular. Therewith we have proved the following lemma.

Lemma 21. *If the TM M is strongly bounded, then* $\mathrm{NF}_R(S)$ *is a regular language for each regular language $S \subseteq \Gamma^*$.*

From Corollary 18 and Lemma 21 we see that $\mathrm{NF}_R(S)$ is regular for each regular language $S \subseteq \Gamma^*$ if and only if M is strongly bounded (see Section 5). Thus, we have the following undecidability result.

Theorem 22. *The following problem is undecidable in general:*

Instance: *A finite, length-reducing, and confluent srs R on Γ.*

Question: *Is* $\mathrm{NF}_R(S)$ *regular for each regular language $S \subseteq \Gamma^*$?*

This generalizes Theorem 9 of [GT95] to signatures containing unary function symbols only and possibly a single constant.

7 Conclusion

It remains open whether the property of preserving regularity is undecidable for the class of finite srss that are length-reducing and confluent. Also it remains the question of whether finite, length-reducing, and confluent systems presenting groups preserve regularity.

However, in the latter case we do at least know the following. If R is a finite, length-reducing, and confluent srs on Σ presenting a group, then there exists a deterministic pushdown automaton P that, given a string $w \in \Sigma^*$ as input, halts with the irreducible descendant w_0 of w mod R on its pushdown store [MO87]. For $L \subseteq \Sigma^*$, let $\mathrm{SC}_P(L)$ denote the language of final stack contents that P can generate from L. Then $\mathrm{SC}_P(L) = \mathrm{NF}_R(L)$. If L is regular, then by a result of Greibach [Gre67] also $\mathrm{SC}_P(L)$ is regular. Thus, in this situation the set of normal forms of each regular language is itself regular.

Acknowledgement. The author wants to thank Klaus Madlener and Paliath Narendran for very fruitful discussions regarding the result presented in Section 6 and for pointing out the papers [Hoo66] and [KTU93].

References

[Ani71] A.V. Anissimov. Group languages. *Cybernetics*, 7:594–601, 1971.

[BO93] R.V. Book and F. Otto. *String-Rewriting Systems*. Springer-Verlag, New York, 1993.

[Bra69] W.J. Brainerd. Tree generating regular systems. *Information and Control*, 14:217–231, 1969.

[CDGV94] J.-L. Coquidé, M. Dauchet, R. Gilleron, and S. Vágvölgyi. Bottom-up tree pushdown automata: classification and connection with rewrite systems. *Theoretical Computer Science*, 127:69–98, 1994.

[Gil91] R. Gilleron. Decision problems for term rewriting systems and recognizable tree languages. In C. Choffrut and M. Jantzen, editors, *Proc. of STACS'91*, Lecture Notes in Computer Science 480, pages 148–159. Springer-Verlag, 1991.

[Gre67] S. Greibach. A note on pushdown store automata and regular systems. *Proc. American Mathematical Society*, 18:263–268, 1967.

[GS84] F. Gécseg and M. Steinby. *Tree Automata*. Akadémiai Kiadó, Budapest, 1984.

[GT95] R. Gilleron and S. Tison. Regular tree languages and rewrite systems. *Fundamenta Informaticae*, 24:157–175, 1995.

[GV97] P. Gyenizse and S. Vágvölgyi. Linear generalized semi-monadic rewrite systems effectively preserve recognizability. *Theoretical Computer Science*, 1997. to appear.

[Hoo66] P.K. Hooper. The undecidability of the Turing machine immortality problem. *J. Symbolic Logic*, 31:219–234, 1966.

[KTU93] A.J. Kfoury, J. Tiuryn, and P. Urzyczyn. The undecidability of the semi-unification problem. *Information and Computation*, 102:83–101, 1993.

[LS77] R.C. Lyndon and P.E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, Berlin, 1977.

[MO87] K. Madlener and F. Otto. Groups presented by certain classes of finite length-reducing string-rewriting systems. In P. Lescanne, editor, *Rewriting Techniques and Applications*, Lecture Notes in Computer Science 256, pages 133–144. Springer-Verlag, Berlin, 1987.

[Ott91] F. Otto. When is an extension of a specification consistent? Decidable and undecidable cases. *Journal of Symbolic Computation*, 12:255–273, 1991.

[Ott96] F. Otto. Preserving regularity and related properties of string-rewriting systems. Mathematische Schriften Kassel 6/96, Universität-GH Kassel, September 1996.

[Sak79] J. Sakarovitch. Syntaxe des langages de Chomsky, essai sur le déterminisme, 1979. Thèse de doctorat d'état de l'université Paris VII.

[Sal88] K. Salomaa. Deterministic tree pushdown automata and monadic tree rewriting systems. *Journal of Computer and System Sciences*, 37:367–394, 1988.

[Sén90] G. Sénizergues. Some decision problems about controlled rewriting systems. *Theoretical Computer Science*, 71:281–346, 1990.