Loughborough University

This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

A survey on the local divisor technique

PLEASE CITE THE PUBLISHED VERSION

https://doi.org/10.1016/j.tcs.2015.07.008

PUBLISHER

© Elsevier

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Diekert, Volker, and Manfred Kufleitner. 2019. “A Survey on the Local Divisor Technique”. figshare. https://hdl.handle.net/2134/31971.

A Survey on the Local Divisor Technique

Volker Diekert Manfred Kufleitner*

University of Stuttgart, FMI
{diekert,kufleitner}@fmi.uni-stuttgart.de

Abstract. Local divisors allow a powerful induction scheme on the size of a monoid. We survey this technique by giving several examples of this proof method. These applications include linear temporal logic, rational expressions with Kleene stars restricted to prefix codes with bounded synchronization delay, Church-Rosser congruential languages, and Simon's Factorization Forest Theorem. We also introduce the notion of a *localizable language class* as a new abstract concept which unifies some of the proofs for the results above.

Keywords. local divisors; factorization forests; bounded synchronization delay; linear temporal logic; Kamp's Theorem

1 Introduction

The notion of a *local divisor* refers to a construction for finite monoids. It appeared in this context first in [4] where it was used by the authors as a tool in the proof that local future temporal logic is expressively complete for Mazurkiewicz traces with respect to first-order logic. The definition of a local divisor is very simple: Let M be a finite monoid and $c \in M$. Then $cM \cap Mc$ is a semigroup, but it fails to be a submonoid unless c is invertible. If c is not invertible then $1 \notin cM \cap Mc$ and, as a consequence, $|cM \cap Mc| < |M|$. The idea is to turn $cM \cap Mc$ into a monoid by defining a new multiplication by $xc \circ cy = xcy$. This is well-defined and $M_c = (cM \cap Mc, \circ, c)$ becomes a monoid where c is the unit. Moreover, if c is not invertible then M_c is a smaller monoid than M; and this makes the construction attractive for induction. (The same idea works for $\{c\} \cup cMc$ and since $\{c\} \cup cMc \subseteq cM \cap Mc$ there is a choice here.) The original definition for a multiplication of type $xc \circ cy = xcy$ was given for associative algebras. It can be traced back to a technical report of Meyberg, [17]. He coined the notion of a

*The second author was supported by the German Research Foundation (DFG) grant DI 435/5-2.

local algebra. Just replace M above by a finite dimensional associative algebra (with a unit element) over a field k. For example, M is the algebra of $n \times n$ matrices over k. If $c \in M$ is not invertible then the vector space $cM \cap Mc$ has at least one dimension less and $(cM \cap Mc, +, \circ, c)$ is again an associative algebra with the unit element c. See also [11] for applications of Meyberg's construction.

Despite (or more accurately *thanks to*) its simplicity, the *local divisor technique* is quite powerful, see *e.g.* [6]. For example, it was used in a new and simplified proof for the Krohn-Rhodes Theorem [9]. Very recently, the construction of local divisors has also been an essential tool in Kuperberg's work on a linear temporal logic for regular cost functions [15]. In [7] we extended a classical result of Schützenberger from finite words to infinite words by showing that *ω-rational expressions with bounded synchronization delay* characterize star-free languages. In 2012 we presented a paper which solved a 25 years old conjecture in formal language theory [8]. We showed that regular languages are Church-Rosser congruential. We come back to this result in more detail below. Our result was obtained in two steps. First, we had to show it for regular group languages, which is very difficult and technical. This part served as a base for induction. The second part uses induction using local divisors. This part is actually easy to explain, it will be done in Section 6.

The outline of the paper is as follows. In Section 3 we give a general framework for the local divisor technique in the context of aperiodic languages (*i.e.*, languages recognized by finite aperiodic monoids). We introduce the notion of *localizable language class* as a new abstract concept.

In the remaining sections we give four applications of the local divisor technique. In Section 4 we apply this technique to linear temporal logic, and in Section 5 it is used for a characterization of the aperiodic languages in terms of restricted rational expressions. In Section 6 we show how to apply the local divisor technique in the context of string rewrite systems. Finally, in Section 7 we give a proof of Simon's Factorization Forest Theorem; the proof is the archetype of how to apply the local divisor technique in arbitrary monoids.

2 Local divisors

We will apply the local divisor techniques mainly to monoids. However, it is instructive to place ourselves first in the slightly more general setting of semigroups. Let $S = (S, \cdot)$ be a finite semigroup. A *divisor* S' of S is a homomorphic image of a subsemigroup. Let $c \in S$ be any element and consider $cS \cap Sc$. We can turn the subset $cS \cap Sc$ into a semigroup by defining a new operation $\circ$ as follows:

$$xc \circ cy = xcy.$$

A direct calculation shows that the operation $\circ$ is well-defined and associative. Hence, $S_c = (cS \cap Sc, \circ)$ is a semigroup. In order to see that S_c is a divisor consider the following subsemigroup $S' = \{x \in S \mid cx \in Sc\}$ of S. Note that $c \in S'$. Define $\varphi : S' \to S_c$ by $\varphi(x) = cx$. It is surjective since $z \in cS \cap Sc$ implies that we can write $z = cx$ with

$x \in S'$. Moreover, $cxy = cx \circ cy$ and S_c is the homomorphic image of S'. Therefore, S_c is a divisor. We call it the *local divisor at* c. We want to use S_c for induction. Therefore we characterize next when $|S_c| < |S|$. Recall that $e \in S$ is called an *idempotent* if $e^2 = e$. For every finite semigroup there is a natural number $\omega \in \mathbb{N}$ such that x^ω is idempotent for every $x \in S$, for instance $\omega = |S|!$. An element y is called a *unit* if it has a left- and right inverse, *i.e.*, if there is a neutral element $1 \in S$ and $xy = yx' = 1$ for some $x, x' \in S$ (and then we have $x = xyx' = x'$). Thus, if S contains a unit y, then it is a monoid with neutral element y^ω. We have the following result.

Proposition 2.1. *Let S be a semigroup and $S_c = (cS \cap Sc, \circ)$ be defined as above.*

(a) *If S is a monoid, then $S_c = (cS \cap Sc, \circ, c)$ is a monoid and S_c is a divisor in terms of monoids, i.e. a homomorphic image of a submonoid S' of S.*

(b) *If c is a unit of S, then $S = \{x \in S \mid cx \in Sc\}$ and $\varphi : S \to S_c$, $x \mapsto cx$ is an isomorphism of monoids.*

(c) *If S is finite and c is not a unit, then $|S_c| < |S|$.*

(d) *If $cxc = cyc$ is idempotent in S_c, then $cxcy$ and $xcyc$ are idempotent in S.*

Proof. (a): Since S is a monoid we have $1 \in S' = \{x \in S \mid cx \in Sc\}$ and S_c is the homomorphic image of the submonoid S'.

(b): Trivial.

(c): If $cS \cap Sc = S$, then we have $cS = S$ and $Sc = S$. This implies that c is a unit. Indeed, we have $c^\omega S = S = Sc^\omega$. For every element $c^\omega x \in S$ we have $c^\omega \cdot c^\omega x = c^\omega x$. Thus, c^ω is neutral and $c^{\omega-1}$ is the inverse of c, *i.e.*, c is a unit. Therefore, if c is not a unit, then $|S_c| < |S|$.

(d) We have $cxcy \cdot cxcy = \big((cxc) \circ (cyc) \circ (cxc)\big) \cdot y = cxc \cdot y$. The last equality uses the fact that $cxc = cyc$ is idempotent in S_c. The claim for $xcyc$ is symmetric. □

Remark 2.2. *Note that $(\{cc\} \cup cSc, \circ)$ is a subsemigroup of $(cS \cap Sc, \circ)$. Moreover, if S is a monoid, then $(\{c\} \cup cSc, \circ, c)$ is a submonoid of $(cS \cap Sc, \circ, c)$. Hence by slight abuse of language, we might call $(\{cc\} \cup cSc, \circ)$ (resp. $(\{c\} \cup cSc, \circ, c)$) a local divisor of S, too. In addition, if $c \in S$ is idempotent, then $(cSc, \circ) = (cSc, \cdot)$ is the usual local monoid at c. The advantage is that $\{cc\} \cup cSc$ (resp. $\{c\} \cup cSc$) might be smaller than $cS \cap Sc$. However, in worst case estimations there is no difference.*

3 Localizable language classes

A *language class* $\mathcal{V}$ assigns to every finite alphabet A a set of languages $\mathcal{V}(A^*) \subseteq 2^{A^*}$. A language class $\mathcal{V}$ is *left-localizable* if for all finite alphabets A and T the following properties hold:

(a) $\emptyset, A^* \in \mathcal{V}(A^*)$.

(b) If $K, L \in \mathcal{V}(A^*)$, then $K \cup L \in \mathcal{V}(A^*)$.

(c) For every $c \in A$, the alphabet $B = A \setminus \{c\}$ satisfies:

1. If $K \in \mathcal{V}(B^*)$, then $K \in \mathcal{V}(A^*)$.
2. If $K \in \mathcal{V}(A^*)$ and $L \in \mathcal{V}(B^*)$, then $KcL \in \mathcal{V}(A^*)$.
3. If $K \in \mathcal{V}(B^*)$ and $L \in \mathcal{V}(A^*)$ with $L \subseteq cA^*$, then $KL \in \mathcal{V}(A^*)$.
4. Suppose $g : B^* \to T$ is a mapping with $g^{-1}(t) \in \mathcal{V}(B^*)$ for all $t \in T$. Moreover, let $\sigma : (cB^*)^* \to T^*$ be defined by $\sigma(cu_1 \cdots cu_k) = g(u_1) \cdots g(u_k)$ for $u_i \in B^*$. If $K \in \mathcal{V}(T^*)$, then $\sigma^{-1}(K) \in \mathcal{V}(A^*)$.

Being *right-localizable* is defined by the right dual of left-localizability. Properties (a), (b) and (c1) are unchanged, but the remaining conditions are replaced by

(c) For every $c \in A$, the alphabet $B = A \setminus \{c\}$ satisfies:

2'. If $K \in \mathcal{V}(B^*)$ and $L \in \mathcal{V}(A^*)$, then $KcL \in \mathcal{V}(A^*)$.

3'. If $K \in \mathcal{V}(A^*)$ with $K \subseteq A^*c$ and $L \in \mathcal{V}(B^*)$, then $KL \in \mathcal{V}(A^*)$.

4'. Suppose $g : B^* \to T$ is a mapping with $g^{-1}(t) \in \mathcal{V}(B^*)$ for all $t \in T$. Moreover, let $\sigma : (B^*c)^* \to T^*$ be defined by $\sigma(u_1c \cdots u_kc) = g(u_1) \cdots g(u_k)$ for $u_i \in B^*$. If $K \in \mathcal{V}(T^*)$, then $\sigma^{-1}(K) \in \mathcal{V}(A^*)$.

A class of languages $\mathcal{V}$ is *localizable* if it is left-localizable or right-localizable.

Theorem 3.1. *If $L \subseteq A^*$ is recognized by a finite aperiodic monoid, then $L \in \mathcal{V}(A^*)$ for every localizable language class $\mathcal{V}$. This means that every localizable language class contains all aperiodic languages.*

Proof. We can assume that $\mathcal{V}$ be left-localizable; the situation with $\mathcal{V}$ being right-localizable is symmetric. Let $h : A^* \to M$ be a homomorphism to a finite aperiodic monoid M. It is enough to show $h^{-1}(p) \in \mathcal{V}(A^*)$ for all $p \in M$. We proceed by induction on $(|M|, |A|)$ with lexicographic order. If $h(A^*) = \{1\}$, then either $h^{-1}(p) = \emptyset$ or $h^{-1}(p) = A^*$; and we are done. Hence, we can assume that there is a letter $c \in A$ with $h(c) \neq 1$. Let $B = A \setminus \{c\}$ and $g : B^* \to M$ be the restriction of h to B^*. For all $p \in M$ we have

$$h^{-1}(p) \;=\; g^{-1}(p) \cup \bigcup_{p\,=\,qrs} g^{-1}(q) \cdot \big(h^{-1}(r) \cap cA^* \cap A^*c\big) \cdot g^{-1}(s) \tag{1}$$

by factoring every word at the first and the last occurrence of c. Induction on the size of the alphabet yields $g^{-1}(p) \in \mathcal{V}(B^*)$ for all $p \in M$. By the closure properties of $\mathcal{V}$, it suffices to show $\big(h^{-1}(r) \cap cA^* \cap A^*c\big) \cdot g^{-1}(s) \in \mathcal{V}(A^*)$ for every $r \in h(c)M \cap Mh(c)$ and $s \in h(B^*)$. Let $T = h(B^*)$. In the remainder of this proof we will use T as a finite alphabet. The mapping $\sigma : (cB^*)^* \to T^*$ is defined by

$$\sigma(cv_1 \cdots cv_k) = g(v_1) \cdots g(v_k)$$

for $v_i \in B^*$, and the homomorphism $f : T^* \to M_c$ to the local divisor $M_c = \big(h(c)M \cap Mh(c), \circ, h(c)\big)$ is defined by

$$f\big(g(v)\big) = h(cvc)$$

for $v \in B^*$. This is well-defined since $h(cvc) = h(c)g(v)h(c)$ only depends on $g(v)$ and not on the word v itself. Consider a word $w = cv_1 \cdots cv_k$ with $k \geq 0$ and $v_i \in B^*$. Then

$$\begin{aligned} f\big(\sigma(w)\big) &= f\big(g(v_1)g(v_2)\cdots g(v_k)\big) \\ &= h(cv_1c) \circ h(cv_2c) \circ \cdots \circ h(cv_kc) \\ &= h(cv_1cv_2 \cdots cv_kc) = h(wc). \end{aligned}$$

Thus, we have $wc \in h^{-1}(r)$ if and only if $w \in \sigma^{-1}\big(f^{-1}(r)\big)$. This shows $h^{-1}(r) \cap cA^* \cap A^*c = \sigma^{-1}\big(f^{-1}(r)\big) \cdot c$ for every $r \in h(c)M \cap Mh(c)$. It follows that

$$\big(h^{-1}(r) \cap cA^* \cap A^*c\big) \cdot g^{-1}(s) = \sigma^{-1}(K) \cdot c \cdot g^{-1}(s)$$

for $K = f^{-1}(r)$. The monoid M_c is aperiodic and $|M_c| < |M|$. Induction on the size of the monoid yields $K \in \mathcal{V}(T^*)$, and induction on the alphabet shows $g^{-1}(t) \in \mathcal{V}(B^*)$ for all $t \in T$. By the closure properties of $\mathcal{V}$ we obtain $\sigma^{-1}(K) \in \mathcal{V}(A^*)$ and $\sigma^{-1}(K) \cdot c \cdot g^{-1}(s) \in \mathcal{V}(A^*)$. This concludes the proof. □

The main application of Theorem 3.1 is to show that some given language class $\mathcal{V}$ contains all aperiodic languages. This can be done by verifying the properties (a), (b), and (c) for $\mathcal{V}$, *i.e.*, by showing that $\mathcal{V}$ is localizable.

4 Linear temporal logic

By Kamp's famous theorem [13], *linear temporal logic* LTL over words has the same expressive power as first-order logic FO[<]. In an algebraic setting, one shows first that every first-order definable language $L \subseteq A^*$ is aperiodic. This is relatively easy and no local divisor technique applies here. In this section we give a simple proof that every aperiodic language is LTL-definable. We give the proof for finite words, only. However, the basic proof techniques generalize to infinite words [5] and also to Mazurkiewicz traces [3].

The syntax of *linear temporal logic* LTL(A) over an alphabet A is defined as follows:

$$\varphi ::= \top \mid a \mid \neg\varphi \mid (\varphi \vee \varphi) \mid \mathsf{X}\varphi \mid (\varphi \,\mathsf{U}\, \varphi)$$

for $a \in A$. The modality X is for "neXt" and U is for "Until". As usual, we omit the bracketing whenever there is no confusion. For the semantics we interpret a word $u = a_1 \cdots a_n$ with $a_i \in A$ as a labeled linear order with positions $\{1, \ldots, n\}$, and position i is labeled by a_i. We write $u, i \models \varphi$ if the word u at position i models φ, and we write

$u, i \not\models \varphi$ if this is not the case. The semantics of LTL(A) is defined by:

$$\begin{array}{lcl}
u, i \models \top & & \text{is always true} \\
u, i \models a & \Leftrightarrow & a_i = a \\
u, i \models \neg\varphi & \Leftrightarrow & u, i \not\models \varphi \\
u, i \models \varphi \vee \psi & \Leftrightarrow & u, i \models \varphi \text{ or } u, i \models \psi \\
u, i \models \mathsf{X}\varphi & \Leftrightarrow & i < n \text{ and } u, i+1 \models \varphi \\
u, i \models \varphi \,\mathsf{U}\, \psi & \Leftrightarrow & \text{there exists } k \in \{i, \ldots, n\} \text{ such that } u, k \models \psi \\
 & & \text{and for all } j \in \{i, \ldots, k-1\} \text{ we have } u, j \models \varphi
\end{array}$$

The formula $\varphi \,\mathsf{U}\, \psi$ holds at position i if there exists a position $k \geq i$ such that ψ holds at k and all positions from i to $k-1$ satisfy φ. A formula φ in LTL(A) defines the language

$$L(\varphi) = \left\{ u \in A^+ \;\middle|\; u, 1 \models \varphi \right\}.$$

This means that when no position is given, then we start at the first position of a nonempty word. We introduce the following macros:

$$\begin{array}{rclcrcl}
\bot & := & \neg\top & & B & := & \bigvee_{b \in B} b \qquad \text{for } B \subseteq A \\
\varphi \wedge \psi & := & \neg(\neg\varphi \vee \neg\psi) & & \mathsf{F}\varphi & := & \top \,\mathsf{U}\, \varphi
\end{array}$$

The macro $\mathsf{F}\varphi$ (for "Future") holds at position i if φ holds at some position $k \geq i$. For $L, K \subseteq A^*$ we define a variant of the Until-modality on languages by

$$K \mathbin{\textcircled{U}} L = \{ vw \in A^* \mid w \in L, \forall\, v = pq \text{ with } q \neq \varepsilon\colon\ qw \in K \}.$$

The language class $\mathcal{LTL}$ resembles the behavior of LTL by using a more global semantics. The languages in $\mathcal{LTL}(A^*)$ are inductively defined by:

- $\emptyset \in \mathcal{LTL}(A^*)$.
- If $K, L \in \mathcal{LTL}(A^*)$ and $a \in A$, then $A^* \setminus L, K \cup L, aL, K \mathbin{\textcircled{U}} L \in \mathcal{LTL}(A^*)$.

The formal connection between $\mathcal{LTL}$ and LTL is given by the following proposition.

Proposition 4.1. *We have $L \in \mathcal{LTL}(A^*)$ if and only if $L \setminus \{\varepsilon\}$ is definable in* LTL(A).

Proof. We first show $L(\varphi) \in \mathcal{LTL}(A^*)$ for every formula $\varphi \in \text{LTL}(A)$. We have $A^* = A^* \setminus \emptyset \in \mathcal{LTL}(A^*)$. For $\varphi := \top$ we have $L(\top) = A^+ = \bigcup_{a \in A} aA^* \in \mathcal{LTL}(A^*)$. For $\varphi := a$ we have $L(a) = aA^* \in \mathcal{LTL}(A^*)$. The construction for negations is $L(\neg\psi) = L(\top) \setminus L(\psi)$, and disjunctions translate into unions. If $\varphi := \mathsf{X}\psi$, then $L(\mathsf{X}\psi) = \bigcup_{a \in A} aL(\psi)$. Finally, if $\varphi := \psi_1 \,\mathsf{U}\, \psi_2$, then $L(\varphi) = L(\psi_1) \mathbin{\textcircled{U}} L(\psi_2)$. It remains to show that $L \cup \{\varepsilon\} \in \mathcal{LTL}(A^*)$ whenever $L \in \mathcal{LTL}(A^*)$. This follows from $\{\varepsilon\} = A^* \setminus \left(\bigcup_{a \in A} aA^* \right)$ and the closure under union.

For the converse, we show that for every language $L \in \mathcal{LTL}(A^*)$ there exists a formula $\varphi_L \in \text{LTL}(A)$ such that $L(\varphi_L) = L \setminus \{\varepsilon\}$. If $L = \emptyset$, then $\varphi_\emptyset = \bot$. Complements translate into negations, and unions translate into disjunctions. If $\varepsilon \notin K$, then the

formula for $L = aK$ is $\varphi_{aK} := a \wedge \mathsf{X}\varphi_K$. If $\varepsilon \in K$, then the formula of $L = aK$ is $\varphi_{aK} := a \wedge (\neg\mathsf{X}\top \vee \mathsf{X}\varphi_K)$. If $L = K_1 \mathbin{Ⓤ} K_2$ for $\varepsilon \notin K_2$, then $\varphi_L := \varphi_{K_1} \mathsf{U} \varphi_{K_2}$. Finally, if $L = K_1 \mathbin{Ⓤ} K_2$ for $\varepsilon \in K_2$, then $\varphi_L := (\varphi_{K_1} \mathsf{U} \varphi_{K_2}) \vee \neg\mathsf{F}\neg\varphi_{K_1}$; the formula $\neg\mathsf{F}\neg\varphi_{K_1}$ says that all positions satisfy φ_{K_1}. □

Proposition 4.2. *The language class $\mathcal{LTL}$ is left-localizable.*

Proof. The properties (a) and (b) are obvious. Let $c \in A$ and $B = A \setminus \{c\}$. The language B^+ is defined by $\neg\mathsf{F}\neg B \in \mathrm{LTL}(A)$ and thus $B^* \in \mathcal{LTL}(A^*)$.

For (c1) let $K, L \in \mathcal{LTL}(B^*)$. By induction we can assume $K, L \in \mathcal{LTL}(A^*)$. This immediately yields $K \cup L, aL, K \mathbin{Ⓤ} L \in \mathcal{LTL}(A^*)$ for all letters a. The set $B^* \setminus L$ can be written as $(A^* \setminus L) \cap B^*$ and hence $B^* \setminus L$ is in $\mathcal{LTL}(A^*)$. This shows $\mathcal{LTL}(B^*) \subseteq \mathcal{LTL}(A^*)$.

For (c2) let $K \in \mathcal{LTL}(A^*)$ and $L \in \mathcal{LTL}(B^*)$. We have $KcL = A^*cL \cap KcB^*$ because the last c in a word is unique. Note that $A^*cL = A^* \mathbin{Ⓤ} cL \in \mathcal{LTL}(A^*)$. It remains to show $KcB^* \in \mathcal{LTL}(A^*)$ by structural induction:

$$\begin{aligned}
(A^* \setminus L')cB^* &= A^*cB^* \setminus L'cB^* \\
(K' \cup L')cB^* &= K'cB^* \cup L'cB^* \\
(aL')cB^* &= a(L'cB^*) \\
(K' \mathbin{Ⓤ} L')cB^* &= (K'cB^*) \mathbin{Ⓤ} (L'cB^*).
\end{aligned}$$

For (c3) let $K \in \mathcal{LTL}(B^*)$ and $L \in \mathcal{LTL}(A^*)$ with $L \subseteq cA^*$. We have $KL = B^*L \cap KcA^*$. Note that $B^*L = BA^* \mathbin{Ⓤ} L \in \mathcal{LTL}(A^*)$. As before, one can easily show $KcA^* \in \mathcal{LTL}(A^*)$ by structural induction. For instance, $(B^* \setminus L')cA^* = B^*cA^* \setminus L'cA^*$ since $L' \subseteq B^*$ and the occurrence of the first c is unique.

For (c4) suppose $g : B^* \to T$ is a mapping with $g^{-1}(t) \in \mathcal{LTL}(B^*)$ for all $t \in T$. Moreover, let $\sigma : (cB^*)^* \to T^*$ be defined by $\sigma(cu_1 \cdots cu_k) = g(u_1) \cdots g(u_k)$ for $u_i \in B^*$. We show $\sigma^{-1}(K) \in \mathcal{LTL}(A^*)$ for every $K \in \mathcal{LTL}(T^*)$ by structural induction on K. For all $K, L \subseteq T^*$ and $t \in T$ we have:

$$\begin{aligned}
\sigma^{-1}(T^*) &= \{\varepsilon\} \cup cA^* \\
\sigma^{-1}(K \setminus L) &= \sigma^{-1}(K) \setminus \sigma^{-1}(L) \\
\sigma^{-1}(K \cup L) &= \sigma^{-1}(K) \cup \sigma^{-1}(L) \\
\sigma^{-1}(tL) &= c \cdot g^{-1}(t) \cdot \sigma^{-1}(L) \\
\sigma^{-1}(K \mathbin{Ⓤ} L) &= \big((\sigma^{-1}(K) \cup BA^*) \mathbin{Ⓤ} \sigma^{-1}(L)\big) \cap \sigma^{-1}(T^*).
\end{aligned}$$

Note that $g^{-1}(t) \cdot \sigma^{-1}(L) \in \mathcal{LTL}(A^*)$ by (c3). □

Together with Theorem 3.1 this leads to the following result.

Corollary 4.3. *If $L \subseteq A^+$ is recognized by a finite aperiodic semigroup, then L is definable in* $\mathrm{LTL}(A)$.

Proof. As a subset of A^*, the language L is recognized by a finite aperiodic monoid. By Theorem 3.1 and Proposition 4.2 we have $L \in \mathcal{LTL}(A^*)$. Since $\varepsilon \notin L$, Proposition 4.1 shows that L is definable in $\mathrm{LTL}(A)$. □

5 Bounded synchronization delay

A fundamental and classical result of Schützenberger from 1965 says that a language is star-free if and only if its syntactic monoid is finite and aperiodic [23]. A language is *star-free* if it can be built from the finite languages by using concatenation and Boolean operations. One can think of the star-free languages as rational languages where the Kleene-star is replaced by complementation. There is another beautiful characterization of the star-free languages due to Schützenberger [24], which seems to be quite overlooked. It characterizes the star-free languages without using complementation, but the inductive definition allows the star-operation on languages K (already belonging to the class) if K is a prefix code with bounded synchronization delay. Since synchronization delay is the main feature in this approach, the class is denoted by SD. The notion of bounded synchronization delay was introduced by Golomb and Gordon [12] and it is an important concept in coding theory.

A language $K \subseteq A^*$ is called *prefix-free* if $u, uv \in K$ implies $u = uv$. A prefix-free language $K \subseteq A^+$ is also called a *prefix code* since every word $u \in K^*$ admits a unique factorization $u = u_1 \cdots u_k$ with $k \geq 0$ and $u_i \in K$. A prefix code K has *synchronization delay* d if for all $u, v, w \in A^*$ we have:

$$\text{if } uvw \in K^* \text{ and } v \in K^d, \text{ then } uv \in K^*.$$

Note that $uv \in K^*$ and $uvw \in K^*$ implies $w \in K^*$ since K is a prefix code. The prefix code K has *bounded synchronization delay* if there is some $d \in \mathbb{N}$ such that K has synchronization delay d. Note that every subset $B \subseteq A$ yields a prefix code with synchronization delay 0. In particular, the sets B are prefix codes of bounded synchronization delay for all $B \subseteq A$.

The intuition behind this concept is the following: Assume a sender emits a stream of code words from K, where K is a prefix code with synchronization delay d. If a receiver misses the beginning of the message, he can wait until he detects a sequence of d code words. Then he can synchronize and decipher the remaining text after these d words.

We inductively define Schützenberger's language class SD:

(a) We have $\emptyset \in \mathrm{SD}(A^*)$ and $\{a\} \in \mathrm{SD}(A^*)$ for all letters $a \in A$.

(b) If $K, L \in \mathrm{SD}(A^*)$, then $K \cup L$, $K \cdot L \in \mathrm{SD}(A^*)$.

(c) If $K \in \mathrm{SD}(A^*)$ is a prefix code with bounded synchronization delay, then $K^* \in \mathrm{SD}(A^*)$.

Note that, unlike the definition of star-free sets, the inductive definition of $\mathrm{SD}(A^*)$ does not use any complementation.

Proposition 5.1. *The language class* SD *is right-localizable.*

Proof. The properties (a), (b), (c1), (c2'), and (c3') are obvious. Let $c \in A$ and $B = A \setminus \{c\}$ and consider the property (c4').

Suppose $g : B^* \to T$ is a mapping with $g^{-1}(t) \in \mathrm{SD}(B^*)$ for all $t \in T$. Moreover, let $\sigma : (cB^*)^* \to T^*$ be defined by $\sigma(u_1 c \cdots u_k c) = g(u_1) \cdots g(u_k)$ for $u_i \in B^*$. We show

$\sigma^{-1}(K) \in \mathrm{SD}(A^*)$ for every $K \in \mathrm{SD}(T^*)$ by structural induction on K:

$$\begin{aligned}\sigma^{-1}(t) &= g^{-1}(t)c\\ \sigma^{-1}(K \cup L) &= \sigma^{-1}(K) \cup \sigma^{-1}(L)\\ \sigma^{-1}(K \cdot L) &= \sigma^{-1}(K) \cdot \sigma^{-1}(L)\\ \sigma^{-1}(K^*) &= \sigma^{-1}(K)^*\,.\end{aligned}$$

It remains to verify that $P = \sigma^{-1}(K)$ is a prefix code of bounded synchronization delay whenever K has this property. Clearly, $\varepsilon \notin P$. To see prefix-freeness, consider $u, uv \in P$. This implies $u \in A^*c$ and hence, $\sigma(uv) = \sigma(u)\sigma(v)$. It follows that $v = \varepsilon$ because K is prefix-free. Finally, suppose K has synchronization delay d. We show that P has synchronization delay $d+1$: Let $uvw \in P^*$ with $v \in P^{d+1}$. Write $v = u'cv'$ with $v' \in P^d$. Note that $v' \in A^*c$. It follows that $\sigma(uv) = \sigma(uu'c)\sigma(v')$ and $\sigma(v') \in K^d$. Thus, $\sigma(uv) \in K^*$. We obtain $uv \in P^*$ as desired. □

Corollary 5.2. *If $L \subseteq A^*$ is recognized by a finite aperiodic semigroup, then $L \in \mathrm{SD}(A^*)$.*

Proof. This is an immediate consequence of Proposition 5.1 and Theorem 3.1. □

6 Church-Rosser congruential languages

The *Word Problem* $\mathrm{WP}(L)$ of a language $L \subseteq A^*$ is the following computational task.

Input: $w \in A^*$.
Question: Do we have $w \in L$?

The following facts are standard in formal language theory.

- If L is regular, then $\mathrm{WP}(L)$ is decidable in real time.
- If L is deterministic context-free, then $\mathrm{WP}(L)$ is decidable in linear time.
- If L is context-free, then $\mathrm{WP}(L)$ is decidable in cubic time.
- If L is context-sensitive, then $\mathrm{WP}(L)$ is decidable in polynomial space, and there are context-sensitive languages such that $\mathrm{WP}(L)$ is PSPACE-complete.

The paper of McNaughton, Narendran, and Otto [16] exploits the following theme: "Go beyond deterministic context-free and keep linear time solvability for the word problem by using Church-Rosser semi-Thue systems."

Before we proceed we need more preliminaries and notation. A *weight* is a homomorphism $\|\cdot\| : A^* \to \mathbb{N}$ such that $\|a\| > 0$ for all letters $a \in A$. The length function is a weight. If the weight $\|\cdot\|$ is given, we say that $(A, \|\cdot\|)$ is a *weighted alphabet*.

A *semi-Thue system* over A is a subset $S \subseteq A^* \times A^*$. The elements of S are called *rules*, and we frequently write $\ell \to r$ for $(\ell, r) \in S$. A system S is called *length-reducing*

(resp. *weight-reducing* for a weight $\|\cdot\|$) if we have $|\ell| > |r|$ (resp. $\|\ell\| > \|r\|$) for all rules $(\ell, r) \in S$. Every system S defines the rewriting relation $\underset{S}{\Longrightarrow} \subseteq A^* \times A^*$ by

$$u \underset{S}{\Longrightarrow} v \text{ if } u = p\ell q,\ v = prq \text{ for some rule } (\ell, r) \in S.$$

By $\overset{*}{\underset{S}{\Longrightarrow}}$ we mean the reflexive and transitive closure of $\underset{S}{\Longrightarrow}$. By $\overset{*}{\underset{S}{\Longleftrightarrow}}$ we mean the symmetric, reflexive, and transitive closure of $\underset{S}{\Longrightarrow}$. We also write $u \overset{*}{\underset{S}{\Longleftarrow}} v$ whenever $v \overset{*}{\underset{S}{\Longrightarrow}} u$. The system S is *confluent* if for all $u \overset{*}{\underset{S}{\Longleftrightarrow}} v$ there is some w such that $u \overset{*}{\underset{S}{\Longrightarrow}} w \overset{*}{\underset{S}{\Longleftarrow}} v$. By $\mathrm{IRR}_S(A^*)$ we denote the set of irreducible words, *i.e.*, the set of words where no left-hand side of a rule occurs as a factor. The relation $\overset{*}{\underset{S}{\Longleftrightarrow}}$ is a congruence, hence the congruence classes $[u]_S = \{v \in A^* \mid u \overset{*}{\underset{S}{\Longleftrightarrow}} v\}$ form a monoid which is denoted by A^*/S. If A^*/S is finite, then we say that S is of *finite index*.

Definition 6.1. *A semi-Thue system $S \subseteq A^* \times A^*$ is called a* Church-Rosser system *if it is length-reducing and confluent. A language $L \subseteq A^*$ is called a* Church-Rosser congruential language *if there is a finite Church-Rosser system S such that L can be written as a finite union of congruence classes $[u]_S$. If in addition A^*/S is of finite index, then $L \subseteq A^*$ is called* strongly Church-Rosser congruential.

The motivation to consider these languages in [16] stems from the following.

Remark 6.2. *Let $S \subseteq A^* \times A^*$ be a weight-reducing system. Then on input $w \in A^*$ of length n we can compute in time $\mathcal{O}(n)$ some word $\widehat{w} \in \mathrm{IRR}(S)$ such that $w \overset{*}{\underset{S}{\Longrightarrow}} \widehat{w}$. In particular, if L is a Church-Rosser congruential language, then its word problem is solvable in linear time.*

Let us consider some examples.

- Let $S = \{aab \to ba, cb \to c\}$. It is Church-Rosser, and hence $L_0 = [ca]_S$ is Church-Rosser congruential. The language L_0 is not context-free since $L_0 \cap ca^*b^* = \{ca^{2^n}b^n \mid n \geq 0\}$. Therefore the class of Church-Rosser congruential languages is not included in the class of context-free languages.
- Let $L_1 = \{a^n b^n \mid n \geq 0\}$. It is Church-Rosser congruential due to $S = \{aabb \to ab\}$ and $L_1 = [ab]_S \cup [\varepsilon]_S$. The monoid A^*/S is infinite because L_1 is not regular. We may also note that $[a^n]_S = \{a^n\}$ for $n \geq 1$, and hence there are infinitely many classes.
- Let $L_2 = \{a^m b^n \mid m \geq n \geq 0\}$. It is deterministic context-free, but not Church-Rosser congruential since a^m must be irreducible for each $m \geq 1$.
- Let $L_3 = \{a, b\}^* a \{a, b\}^*$. It is strongly Church-Rosser congruential due to $S = \{aa \to a,\ b \to \varepsilon\}$, $L_3 = [a]_S$, and $A^*/S = \{[\varepsilon]_S, [a]_S\}$.

- Let $L_4 = (ab)^*$ and $S = \{aba \to a\}$. The system S is Church-Rosser and $L_4 = [ab]_S \cup [\varepsilon]_S$. However, A^*/S is infinite although L_4 is regular. Therefore S does not show that L_4 is strongly Church-Rosser congruential. However, choosing $T = \{aaa \to aa,\ aab \to aa,\ baa \to aa, bbb \to aa,\ bba \to aa,\ abb \to aa, aba \to a,\ bab \to b\}$, we obtain $L_4 = [ab]_T \cup [\varepsilon]_T$ and A^*/T has 7 elements, only. Hence, L_4 is indeed strongly Church-Rosser congruential.

The languages L_0 and L_2 show that the classes of (deterministic) context-free languages and Church-Rosser congruential languages are incomparable. Therefore in [16] a weaker notion of Church-Rosser languages has been considered, too. The new class contained all Church-Rosser congruential languages as well as all deterministic context-free languages; and their word problem remains decidable in linear time. We do not go into details, but focus on the following conjecture dating back to 1988.

Conjecture 6.3 ([16]). *Every regular language is Church-Rosser congruential.*

After some significant progress on this conjecture in [18, 19, 20, 21, 22] there was stagnation. It was announced in 2003 by Reinhardt and Thérien in [22] that Conjecture 6.3 is true for all regular languages where the syntactic monoid is a group. However, the manuscript has never been published as a refereed paper and there are some flaws in its presentation. Let us continue with some examples which show that this statement is far from being trivial even for finite cyclic groups. It shows that a major difficulty is the number of generators.

- Let $L_5 = \{w \in a^* \mid |w| \equiv 0 \bmod 3\}$. Then $S = \{aaa \to \varepsilon\}$ shows that L_5 is strongly Church-Rosser congruential.
- Let $L_6 = \{w \in \{a,b\}^* \mid |w| \equiv 0 \bmod 3\}$. We have $L_6 = [\varepsilon]_S$ with respect to the system $S = \{u \to \varepsilon \mid |u| = 3\}$. But S is not confluent, as we can see from $a \underset{S}{\Longleftarrow} aabb \underset{S}{\Longrightarrow} b$. The smallest system (we are aware of) showing that L_6 is Church-Rosser congruential is rather large. We may choose $T = \{aaa \to \varepsilon,\ baab \to b,\ (ba)^3b \to b\} \cup \{bb\,u\,bb \to b^{|u|+1} \mid 1 \le |u| \le 3\}$. The language L_6 is a union of elements in A^*/T, and A^*/T contains 272 elements with the longest irreducible word having length 16.

The solution of Conjecture 6.3 is a typical example for the principle of *loading the induction*: Proving a more general statement is sometimes easier because a stronger inductive assumption can be used. Conjecture 6.3 speaks about Church-Rosser congruential languages. First, we replace the existence of a finite Church-Rosser system by starting with an arbitrary weighted alphabet $(A, \|\cdot\|)$ and we consider only finite confluent systems $S \subseteq A^* \times A^*$ of finite index which are weight-reducing for the given weight; such a weight-reducing version of a Church-Rosser system is called a *weighted Church-Rosser system*. Second, we switch to a purely algebraic statement. We say that a homomorphism $h : A^* \to M$ *factorizes through a semi-Thue system* S if $u \overset{*}{\underset{S}{\Longleftrightarrow}} v$ implies $h(u) = h(v)$.

Proposition 6.4 ([8, 10]). *Let $(A, \|\cdot\|)$ be a weighted alphabet and let $h : A^* \to M$ be a homomorphism to a finite monoid M. Assume for every weighted alphabet $(B, \|\cdot\|)$ and*

every group G which divides M, every homomorphism $g : B^* \to G$ factorizes through a finite weighted Church-Rosser system of finite index. Then h factorizes through a finite weighted Church-Rosser system S of finite index.

Proof. The proof is by induction on $(|M|, |A|)$ with lexicographic order. If $h(A^*)$ is a finite group, then the claim follows from the assumption. If $h(A^*)$ is not a group, then there exists $c \in A$ such that $h(c)$ is not a unit. Let $B = A \setminus \{c\}$. By induction on the size of the alphabet there exists a weighted Church-Rosser system R for the restriction $h : B^* \to M$. Let

$$K = \mathrm{IRR}_R(B^*)c.$$

We consider the prefix code K as a weighted alphabet. The weight of a letter $uc \in K$ is the weight $\|uc\|$ when read as a word over the weighted alphabet $(A, \|\cdot\|)$. Let $M_c = h(c)M \cap Mh(c)$ be the local divisor of M at $h(c)$. We let $g : K^* \to M_c$ be the homomorphism induced by $g(uc) = h(cuc)$ for $uc \in K$. By induction on the size of the monoid there exists a weighted Church-Rosser system $T \subseteq K^* \times K^*$ for g. Suppose $g(\ell) = g(r)$ for $\ell, r \in K^*$ and let $\ell = u_1 c \cdots u_j c$ and $r = v_1 c \cdots v_k c$ with $u_i, v_i \in \mathrm{IRR}_R(B^*)$. Then

$$\begin{aligned} h(c\ell) &= h(cu_1c) \circ \cdots \circ h(cu_jc) \\ &= g(u_1c) \circ \cdots \circ g(u_jc) \\ &= g(\ell) = g(r) = h(cr). \end{aligned}$$

This means that every T-rule $\ell \to r$ yields a h-invariant rule $c\ell \to cr$. We can transform the system $T \subseteq K^* \times K^*$ for g into a system $T' \subseteq A^* \times A^*$ for h by

$$T' = \{c\ell \to cr \in A^* \times A^* \mid \ell \to r \in T\}.$$

Since T is confluent and weight-reducing over K^*, the system T' is confluent and weight-reducing over A^*. Combining R and T' leads to $S = R \cup T'$. The left sides of a rule in R and a rule in T' cannot overlap, and hence S is confluent, see *e.g.* [1, Theorem 1.1.13]. Therefore, S is a weighted Church-Rosser system such that h factorizes through A^*/S. Suppose that every word in $\mathrm{IRR}_T(K^*)$ has length at most k. Here, the length is over the alphabet K. Similarly, let every word in $\mathrm{IRR}_R(B^*)$ have length at most m. Then

$$\mathrm{IRR}_S(A^*) \subseteq \{u_0cu_1 \cdots cu_{k'+1} \mid u_i \in \mathrm{IRR}_R(B^*),\ k' \leq k\}$$

and every word in $\mathrm{IRR}_S(A^*)$ has length at most $(k+2)m+k+1$. In particular, $\mathrm{IRR}_S(A^*)$ and A^*/S are finite. □

Let $(B, \|\cdot\|)$ be a weighted alphabet. If $G = \{1\}$ is trivial, then the naïve system $T = \{b \to \varepsilon \mid b \in B\}$ is a weighted Church-Rosser system of finite index (the size of B^*/T is one) such that every homomorphism $g : B^* \to G$ factors through T. Since every group divisor of an aperiodic monoid is trivial, we obtain the following corollary.

Corollary 6.5 ([10])**.** *Every aperiodic language is strongly Church-Rosser congruential.*

In order to prove Conjecture 6.3 it remains to show that every homomorphism to a finite group factorizes through a finite weighted Church-Rosser system of finite index, which is done in [8]. Surprisingly, the result for non-cyclic simple groups is a lot easier than for cyclic or non-simple groups.

7 Factorization forests

In the following let M be a finite monoid and A be a finite alphabet. A *factorization forest* of a homomorphism $f : A^* \to M$ is a function d which maps every word w with length $|w| \geq 2$ to a factorization $d(w) = (w_1, \ldots, w_n)$ with $w = w_1 \cdots w_n$ such that

- $w_i \neq 1$ for all $1 \leq i \leq n$, and
- $n \geq 3$ implies that $f(w_1) = \cdots = f(w_n)$ is idempotent in M.

By successive factorization, every non-empty word can be visualized as a tree where the leaves are labeled with letters. Thus, d defines a *factorization tree* for each word w. The *height* $h(w)$ of a word w is defined as

$$h(w) = \begin{cases} 0 & \text{if } |w| \leq 1, \\ 1 + \max\{h(w_1), \ldots, h(w_n)\} & \text{if } d(w) = (w_1, \ldots, w_n). \end{cases}$$

The height of a factorization forest d is the supremum of $\{h(w) \mid w \in A^*\}$. The famous Factorization Forest Theorem of Simon says that every homomorphism $f : A^* \to M$ has a factorization forest of height $\mathcal{O}(|M|)$, see [25]. The original proof of Simon was rather technical. A simplified proof with a worse bound based on the Krohn-Rhodes decomposition was found by Simon in [26]. Later, improved bounds were found using Green's relations [2, 14]. However, in many cases it is enough to know that there is a factorization forest of bounded height, but the actual bound is not important. Based on local divisors, we give a new proof for the existence of such a bound.

Theorem 7.1 (Simon)**.** *Let M be a finite monoid. There is a constant $h(|M|)$ such that every homomorphism $f : A^* \to M$ has a factorization forest of height at most $h(|M|)$.*

We give the proof of Theorem 7.1 at the end of this section. The case where M is a finite group G is rather simple and nicely exposed in [2]. For convenience we repeat the argument.

Proposition 7.2 ([2])**.** *Let G be a finite group. Every homomorphism $f : A^* \to G$ has a factorization forest of height at most $3\,|G|$.*

Proof. Let $a_1 \cdots a_n \in A^*$. The basic idea is to perform an induction on the size of the *prefix set*

$$P(a_1 \cdots a_n) = \{f(a_1 \cdots a_i) \in G \mid 1 \leq i < n\}.$$

By induction on the size of the prefix set, we show that there is a factorization forest d such that the height of $a_1 \cdots a_n$ is at most $3\,|P(a_1 \cdots a_n)|$. If $P(a_1 \cdots a_n) = \emptyset$, then $n \leq 1$

and we are done. Thus let $P(a_1 \cdots a_n) \neq \emptyset$. Choose some maximal nonempty subset $\{i_1, \ldots, i_t\}$ of $\{1, \ldots, n-1\}$ such that all prefixes $a_1 \cdots a_{i_j}$ give the same group element p under f. Let $i_0 = 0$ and $i_{t+1} = n$. Consider the $t+1$ factors $v_j = a_{i_{j-1}+1} \cdots a_{i_j}$. The word $a_1 \cdots a_n$ factorizes as $v_1 \cdots v_{t+1}$. We have $f(v_2) = \cdots = f(v_t) = 1$. Thus we can define

$$\begin{aligned} d(v_1 \cdots v_{t+1}) &= (v_1 \cdots v_t, v_{t+1}), \\ d(v_1 \cdots v_t) &= (v_1, v_2 \cdots v_t) \qquad \text{if } t \geq 2, \\ d(v_2 \cdots v_t) &= (v_2, \ldots, v_t) \qquad \text{if } t \geq 3. \end{aligned}$$

For applying induction on the words v_i, it remains to show that each prefix set $P(v_j)$ is smaller than $P(a_1 \cdots a_n)$. The set $P(a_1 \cdots a_{i_1})$ is smaller than $P(a_1 \cdots a_n)$ since the prefix $p = f(a_1 \cdots a_{i_1})$ does not occur anymore. For $2 \leq j \leq t+1$ we have

$$p \cdot P(a_{i_{j-1}+1} \cdots a_{i_j}) \subseteq P(a_1 \cdots a_n) \setminus \{p\}\,.$$

The result follows because the translation by any group element is injective. □

For a letter $c \in A$ we write M_c for the local divisor $f(c)M \cap Mf(c)$ of M at $f(c)$. The proof of the following lemma gives an algorithm for lifting a factorization forest of M_c to the original monoid M.

Lemma 7.3. *Let $f : A^* \to M$ be a homomorphism to a finite monoid M, let $c \in A$, and let $g : A^* \to M_c$ be the homomorphism to the local divisor M_c of M at $f(c)$ defined by $g(b) = f(cbc)$ for $b \in A$. If $w_c = b_1 \cdots b_k$ with $b_i \in A$ has a factorization tree of height h for g, then $w = cb_1 \cdots cb_k$ has a factorization tree of height at most $4\,|M|\,h + 1$ for f.*

Proof. Let d_c be the factorization forest for g. The proof is by induction on the height of w_c. If $d_c(w_c) = (b_1 \cdots b_i,\ b_{i+1} \cdots b_k)$, then we let

$$d(w) = (cb_1 \cdots cb_i,\ cb_{i+1} \cdots cb_k).$$

Next we treat the case

$$d_c(w_c) = (u_1, \ldots, u_\ell) \tag{2}$$

with $\ell \geq 3$. In this situation $g(u_1) = \cdots = g(u_\ell)$ is idempotent in M_c. Each u_s is an element in A^* of the form $b_{i_s} \cdots b_{i_{s+1}-1}$, and we let $v_s = b_{i_s} c b_{i_s+1} \cdots c b_{i_{s+1}-1}$. Note that $w = cv_1 \cdots cv_\ell$ and $g(u_s) = f(cv_sc)$. Let

$$T(w) = \{f(v_s) \in M \mid f(v_{s'}) = f(v_s) \text{ for some } 1 \leq s' < s < \ell\}$$

be the elements with at least two occurrences. By induction on the size of $T(w)$ we translate the factorization in (2) into a factorization tree for w. If $T(w) = \emptyset$, then a tree of height $|M|$ is sufficient (in fact, even $\log |M| + 1$ would suffice). Let now $T(w) \neq \emptyset$. We choose some maximal subset $\{j_1, \ldots, j_t\}$ of $\{2, \ldots, \ell\}$ such that we have both $f(v_{j_1}) = \cdots = f(v_{j_t})$ and $j_{t'} + 1 < j_{t'+1}$ for all $1 \leq t' < t$. If $j_t \neq \ell$, we can write

$$w = cw_1 cv_{j_1} \cdots cw_t cv_{j_t} cw_{t+1}$$

with $w_i \neq 1$ and $f(v_{j_1}) \notin T(w_i)$. Note that

$$f(cw_1c) = \cdots = f(cw_tc) = f(cv_{j_1}c) = \cdots = f(cv_{j_t}c),$$

and this element is idempotent in M_c. The case $j_t = \ell$ is similar, but without the factor cw_{t+1} at the end. We have $f(cw_1cv_{j_1}) = \cdots = f(cw_tcv_{j_t})$, and by Proposition 2.1(d) this element is idempotent in M. Therefore, we can set

$$\begin{aligned} d(w) &= (cw_1cv_{j_1} \cdots cw_tcv_{j_t}, cw_{t+1}), \\ d(cw_1cv_{j_1} \cdots cw_tcv_{j_t}) &= (cw_1cv_{j_1}, \ldots, cw_tcv_{j_t}) && \text{if } t \geq 2, \\ d(cw_icv_{j_i}) &= (cw_i, cv_{j_i}) && \text{for } 1 \leq i \leq t. \end{aligned}$$

In total, we need 3 steps for every element in $T(w)$ (which in total adds at most $3\,|M|$ to the height); and then there are at most $|M|$ steps after $T(w)$ has become empty. Thus with a tree of height at most $4\,|M|$ we can simulate the factorization in equation (2). After simulating every factorization of d_c, we need one additional factorization for $d(cb_i) = (c, b_i)$. □

Proof of Theorem 7.1. If different letters are mapped to the same element in M, we can identify these letters without changing the height. Thus, we may assume $|A| \leq |M|$. The proof is by induction on $(|M|, |A|)$ with lexicographic order. Consider a word $w = a_1 \cdots a_n$ with $a_i \in A$. If all $f(a_i)$ are units, then w is mapped to a subgroup G of M, and we are done by Proposition 7.2. Therefore we may assume that w has some letter c such that $f(c)$ is not a unit. The word w admits a factorization

$$w = w_0cw_1cw_2 \cdots cw_k,$$

where c does not occur in any w_i for $0 \leq i \leq k$. By induction on the alphabet size of w, there exists a factorization tree of small height for each w_i. This allows us to treat each factor w_i as a letter. Let $b_i = f(w_i)$ and $A' = \{b_1, \ldots, b_k, f(c)\}$, let $f' : A'^* \to M$ be the homomorphism induced by the inclusion $A' \subseteq M$. If $w_i = 1$ is empty, then $b_i = 1$ is the neutral element of M, and this element is a *letter* in A' (not the empty word). Let $g : A'^* \to M_c$ be the homomorphism to the local divisor $M_c = f(c)M \cap Mf(c)$ of M at $f(c)$ defined by $g(b) = f(c)bf(c)$ for $b \in A'$. We have $|M_c| < |M|$ and hence by induction on the size of the monoid, there exists a factorization forest d_c for g of bounded height. By Lemma 7.3 the factorization tree for $w_c = b_1 \cdots b_k$ with respect to g can be translated into a factorization tree for $w' = f(c)b_1 \cdots f(c)b_k$ with respect to f', and the bound on the height of this tree only depends on $|M|$. Combining the tree for w' with the trees for the w_i yields a factorization tree for w such that its height only depends on M, but not on the length of w. □

Acknowledgements

We thank the anonymous reviewers for their numerous suggestions which significantly improved the presentation.

References

[1] R. Book and F. Otto. *String-Rewriting Systems*. Springer-Verlag, 1993.

[2] J. Chalopin and H. Leung. On factorization forests of finite height. *Theoretical Computer Science*, 310(1-3):489–499, 2004.

[3] V. Diekert and P. Gastin. LTL is expressively complete for Mazurkiewicz traces. *Journal of Computer and System Sciences*, 64:396–418, 2002.

[4] V. Diekert and P. Gastin. Pure future local temporal logics are expressively complete for Mazurkiewicz traces. *Information and Computation*, 204:1597–1619, 2006. Conference version in LATIN 2004, LNCS 2976, 170–182, 2004.

[5] V. Diekert and P. Gastin. First-order definable languages. In J. Flum, E. Grädel, and Th. Wilke, editors, *Logic and Automata: History and Perspectives*, Texts in Logic and Games, pages 261–306. Amsterdam University Press, 2008.

[6] V. Diekert, P. Gastin, and M. Kufleitner. A survey on small fragments of first-order logic over finite words. *International Journal of Foundations of Computer Science*, 19:513–548, 2008. Special issue DLT 2007.

[7] V. Diekert and M. Kufleitner. Omega-rational expressions with bounded synchronization delay. *Theory Comput. Syst.*, 2015.

[8] V. Diekert, M. Kufleitner, K. Reinhardt, and T. Walter. Regular languages are Church-Rosser congruential. In A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer, editors, *International Colloquium Automata, Languages and Programming (ICALP) 2012, Conference Proceedings, Part II*, volume 7392 of *Lecture Notes in Computer Science*, pages 177–188. Springer-Verlag, 2012.

[9] V. Diekert, M. Kufleitner, and B. Steinberg. The Krohn-Rhodes theorem and local divisors. *Fundamenta Informaticae*, 116(1-4):65–77, 2012.

[10] V. Diekert, M. Kufleitner, and P. Weil. Star-free languages are Church-Rosser congruential. *Theoretical Computer Science*, 454:129–135, 2012.

[11] A. Fernández López and M. Tocón Barroso. The local algebras of an associative algebra and their applications. In J. Misra, editor, *Applicable Mathematics in the Golden Age*, pages 254–275. Narosa, 2002.

[12] S. W. Golomb and B. Gordon. Codes with bounded synchronization delay. *Information and Control*, 8(4):355–372, 1965.

[13] J. A. W. Kamp. *Tense Logic and the Theory of Linear Order*. PhD thesis, University of California, Los Angeles (California), 1968.

[14] M. Kufleitner. The height of factorization forests. In *MFCS*, volume 5162 of *Lecture Notes in Computer Science*, pages 443–454. Springer-Verlag, 2008.

[15] D. Kuperberg. Linear temporal logic for regular cost functions. *Logical Methods in Computer Science*, 10:1–37, 2014.

[16] R. McNaughton, P. Narendran, and F. Otto. Church-Rosser Thue systems and formal languages. *J. ACM*, 35(2):324–344, 1988.

[17] K. Meyberg. Lectures on algebras and triple systems. Technical report, University of Virginia, Charlottesville, 1972.

[18] P. Narendran. *Church-Rosser and related Thue systems.* PhD thesis, Dept. of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY, USA, 1984.

[19] G. Niemann. *Church-Rosser Languages and Related Classes.* Kassel University Press, 2002. PhD thesis.

[20] G. Niemann and F. Otto. The Church-Rosser languages are the deterministic variants of the growing context-sensitive languages. *Inf. Comput.*, 197:1–21, 2005.

[21] G. Niemann and J. Waldmann. Some regular languages that are Church-Rosser congruential. In *DLT'01, Proceedings*, volume 2295 of *LNCS*, pages 330–339. Springer, 2002.

[22] K. Reinhardt and D. Thérien. Some more regular languages that are Church Rosser congruential. In *13. Theorietag, Automaten und Formale Sprachen, Herrsching, Germany*, pages 97–103, 2003.

[23] M. P. Schützenberger. On finite monoids having only trivial subgroups. *Inf. Control*, 8:190–194, 1965.

[24] M. P. Schützenberger. Sur certaines opérations de fermeture dans les langages rationnels. In *Symposia Mathematica, Vol.XV (Convegno di Informatica Teorica, INDAM, Roma, 1973)*, pages 245–253. Academic Press, London, 1975.

[25] I. Simon. Factorization forests of finite height. *Theoretical Computer Science*, 72(1):65–94, 1990.

[26] I. Simon. A short proof of the factorization forest theorem. In M. Nivat and A. Podelski, editors, *Tree Automata and Languages*, pages 433–438. Elsevier, 1992.