Decidability of Trajectory-Based Equations*

Technical Report 2003-472

Michael Domaratzki and Kai Salomaa
School of Computing, Queen's University
Kingston, ON Canada K7L 3N6
email: {domaratz,ksalomaa}@cs.queensu.ca

Abstract

We consider the decidability of existence of solutions to language equations involving the operations of shuffle and deletion along trajectories. These operations generalize the operations of concatenation, insertion, shuffle, quotient, sequential and scattered deletion, as well as many others. Our results are constructive in the sense that if a solution exists, it can be effectively represented. We show both positive and negative decidability results.

1 Introduction

Work on language equations is one of the core areas of formal language theory [10]. Much of the classical work deals with equations over the Boolean operations, concatenation and Kleene closure. Recent research [2, 5, 7, 8] has investigated the question of decidability of existence of solutions to equations of the form $X_1 \diamond X_2 = X_3$, where $\diamond$ is a binary operation on languages, and some of X_1, X_2, X_3 are fixed languages and some are unknowns.

As a particular case of the above type of equations we get the shuffle decomposition problem for regular languages, that is, the question whether a given regular language can be written as a shuffle of two languages in a non-trivial way. In spite of its apparent simplicity the question remains still open for general regular languages [2, 5]. The decomposition of regular languages with respect to catenation is known to be decidable [9, 13].

In this paper, we focus on operations $\diamond$ which are taken from the class of operations defined by shuffle on trajectories [11]. Shuffle on trajectories provides a unifying framework for studying various language composition operations. The complementary notion of deletion along trajectories introduced by the first author [3] provides, in the sense of Kari [7], the inverse of shuffle on trajectories and makes it possible to attack in a systematic way questions of decidability of existence of solutions to equations involving shuffle on

*Research supported in part by NSERC.

trajectories. Some positive decidability results have already been completed by the first author [3].

We establish for certain classes of trajectories the decidability of the existence of a decomposition for a given regular language. However, our results leave open the question for the trajectory $(0+1)^*$ corresponding to ordinary shuffle [2, 5]. Also we show that for given regular languages L_1, L_2, and R we can decide whether or not there exists a trajectory T such that $L_1 \sqcup\!\sqcup_T L_2 = R$, where $\sqcup\!\sqcup_T$ denotes shuffle along the trajectory T. To conclude we provide undecidability results for equations involving one or two variables.

2 Definitions and Preliminary Results

For additional background in formal languages and automata theory, please see Yu [14]. Let Σ be a finite set of symbols, called *letters*. Then Σ^* is the set of all finite sequences of letters from Σ, which are called *words*. The empty word ϵ is the empty sequence of letters. The *length* of a word $w = w_1 w_2 \cdots w_n \in \Sigma^*$, where $w_i \in \Sigma$, is n, and is denoted $|w|$. Note that ϵ is the unique word of length 0. A *language* L is any subset of Σ^*. By $\overline{L}$, we mean $\Sigma^* - L$, the complement of L. If L is a language over alphabet Σ, we denote by $\mathrm{alph}(L)$ the set of all symbols of Σ occurring in words of L ($\mathrm{alph}(L) \subseteq \Sigma$).

A *deterministic finite automaton* (DFA) is a five-tuple $M = (Q, \Sigma, \delta, q_0, F)$ where Q is the finite set of states, Σ is the alphabet, $\delta : Q \times \Sigma \to \Sigma$ is the transition function, $q_0 \in Q$ is the distinguished start state, and $F \subseteq Q$ is the set of final states. We extend δ to $Q \times \Sigma^*$ in the usual way. A word $w \in \Sigma^*$ is accepted by M if $\delta(q_0, w) \in F$. The *language accepted* by M, denoted $L(M)$ is the set of all words accepted by M. A language is called *regular* if it is accepted by some DFA. A DFA $M = (Q, \Sigma, \delta, q_0, F)$ is *complete* if $\delta(q, a)$ is defined for all $(q, a) \in Q \times \Sigma$.

A *nondeterministic finite automaton* (NFA) is a five-tuple $M = (Q, \Sigma, \delta, q_0, F)$ where Q, Σ, q_0 and F are as in the deterministic case, while $\delta : Q \times (\Sigma \cup \epsilon) \to 2^Q$ is the nondeterministic transition function. Again, δ is extended to $Q \times \Sigma^*$ in the natural way. A word w is accepted by M if $\delta(q_0, w) \cap F \neq \emptyset$. It is known that the language accepted by an NFA is regular.

We recall the definition of shuffle on trajectories, originally given by Mateescu *et al.* [11]. Shuffle on trajectories is defined by first defining the shuffle of two words x and y over an alphabet Σ on a trajectory t, which is simply a word in $\{0, 1\}^*$. We denote the shuffle of x and y along the trajectory t by $x \sqcup\!\sqcup_t y$.

If $x = ax'$ and $y = by'$ (with $a, b \in \Sigma$) then if $t = et'$ (with $e \in \{0, 1\}$), we have that

$$x \sqcup\!\sqcup_{et'} y = \begin{cases} a(x' \sqcup\!\sqcup_{t'} by') & \text{if } e = 0; \\ b(ax' \sqcup\!\sqcup_{t'} y') & \text{if } e = 1. \end{cases}$$

If $x = ax'$ $(a \in \Sigma)$ and $y = \epsilon$, then

$$x \sqcup\!\sqcup_{et'} \epsilon = \begin{cases} a(x' \sqcup\!\sqcup_{t'} \epsilon) & \text{if } e = 0; \\ \emptyset & \text{otherwise.} \end{cases}$$

If $x = \epsilon$ and $y = by'$ ($b \in \Sigma$), then

$$\epsilon \sqcup\!\sqcup_{et'} y = \begin{cases} b(\epsilon \sqcup\!\sqcup_{t'} y') & \text{if } e = 1; \\ \emptyset & \text{otherwise.} \end{cases}$$

If $x = y = \epsilon$, then $\epsilon \sqcup\!\sqcup_t \epsilon = \epsilon$ if $t = \epsilon$ and $\emptyset$ otherwise. Finally, $x \sqcup\!\sqcup_\epsilon y = \emptyset$ if $\{x, y\} \neq \{\epsilon\}$.

We extend shuffle on trajectories to sets $T \subseteq \{0,1\}^*$ of trajectories as follows:

$$x \sqcup\!\sqcup_T y = \bigcup_{t \in T} x \sqcup\!\sqcup_t y.$$

Further, for $L_1, L_2 \subseteq \Sigma^*$, we define

$$L_1 \sqcup\!\sqcup_T L_2 = \bigcup_{\substack{x \in L_1 \\ y \in L_2}} x \sqcup\!\sqcup_T y.$$

We now give the definition of *deletion along trajectories* [3], which models deletion operations controlled by a set of trajectories. Let $x, y \in \Sigma^*$ be words with $x = ax'$, $y = by'$ ($a, b \in \Sigma$). Let t be a word over $\{i, d\}$ such that $t = et'$ with $e \in \{i, d\}$. then we define $x \rightsquigarrow_t y$ as follows:

$$x \rightsquigarrow_t y = \begin{cases} a(x' \rightsquigarrow_{t'} by') & \text{if } e = i; \\ x' \rightsquigarrow_{t'} y' & \text{if } e = d \text{ and } a = b; \\ \emptyset & \text{otherwise.} \end{cases}$$

Also,

$$x \rightsquigarrow_t \epsilon = \begin{cases} a(x' \rightsquigarrow_{t'} \epsilon) & \text{if } e = i; \\ \emptyset & \text{otherwise.} \end{cases}$$

Further, $\epsilon \rightsquigarrow_t y = \epsilon$ if $t = y = \epsilon$. Otherwise, $\epsilon \rightsquigarrow_t y = \emptyset$. Finally, $x \rightsquigarrow_\epsilon y = \emptyset$ if $x \neq \epsilon$.

Let $T \subseteq \{i, d\}^*$. Then

$$x \rightsquigarrow_T y = \bigcup_{t \in T} x \rightsquigarrow_t y.$$

We extend this to languages as expected: Let $L_1, L_2 \subseteq \Sigma^*$ and $T \subseteq \{i, d\}^*$. Then

$$L_1 \rightsquigarrow_T L_2 = \bigcup_{\substack{x \in L_1 \\ y \in L_2}} x \rightsquigarrow_T y.$$

Note that $\rightsquigarrow_T$ is neither an associative nor a commutative operation on languages, in general. For the closure properties of $\rightsquigarrow_T$, please see [3].

Given two binary word operations $\diamond, \star : (\Sigma^*)^2 \to 2^{\Sigma^*}$, we say that $\diamond$ is a *left-inverse* of $\star$ [7, Defn. 4.5] if, for all $u, v, w \in \Sigma^*$, $w \in u \star v \iff u \in w \diamond v$. We say that $\diamond$ is a *right-inverse* of $\star$ [7, Defn. 4.1] if, for all $u, v, w \in \Sigma^*$, $w \in u \star v \iff v \in u \diamond w$.

Let $\tau : \{0,1\}^* \to \{i, d\}^*$ be the morphism given by $\tau(0) = i$ and $\tau(1) = d$. The following result will prove useful [3]:

Theorem 2.1 *Let $T \subseteq \{0,1\}^*$ be a set of trajectories. Then $\sqcup\!\sqcup_T$ and $\rightsquigarrow_{\tau(T)}$ are left-inverses of each other.*

Similarly, let $\pi : \{0,1\}^* \to \{i,d\}^*$ be the morphism given by $\pi(0) = d$ and $\pi(1) = i$. Given an operation $\diamond$, let $\diamond^R$ be the operation defined by $x \diamond^R y = y \diamond x$ for all $x, y \in \Sigma^*$.

Theorem 2.2 *Let $T \subseteq \{0,1\}^*$ be a set of trajectories. Then $\sqcup\!\sqcup_T$ and $(\rightsquigarrow_{\pi(T)})^R$ are right-inverses of each other.*

The following results of Kari [7, Thms. 4.2 and 4.6] allow us to find solutions to equations involving shuffle on trajectories.

Theorem 2.3 *Let L, R be languages over Σ and $\diamond, \star$ be two binary word operations, which are left-inverses to each other. If the equation $X \diamond L = R$ has a solution $X \subseteq \Sigma^*$, then the language $R' = \overline{\overline{R} \star L}$ is also a solution of the equation. Moreover, R' is a superset of all other solutions of the equation.*

Theorem 2.4 *Let L, R be languages over Σ and $\diamond, \star$ be two binary word operations, which are right-inverses to each other. If the equation $L \diamond X = R$ has a solution $X \subseteq \Sigma^*$, then the language $R' = \overline{L \star \overline{R}}$ is also a solution of the equation. Moreover, R' is a superset of all other solutions of the equation.*

3 Decidability of Shuffle Decompositions

Say that a language L has a *non-trivial shuffle decomposition* with respect to a set of trajectories $T \subseteq \{0,1\}^*$ if there exist $X_1, X_2 \neq \{\epsilon\}$ such that $L = X_1 \sqcup\!\sqcup_T X_2$.

In this section, we are concerned with giving a class of sets of trajectories $T \subseteq \{0,1\}^*$ such that it is decidable, given a regular language R, whether R has a non-trivial shuffle decomposition with respect to T. For $T = (0+1)^*$, this is an open problem [2, 5]. While we do not settle this open problem, we establish a non-trivial generalization of the results of Kari and Kari and Thierrin [6, 7, 8, 9], which leads to a large class of examples of trajectories where the shuffle problem can be proven to be decidable.

A language $L \subseteq \Sigma^*$ is *bounded* if there exist $w_1, w_2, \ldots, w_n \in \Sigma^*$ such that $L \subseteq w_1^* w_2^* \cdots w_n^*$. Say that L is *letter-bounded* if $w_i \in \Sigma$ for all $1 \leq i \leq n$.

We now define a class of letter-bounded sets of trajectories, called *i-regular* sets of trajectories, which will have strong closure properties. In particular, we can delete, along an i-regular set of letter-bounded trajectories, any language from a regular language and the resulting language will be regular. This will allow us to solve the corresponding decidability problems related to the shuffle decomposition.

Let Δ_m be the alphabet $\Delta_m = \{\#_1, \#_2, \ldots, \#_m\}$ for any $m \geq 1$. We define a class of regular substitutions from $(d + \Delta_m)^*$ to $2^{(i+d)^*}$, denoted $\mathfrak{S}_m$, as follows: a regular substitution $\varphi : (d + \Delta_m)^* \to 2^{(i+d)^*}$ is in $\mathfrak{S}_m$ if both

(a) $\varphi(d) = \{d\}$; and

(b) $\varphi(\#_j) \subseteq i^*$ for all $1 \le j \le m$.

For all $m \ge 1$, we also define a class of languages over the alphabet $d + \Delta_m$, denoted $\mathfrak{T}_m$, as the set of all languages $T \subseteq \#_1 d^* \#_2 d^* \cdots \#_{m-1} d^* \#_m$. Define the class of trajectories $\mathfrak{I}$ as follows:

$$\mathfrak{I} = \{T \subseteq \{i,d\}^* \ : \ \exists m \ge 1, T_m \in \mathfrak{T}_m, \varphi \in \mathfrak{S}_m \text{ such that } T = \varphi(T_m)\}.$$

If $T \in \mathfrak{I}$, we say that T is *i-regular*. As we shall see, the condition that T be i-regular is sufficient for showing that $R \rightsquigarrow_T L$ is regular for all regular languages R and all languages L.

We will require the following result of Ginsburg and Spanier [4] on bounded regular languages:

Theorem 3.1 *Let $L \subseteq w_1^* w_2^* \cdots w_k^*$ be a regular language. Then there exist constants $N, b_{i,j}, c_{i,j}$ such that $1 \le i \le N$ and $1 \le j \le k$ such that*

$$L = \bigcup_{i=1}^{N} w_1^{b_{i,1}} (w_1^{c_{i,1}})^* \cdots w_k^{b_{i,k}} (w_k^{c_{i,k}})^*. \tag{3.1}$$

Theorem 3.2 *Let $T \in \mathfrak{I}$. Then for all regular languages R and all languages L, $R \rightsquigarrow_T L$ is a regular language.*

Proof. Let $T \in \mathfrak{I}$. Let $m \ge 1$, $T' \in \mathfrak{T}_m$ and $\varphi \in \mathfrak{S}_m$ be such that $T = \varphi(T')$. Then we define $K(T) \subseteq \mathbb{N}^{m-1}$ as

$$K(T) = \{(j_1, \ldots, j_{m-1}) \ : \ \#_1 d^{j_1} \#_2 d^{j_2} \cdots \#_{m-1} d^{j_{m-1}} \#_m \in T'\}.$$

As φ is a regular substitution, $\varphi(\#_j) \subseteq i^*$ is a bounded regular language for all $1 \le j \le m$. By Theorem 3.1, let $N_j \ge 1$ and $a_r^{(j)}, b_r^{(j)}$ for $1 \le j \le m$ and $1 \le r \le N_j$ be such that

$$\varphi(\#_j) = \bigcup_{r=1}^{N_j} i^{a_r^{(j)}} (i^{b_r^{(j)}})^*$$

for all $1 \le j \le m$. We may assume that $N_j = 1$ for all $1 \le j \le m$, since we may establish the result for $N_j > 1$ by proving the result for $N_j = 1$ and noting the fact that $R \rightsquigarrow_{T_1 \cup T_2} L_1 = (R \rightsquigarrow_{T_1} L_1) \cup (R \rightsquigarrow_{T_2} L_1)$. Thus, we let a_j, b_j be defined so that $\varphi(\#_j) = i^{a_j} (i^{b_j})^*$ for all $1 \le j \le m$. Let $I_j = \{a_j + n b_j \ : \ n \ge 0\}$.

Let R be regular and L be arbitrary. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting R. For all $q_j, q_k \in Q$, let $R(q_j, q_k) = L((Q, \Sigma, \delta, q_j, \{q_k\}))$. For $I \subseteq \mathbb{N}$, let $R'_I(q_j, q_k) = R(q_j, q_k) \cap \{x \ : \ |x| \in I\}$.

We now define the set $Q_R(T, L) \subseteq Q^{2m-2}$:

$$Q_R(T,L) = \{(q_j)_{j=1}^{2m-2} \ : \ \exists (k_j)_{j=1}^{m-1} \in K(T) \text{ such that } L \cap \prod_{\ell=1}^{m-1} R'_{\{k_\ell\}}(q_{2\ell-1}, q_{2\ell}) \ne \emptyset\}. \tag{3.2}$$

We now claim that

$$R \rightsquigarrow_T L = \bigcup_{\substack{(q_j)_{j=1}^{2m-2} \in Q_R(T,L) \\ q_f \in F}} \left(\prod_{\ell=1}^{m-1} R'_{I_\ell}(q_{2(\ell-1)}, q_{2\ell-1}) \right) \cdot R'_{I_m}(q_{2m-2}, q_f). \tag{3.3}$$

Let $x \in R \rightsquigarrow_T L$. Then we can write $x = x_1 x_2 \cdots x_m$ such that there exists some $z = z_1 z_2 \cdots z_{m-1} \in L$ such that $y = x_1 z_1 x_2 z_2 \cdots x_{m-1} z_{m-1} x_m \in R$. Further, by the conditions on T, $(|z_j|)_{j=1}^{m-1} \in K(T)$ and $|x_j| \in I_j$ for all $1 \le j \le m$. We let $q \overset{x}{\vdash} q'$ denote the fact that $\delta(q, x) = q'$ in M. As $y \in R$, there are some $q_1, q_2, \ldots, q_{2m-2}, q_f \in Q$ such that

$$q_0 \overset{x_1}{\vdash} q_1 \overset{z_1}{\vdash} q_2 \overset{x_2}{\vdash} \cdots \overset{x_{m-1}}{\vdash} q_{2m-3} \overset{z_{m-1}}{\vdash} q_{2m-2} \overset{x_m}{\vdash} q_f$$

and $q_f \in F$. Then $z_j \in R'_{\{|z_j|\}}(q_{2j-1}, q_{2j})$ for all $1 \le j \le m-1$, $x_j \in R'_{I_j}(q_{2(j-1)}, q_{2j-1})$ for all $1 \le j \le m-1$ and $x_m \in R'_{I_m}(q_{2m-2}, q_f)$. Further, note that

$$z \in L \cap \prod_{\ell=1}^{m-1} R'_{\{|z_\ell|\}}(q_{2\ell-1}, q_{2\ell}).$$

We conclude that $(q_1, q_2, \ldots, q_{2m-2}) \in Q_R(T, L)$, as $(|z_j|)_{j=1}^{m-1} \in K(T)$, and thus x is contained in the right-hand side of (3.3).

For the reverse inclusion, let $(q_1, q_2, \ldots, q_{2m-2}) \in Q_R(T, L)$ and $q_f \in F$. Let $(k_1, \ldots, k_{m-1}) \in K(T)$ be the $(m-1)$-tuple which witnesses $(q_i)_{i=1}^{2m-2}$'s membership in $Q_R(T, L)$. Then we show that $(\prod_{\ell=1}^{m-1} R'_{I_\ell}(q_{2(\ell-1)}, q_{2\ell})) R'_{I_m}(q_{2m-2}, q_f) \subseteq R \rightsquigarrow_T L$.

Let $z_j \in R'_{\{k_j\}}(q_{2j-1}, q_{2j})$ for all $1 \le j \le m-1$ be such that $z = z_1 \cdots z_{m-1} \in L$. Such z_j exist by definition of $Q_R(T, L)$. Let $x_j \in R'_{I_j}(q_{2(j-1)}, q_{2j-1})$ for all $1 \le j \le m-1$, and $x_m \in R'_{I_m}(q_{2m-2}, q_f)$. Then

$$q_0 \overset{x_1}{\vdash} q_1 \overset{z_1}{\vdash} q_2 \overset{x_2}{\vdash} \cdots \overset{x_{m-1}}{\vdash} q_{2m-3} \overset{z_{m-1}}{\vdash} q_{2m-2} \overset{x_m}{\vdash} q_f.$$

Thus, $y = x_1 z_1 \cdots x_{m-1} z_{m-1} x_m \in R$. Further, the length considerations are met by definition of I_j and $(k_j)_{j=1}^{m-1} \in K(T)$. Thus $x \in y \rightsquigarrow_T z \subseteq R \rightsquigarrow_T L$.

Thus, since $Q_R(T, L)$ is finite, $R \rightsquigarrow_T L$ is a finite union of regular languages, and thus is regular. ■

We note that if T is not letter-bounded, it may define an operation which does not preserve regularity in the sense of Theorem 3.2. In particular, we note that for $T = (di)^*$,

$$(a^2)^*(b^2)^* \rightsquigarrow_T \{a^n b^n \ : \ n \ge 0\} = \{a^n b^n \ : \ n \ge 0\},$$

a non-regular context-free language (CFL). For $T = (i + d)^*$, we have that

$$((ab)^* \# (ab)^* \rightsquigarrow_T \{a^n \# b^n \ : \ n \ge 0\}) \cap b^* a^* = \{b^n a^n \ : \ n \ge 0\}.$$

Further, if T is letter-bounded but not i-regular, then T may not preserve regularity. For example, let $T = \{i^n d i^n \ : \ n \ge 0\}$. Then $a^* b c^* \rightsquigarrow_T \{b\} = \{a^n c^n \ : \ n \ge 0\}$.

As an example of Theorem 3.2, consider $T = \{d^n i^m d^n \ : \ n, m \geq 0\}$. It is easily verified that $T \in \mathfrak{I}$. Thus, the language $R \rightsquigarrow_T L$ is regular for all regular languages R and all languages L. For any language $L \subseteq \Sigma^*$, define $sq(L) = \{x^2 \ : \ x \in L\}$. Consider then that

$$R \rightsquigarrow_T sq(L) = \{w \ : vwv \in R, v \in L\}.$$

This precisely defines the *middle-quotient* operation, which has been investigated by Meduna [12] for linear CFLs. Let $R \,|\, L$ denote the middle quotient of R by L, i.e., $R \,|\, L = R \rightsquigarrow_T sq(L)$. Thus, we can immediately conclude the following result, which was not considered by Meduna:

Theorem 3.3 *Given a regular language R and arbitrary language L, the language $R \,|\, L$ is regular.*

We now return to letter-bounded sets of trajectories. Clearly, every letter-bounded set of trajectories which is regular is also i-regular (consider Theorem 3.1). Thus, we have the following corollary of Theorem 3.2.

Corollary 3.4 *Let $T \subseteq \{i, d\}^*$ be a letter-bounded regular set of trajectories. Then for all regular languages R, there are only finitely many regular languages L' such that $L' = R \rightsquigarrow_T L$ for some language L. Furthermore, this finite set of regular languages can be effectively constructed, given an effective construction for T and R.*

Proof. Let R be a regular language accepted by a DFA $M = (Q, \Sigma, \delta, q_0, F)$. Let $T \subseteq (i^* d^*)^m i^*$ for some $m \geq 0$. By (3.3), we know that if $Q_R(T, L) = Q_R(T, L')$, then $R \rightsquigarrow_T L = R \rightsquigarrow_T L'$.

Note that, for all $L \subseteq \Sigma^*$, $Q_R(T, L) \subseteq Q^{2m-2}$. As Q^{2m-2} is a finite set, there are only finitely many languages of the form $R \rightsquigarrow_T L$. This set can be effectively obtained by considering all possible choices of sets $Q' \subseteq Q^{2m-2}$, and constructing the (effective) regular language from (3.3) with $Q' = Q_R(T, L)$ (duplicates may also then be removed, as we can effectively compare the resulting (effectively) regular languages). ■

Theorem 3.5 *Let $T \subseteq \{0, 1\}^*$ be a letter-bounded regular set of trajectories. Let R be a regular language over an alphabet Σ. Then there exists a natural number $n \geq 1$ such that there are n distinct regular languages Y_i with $1 \leq i \leq n$ such that for any $L \subseteq \Sigma^*$ the following are equivalent:*

(a) there exists a solution $Y \subseteq \Sigma^*$ to the equation $L \sqcup\!\sqcup_T Y = R$;

(b) there exists an index i with $1 \leq i \leq n$ such that $L \sqcup\!\sqcup_T Y_i = R$.

The languages Y_i can be effectively constructed, given effective constructions for T and R. Further, if Y is a solution to $L \sqcup\!\sqcup_T Y = R$, then there is some $1 \leq i \leq n$ such that $Y \subseteq Y_i$.

Proof. Let T, R be given. Then consider the finite set $\mathcal{S}_L(T, R) = \{\overline{(\overline{R} \rightsquigarrow_{\pi(T)} L)} : L \subseteq \Sigma^*\}$. This set is finite by Corollary 3.4. Let L be arbitrary. Thus, if $L \sqcup\!\sqcup_T Y = R$, then $Y \subseteq X$ for some $X \in \mathcal{S}_L(T, R)$ by Theorems 2.2 and 2.4. Thus, (a) implies (b). The implication (b) implies (a) is trivial. ■

The symmetric result also holds:

Theorem 3.6 *Let $T \subseteq \{0,1\}^*$ be a letter-bounded regular set of trajectories. Let R be a regular language over an alphabet Σ. Then there exists a natural number $n \geq 1$ such that there are n distinct regular languages Z_i with $1 \leq i \leq n$ such that for any $L \subseteq \Sigma^*$ the following are equivalent:*

(a) there exists a solution $Z \subseteq \Sigma^*$ to the equation $Z \sqcup\!\sqcup_T L = R$;

(b) there exists an index i with $1 \leq i \leq n$ such that $Z_i \sqcup\!\sqcup_T L = R$.

The languages Z_i can be effectively constructed, given effective constructions for T and R. Further, if Z is a solution to $Z \sqcup\!\sqcup_T L = R$, then there is some $1 \leq i \leq n$ such that $Z \subseteq Z_i$.

Theorem 3.7 *Let $T \subseteq \{0,1\}^*$ be a letter-bounded regular set of trajectories. Then given a regular language R, it is decidable whether there exist X_1, X_2 such that $X_1 \sqcup\!\sqcup_T X_2 = R$.*

Proof. Let $\mathcal{S}_L(T,R) = \{Y_i\}_{i=1}^{n_1}$ be the set of languages described by Theorem 3.5 and, analogously, let $\mathcal{T}_L(T,R) = \{Z_i\}_{i=1}^{n_2}$ be the set of languages described by Theorem 3.6.

We now note the result follows since if $X_1 \sqcup\!\sqcup_T X_2 = R$ has a solution, it also has a solution in $\mathcal{S}_L(T,R) \times \mathcal{T}_L(T,R)$, since $\sqcup\!\sqcup_T$ is monotone. Thus, we simply test all the finite (non-trivial) pairs in $\mathcal{S}_L(T,R) \times \mathcal{T}_L(T,R)$ for the desired equality. ■

This result was known for catenation, $T = 0^*1^*$ (see, e.g., Kari and Thierrin [9]). However, it also holds for, e.g., the following operations: insertion ($0^*1^*0^*$), k-insertion ($0^*1^*0^{\leq k}$ for fixed $k \geq 0$), and bi-catenation ($1^*0^* + 0^*1^*$).

3.1 1-thin sets of trajectories

Recall that a language L is *1-thin* if $|L \cap \Sigma^n| \leq 1$ for all $n \geq 0$. We now prove that if $T \subseteq \{0,1\}^*$ is a fixed 1-thin set of trajectories, given R regular, it is decidable whether R has a non-trivial shuffle decomposition with respect to T.

Define the right-useful solutions to $L \sqcup\!\sqcup_T X = R$ as

$$use_T^{(r)}(X;L) = \{x \in X \ : \ L \sqcup\!\sqcup_T x \neq \emptyset\}. \tag{3.4}$$

The left-useful solutions, denoted $use_T^{(\ell)}(X;L)$, are defined similarly for the equation $X \sqcup\!\sqcup_T L = R$.

Theorem 3.8 *Let $T \subseteq \{0,1\}^*$ be a 1-thin set of trajectories. Given a regular language R, the existence of X_1, X_2 such that $R = X_1 \sqcup\!\sqcup_T X_2$ is decidable.*

Proof. Let

$$\begin{aligned} L_1 &= R \rightsquigarrow_{\tau(T)} \Sigma^* \\ L_2 &= R \rightsquigarrow_{\pi(T)} \Sigma^*. \end{aligned}$$

Then we claim that

$$\exists X_1, X_2 \text{ such that } R = X_1 \sqcup\!\sqcup_T X_2 \iff L_1 \sqcup\!\sqcup_T L_2 = R. \tag{3.5}$$

The right-to-left implication is trivial. To prove the reverse implication, we first show that if $X_1 \sqcup\!\sqcup_T X_2 = R$, then $use_T^{(\ell)}(X_1; X_2) \subseteq L_1$ and $use_T^{(r)}(X_2; X_1) \subseteq L_2$.

We show only that $use_T^{(\ell)}(X_1; X_2) \subseteq L_1$. The other inclusion is proven similarly. Let $x \in use_T^{(\ell)}(X_1; X_2)$. Then there is some $y \in X_2$ such that $x \sqcup\!\sqcup_T y \neq \emptyset$. As $X_1 \sqcup\!\sqcup_T X_2 = R$, we must have that for all $z \in x \sqcup\!\sqcup_T y$, $z \in R$. Thus, by Theorem 2.1, $x \in z \rightsquigarrow_{\tau(T)} y \subseteq L_1$. The inclusion is proven. Thus,

$$R = X_1 \sqcup\!\sqcup_T X_2 = use_T^{(\ell)}(X_1; X_2) \sqcup\!\sqcup_T use_T^{(r)}(X_2; X_1) \subseteq L_1 \sqcup\!\sqcup_T L_2.$$

To conclude the proof, we need only establish the inclusion $L_1 \sqcup\!\sqcup_T L_2 \subseteq R$.

Let $x \in L_1$. Thus, there exists $\alpha \in R$, $\beta \in \Sigma^*$ and $t \in T$ such that $x \in \alpha \rightsquigarrow_t \beta$. Thus, $\{\alpha\} = x \sqcup\!\sqcup_t \beta$. Now, as $\alpha \in R = X_1 \sqcup\!\sqcup_T X_2$, there is some $x_1 \in X_1$, $x_2 \in X_2$ and $t' \in T$ such that $\{\alpha\} = x_1 \sqcup\!\sqcup_{t'} x_2$.

Consider now that $|t| = |\alpha| = |t'|$. As T is 1-thin, this implies that $t = t'$. Thus,

$$x \sqcup\!\sqcup_t \beta = x_1 \sqcup\!\sqcup_t x_2,$$

or, $x \in (x_1 \sqcup\!\sqcup_t x_2) \rightsquigarrow_{\tau(t)} \beta$, from which it is clear that $x = x_1$ and $x_2 = \beta$. Thus, $x \in X_1$. A similar argument establishes that $L_2 \subseteq X_2$. Thus, we have established that $R = L_1 \sqcup\!\sqcup_T L_2$ and (3.5) holds. ■

We note that Theorem 3.7 and Theorem 3.8 do not apply to the following sets of trajectories. Thus, to our knowledge, the question of the decidability of the existence of solutions to $R = X_1 \sqcup\!\sqcup_T X_2$ for regular R is still open in the following cases (for details on literal and initial literal shuffle, see Berard [1]):

(a) arbitrary shuffle: $T = (0 + 1)^*$;

(b) literal shuffle: $T = (0^* + 1^*)(01)^*(0^* + 1^*)$;

(c) initial literal shuffle: $T = (01)^*(0^* + 1^*)$.

3.2 Solving Quadratic Equations

Let $T \subseteq \{0, 1\}^*$ be a letter-bounded regular set of trajectories. We can also consider solutions X to the equation $X \sqcup\!\sqcup_T X = R$, for regular languages R. This is a generalization of a result due to Kari and Thierrin [8].

Theorem 3.9 *Fix a letter-bounded regular set of trajectories T. Then it is decidable whether there exists a solution X to the equation $X \sqcup\!\sqcup_T X = R$ for a given regular language R.*

Proof. Let $\mathcal{S}_L(T, R)$ be the set of languages described by Theorem 3.5, and, analogously, let $\mathcal{T}_L(T, R)$ be the set of languages described by Theorem 3.6.

Assume the equation $X \sqcup\!\sqcup_T X = R$ has a solution. Then we claim that it also has a regular solution. Let X be a language such that $X \sqcup\!\sqcup_T X = R$. Then, in particular, X is a solution to the equation $X \sqcup\!\sqcup_T Y = R$, where X is fixed and Y is a variable. Thus, by Theorem 3.5, there is some regular language $Y_i \in \mathcal{S}_X(T, R)$ such that $X \sqcup\!\sqcup_T Y_i = R$. Further, $X \subseteq Y_i$. Analogously, considering the equation $X \sqcup\!\sqcup_T Y_i = R$, $X \subseteq Z_j$ for some regular language $Z_j \in \mathcal{T}_{Y_i}(T, R)$. Thus, $X \subseteq Y_i \cap Z_j$, and $Z_j \sqcup\!\sqcup_T Y_i = R$.

Let $X_0 = Y_i \cap Z_j$. Then note that $R = X \sqcup\!\sqcup_T X \subseteq X_0 \sqcup\!\sqcup_T X_0 \subseteq Z_j \sqcup\!\sqcup_T Y_i = R$. The inclusion follows by the monotonicity of $\sqcup\!\sqcup_T$. Thus, $X_0 \sqcup\!\sqcup_T X_0 = R$. By construction, X_0 is regular.

Thus, to decide whether there exists X such that $X \sqcup\!\sqcup_T X = R$, we construct the set

$$\mathcal{U}_X(T, R) = \{Y_i \cap Z_j \ : \ Y_i \in \mathcal{S}_X(T, R), Z_j \in \mathcal{T}_{Y_i}(T, R)\},$$

and test each language for equality. If a solution exists, we answer yes. Otherwise, we answer no. ■

4 Existence of Trajectories

In this section, we consider the following problem: given languages L_1, L_2 and R, does there exist a set of trajectories T such that $L_1 \sqcup\!\sqcup_T L_2 = R$? We prove this to be decidable when L_1, L_2, R are regular languages.

Theorem 4.1 *Let $L_1, L_2, R \subseteq \Sigma^*$ be regular languages. Then it is decidable whether there exists a set $T \subseteq \{0,1\}^*$ of trajectories such that $L_1 \sqcup\!\sqcup_T L_2 = R$.*

Proof. Let

$$T_0 = \{t \in \{0,1\}^* \ : \ \forall x \in L_1, y \in L_2, x \sqcup\!\sqcup_t y \subseteq R\}. \tag{4.6}$$

Note that the following are equivalent definitions of T_0:

$$\begin{aligned} T_0 &= \{t \in \{0,1\}^* \ : \ \forall x \in L_1, y \in L_2, (x \sqcup\!\sqcup_t y \neq \emptyset \Rightarrow x \sqcup\!\sqcup_t y \subseteq R)\}; && (4.7) \\ T_0 &= \{t \in \{0,1\}^* \ : \ \forall x \in L_1 \cap \Sigma^{|t|_0}, y \in L_2 \cap \Sigma^{|t|_1}, (x \sqcup\!\sqcup_t y \subseteq R)\}. && (4.8) \end{aligned}$$

Then we claim that

$$\exists T \subseteq \{0,1\}^* \text{ such that } (L_1 \sqcup\!\sqcup_T L_2 = R) \iff L_1 \sqcup\!\sqcup_{T_0} L_2 = R.$$

The right-to-left implication is trivial. Assume that there is some $T \subseteq \{0,1\}^*$ such that $L_1 \sqcup\!\sqcup_T L_2 = R$. Let $t \in T$. Then for all $x \in L_1$ and $y \in L_2$, $x \sqcup\!\sqcup_t y \subseteq L_1 \sqcup\!\sqcup_T L_2 = R$. Thus, $t \in T_0$ by definition, and $T_0 \supseteq T$.

Thus, note that $R = L_1 \sqcup\!\sqcup_T L_2 \subseteq L_1 \sqcup\!\sqcup_{T_0} L_2$. It remains to establish that $L_1 \sqcup\!\sqcup_{T_0} L_2 \subseteq R$. But this is clear from the definition of T_0. Thus $L_1 \sqcup\!\sqcup_{T_0} L_2 = R$ and the claim is established.

We now establish that T_0 is regular and effectively constructible; to do this, we establish instead that $\overline{T_0} = \{0,1\}^* - T_0$ is regular.

Let $M_j = (Q_j, \Sigma, \delta_j, q_j, F_j)$ be a complete DFA accepting L_j for $j = 1, 2$. Let $M_r = (Q_r, \Sigma, \delta_r, q_r, F_r)$ be a complete DFA accepting R. Define an NFA $M = (Q, \{0,1\}, \delta, q_0, F)$ where $Q = Q_1 \times Q_2 \times Q_r$, $q_0 = [q_1, q_2, q_r]$, $F = F_1 \times F_2 \times (Q_r - F_r)$, and δ is defined as follows:

$$\begin{aligned} \delta([q_j, q_k, q_\ell], 0) &= \{[\delta_1(q_j, a), q_k, \delta_r(q_\ell, a)] \ : \ a \in \Sigma\} \quad \forall [q_j, q_k, q_\ell] \in Q_1 \times Q_2 \times Q_r, \\ \delta([q_j, q_k, q_\ell], 1) &= \{[q_j, \delta_2(q_k, a), \delta_r(q_\ell, a)] \ : \ a \in \Sigma\} \quad \forall [q_j, q_k, q_\ell] \in Q_1 \times Q_2 \times Q_r. \end{aligned}$$

Then we note that δ has the following property: for all $t \in \{0,1\}^*$,

$$\delta([q_1, q_2, q_r], t) = \{[\delta(q_1, x), \delta(q_2, y), \delta(q_r, x \sqcup\!\sqcup_t y)] \ : \ x, y \in \Sigma^*, |x| = |t|_0, |y| = |t|_1\}.$$

By (4.8), if $t \in \overline{T_0}$ there is some $x, y \in \Sigma^*$ such that $x \in L_1$, $y \in L_2$, $|x| = |t|_0$, $|y| = |t|_1$ but $x \sqcup\!\sqcup_t y \cap \overline{R} \neq \emptyset$. This is exactly what is reflected by the choice of F. Thus, $L(M) = \overline{T_0}$.

Thus, as T_0 is effectively regular, to determine whether there exists T such that $L_1 \sqcup\!\sqcup_T L_2 = R$, we construct T_0 and test $L_1 \sqcup\!\sqcup_{T_0} L_2 = R$. ■

Note that the proof of Theorem 4.1 is similar in theme to the proofs of, e.g., Kari [7, Thm. 4.2, Thm. 4.6]: they each construct a maximal solution to an equation, and that solution is regular. The maximal solution is then tested as a possible solution to the equation to determine if any solutions exist. However, unlike the results of Kari, Theorem 4.1 does not use the concept of an inverse operation.

We can also repeat Theorem 4.1 for the case of deletion along trajectories. The results are identical, with the proof following by the substitution of $T_0 = \{t \in \{i,d\}^* \ : \ \forall x \in L_1, y \in L_2, x \rightsquigarrow_t y \subseteq R\}$. The proof that T_0 is regular differs slightly from that above; we leave the construction to the reader. Thus, we have the following result:

Theorem 4.2 *Let $L_1, L_2, R \subseteq \Sigma^*$ be regular languages. Then it is decidable whether there exists a set $T \subseteq \{i,d\}^*$ of trajectories such that $L_1 \rightsquigarrow_T L_2 = R$.*

5 Undecidability Results

We now demonstrate some undecidability results relating to equations involving shuffle on trajectories.

5.1 Undecidability of One-Variable Equations

Recall that a set $T \subseteq \{0,1\}^*$ is said to be *complete* if $\alpha \sqcup\!\sqcup_T \beta \neq \emptyset$ for all $\alpha, \beta \in \Sigma^*$. Say that a set $T \subseteq \{0,1\}^*$ of trajectories if *left-preserving* (resp., *right-preserving*) if $T \supseteq 0^*$ (resp., $T \supseteq 1^*$). Note that if T is complete, then it is both left- and right-preserving.

Let $\Pi_0, \Pi_1 \ : \ \{0,1\}^* \to \{0,1\}^*$ be the projections given by $\Pi_0(0) = 0, \Pi_0(1) = \epsilon$ and $\Pi_1(1) = 1, \Pi_1(0) = \epsilon$. We say that $T \subseteq \{0,1\}^*$ is *left-enabling* (resp., *right-enabling*) if $\Pi_0(T) = 0^*$ (resp., $\Pi_1(T) = 1^*$).

In this section, we examine undecidability of the existence of solutions of equations involving context-free languages. Namely, we show that:

Theorem 5.1 *Fix $T \subseteq \{0,1\}^*$ to be a regular set of left-enabling (resp., right-enabling) trajectories. For a given CFL L and regular language R, it is undecidable whether or not $L \sqcup_T X = R$ (resp., $X \sqcup_T L = R$) has a solution X.*

Proof. Let T be left-enabling. Let Σ be an alphabet of size at least two and let $\#, \$ \notin \Sigma$. Let $R = (\Sigma^+ + \#^+) \sqcup_T \$^*$. By the closure properties of $\sqcup_T$, and the fact that T is regular, R is a regular language. Let $L \subseteq \Sigma^+$ be an arbitrary CFL and $L_\# = L + \#^+$. We claim that

$$L_\# \sqcup_T X = R \text{ has a solution } \iff L = \Sigma^+. \tag{5.9}$$

This will establish the result, since it is undecidable whether an arbitrary CFL $L \subseteq \Sigma^+$ satisfies $L = \Sigma^+$.

First, if $L = \Sigma^+$, then note that $X = \$^*$ is a solution for (5.9). Second, assume that X is a solution for (5.9). It is clear that for all X,

$$L \sqcup_T X = R \iff L \sqcup_T use_T^{(r)}(X; L) = R, \tag{5.10}$$

where $use_T^{(r)}(X, L)$ is defined by (3.4). Thus, we will focus on useful solutions to the equations $L \sqcup_T X = R$.

Now, we note that, assuming that $use_T^{(r)}(X, L_\#)$ is a solution to (5.9), words in $use_T^{(r)}(X, L_\#)$ cannot contain words with symbols from Σ, because words in R do not contain words with both Σ and $\#$.

In particular, let $x \in use_T^{(r)}(X, L_\#)$. Then there exists a $y \in L_\#$ (in particular, $y \neq \epsilon$) such that $y \sqcup_T x \neq \emptyset$. Consider the word $\#^{|y|}$. As y and $\#^{|y|}$ have the same length, we must have that $\#^{|y|} \sqcup_T x \neq \emptyset$.

Consider any $z \in \#^{|y|} \sqcup_T x$. As $|y| \neq 0$, $|z|_\# > 0$. As $L_\# \sqcup_T X = R$, we must have that $z \in (\Sigma^+ + \#^+) \sqcup_T \$^*$. Thus, $z \in (\# + \$)^+$, and consequently, $x \in (\# + \$)^*$. Thus, $use_T^{(r)}(X, L_\#) \subseteq (\# + \$)^*$.

Let $\Pi_\Sigma : (\Sigma + \{\#, \$\})^* \to \Sigma^*$ be the projection onto Σ. Now as T is left-enabling, note that $\Pi_\Sigma(R) = \Sigma^+$, by definition of $R = (\Sigma^+ + \#^+) \sqcup_T \$^*$. Thus,

$$\begin{aligned}
\Sigma^+ &= \Pi_\Sigma(R) = \Pi_\Sigma(L_\# \sqcup_T X) \\
&= \Pi_\Sigma(L_\# \sqcup_T use_T^{(r)}(X, L_\#)) \subseteq \Pi_\Sigma(L_\# \sqcup_T (\# + \$)^*) \\
&= \Pi_\Sigma((L + \#^+) \sqcup_T (\# + \$)^*) = \Pi_\Sigma((L \sqcup_T (\# + \$)^*) + (\#^+ \sqcup_T (\# + \$)^*)) \\
&= \Pi_\Sigma(L \sqcup_T (\# + \$)^*) \\
&= L \subseteq \Sigma^+.
\end{aligned}$$

The last equality is valid since T is left-enabling. Thus, for all $x \in L$, there is some $j \geq 0$ such that $x \sqcup_T \#^j \neq \emptyset$. We conclude that $L = \Sigma^+$, and thus, by (5.9), the result follows.

The proof in the case that T is right-enabling is similar. ∎

We can give an incomparable result which removes the condition that T must be regular, but must strengthen the conditions on words in T. Namely, T must be left-preserving rather than left-enabling:

Theorem 5.2 *Fix $T \subseteq \{0,1\}^*$ to be a set of left-preserving (resp., right-preserving) trajectories. Given a CFL L and a regular language R, it is undecidable whether there exists a language X such that $L \sqcup\!\sqcup_T X = R$ (resp., $X \sqcup\!\sqcup_T L = R$).*

Proof. Let T be left-preserving (the proof when T is right-preserving is similar). Again, if X is a solution, let

$$use_T^{(r)}(X; L) = \{x \in X \ : \ L \sqcup\!\sqcup_T x \neq \emptyset\}.$$

It is clear that for all X,

$$L \sqcup\!\sqcup_T X = R \iff L \sqcup\!\sqcup_T use_T^{(r)}(X; L) = R.$$

Thus, we will focus on useful solutions to our equation.

Let Σ be our alphabet and $\# \notin \Sigma$. Let $L_\# = L + \#^+$. Note that $\epsilon \notin L_\#$. We claim that $(L_\#) \sqcup\!\sqcup_T X = \Sigma^+ + \#^+$ iff $L = \Sigma^+$ and $use_T^{(r)}(X; L_\#) = \{\epsilon\}$.

First, assume that $L = \Sigma^+$ and $use_T^{(r)}(X, L_\#) = \{\epsilon\}$. Then $L_\# = \Sigma^+ + \#^+$ and

$$\begin{aligned} L_\# \sqcup\!\sqcup_T X &= L_\# \sqcup\!\sqcup_T use_T^{(r)}(X; L_\#) \\ &= (\Sigma^+ + \#^+) \sqcup\!\sqcup_T \{\epsilon\} \\ &= (\Sigma^+ + \#^+), \end{aligned}$$

since $T \supseteq 0^*$.

Now, assume that $L_\# \sqcup\!\sqcup_T X = \Sigma^+ + \#^+$.

Let $x \in use_T^{(r)}(X; L_\#)$. Then there exists $y \in L_\#$ $(y \neq \epsilon)$ such that $y \sqcup\!\sqcup_T x \neq \emptyset$. Consider $\#^{|y|}$. As $|y| = |\#^{|y|}|$, we must have that $\#^{|y|} \sqcup\!\sqcup_T x \neq \emptyset$.

For all $z \in \#^{|y|} \sqcup\!\sqcup_T x$, as $|y| \neq 0$, $|z|_\# > 0$. Further, $z \in \Sigma^+ + \#^+$. Thus, we must have that $x \in \#^+$ or $x = \epsilon$; i.e., $x \in \#^*$. Thus, $use_T^{(r)}(X; L_\#) \subseteq \#^*$.

We now show that $\epsilon \in use_T^{(r)}(X; L_\#)$. As $L_\# \sqcup\!\sqcup_T use_T^{(r)}(X; L_\#) = \Sigma^+ + \#^+$, for all $y \in \Sigma^+$, there exists $\alpha \in L_\#$ and $\beta \in use_T^{(r)}(X; L_\#)$ such that $y \in \alpha \sqcup\!\sqcup_T \beta$. If $\beta \neq \epsilon$, then $|y|_\# > 0$. Thus $\alpha = y$, and $\beta = \epsilon \in use_T^{(r)}(X; L_\#)$. This also demonstrates that $\Sigma^+ \subseteq L_\#$, which implies that $L = \Sigma^+$.

It remains to show that $use_T^{(r)}(X; L_\#) = \{\epsilon\}$. Let $\#^i \in use_T^{(r)}(X; L_\#)$ for some $i > 0$. Then, there is some $y \in L_\# = \Sigma^+ + \#^+$ such that $y \sqcup\!\sqcup_T \#^i \neq \emptyset$.

If $y \in \Sigma^+$, then for all $z \in y \sqcup\!\sqcup_T \#^i$, $|z|_\Sigma, |z|_\# > 0$, which contradicts that $z \in \Sigma^+ + \#^+$, since $L_\# \sqcup\!\sqcup_T X = \Sigma^+ + \#^+$.

Thus, $y \in \#^+$. But then there exists some $y' \in \Sigma^+$ such that $|y| = |y'|$, and $y' \in L_\#$ as well. We are reduced to the first case with y' and $\#^i$, and our assumption that $\#^i \in use_T^{(r)}(X; L_\#)$ is therefore false.

We have established that a (useful) solution to the equation

$$(L + \#^+) \sqcup\!\sqcup_T X = (\Sigma^+ + \#^+)$$

exists iff $L = \Sigma^+$. Therefore, the existence of such solutions must be undecidable. ■

Note that as $use_T^{(r)}(X; L_\#) = \{\epsilon\}$, Theorem 5.2 remains undecidable even if the required (useful) language is required to be a singleton.

We also note that if R and L are interchanged in the equations of the statements of Theorem 5.2 or Theorem 5.1, the corresponding problems are still undecidable. The proofs are trivial, and are left to the reader.

5.2 Undecidability of Shuffle Decompositions

It has been shown [2] that it is undecidable whether a context-free language has a nontrivial shuffle decomposition with respect to the trajectory $\{0,1\}^*$. Here we extend this result for arbitrary complete regular sets trajectories.

If T is a complete set of trajectories, then any language L has decompositions $L \sqcup\!\sqcup_T \{\varepsilon\}$ and $\{\varepsilon\} \sqcup\!\sqcup_T L$. Below we exclude these trivial decompositions; all other decompositions of L are said to be nontrivial.

Theorem 5.3 *Let T be any fixed complete regular set of trajectories. For a given context-free language L it is undecidable whether or not there exist languages $X_1, X_2 \neq \{\varepsilon\}$ such that $L = X_1 \sqcup\!\sqcup_T X_2$.*

Proof. Let $I = (u_1, \ldots, u_k; v_1, \ldots, v_k)$, $k \geq 1$, $u_i, v_i \in \Sigma^*$, $i = 1, \ldots, k$, be an arbitrary instance of the Post Correspondence Problem (PCP). We construct a context-free language $L(I)$ such that $L(I)$ has a nontrivial decomposition along the set of trajectories T if and only if the instance I does not have a solution.

Choose

$$\Omega = \Sigma \cup \{a, b, \#, \flat_1, \flat_2, \natural_1, \natural_2, \$_1, \$_2\},$$

where $\{a, b, \#, \flat_1, \flat_2, \natural_1, \natural_2, \$_1, \$_2\} \cap \Sigma = \emptyset$. Let

$$L_0 = [\flat_1^+(\Sigma \cup \{a, b, \#\})^* \natural_1^+ \cup \flat_2^+(\Sigma \cup \{a, b, \#\})^* \natural_2^+] \sqcup\!\sqcup_T (\$_1^+ \cup \$_2^+). \tag{5.11}$$

Define

$$\begin{aligned} L_1' \;\; = \;\; & \{ab^{i_1} \cdots ab^{i_m} \# u_{i_m} \cdots u_{i_1} \# \mathrm{rev}(v_{j_1}) \cdots \mathrm{rev}(v_{j_n}) \# b^{j_n} a \cdots b^{j_1} a \;\mid \\ & i_1, \ldots, i_m, j_1, \ldots, j_n \in \{1, \ldots, k\},\; m, n \geq 1\} \end{aligned}$$

and let

$$L_1 = L_0 - [\flat_1^+ L_1' \natural_1^+ \sqcup\!\sqcup_T \$_2^+].$$

Using the fact that T is regular, it is easy to see that a nondeterministic pushdown automaton $\mathbf{M}$ can verify that a given word is not in $\flat_1^+ L_1' \natural_1^+ \sqcup\!\sqcup_T \$_2^+$. On input w, using the finite state control $\mathbf{M}$ keeps track of the unique trajectory t (if it exists) such that $w \leadsto_{\tau(t)} \$_2^* \in \flat_1^+(\Sigma \cup \{a, b, \#\})^* \natural_1^+$ and $w \leadsto_{\pi(t)} \flat_1^+(\Sigma \cup \{a, b, \#\})^* \natural_1^+ \in \$_2^*$. If $t \notin T$, $\mathbf{M}$ accepts. Also if t does not exist, $\mathbf{M}$ accepts. Using the stack $\mathbf{M}$ can verify that $w \leadsto_{\tau(t)} \$_2^* \notin \flat_1^+ L_1' \natural_1^+$ by guessing where the word violates the definition of L_1'. Note that this verification can be interleaved with the computation checking whether t is in T. Since L_0 is regular, it follows that L_1 is context-free.

Define

$$L_2' = \{ab^{i_1} \cdots ab^{i_m}\#w\#\mathrm{rev}(w)\#b^{i_m}a \cdots b^{i_1}a \mid w \in \Sigma^*, i_1, \ldots, i_m \in \{1, \ldots, k\},\ m \geq 1\}$$

and let

$$L_2 = L_0 - [\flat_1^+ L_2' \natural_1^+ ⧢_T \$_2^+].$$

As above it is seen that L_2 is context-free. It follows that also the language

$$L(I) = L_1 \cup L_2 = L_0 - [\flat_1^+(L_1' \cap L_2')\natural_1^+ ⧢_T \$_2^+] \tag{5.12}$$

is context-free.

First consider the case where the PCP instance I does not have a solution. Now $L_1' \cap L_2' = \emptyset$ and (5.11) gives a nontrivial decomposition for $L(I) = L_0$ along the set of trajectories T.

Secondly, consider the case where the PCP instance I has a solution. This means that there exists a word

$$w_0 \in L_1' \cap L_2'. \tag{5.13}$$

For the sake of contradiction we assume that we can write

$$L(I) = X_1 ⧢_T X_2, \tag{5.14}$$

where $X_1, X_2 \neq \{\varepsilon\}$.

We establish a number of properties that the languages X_1 and X_2 must necessarily satisfy. We claim that it is not possible that

$$\mathrm{alph}(X_1) \cap \{\flat_i, \natural_i\} \neq \emptyset \text{ and } \mathrm{alph}(X_2) \cap \{\flat_j, \natural_j\} \neq \emptyset \tag{5.15}$$

where $\{i, j\} = \{1, 2\}$. If the above relations would hold, then the completeness of T would imply that $X_1 ⧢_T X_2$ has some word containing a symbol of $\{\flat_1, \natural_1\}$ and a symbol of $\{\flat_2, \natural_2\}$. This is impossible since $X_1 ⧢_T X_2 \subseteq L_0$.

Denote

$$\Phi = \{\flat_1, \flat_2, \natural_1, \natural_2\}.$$

Since $L(I)$ has both words that contain symbols $\flat_1$, $\natural_1$ and words that contain symbols $\flat_2$, $\natural_2$, by (5.15) the only possibility is that all the symbols of Φ "come from" one of the components X_1 and X_2. We assume in the following that

$$\mathrm{alph}(X_2) \cap \Phi = \emptyset. \tag{5.16}$$

This can be done without loss of generality since the other case is completely symmetric. (We can just interchange the symbols 0 and 1 in T.)

Next we show that

$$\mathrm{alph}(X_2) \cap (\Sigma \cup \{a, b, \#\}) = \emptyset. \tag{5.17}$$

Let $\Pi_\Phi : \Omega^* \to \Phi^*$ be the projection onto Φ. Since $\Pi_\Phi(L(I)) = \flat_1^+\natural_1^+ \cup \flat_2^+\natural_2^+$ and X_2 does not contain any symbols of Φ, it follows that $\Pi_\Phi(X_1) = \flat_1^+\natural_1^+ \cup \flat_2^+\natural_2^+$. Thus if (5.17) would not hold, the completeness of T would imply that $X_1 ⧢_T X_2$ contains words where a symbol of $\Sigma \cup \{a, b, \#\}$ occurs before a symbol of $\{\flat_1, \flat_2\}$ or after a symbol of $\{\natural_1, \natural_2\}$. Hence (5.17) holds.

Since $X_2 \neq \{\varepsilon\}$, the equations (5.16) and (5.17) imply that

$$\mathrm{alph}(X_2) \cap \{\$_1, \$_2\} \neq \emptyset.$$

Since $L(I)$ has words with symbols $\$_1$, other words with symbols $\$_2$, and no words containing both symbols $\$_1$, $\$_2$, using again the completeness of T it follows that

$$\mathrm{alph}(X_1) \cap \{\$_1, \$_2\} = \emptyset. \tag{5.18}$$

Now consider the word $w_0 \in L_1' \cap L_2'$ given by (5.13). We have $\flat_i w_0 \natural_i ⧢_T \$_i \subseteq L_i$, $i = 1, 2$, and let $u_i \in \flat_i w_0 \natural_i ⧢_T \$_i$, $i = 1, 2$, be arbitrary. We can write

$$u_i = x_{i,1} ⧢_{t_i} x_{i,2}, \text{ such that } x_{i,j} \in X_j, t_i \in T, i = 1, 2, j = 1, 2.$$

By (5.16), (5.17) and (5.18) we have

$$X_1 \subseteq (\Phi \cup \Sigma \cup \{a, b, \#\})^* \text{ and } X_2 \subseteq \{\$_1, \$_2\}^*$$

and hence

$$x_{i,1} = \flat_i w_0 \natural_i, \quad x_{i,2} = \$_i, \quad i = 1, 2.$$

Now $x_{1,1} ⧢_{t_1} x_{2,2} \subseteq X_1 ⧢_T X_2$ is equal to the word $\flat_1 w_0 \natural_1 ⧢_{t_1} \$_2$, and it is not in $L(I)$ by the choice of w_0 and (5.12). This contradicts (5.14). ■

In the proof of Theorem5.3, whenever the CFL has a nontrivial decomposition along the set of trajectories T, it has a decomposition where the component languages are, in fact, regular. This gives the following:

Corollary 5.4 *Let T be any fixed complete regular set of trajectories. For a given context-free language L it is undecidable whether or not*

(a) there exist regular languages $X_1, X_2 \neq \{\varepsilon\}$ such that $L = X_1 ⧢_T X_2$.

(b) there exist context-free languages $X_1, X_2 \neq \{\varepsilon\}$ such that $L = X_1 ⧢_T X_2$.

References

[1] Berard, B. Literal Shuffle. *Theor. Comp. Sci. 51*, (1987), 281–299.

[2] Câmpeanu, C., Salomaa, K., and Vágvölgyi, S. Shuffle decompositions of regular languages. *Int. J. Found. Comp. Sci. 13*, 6 (2002), 799–816.

[3] Domaratzki, M. Deletion along trajectories. Tech. Rep. 2003–464, School of Computing, Queen's University at Kingston, 2003. Available electronically at http://www.cs.queensu.ca/TechReports/.

[4] Ginsburg, S., and Spanier, E. H. Bounded regular sets. *Proc. Amer. Math. Soc. 17* (1966), 1043–1049.

[5] Ito, M. Shuffle decomposition of regular languages. *J. Univ. Comp. Sci. 8*, 2 (2002), 257–259.

[6] Kari, L. Deletion operations: Closure properties. *Intl. J. Comp. Math. 52* (1994), 23–42.

[7] Kari, L. On language equations with invertible operations. *Theor. Comp. Sci. 132* (1994), 129–150.

[8] Kari, L., and Thierrin, G. k-catenation and applications: k-prefix codes. *J. Inf. Opt. Sci. 16*, 2 (1995), 263–276.

[9] Kari, L., and Thierrin, G. Maximal and minimal solutions to language equations. *J. Comp. Sys. Sci. 53* (1996), 487–496.

[10] Leiss, E. *Language Equations.* Springer Monographs in Computer Science, 1999.

[11] Mateescu, A., Rozenberg, G., and Salomaa, A. Shuffle on trajectories: Syntactic constraints. *Theor. Comp. Sci. 197* (1998), 1–56.

[12] Meduna, A. Middle quotients of linear languages. *Intl. J. Comp. Math. 71* (1999), 319–335.

[13] Salomaa, A., and Yu, S. On the decomposition of finite languages. In *Developments in Language Theory* (1999), G. Rozenberg and W. Thomas, Eds., pp. 22–31.

[14] Yu, S. Regular languages. In *Handbook of Formal Languages, Vol. I*, G. Rozenberg and A. Salomaa, Eds. Springer-Verlag, 1997, pp. 41–110.