Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Computer Science 332 (2005) 179–198
Theoretical Computer Science
www.elsevier.com/locate/tcs

Mixed languages

Jean Berstel[a], Luc Boasson[b,*], Michel Latteux[c]

[a] *Institut Gaspard Monge (IGM), Université de Marne-la-Vallée, 5, boulevard Descartes, 77454 Marne-la-Vallée Cedex 2, France*

[b] *Laboratoire d'informatique algorithmique: fondements et applications (LIAFA), Université Denis-Diderot (Paris VII), 2, place Jussieu, 75251 Paris Cedex 05, France*

[c] *Laboratoire d'informatique fondamentale de Lille (LIFL), Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France*

Received 2 May 2003; received in revised form 2 May 2004; accepted 19 August 2004
Communicated by G. Rozenberg

Abstract

Let $T = A \cup B \cup C$ be an alphabet that is partitioned into three subalphabets. The *mixing product* of a word g over $A \cup B$ and of a word d over $A \cup C$ is the set of words w over T such that its projection onto $A \cup B$ gives g and its projection onto $A \cup C$ gives d.

Let R be a regular language over T such that $xbcy$ is in R if and only if $xcby$ is in R for any two letters b in B and c in C. In other words, R is commutative over B and C. Is this property "structural" in the sense that R can then be obtained as a mixing product of a regular language over $A \cup B$ and of a regular language over $A \cup C$?

This question has a rather easy answer, but there are many cases where the answer is negative. A more interesting question is whether R can be represented as a finite union of mixed products of regular languages. For the moment, we do not have an answer to this question. However, we prove that it is decidable whether, for a given k, the language R is a union of at most k mixed products of regular languages.

Résumé

Soit $T = A \cup B \cup C$ un alphabet partitionné en trois sous-alphabets. Le *mélange* d'un mot g sur $A \cup B$ et d'un mot d sur $A \cup C$ est l'ensemble des mots w sur T dont la projection sur $A \cup B$ donne le mot g et sur $A \cup C$ donne le mot d.

*Corresponding author.
E-mail address: boasson@liafa.jussieu.fr (L. Boasson).

doi:10.1016/j.tcs.2004.08.016

Soit R un langage rationnel sur T tel que *xbcy* est dans R si et seulement si *xcby* est dans R pour deux lettres quelconques $b \in B$ et $c \in C$. En d'autres termes, R est commutatif sur B et C. Est-ce que cette propriété est "structurelle", c'est-à-dire peut-on alors obtenir R comme mélange d'un langage rationnel sur $A \cup B$ et d'un langage rationnel sur $A \cup C$?

Cette question a une réponse plutôt facile, mais il existe de trop nombreux cas où la réponse est négative. Une question plus intéressante est de savoir si on peut représenter R comme une union finie de mélanges de langages rationnels. Pour l'instant, nous n'avons pas de réponse à cette question. En revanche, nous montrons qu'il est décidable, pour un entier k donné, si R est union d'au plus k mélanges de langages rationnels.

Keywords: Formal languages; Theory of traces; Synchronized product

1. Introduction

This paper is concerned with a special case of a problem about trace languages, that we address in a particular setting, and for which we give a partial answer. We first state the problem and then sketch its relation to trace languages.

Let T be an alphabet that is partitioned into three pairwise disjoint alphabets A, B, and C, so that

$$T = A \cup B \cup C \quad A, B, C \text{ pairwise disjoint.} \tag{1}$$

Consider a regular language R that is (B, C)-*commutative*, i.e., satisfies

$$zbcy \in R \iff zcby \in R$$

for all letters $b \in B$, $c \in C$ and words z, y. One may ask whether R can be built up by "mixing" regular languages G over $A \cup B$ and D over $A \cup C$. To be more precise, denote by $\pi_B : T^* \to (A \cup B)^*$, and $\pi_C : T^* \to (A \cup C)^*$ the projections from T^* onto $(A \cup B)^*$ and $(A \cup C)^*$ respectively, and define the *mixing product* of two words $u \in (A \cup B)^*$ and $v \in (A \cup C)^*$ by

$$u \uparrow v = \pi_B^{-1}(u) \cap \pi_C^{-1}(v)\,.$$

These products extend to set as usual by

$$G \uparrow D = \bigcup_{g \in G, d \in D} g \uparrow d = \pi_B^{-1}(G) \cap \pi_C^{-1}(D)$$

for $G \subset (A \cup B)^*$ and $D \subset (A \cup C)^*$. The question can be stated more formally as follows: if R is (B, C)-commutative, does there exist G and D such that $R = G \uparrow D$. This is easily answered, as well shall see. A more interesting question is: is it possible to write R as a finite union of sets $G_i \uparrow D_i$. We do not know whether this problem is decidable. However, we prove that, given an integer $k \geqslant 1$, it is decidable where R can be written as a union of at most k sets $G_i \uparrow D_i$.

The general framework is that of free partially commutative monoids and of trace languages (see e.g., [1]). Such a free partially commutative monoid $M(T, I)$ over the alphabet

T is defined by an independence relation $I \subset T \times T$. In our case, letters in B commute with letters in C, so $I = B \times C \cup C \times B$. Languages we call (B, C)-commutative are precisely *trace languages*, that is subsets of T^* that are inverse homomorphic images of subsets in $M(T, I)$ by the canonical homomorphism. A trace language that is regular is the inverse homomorphic image of a recognizable subset of $M(T, I)$. A famous theorem of Zielonka [3] shows that recognizable trace languages are precisely those recognized by asynchronous automata.

Duboc [2] considers mixing products of languages (that we defined above in our special setting) and she called weakly mixing those languages that are finite unions of mixing languages (we will call them *mixing* for short later). She observed that regular trace languages are not always weakly mixing, but she proved that every regular trace language is the homomorphic image of some weakly mixing language.

The problem we address can be stated in general as follows: given a regular trace language, is it decidable if it is mixing (weakly mixing)? We consider only the very simple case of the special independence relation given before, and give only a partial answer.

For more motivation, let us consider an automata-theoretic approach. Consider two automata $\mathcal{B}$ and $\mathcal{C}$ over $A \cup B$ and $A \cup C$ respectively. Transform automaton $\mathcal{B}$ by adding loops labelled by all letters in C to each state, and similarly for $\mathcal{C}$. This gives automata $\bar{\mathcal{B}}$ and $\bar{\mathcal{C}}$ over T. The direct product $\bar{\mathcal{B}} \times \bar{\mathcal{C}}$ is called the mixing product by Duboc [2]. In $\bar{\mathcal{B}} \times \bar{\mathcal{C}}$, choose a set F of final states, and then minimize the automaton. Call the resulting minimal automaton $\mathcal{A}$. The language recognized by $\mathcal{A}$ is

$$L(\mathcal{A}) = L(\bar{\mathcal{B}} \times \bar{\mathcal{C}}) = \bigcup_{(g,d)\in F} L(g, d),$$

where $L(g, d)$ denotes the language recognized by taking (g, d) as the unique final state. Then

$$L(g, d) = L_{\mathcal{B}}(g) \uparrow L_{\mathcal{C}}(d),$$

where $L_{\mathcal{B}}(g)$ is the language recognized in $\mathcal{B}$ with the unique final state g and similarly for $L_{\mathcal{C}}(d)$. This shows that $L(\mathcal{A})$ is a union of $\mathrm{Card}(F)$ mixed languages. However, it may happen, as in the example we give now, that the number α of final states in the minimal automaton $\mathcal{A}$ is strictly less than the size of F, so that the mixing decomposition cannot be "read" from the form of $\mathcal{A}$. In fact, we do not know of an upper bound for $\mathrm{Card}(F)$ expressed as a function of α.

Example 1.1. Let $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, let W be the set of words of even length over $\{b, c\}$ and set $R = aW$. This language is recognized by the automaton $\mathcal{A}$ of Fig. 1. On the other hand, consider the automata $\mathcal{B}$ and $\mathcal{C}$ of Fig. 2. No final states are specified. Adding loops on states gives the automata of Fig. 3. In the (accessible part) of the direct product of these automata we choose final states $\underline{1}\bar{1}$ and $\underline{2}\bar{2}$ (see Fig. 4).

The language recognized is therefore

$$R = \left(a(b^2)^* \uparrow a(c^2)^*\right) \cup \left(ab(b^2)^* \uparrow ab(c^2)^*\right) .$$

Minimizing the product automaton yields the automaton of Fig. 1 with a unique final state, and R is easily shown not to be representable as a unique mixing of two languages.

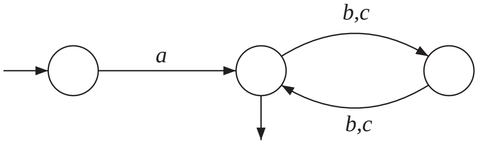

Fig. 1. Minimal automaton recognizing R.

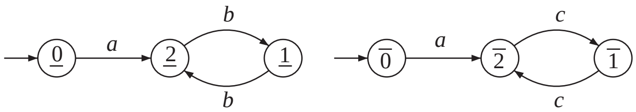

Fig. 2. Automata $\mathcal{B}$ (on the left) and $\mathcal{C}$ (on the right).

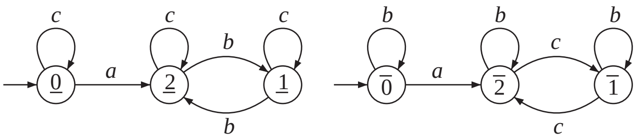

Fig. 3. Automata $\bar{\mathcal{B}}$ and $\bar{\mathcal{C}}$.

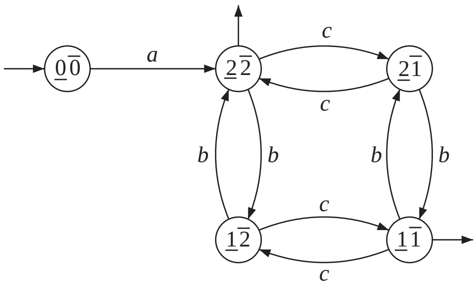

Fig. 4. The direct product of $\bar{\mathcal{B}}$ and $\bar{\mathcal{C}}$.

The automata-theoretic description seems not to lead directly to an answer to our question. We therefore consider in the sequel a language-theoretic approach.

The paper is organized as follows: the next section contains some notation. In Section 3, we prove the first result we announced (Proposition 3.4), namely that it is decidable whether a language is strongly mixing. Sections 4 and 5 contain some preliminary results and examples on k-mixing languages. The basic construction for answering the question whether a language is k-mixing is presented in Section 6. It relates mixing to a kind of syntactic notion called the *index*: the index is, roughly speaking, the maximum number of classes of traces that compose the inverse image of a skeleton. The construction, proved in Proposition 6.5 is in fact a semi-algorithm in the sense that it yields only a bound between k and 4^k. This proposition heavily relies on a surprising result (Lemma 6.6) showing that a certain language is regular. In Section 7, a second algorithm is presented that shows how a decomposition into mixing languages can be splitted and recomposed into smaller ones, yielding the answer to our question (Theorem 7.8).

2. Mixing product

Recall that T denotes an alphabet that is partitioned into three pairwise disjoint alphabets A, B, and C, so that

$$T = A \cup B \cup C \quad A, B, C \text{ pairwise disjoint.} \tag{2}$$

We denote by $\pi : T^* \to A^*$, $\pi_B : T^* \to (A \cup B)^*$, and $\pi_C : T^* \to (A \cup C)^*$ the three projections from T^* onto A^*, $(A \cup B)^*$ and $(A \cup C)^*$, respectively. Clearly, $\pi = \pi_B \circ \pi_C = \pi_C \circ \pi_B$. Observe also that if $u \in (A \cup B)^*$, then $\pi_B^{-1} = u \sqcup\!\sqcup C^*$, where $\sqcup\!\sqcup$ denotes the shuffle operation. The projection $\pi(w)$ of a word $w \in T^*$ is called the *skeleton* of w.

The *mixing product* of two words $u \in (A \cup B)^*$ and $v \in (A \cup C)^*$ is defined by

$$u \uparrow v = \pi_B^{-1}(u) \cap \pi_C^{-1}(v)$$

There are other notations for this product: Zielonka [3] writes $u \| v$ and Duboc [2] uses still another notation.

Example 2.1. Let $A = \{a\}$, $B = \{b\}$, $C = \{c\}$. Then $ab \uparrow ac = \{abc, acb\}$, and $aba \uparrow ac = \emptyset$.

If the alphabet A is empty, then the mixing product is merely the shuffle. Observe that $u \uparrow v \neq \emptyset$ if and only if $\pi_C(u) = \pi_B(v)$ or equivalently if and only if u and v have the same skeleton. Observe also that if $(u \uparrow v) \cap (u' \uparrow v') \neq \emptyset$, then $u = u'$ and $v = v'$. Indeed, if $w \in (u \uparrow v) \cap (u' \uparrow v')$, then $\pi_B(w) = u$ because $w \in u \uparrow v$ and similarly $\pi_B(w) = u'$. Symmetrically, $v = v'$.

3. Strongly mixing languages

As usual, the mixing product is extended to sets of words as follows. Let $G \subset (A \cup B)^*$ and $D \subset (A \cup C)^*$. Then

$$G \uparrow D = \bigcup_{g \in G, d \in D} g \uparrow d = \pi_B^{-1}(G) \cap \pi_C^{-1}(D).$$

Observe that the union is over all pairs $(g, d) \in G \times D$, but that the pair (g, d) has a non empty contribution to the union only if $\pi_C(g) = \pi_B(d)$.

A language L over T is *strongly mixing* if there exist languages G over $A \cup B$ and D over $A \cup C$ such that $L = G \uparrow D$.

Lemma 3.1. *A language $L \subset T^*$ is strongly mixing if and only if $L = \pi_B(L) \uparrow \pi_C(L)$. If $L = G \uparrow D$ then $G \supset \pi_B(L)$ and $D \supset \pi_C(L)$.*

Proof. Assume L is strongly mixing, $L = G \uparrow D$. Since $L \subset \pi_B^{-1}(G)$, one has $\pi_B(L) \subset G$ and similarly $\pi_C(L) \subset D$. This shows the second part, and also the inclusion $\pi_B(L) \uparrow \pi_C(L) \subset L$. On the other hand, the inclusion $L \subset \pi_B^{-1}(\pi_B(L)) \cap \pi_C^{-1}(\pi_C(L))$ always holds, so $L \subset \pi_B(L) \uparrow \pi_C(L)$. □

Observe that equality $G = \pi_B(L)$ holds in the lemma if, for every $g \in G$, there is $d \in D$ such that $g \uparrow d \neq \emptyset$. Indeed, in this case, let $w \in L$ be in $g \uparrow d$. Then $\pi_B(w) = g$, so $g \in \pi_B(L)$.

Example 3.2. Let $A = \{a\}$, $B = \{b\}$ and $C = \{c\}$. The language $K = \{aw \mid w \in \{b,c\}^*, |w|_b = 1, |w|_c > 0\}$ is strongly mixing. Indeed, one has $K = ab \uparrow ac^+$. The language can also be written for instance as $K = (ab \cup aa^+) \uparrow ac^+$, since $aa^+ \uparrow ac^+ = \emptyset$.

Example 3.3. The language $R = \{aw \mid w \in \{b,c\}^*, |w| \text{ even}\}$ is not strongly mixing, since $\pi_B(R) = ab^*$ and $\pi_C(R) = ac^*$, and $R \neq ab^* \uparrow ac^*$.

Proposition 3.4. *Given a regular language R over T, it is decidable whether R is strongly mixing. Moreover, if R is strongly mixing, then it is the mixing product of two regular languages.*

Proof. In order to check whether R is strongly mixing, it suffices to compute the regular languages $G = \pi_B(R)$ and $D = \pi_C(R)$ and to check whether $R = \pi_B^{-1}(G) \cap \pi_C^{-1}(D)$. All these computations are effective because the languages involved are regular. Clearly, the language R is strongly mixing if and only if the equality holds. □

We have the following closure property.

Proposition 3.5. *The intersection of two strongly mixing languages is again strongly mixing.*

Proof. Let $L = G \uparrow D = \pi_B^{-1}(G) \cap \pi_C^{-1}(D)$ and $L' = G' \uparrow D' = \pi_B^{-1}(G') \cap \pi_C^{-1}(D')$. Then $L \cap L' = \pi_B^{-1}(G \cap G') \cap \pi_C^{-1}(D \cap D') = (G \cap G') \uparrow (D \cap D')$. □

4. Mixing languages

A language L is *k-mixing* for some integer k if there exist k strongly mixing languages $L_1, \ldots, L_k$ such that $L = L_1 \cup \cdots \cup L_k$. The language L is *mixing* if it is k-mixing for some k. Clearly, 1-mixing languages are precisely the strongly mixing languages. Since the empty set is strongly mixing, any k-mixing language is also k'-mixing for $k' > k$.

Example 4.1. The language $R = \{aw \mid w \in \{b,c\}^*, |w| \text{ even}\}$ of Example 3.3 is 2-mixing, since

$$R = \left(a(b^2)^* \uparrow a(c^2)^*\right) \cup \left(ab(b^2)^* \uparrow ab(c^2)^*\right).$$

Example 4.2. Let again $A = \{a\}$, $B = \{b\}$, $C = \{c\}$. Let W be the set of words of even length over $\{b,c\}$. We show that the language $R = (aW)^+$ is not mixing. For this, we consider, for each $n \geqslant 1$, the words $(abc)^i(ab^2c^2)^{n-i}$ for $0 \leqslant i \leqslant n$. All these words are in R.

Assume R is k-mixing. Then, if $n \geqslant k$, there are two distinct words $w = (abc)^i(ab^2c^2)^{n-i}$ and $w' = (abc)^j(ab^2c^2)^{n-j}$ (with $i < j$) which are in the same strongly mixing subset of R, say $w, w' \in G \uparrow D \subset R$. It follows that $(ab)^i(ab^2)^{n-i}, (ab)^j(ab^2)^{n-j} \in G$ and $(ac)^i(ac^2)^{n-i}, (ac)^j(ac^2)^{n-j} \in D$. But then $(ab)^i(ab^2)^{n-i} \uparrow (ac)^j(ac^2)^{n-j} \subset R$, and this shows in particular that the word $(abc)^i(ab^2c)^{j-i}(ab^2c^2)^{n-j}$ is in R, a contradiction.

In the sequel, we shall prove that, given a regular set R over T and an integer k, it is decidable whether R is k-mixing. It remains open whether it is decidable that a regular language R is mixing. In other words, we are able to answer the question for a fixed k, but we do not know the answer if k is not fixed.

As a simple consequence of Proposition 3.5, we have

Proposition 4.3. *If L is k-mixing and L' is k'-mixing, then $L \cap L'$ is $k \cdot k'$-mixing.*

Corollary 4.4. *If L is a k-mixing language, then $T^* \setminus L$ is 2^k-mixing.*

Proof. If $L = G \uparrow D$ is strongly mixing, then $T^* \setminus L = (\bar{G} \uparrow (A \cup C)^*) \cup ((A \cup B)^* \uparrow \bar{D})$, where $\bar{G} = (A \cup B)^* \setminus G$ and $\bar{D} = (A \cup C)^* \setminus D$, showing that $T^* \setminus L$ is 2-mixing. Next, if $L = L_1 \cup \cdots \cup L_k$ with $L_1, \ldots, L_k$ strongly mixing, then $T^* \setminus L = (T^* \setminus L_1) \cap \cdots \cap (T^* \setminus L_k)$, and the result follows from the preceding proposition. $\square$

We introduce now a running example that will be used repeatedly to illustrate the arguments developed in this paper.

Example 4.5. Let $A = \{a\}$, $B = \{b\}$ and $C = \{c\}$ and consider the three languages

$$\begin{aligned}
R_1 &= ab \uparrow ac^+ = \{aw \mid w \in \{b,c\}^*, |w|_b = 1, |w|_c > 0\} \\
R_2 &= a(b^2)^+ \uparrow a(c^2)^+ \\
&= \{aw \mid w \in \{b,c\}^*, |w|_b \geqslant 2, |w|_c \geqslant 2, |w|_b \equiv |w|_c \equiv 0 \bmod 2\} \\
R_3 &= ab^3(b^2)^* \uparrow ac(c^2)^* \\
&= \{aw \mid w \in \{b,c\}^*, |w|_b \geqslant 3, |w|_c \geqslant 1, |w|_b \equiv |w|_c \equiv 1 \bmod 2\}
\end{aligned}$$

Set $R = R_1 \cup R_2 \cup R_3$. All words in R have the same skeleton a. By construction, the language R is 3-mixing. It is not strongly mixing. Indeed $\pi_B(R) = ab^+$, $\pi_C(R) = ac^+$, and for instance abc^2 is in $ab^+ \uparrow ac^+$ and is not in R. However, R is 2-mixing, because

$$R = ab(b^2)^* \uparrow ac(c^2)^* \cup (ab \cup a(b^2)^+) \uparrow a(c^2)^+. \qquad \square$$

5. Preliminary results

Consider the mapping

$$\mu : T^* \to (A \cup B)^* \times (A \cup C)^*$$

defined by

$$\mu(w) = (\pi_B(w), \pi_C(w)).$$

For $u \in (A \cup B)^*$, $v \in (A \cup C)^*$, one has $\mu^{-1}(u, v) = u \uparrow v$. If $G \subset (A \cup B)^*$ and $D \subset (A \cup C)^*$, then $G \uparrow D = \mu^{-1}(G \times D)$. However, since $g \uparrow d = \emptyset$ if $\pi(g) \neq \pi(d)$, it is natural to consider the "diagonal" composed of pairs of words with the same skeleton

$$\Delta = \{(u, v) \in (A \cup B)^* \times (A \cup C)^* \mid \pi_C(u) = \pi_B(v)\}.$$

This is a rational relation, and

$$G \uparrow D = \mu^{-1}(G \times D \cap \Delta), \qquad \mu(G \uparrow D) = G \times D \cap \Delta.$$

A set $X \subset T^*$ is called (B, C)-*commutative* if

$$zbcy \in X \iff zcby \in X$$

for all letters $b \in B$, $c \in C$ and words z, y. It is easily seen that this property is decidable for regular languages. Indeed, one may consider the minimal automaton of a regular languages X. Then X is (B, C)-commutative if and only if, whenever there are edges $p \xrightarrow{b} q$ and $q \xrightarrow{c} r$, there exist also edges $p \xrightarrow{c} q'$ and $q' \xrightarrow{b} r$ in this automaton.

Observe that, X is (B, C)-commutative if and only if $\mu^{-1}(\mu(X)) = X$. If X is (B, C)-commutative, then $u \uparrow v \cap X \neq \emptyset \implies u \uparrow v \subset X$. Indeed, let $w \in u \uparrow v$. Then $u \uparrow v = \mu^{-1}(\mu(w)) \subset \mu^{-1}(\mu(X)) = X$.

For any regular set R over T, the set $\mu(R)$ is a rational relation. The relation $\mu(R)$ defines two reciprocal rational transductions $\beta_R : (A \cup B)^* \to (A \cup C)^*$ and $\gamma_R : (A \cup C)^* \to (A \cup B)^*$ by

$$\beta_R(u) = \{v \in (A \cup C)^* \mid (u, v) \in \mu(R)\}$$

and

$$\gamma_R(v) = \{u \in (A \cup B)^* \mid (u, v) \in \mu(R)\}.$$

Clearly, $v \in \beta_R(u)$ if and only if $u \in \gamma_R(v)$. Observe that

$$\beta_R(u) = \{v \in (A \cup C)^* \mid (u \uparrow v) \cap R \neq \emptyset\} = \pi_C(\pi_B^{-1}(u) \cap R).$$

Moreover, if R is (B, C)-commutative, then

$$\beta_R(u) = \{v \in (A \cup C)^* \mid u \uparrow v \neq \emptyset \text{ and } u \uparrow v \subset R\}.$$

Symmetric expressions hold for γ_R. If R is regular, then each of these sets is regular and effectively computable. If R is strongly mixing, i.e., if $R = G \uparrow D$, then for $u \in G$, one has

$$\beta_R(u) = \{v \mid (u, v) \in G \times D \cap \Delta\} = D \cap \pi_B^{-1}(\pi_C(u)) \,. \tag{3}$$

Example 5.1. In the previous example, the transductions β_R and γ_R are readily computed. One has

$$\begin{aligned}
\beta_R(ab) &= ac^+, & \gamma_R(ac^{2n}) &= ab \cup a(b^2)^+ \quad n \geqslant 1,\\
\beta_R(ab^{2n}) &= a(c^2)^+ \quad n \geqslant 1, & \gamma_R(ac^{2n+1}) &= ab(b^2)^* \quad n \geqslant 0.\\
\beta_R(ab^{2n+1}) &= ac(c^2)^* \quad n \geqslant 1, & &
\end{aligned}$$

We now consider the nuclear equivalences associated to these transductions. Given $u, u' \in (A \cup B)^*$, we set $u \sim_{\beta,R} u'$ if and only if $\beta_R(u) = \beta_R(u')$, and symmetrically, given $v, v' \in (A \cup C)^*$, we set $v \sim_{\gamma,R} v'$ if and only if $\gamma_R(v) = \gamma_R(v')$. We denote the equivalence class of u by $[u]_{\beta,R}$ and the equivalence class of v by $[v]_{\gamma,R}$.

Finally, we define an equivalence $\sim_R$ over T^*, called the *R-equivalence* by setting, for two words w and w' over T, $w \sim_R w'$ if and only if $\pi_B(w) \sim_{\beta,R} \pi_B(w')$ and $\pi_C(w) \sim_{\gamma,R} \pi_C(w')$. The equivalence class of w for the R-equivalence is denoted by $[w]_R$. It is convenient to set $[K]_\sim = \bigcup_{x \in K} [x]_\sim$, where $[x]_\sim$ denotes the equivalence class of x for some equivalence $\sim$. It is easily checked that

$$[w]_R = [\pi_B(w)]_{\beta,R} \uparrow [\pi_C(w)]_{\gamma,R} \,.$$

More generally, one has the implications

$$G \uparrow D \subset M \;\Rightarrow\; [G]_{\beta,R} \uparrow [D]_{\gamma,R} \subset [M]_R$$

and also

$$M = G \uparrow D \;\Rightarrow\; [M]_R = [G]_{\beta,R} \uparrow [D]_{\gamma,R} \,. \tag{4}$$

Example 5.2. Let us continue our example. The equivalence $\sim_{\beta,R}$ has three classes contained in $\pi_B(R)$, namely $[ab]_{\beta,R} = \{ab\}$, $[ab^2]_{\beta,R} = a(b^2)^+$ and $[ab^3]_{\beta,R} = ab^3(b^2)^*$. Similarly, there are two equivalence classes for $\sim_{\gamma,R}$, namely $[ac]_{\gamma,R} = ac(c^2)^*$ and $[ac^2]_{\gamma,R} = a(c^2)^+$. The language R is saturated for R-equivalence, and it is the union of four equivalence classes:

$$\begin{aligned}
R \;=\; & [ab]_{\beta,R} \uparrow [ac]_{\gamma,R} \cup [ab]_{\beta,R} \uparrow [ac^2]_{\gamma,R}\\
\cup\; & [ab^2]_{\beta,R} \uparrow [ac^2]_{\gamma,R} \cup [ab^3]_{\beta,R} \uparrow [ac]_{\gamma,R}
\end{aligned} \tag{5}$$

giving yet another decomposition of R.

Also, if R is (B, C)-commutative, then R is saturated for R-equivalence, that is $[R]_R = R$, for if $x \in R$ and $x' \sim_R x$, then setting $u = \pi_B(x)$, $u' = \pi_B(x')$ and $v = \pi_C(x)$, $v' = \pi_C(x')$, one has $u \uparrow v \subset R$ because R is (B, C)-commutative, and $u' \uparrow v \in R$ because $u \sim_{\beta,R} u'$ and $u' \uparrow v' \in R$ because $v \sim_{\gamma,R} v'$, and thus $x' \in R$.

6. The basic construction

The R-equivalence introduced at the end of the previous section provides a property of k-mixing languages that will be used as a test for termination in the construction we will describe now.

A set X of words over T is *monoskeletal* if all words in X have the same skeleton. Thus, a subset X of R is monoskeletal if and only if it is a subset of a set $R \cap \pi^{-1}(s)$, for some $s \in \pi(R)$. Each R-equivalence class is monoskeletal because R equivalent words have the same skeleton. Thus, each set $R \cap \pi^{-1}(s)$ is saturated for R-equivalence. The *index* of $R \cap \pi^{-1}(s)$ is the number of R-equivalence classes it contains. More generally, for a subset X of R, the index of the subset $X \cap \pi^{-1}(s)$ is the number of R-equivalence classes that $X \cap \pi^{-1}(s)$ intersects and the index of X is the maximum of the indices of the sets $X \cap \pi^{-1}(s)$, where s ranges over the skeletons of R.

A set $R \subset T^*$ has *index* k if any monoskeletal subset X of R has index at most k, that is intersects at most k distinct R-equivalence classes. In other words, R has index k if any monoskeletal subset of R composed of at least $k+1$ words contains two distinct R-equivalent words.

Proposition 6.1. *If a language R is k-mixing, then it has index* 4^k.

Proof. Set

$$R = R_1 \cup \cdots \cup R_k, \quad R_i = G_i \uparrow D_i \,.$$

Consider a word $w \in R$. Define the B-index set of w by $\mathrm{Ind}_B(w) = \{i \mid \pi_B(w) \in G_i\}$. It follow from (3) that

$$\beta_R(\pi_B(w)) = \bigcup_{i \in \mathrm{Ind}_B(w)} D_i \cap \pi_B^{-1}(\pi(w)) \,.$$

This shows that if $w, w' \in R$ have same skeleton and same B-index set, then $\pi_B(w) \sim_{\beta,R} \pi_B(w')$. Symmetrically, one defines the C-index set of w by $\mathrm{Ind}_C(w) = \{i \mid \pi_C(w) \in D_i\}$, and one shows that if w, w' have same skeleton and same C-index set, then $\pi_C(w) \sim_{\gamma,R} \pi_C(w')$. Thus, if w and w' have the same skeleton and the same pair of index sets, they are R-equivalent. Clearly, there are at most 4^k pairs of index sets. Thus, if one takes any monoskeletal set of at least $4^k + 1$ words, two among them are R-equivalent. This shows that R has index 4^k. $\square$

Example 6.2. Let $A = \{a\}$, $B = \{b\}$, $C = \{c\}$. Let W be the set of words of even length over $\{b, c\}$, and consider again the language $R = (aW)^+$ of Example 4.2.

The argument used before to show that R has infinite index can be rewritten as follows. We observe that if R is k-mixing, then by the previous proposition, the number of R-equivalence classes for each skeleton is bounded by 4^k. However, two words $w = (abc)^i(ab^2c^2)^{n-i}$ and $w' = (abc)^j(ab^2c^2)^{n-j}$ (with $i < j$) have the same skeleton a^n and are not R-equivalent, since otherwise $(ab)^i(ab^2)^{n-i} \uparrow (ac)^j(ac^2)^{n-j} \subset R$. Thus the number of R-equivalence classes for the skeleton a^n is at least $n+1$.

We now prove a weak converse of the previous property.

Theorem 6.3. *Let R be a (B, C)-commutative regular language over T. If R has index k, then R is k-mixing.*

This is an immediate consequence of the next proposition which gives a more precise description of the construction used in the proof. For this, we introduce some additional notions. A *k*-mixing *decomposition* of *R*, or a *k*-decomposition for short, is a decomposition

$$R = R_1 \cup \cdots \cup R_k, \tag{6}$$

where each R_i is strongly mixing. This decomposition is called *regular* if each R_i is regular. It is called *basic* if each R_i is saturated for $\sim_R$ and if any two words in R_i with the same skeleton are *R*-equivalent. In other words, a decomposition is basic if each R_i has index 1.

The second condition deserves some comment. Let $w, w' \in R$. If $w \sim_R w'$, then $\pi(w) = \pi(w')$, but the converse need not to be true. So what is required is precisely that the converse holds for each component set R_i, that is, for each $w, w' \in R_i$, $\pi(w) = \pi(w') \Longrightarrow w \sim_R w'$.

Example 6.4. In our running example, all words in *R* have the same skeleton, and the *R*-equivalence has four classes. The 4-decomposition (5) is basic and regular.

Proposition 6.5. *Let R be a (B, C)-commutative regular language over T. If R has index k, then R has a k-decomposition that is regular and basic.*

Proof. The proof is constructive. Starting with the regular language *R*, we first choose a regular language $K \subset R$ such that $\pi(K) = \pi(R)$ and π is injective on *K*. In other words, two distinct words in *K* have different skeletons. The cross-section theorem ensures that a regular language *K* with these properties can be effectively constructed. Of course, *K* is not (B, C)-commutative in general.

Set $G'_1 = \pi_B(K)$ and $D'_1 = \pi_C(K)$. Then $K \subset G'_1 \uparrow D'_1 \subset R$ because *R* is (B, C)-commutative. We now consider the saturation of G'_1 for $\sim_{\beta,R}$ and of D'_1 for $\sim_{\gamma,R}$. Set $G_1 = [G'_1]_{\beta,R}$ and $D_1 = [D'_1]_{\gamma,R}$. Then

$$G'_1 \uparrow D'_1 \subset G_1 \uparrow D_1 \subset R.$$

The first inclusion is clear. The second follows from the fact that *R* is (B, C)-commutative, and so $[R]_R = R$. Observe that $G_1 \uparrow D_1$ is basic. Indeed, it is saturated for $\sim_R$ by construction, and any two words with the same skeleton are *R*-equivalent to the only word in *K* having this skeleton, and so are *R*-equivalent.

We prove in a separate lemma (Lemma 6.6 below) that G_1 and D_1 are regular languages. Taking this for granted, one gets a regular language $G_1 \uparrow D_1$ contained in *R*. If $G_1 \uparrow D_1 = R$, the language *R* is strongly mixing and therefore 1-mixing. Otherwise, set $R_1 = G_1 \uparrow D_1$.

$\bar{R}_1 = R \setminus R_1$. Since both R and R_1 are regular and (B, C)-commutative, the language $\bar{R}_1$ also is regular and (B, C)-commutative. Moreover, every skeleton of $\bar{R}_1$ is a skeleton of R, and is also a skeleton of R_1 because R and R_1 have the same sets of skeletons.

We now repeat the same construction on $\bar{R}_1$: we choose a cross-section of $\bar{R}_1$ that is injective for π, we build G'_2, D'_2. Saturation is always with respect to the initial language R. This yields regular languages G_2 over $A \cup B$ and D_2 over $A \cup C$ such that $R_2 = G_2 \uparrow D_2 \subset \bar{R}_1$. Set $\bar{R}_2 = \bar{R}_1 \setminus R_2 = R \setminus (R_1 \cup R_2)$. Again, $\bar{R}_2$ may or may not be empty. This construction is repeated at most k times. Observe that the languages $R_i = G_i \uparrow D_i$ are pairwise disjoint.

We prove that there is an integer $\ell \leqslant k$ such that

$$R = R_1 \cup \cdots \cup R_\ell$$

showing that R is ℓ-mixing and thus also k-mixing. Arguing by contradiction, assume that the claim is false. Then $R \neq R_1 \cup \cdots \cup R_\ell$ for $1 \leqslant \ell \leqslant k$. Repeating the construction once more, we get regular languages G_{k+1} and D_{k+1} such that, setting $R_{k+1} = G_{k+1} \uparrow D_{k+1}$,

$$R \supset R_1 \cup \cdots \cup R_k \cup R_{k+1}\,.$$

Choose $k+1$ words $w_1, \ldots, w_{k+1}$ such that $w_i \in R_i \setminus (R_1 \cup \cdots \cup R_{i-1})$ and $\pi(w_1) = \cdots = \pi(w_{k+1})$. This is possible because the skeletons of R_i are skeletons of R_{i-1}. The set $W = \{w_1, \ldots, w_{k+1}\}$ is monoskeletal. Since R has index k, there are two R-equivalent words in this set, say $w_i \sim_R w_j$ with $i < j$. Then $\pi_B(w_i) \sim_{\beta,R} \pi_B(w_j)$ and $\pi_C(w_i) \sim_{\gamma,R} \pi_C(w_j)$. Since G_i is saturated for the equivalence $\sim_{\beta,R}$ and $\pi_B(w_i) \in G_i$, it follows that $\pi_B(w_j) \in G_i$ and similarly $\pi_C(w_j) \in D_i$. Thus, $w_j \in G_i \uparrow D_i = R_i$, a contradiction. □

Lemma 6.6. *Let G' be a regular language over $A \cup B$ such that π_C is injective on G'. The language $G = [G']_{\beta,R}$ obtained by saturating G' for the equivalence relation $\sim_{\beta,R}$ is an effectively computable regular language.*

Proof. The proof is in two steps. We first show that, for a word u over $A \cup B$, the equivalence class $[u]_{\beta,R}$ is a regular language. This is done by giving two rational transductions for the complement of the language. In a second step, these transductions are used to give a regular expression for the complement of G (see Eq. (7)). The injectivity of π_C plays a central role in the last argument.

Let u be a word over $A \cup B$. Then $[u]_{\beta,R} = \{\bar{u} \mid \beta_R(\bar{u}) = \beta_R(u)\}$. We first show that $[u]_{\beta,R}$ is a regular language for each u. For this, we show that the set

$$L(u) = \pi_C^{-1}(\pi_C(u)) \setminus [u]_{\beta,R}$$

is a regular language. Since $L(u) = \{\bar{u} \in \pi_C^{-1}(\pi_C(u)) \mid \beta_R(\bar{u}) \neq \beta_R(u)\}$ it is composed of two sets, namely those $\bar{u}$ for which $\beta_R(u)$ contains words not in $\beta_R(\bar{u})$, and those $\bar{u}$ for which $\beta_R(\bar{u})$ contains words not in $\beta_R(u)$. Thus, the set $L(u)$ is the union of two, not necessarily disjoint languages

$$L_1(u) = \{\bar{u} \in \pi_C^{-1}(\pi_C(u)) \mid \beta_R(u) \setminus \beta_R(\bar{u}) \neq \emptyset\}$$

and

$$L_2(u) = \{\bar{u} \in \pi_C^{-1}(\pi_C(u)) \mid \beta_R(\bar{u}) \setminus \beta_R(u) \neq \emptyset\}\,.$$

We claim that

$$L_1(u) = \gamma_{T^*\setminus R}(\beta_R(u)) \quad \text{and} \quad L_2(u) = \gamma_R(\beta_{T^*\setminus R}(u))\,.$$

Consider indeed $\bar{u} \in L_1(u)$. Then $\pi_C(\bar{u}) = \pi_C(u)$ and there exists $v \in \beta_R(u)$ such that $v \notin \beta(\bar{u})$. From the first relation, it follows that $\pi_B(v) = \pi_C(u)$ and so also $\pi(v) = \pi_C(\bar{u})$. Thus $\bar{u} \uparrow v \neq \emptyset$, and since $v \notin \beta_R(\bar{u})$, one has $\bar{u} \uparrow v \subset T^* \setminus R$ which means that $v \in \beta_{T^*\setminus R}(\bar{u})$ or equivalently $\bar{u} \in \gamma_{T^*\setminus R}(v)$. Conversely, if $\bar{u} \in \gamma_{T^*\setminus R}(\beta_R(u))$, then there is a word v over $A \cup C$ such that $v \in \beta_R(u)$ and $\bar{u} \in \gamma_{T^*\setminus R}(v)$. The second relation means that $v \in \beta_{T^*\setminus R}(\bar{u})$ which in turn shows that $\bar{u} \uparrow v \subset T^* \setminus R$. Again, $\bar{u} \uparrow v \neq \emptyset$ because $u \uparrow v \neq \emptyset$. Thus $v \notin \beta_R(\bar{u})$. The proof of the second relation is symmetric. This shows that $[u]_{\beta,R}$ is regular.

We now turn to the second step. The sets $L_1 = \bigcup_{u\in G'} L_1(u)$ and $L_2 = \bigcup_{u\in G'} L_2(u)$ are regular because

$$L_1 = \gamma_{T^*\setminus R}(\beta_R(G')) \quad \text{and} \quad L_2 = \gamma_R(\beta_{T^*\setminus R}(G'))\,.$$

Thus the language

$$L = L_1 \cup L_2 = \bigcup_{u\in G'} \pi_C^{-1}(\pi_C(u)) \setminus [u]_{\beta,R}$$

is regular. Since π_C is injective on G', each set $\pi_C^{-1}(\pi_C(u))$ for $u \in G'$ contains exactly one class for $\sim_{\beta,R}$, namely $[u]_{\beta,R}$. In other words, $\pi_C^{-1}(\pi_C(u)) \cap [u']_{\beta,R} = \emptyset$ for $u' \neq u$, $u' \in G'$. This implies that

$$L = \bigcup_{u\in G'} \pi_C^{-1}(\pi_C(u)) \Bigg\backslash \bigcup_{u\in G'} [u]_{\beta,R} \tag{7}$$

Thus, $L = \pi_C^{-1}(\pi_C(G')) \setminus G$, showing that G is regular. $\square$

Example 6.7. Let us perform the construction of the proof of Proposition 6.5 on our running example. Since R is monoskeletal, any word in R is a candidate for the language K. So take $K = \{abc\}$. Then $G'_1 = \{ab\}$, $D'_1 = \{ac\}$, and $G_1 = \{ab\}$, $D_1 = ac(c^2)^*$ and $R_1 = ab \uparrow ac(c^2)^*$. Since $R_1 \neq R$, we continue the construction. Take for instance abc^2 in $R \setminus R_1$. One gets $G_2 = \{ab\}$, $D_2 = a(c^2)^+$, and $R_2 = ab \uparrow a(c^2)^+$. Still $R_1 \cup R_2$ is not covered. Take $w = ab^2c^2$. Then $R_3 = a(b^2)^+ \uparrow a(c^2)^+$. Take $ab^3c \in R \setminus (R_1 \cup R_2 \cup R_3)$. This gives a language $R_4 = ab^3(b^2)^* \uparrow ac(c^2)^*$, and $R = R_1 \cup R_2 \cup R_3 \cup R_4$. In fact, this decomposition is precisely that of Eq. (5). Observe that the languages R_i do *not* correspond to the languages R_i of Example 4.5.

It might be interesting to consider another example, with infinitely many skeletons.

Example 6.8. Set $A = \{a\}$, $B = \{b\}$, and $C = \{c\}$. Let W be the set of words of even length over $\{b, c\}$, and let W_0 (W_1) be the set of words in W having an even (odd) number of b's and of c's. The language we consider is

$$R = \{aw_1aw_2 \cdots aw_n \mid n \geqslant 1, w_1w_2 \cdots w_n \in W\} .$$

Of course, the set of skeletons is $\pi(R) = a^+$. We perform the construction of the proof of Proposition 6.5. We start with a first cross-section $K = a^+$. Clearly, $G'_1 = D'_1 = K$, and $G_1 = [G'_1]_{\beta,R} = \{au_1au_2 \cdots au_n \mid n \geqslant 1, |u_1u_2 \cdots u_n| \equiv 0 \bmod 2\}$, and $D_1 = [D'_1]_{\gamma,R} = \{av_1av_2 \cdots av_n \mid n \geqslant 1, |v_1v_2 \cdots v_n| \equiv 0 \bmod 2\}$. It follows that $R_1 = G_1 \uparrow D_1 = \{aw_1aw_2 \cdots aw_n \mid n \geqslant 1, w_1w_2 \cdots w_n \in W_0\}$. Next, consider a second cross-section $K_2 = a^+bc$. Then $G'_2 = a^+b$ and $D'_2 = a^+c$, and $G_2 = [G'_2]_{\beta,R} = \{au_1au_2 \cdots au_n \mid n \geqslant 1, |u_1u_2 \cdots u_n| \equiv 1 \bmod 2\}$, and similarly for $D_2 = [D'_2]_{\gamma,R}$. It follows that $R_2 = G_2 \uparrow D_2 = \{aw_1aw_2 \cdots aw_n \mid n \geqslant 1, w_1w_2 \cdots w_n \in W_1\}$. Since $R = R_1 \cup R_2$, the language R is 2-mixing.

Observe that the choice of the cross-section may change the decomposition that is obtained. For $i, j = 0, 1$, define

$$V_{i,j} = \{aw_1aw_2 \cdots aw_n \mid n \equiv i \bmod 2, w_1w_2 \cdots w_n \in W_j\} .$$

The languages R_1 and R_2 of the previous 2-decomposition are

$$R_1 = V_{0,0} \cup V_{1,0}, \qquad R_2 = V_{0,1} \cup V_{1,1} .$$

Consider now for instance the language $K' = (a^2)^+ \cup (a^2)^*abc$. This is a cross-section of R. The corresponding projections are $G'_1 = (a^2)^+ \cup (a^2)^*ab$ and $D'_1 = (a^2)^+ \cup (a^2)^*ac$, and a first component of the decomposition of $R = S_1 \cup S_2$ is $S_1 = V_{0,0} \cup V_{1,1}$. The second component is obtained by exchanging even and odd's: $S_2 = V_{1,0} \cup V_{0,1}$.

At this point, we are able to check only partially whether a language R is k-mixing. We proceed as follows:

1. First, we check whether R is (B, C)-commutative. If it is not, then it is not mixing.
2. We use at most 4^k steps of the construction given in the proof of theorem 6.3.
   (a) If the construction stops before at most k steps, we know that R is k-mixing.
   (b) If the construction does not stop after 4^k steps, we know that R is not k-mixing by Proposition 6.1.

However, if the construction stops between k and 4^k steps, we do not (yet) know whether R is k-mixing or not. We will show in Section 7 that in this case, it is decidable whether R *is* k-mixing or not.

6.1. *A second example*

Let us consider again $A = \{a\}$, $B = \{b\}$, and $C = \{c\}$, and consider the language

$$R = a^+ \cup b^+a^+c^* .$$

It is easily seen that R is 2-mixing since

$$R = (a^+ \uparrow a^+) \cup (b^+a^+ \uparrow a^+c^*) .$$

It is also easily seen that R is not 1-mixing. Indeed $\pi_B(R) = b^*a^+$, $\pi_C(R) = a^+c^*$, and $\pi_B(R) \uparrow \pi_C(R) \neq R$ since $\pi_B(R) \uparrow \pi_C(R) \supset a^+c^*$.

Let us compute the R-equivalence classes. For this, we consider first the three words a, ba, and bac. One gets, for the word a,

$$\beta_R(a) = \{a\},\ [a]_{\beta,R} = \{a\},\ \gamma_R(a) = b^*a,\ [a]_{\gamma,R} = \{a\}\,,$$

for the word ba, one gets

$$\beta_R(ba) = ac^*,\ [ba]_{\beta,R} = b^+a\,,$$

and for the word bac

$$\gamma_R(ac) = b^+a,\ [ac]_{\gamma,R} = ac^+\,.$$

This shows that the R-equivalence classes of the words a, ba, and bac are different. In fact, it is now easy to see that $[a]_R = \{a\}$, $[ba]_R = b^+a$, and $[bac]_R = b^+ac^+$. This holds also for words containing more than one letter a. So finally, for all $n \geqslant 1$,

$$[a^n]_R = \{a^n\}, \quad [ba^n]_R = b^+a^n, \quad [ba^nc]_R = b^+a^nc^+\,.$$

Let us apply the construction of Theorem 6.3.

We start with a first cross-section $K_1 = a^+$. Then $G_1 = a^+$ and $D_1 = a^+$, so $R_1 = G_1 \uparrow D_1 = a^+$. The remaining set is $\bar{R}_1 = R \setminus R_1 = b^+a^+c^*$.

Consider next the cross-section $K_2 = ba^+$. Then $G_2 = b^+a^+$, $D_2 = a^+$ and $R_2 = G_2 \uparrow D_2 = b^+a^+$. The remaining set is $\bar{R}_2 = R \setminus (R_1 \cup R_2) = b^+a^+c^+$.

Consider the cross-section $K_3 = ba^+c$. Then $G_3 = b^+a^+$, $D_3 = a^+c^+$ and $R_3 = b^+a^+c^+$. Thus we get the basic 3-decomposition $R = R_1 \cup R_2 \cup R_3$ with

$$\begin{aligned} R_1 &= a^+, \\ R_2 &= b^+a^+, \\ R_3 &= b^+a^+c^+. \end{aligned} \tag{8}$$

Assume now that we start the construction by choosing another initial cross-section, namely $K_1' = b(a^2)^*a \cup b(a^2)^+c$ instead of the set K_1. Then $G_1 = b^+a^+$ and $D_1 = (a^2)^*a \cup (a^2)^+c^+$. Thus one gets $R_1 = G_1 \uparrow D_1 = b^+(a^2)^*a \cup b^+(a^2)^+c^+$, and $\bar{R}_1 = a^+ \cup b^+(a^2)^+ \cup b^+(a^2)^*ac^+$.

Take now the cross-section $K_2' = b(a^2)^+ \cup a(a^2)^*$. Then $G_2 = b^+(a^2)^+ \cup a(a^2)^*$, $D_2 = a^+$, so $R_2 = G_2$ and $\bar{R}_2 = b^+(a^2)^*ac^+ \cup (a^2)^+$.

Finally, we take the cross-section $K_3' = b(a^2)^*a \cup (a^2)^+$. Then we obtain $G_3 = b^+(a^2)^*a \cup (a^2)^+$, again $D_3 = a^+$, and $R_3 = G_3 \uparrow D_3 = \bar{R}_2$. This yields another basic 3-decomposition $R = R_1 \cup R_2 \cup R_3$ with

$$\begin{aligned} R_1 &= b^+(a^2)^*a \cup b^+(a^2)^+c^+, \\ R_2 &= b^+(a^2)^+ \cup (a^2)^*a, \\ R_3 &= b^+(a^2)^*ac^+ \cup (a^2)^+. \end{aligned} \tag{9}$$

7. Complement to the basic construction

In this section, we show that it is decidable whether a regular language R is k-mixing, provided we know a basic regular t-mixing decomposition of R for some t with $k < t \leqslant 4^k$. For this, we show that if R is k-mixing, then there exists a k-decomposition of R that is obtained by gluing together parts of a basic t-decomposition. We moreover show that this k-decomposition can be chosen among a finite number of candidates, proving thus the decidability.

We start with some elementary properties of monoskeletal languages. Recall that a set $K \subset T^*$ is monoskeletal if the set $\pi(K)$ of its skeletons is a singleton. A useful property of monoskeletal languages is that the union of monoskeletal, strongly mixable languages with distinct skeletons is again strongly mixable. More precisely, consider a set $S \subset A^*$ and, for each $s \in S$, monoskeletal languages $G(s)$ over $A \cup B$ and $D(s)$ over $A \cup C$ with skeleton s. Then

$$\bigcup_{s\in S} G(s) \uparrow D(s) = \left(\bigcup_{s\in S} G(s)\right) \uparrow \left(\bigcup_{s\in S} D(s)\right) \tag{10}$$

because $G(s) \uparrow D(s') = \emptyset$ for $s \neq s'$.

Given an arbitrary $K \subset T^*$, and a skeleton $s \in \pi(K)$, the set $K \cap \pi^{-1}(s)$ is monoskeletal by construction. The skeleton s is *simple* for K if $s \in \pi(K)$ and $K \cap \pi^{-1}(s)$ is strongly mixing.

Set $K_B = \pi_B(K)$ and $K_C = \pi_C(K)$. It is easy to check that, for each $s \in \pi(K)$,

$$(K_B \uparrow K_C) \cap \pi^{-1}(s) = (K_B \cap \pi_C^{-1}(s)) \uparrow (K_C \cap \pi_B^{-1}(s))\,.$$

Next, s is simple if and only if

$$K \cap \pi^{-1}(s) = \pi_B(K \cap \pi^{-1}(s)) \uparrow \pi_C(K \cap \pi^{-1}(s))\,.$$

Since $\pi_B(K \cap \pi^{-1}(s)) = K_B \cap \pi_C^{-1}(s)$ (and similarly for the other term), it follows that s is simple if and only if

$$K \cap \pi^{-1}(s) = (K_B \uparrow K_C) \cap \pi^{-1}(s)\,. \tag{11}$$

If $X \subset A^*$ is a set of simple skeletons for K, then $K \cap \pi^{-1}(X)$ is strongly mixing, because

$$K \cap \pi^{-1}(X) = \bigcup_{s\in X} K \cap \pi^{-1}(x) = (K_B \cap \pi_C^{-1}(X)) \uparrow (K_C \cap \pi_B^{-1}(X))\,.$$

Lemma 7.1. *Let $K \subset T^*$ be a regular language. The set $S(K)$ of simple skeletons of K is an effectively computable, regular subset of A^*.*

Proof. Set $K_B = \pi_B(K)$ and $K_C = \pi_C(K)$, and set $L = (K_B \uparrow K_C) \setminus K$. For each $s \in \pi(K)$, one has

$$K_B \uparrow K_C \cap \pi^{-1}(s) = (K \cap \pi^{-1}(s)) \cup (L \cap \pi^{-1}(s))\,.$$

In view of (11), $s \in \pi(K)$ is simple if and only if $L \cap \pi^{-1}(s) = \emptyset$, that is if and only if $s \in \pi(K) \setminus \pi(L)$. This shows that $S(K) = \pi(K) \setminus \pi(L)$ and proves the lemma. $\square$

In the sequel, we consider a regular language R over T that admits a basic t-decomposition

$$R = R_1 \cup \cdots \cup R_t,$$

where the R_i are strongly mixing. Set $N = \{1, \ldots, t\}$. A *k-cover* for N is a set $H = \{I_1, \ldots, I_k\}$ of k subsets $I_1, \ldots, I_k$ of N such that $I_1 \cup \cdots \cup I_k = N$. For any subset I of N, we set $R_I = \cup_{i \in I} R_i$. To any k-cover H of N, we associate the regular set

$$S(H) = S(R_{I_1}) \cap \cdots \cap S(R_{I_k}) . \tag{12}$$

Thus $s \in S(H)$ if and only if $s \in \pi(R)$ and each $R_{I_j} \cap \pi^{-1}(s)$ is strongly mixing.

Lemma 7.2. *For each k-cover H, the language $R \cap \pi^{-1}(S(H))$ is a regular k-mixing language.*

Proof. Set $H = \{I_1, \ldots, I_k\}$. In view of Eq. (12) and Lemma 7.1, the language $R \cap \pi^{-1}(S(H))$ is indeed regular.

Since $S(H)$ is a set of simple skeletons for each R_{I_j}, each language $K_j = R_{I_j} \cap \pi^{-1}(S(H))$ is strongly mixing, so $K_1 \cup \cdots \cup K_k$ is k-mixing. Next, $R = R_{I_1} \cup \cdots \cup R_{I_k}$ because H is a k-cover. Thus,

$$R \cap \pi^{-1}(S(H)) = K_1 \cup \cdots \cup K_k .$$

This proves the lemma. $\square$

The same result holds for several k-covers

Lemma 7.3. *Let $H_1, \ldots, H_n$ be k-covers of N. Then the union of the languages $R \cap \pi^{-1}(S(H_i))$, for $i = 1, \ldots, n$, is k-mixing.*

Proof. The union is

$$R \cap \pi^{-1}(S(H_1) \cup \cdots \cup S(H_n)) .$$

Set $S_1 = S(H_1)$, and $S_i = S(H_i) \setminus (S(H_1) \cup \cdots \cup S(H_{i-1}))$ for $i = 2, \ldots, n$. Then $S_1, \ldots, S_n$ are pairwise disjoint and $S(H_1) \cup \cdots \cup S(H_n) = S_1 \cup \cdots \cup S_n$. Each of the sets $R \cap \pi^{-1}(S_i)$ is k-mixing, and since the union is now over disjoint sets of skeletons, it is again k-mixing. $\square$

A set $H_1, \ldots, H_n$ of k-covers is *complete* for R if $\pi(R) = S(H_1) \cup \cdots \cup S(H_n)$, that is if every skeleton is in at least one of the sets $S(H_i)$.

Proposition 7.4. *If R has a complete set of k-covers, then R is k-mixing.*

Proof. Let $H_1, \ldots, H_n$ be a complete set of k-covers. By the previous lemma, the union of the sets $R \cap \pi^{-1}(S(H_i))$ is k-mixing. This union is $R \cap \pi^{-1}(\cup S(H_i)) = R \cap \pi^{-1}(\pi(R)) = R$. $\square$

Example 7.5. Let us illustrate the preceding proposition with the 3-decomposition (9). We consider the two 3-covers $H_1 = \{\{1,3\},\{2\}\}$ and $H_2 = \{\{1\},\{2,3\}\}$. Consider the first one. Then $R_{\{1,3\}} = R_1 \cup R_3 = b^+(a^2)^*ac^* \cup b^+(a^2)^+c^+ \cup (a^2)^+$ and it follows that $S(R_{\{1,3\}}) = (a^2)^*a$. Assume indeed that a^n is a simple skeleton of $R_{\{1,3\}}$ for some even integer n. Then $R_{\{1,3\}} \cap \pi^{-1}(a^n)$ must be the mixing product of its projections, that is must be equal to $b^*a^n \uparrow a^n c^*$, and this does not hold. Clearly, $S(R_3) = a^+$ because R_3 is strongly mixing. So $S(H_1) = (a^2)^*a$. A similar computation shows that $R_{\{2,3\}} = b^*(a^2)^* \cup b^+(a^2)^*ac^+ \cup (a^2)^*a$ and that $S(R_{\{2,3\}}) = (a^2)^+$. So $S(H_2) = (a^2)^+$, and the set H_1, H_2 is a complete set of 3-covers. According to the construction given in the previous proof, it suffices to compute the union of the languages $R \cap \pi^{-1}((a^2)^*a)$ and $R \cap \pi^{-1}((a^2)^+)$. One gets $R \cap \pi^{-1}((a^2)^*a) = b^+(a^2)^*ac^* \cup (a^2)^*a$ and $R \cap \pi^{-1}((a^2)^+) = b^+(a^2)^+c^+ \cup b^*(a^2)^+$ and finally $R = b^*a^+ \cup b^+a^+c^+$.

Conversely, one has the following.

Proposition 7.6. *If R is k-mixing, then for any basic t-decomposition $R = R_1 \cup \cdots \cup R_t$, there exists a complete set of k-covers.*

Proof. If R is k-mixing, then

$$R = M_1 \cup \cdots \cup M_k$$

with $M_1, \ldots, M_k$ strongly mixing. Since R is saturated for R-equivalence, we may assume that each M_j is saturated, i.e., $M_j = [M_j]_R$. Assume another, basic t-decomposition

$$R = R_1 \cup \cdots \cup R_t$$

exists. Set $N = \{1, \ldots, t\}$. Let $s \in \pi(R)$ be a skeleton. For each $j \in \{1, \ldots, k\}$, consider the set $I'_j \subset N$ of integers $i \in N$ such that $M_j \cap \pi^{-1}(s) \cap R_i \neq \emptyset$. Clearly,

$$\bigcup_{j=1}^{k} I'_j = \{i \in N \mid R_i \cap \pi^{-1}(s) \neq \emptyset\}\,.$$

This set may be a strict subset of N, so that $\{I'_1, \ldots, I'_k\}$ is not necessarily a k-cover. Define $I_j = I'_j \cup \{i \in N \mid R_i \cap \pi^{-1}(s) = \emptyset\}$. Then $H(s) = \{I_1, \ldots, I_k\}$ is a k-cover. In this way, we associated a k-cover $H(s)$ to each skeleton s. We claim that $s \in S(H(s))$. Assume this for granted. Then,

$$\pi(R) = \bigcup_{s \in \pi(R)} S(H(s))\,. \tag{13}$$

Observe that there are only finitely many k-covers. Thus, the union on the right-hand side of (13) is finite, showing that the finite set $(S(H(s)))_{s \in \pi(K)}$ is complete.

It remains to prove the claim, namely that $s \in S(H(s))$ for each skeleton $s \in \pi(K)$. This is equivalent to show that $s \in S(R_{I_j})$ for $j = 1, \ldots, k$, which in turn means that the set $(\bigcup_{i \in I_j} R_i) \cap \pi^{-1}(s)$ is strongly mixing. For this, it suffices to show that

$$\left(\bigcup_{i \in I_j} R_i\right) \cap \pi^{-1}(s) = M_j \cap \pi^{-1}(s) \,. \tag{14}$$

If $w \in M_j \cap \pi^{-1}(s)$, then $w \in R_1 \cup \cdots \cup R_t$, so $w \in R_i$ for some i. This index i is in I'_j. Conversely, let $w \in (\bigcup_{i \in I_j} R_i) \cap \pi^{-1}(s)$. Then $w \in R_i \cap \pi^{-1}(s)$ for some $i \in I'_j$ (because $R_i \cap \pi^{-1}(s) = \emptyset$ for $i \in I_j \setminus I'_j$). Consider any word $z \in M_j \cap R_i \cap \pi^{-1}(s)$. Then $\pi(z) = \pi(w) = s$, and because the languages R_i are basic, one has $z \sim_R w$. Since $z \in M_j$ and M_j is saturated for the R-equivalence, also $w \in M_j$. Thus, $w \in M_j \cap \pi^{-1}(s)$. This completes the proof. □

Proposition 7.7. *Given a t-decomposition of R, and an integer k, it is decidable whether a complete set of k-covers exists.*

Proof. Any k-cover $H = \{I_1, \ldots, I_k\}$ of $N = \{1, \ldots, t\}$ yields a language $S(H)$. This language is regular and effectively computable by Lemma 7.1. There are only finitely many k-covers, so only finitely many $S(H)$. It suffices to test whether their union is equal to $\pi(R)$. □

Theorem 7.8. *Let R be a regular language over T. Given an integer k, it is decidable whether R is k-mixing.*

Proof. The algorithm goes as follows.

1. Check whether R is (B, C)-commutative. If not, R is not mixing.
2. Try to construct a basic representation of R by the method given in the proof of Proposition 6.5.
   (a) If the construction succeeds in at most k steps, R is k-mixing.
   (b) If the construction fails after 4^k steps, then R is not k-mixing.
   (c) If the construction succeeds in t steps with $k < t < 4^k$, go to the next step.
3. Check whether a complete set k-covers exists for the t-decomposition of the previous step. This is done by simple (but time-consuming!) computation of the finitely many k-covers enumeration of all k-cover, and by trying all combinations. If a complete set exists, then R is k-mixing, otherwise it is not. □

Let us mention some additional facts.

Proposition 7.9. *If $R = M_1 \cup \cdots \cup M_k$ is any k-mixing decomposition of a regular language R, then the R-equivalence closures $[M_1]_R, \ldots, [M_k]_R$ are regular languages.*

Proof. If $R = M_1 \cup \cdots \cup M_k$, then $R = [M_1]_R \cup \cdots \cup [M_k]_R$, so we may assume that $M_j = [M_j]_R$ for $j = 1, \ldots, k$. With the notation of the proof of Proposition 7.6, assume that there exist k-covers $H_1, \ldots, H_n$ that are complete for some basic regular t-decomposition, so that

$$\pi(R) = S(H_1) \cup \cdots \cup S(H_n)$$

according to Eq. (13). Given $m \in \{1, \ldots, n\}$, let $H_m = \{I_1, \ldots, I_k\}$. In view of Eq. (14), one has

$$M_j \cap \pi^{-1}(S(H_m)) = \left(\bigcup_{i \in I_j} R_i \right) \cap \pi^{-1}(S(H_m)) \,.$$

This set is regular by Lemma 7.1. Since $M_j = \bigcup_{m=1}^{n} M_j \cap \pi^{-1}(S(H_m))$, the proposition is proved. $\square$

Corollary 7.10. *If R is a k-mixing regular language R, then R has a regular k-mixing decomposition.*

Proof. Indeed, if $R = M_1 \cup \cdots \cup M_k$, then $R = [M_1]_R \cup \cdots \cup [M_k]_R$ and $[M_1]_R, \ldots, [M_k]_R$ are regular languages. $\square$

8. Concluding remarks

We have shown that for a given integer k, it is decidable whether a regular language R is k-mixing. It still remains an open question if one can decide whether R is k-mixing for some k.

The case we have studied here is a partition of the alphabet T into 3 subalphabet. This is a special case of a more general case, namely a partition into $m + 1$ subsets $T = A \cup B_1 \cup \cdots \cup B_m$, where closure under permutation of letters from different subalphabet B_i is permitted. The question whether our result extend to this case is open.

References

[1] V. Diekert, G. Rozenberg (Eds.), The Book of Traces, World Scientific, Singapore, 1995.
[2] C. Duboc, Mixed product and asynchronous automata, Theoret. Comput. Sci. 1986.
[3] W. Zielonka, Notes on finite asynchronous automata, Theoret. Inform. Appl. 21 (2) (1987) 99–135.