Closure properties of synchronized relations

María Emilia Descotte, Diego Figueira, Santiago Figueira

Closure properties of synchronized relations

María Emilia Descotte
LaBRI, Université de Bordeaux, France

Diego Figueira
CNRS & LaBRI, Université de Bordeaux, France

Santiago Figueira
CONICET & Universidad de Buenos Aires, Argentina

Abstract

A standard approach to define k-ary word relations over a finite alphabet $\mathbb{A}$ is through k-tape finite state automata that recognize regular languages L over $\{1,\ldots,k\} \times \mathbb{A}$, where (i,a) is interpreted as reading letter a from tape i. Accordingly, a word $w \in L$ denotes the tuple $(u_1,\ldots,u_k) \in (\mathbb{A}^*)^k$ in which u_i is the projection of w onto i-labelled letters. While this formalism defines the well-studied class of rational relations, enforcing restrictions on the reading regime from the tapes, which we call *synchronization*, yields various sub-classes of relations. Such synchronization restrictions are imposed through regular properties on the projection of the language L onto $\{1,\ldots,k\}$. In this way, for each regular language $C \subseteq \{1,\ldots,k\}^*$, one obtains a class $\textsc{Rel}(C)$ of relations. Synchronous, Recognizable, and Length-preserving rational relations are all examples of classes that can be defined in this way.

We study basic properties of these classes of relations, in terms of closure under intersection, complement, concatenation, Kleene star and projection. We characterize the classes with each closure property. For the binary case ($k = 2$) this yields effective procedures.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory

Keywords and phrases synchronized word relations, rational, closure, characterization, intersection, complement, Kleene star, concatenation

Acknowledgements Work supported by ANR project DELTA, grant ANR-16-CE40-0007, grant PICT-2016-0215, and LIA INFINIS.

1 Introduction

We study relations of finite words, that is, sets $R \subseteq (\mathbb{A}^*)^k$ for a finite alphabet $\mathbb{A}$ and $k \in \mathbb{N}$, where $(\mathbb{A}^*)^k$ is the cartesian product of k copies of $\mathbb{A}^*$. The study of these relations dates back to the works of Büchi, Elgot, Mezei, and Nivat in the 1960s [11, 16, 26], with much subsequent work done later (*e.g.*, [7, 13]). Most of the investigations focused on extending the standard notion of regularity from languages to relations. This effort has followed the long-standing tradition of using equational, operational, and descriptive formalisms — that is, finite monoids, automata, and regular expressions — for describing relations, and gave rise to three different classes of relations: *Recognizable*, *Automatic* (*a.k.a. Regular* [7] or *Synchronous* [21, 13]), and *Rational*.

The above classes of relations can be seen as three particular examples of a much larger (in fact infinite) range of possibilities, where relations are described by special languages over extended alphabets, called *synchronizing languages* [19]. Intuitively, the idea is to describe a k-ary relation by means of a k-tape automaton with k heads, one for each tape, which can move independently of one another. In the basic framework of synchronized relations, one lets each head of the automaton either move right or stay in the same position. In addition, one can constrain the possible sequences of head motions by a suitable regular language $C \subseteq \{1, \ldots, k\}^*$. In this way, each regular language $C \subseteq \{1, \ldots, k\}^*$ induces a class of k-ary relations, denoted $\textsc{Rel}(C)$, which is contained in the class *Rational* (due to Nivat's Theorem [26]). For example, on binary relations, the classes *Recognizable*, *Automatic*, and *Rational* are captured, respectively, by the languages $C_{\text{Rec}} = \{1\}^* \cdot \{2\}^*$, $C_{\text{Aut}} = \{12\}^* \cdot \{1\}^* \cup \{12\}^* \cdot \{2\}^*$, and $C_{\text{Rat}} = \{1, 2\}^*$. Roughly speaking, any other class that can be defined through the 'tape behavior' of a multi-tape automaton will be also captured by this framework. Other examples include length-preserving, or α-synchronous relations [12]. However, it should be noted that other well-known subclasses of rational relations, such as deterministic or functional relations, are not captured by the notion of synchronization. In general, the correspondence between a language $C \subseteq \{1, \ldots, k\}^*$ and the induced class $\textsc{Rel}(C)$ of synchronized relations is not one-to-one: it may happen that different languages C, D induce the same class of synchronized relations. The problem of when two classes of synchronized relations coincide, and when one is contained in the other has been only recently solved for the case of binary relations [14], while the case for arbitrary k-ary relations remains open. In this work we identify, among the infinitely many synchronized classes of relations, which are those with good closure properties, in terms of paradigmatic operations such as intersection, complement, concatenation, projection, or Kleene star.

Motivation

The motivation for identifying and studying well-behaved classes of word relations, besides its intrinsic interest within formal language theory, stems from various areas. One motivation comes from verification of safety and liveness properties of parameterized systems, where relations describe transitions [1, 10, 24, 28]. Another one arises from the study of Automatic Structures [8], where word languages and relations are used to describe infinite structures, and good closure properties are necessary to obtain effective model checking of logics. Another example is the study of formal models underlying IBM's tools for text extraction into a relational model [17]; where several classes of relations emerge (some outside *Rational*) with differing closure properties. Yet another comes from graph databases, which are actively studied as a suitable model for RDF data, social networks data, and others [2]. Paths in graph databases are described by their labels and hence they are abstracted as finite words. These paths need to be compared, for instance, for their degree of similarity, edit distance, or other relations [3, 5, 25]. As a concrete link with the present work we consider CRPQs —a basic query language for graph-structured data. As it was shown in [4], allowing rational relations in CRPQs turns the query evaluation problem undecidable. There have therefore been efforts towards finding subclasses of *Rational* relations that preserve decidability for CRPQs (*e.g.* [5, 18, 6]), often exploiting an effective closure under intersection on the underlying subclass of relations. Part of our motivation for studying closure under intersection stems from our ambition, as future work, to characterize all synchronized classes of relations that can be added to CRPQs while preserving decidability.

Contribution

Our main contribution is a characterization for each of the studied closure properties, the main results can be summarized as follows.

▷ Theorem. For every regular $C \subseteq \mathbb{2}^*$, it is decidable whether $\text{Rel}(C)$ is closed under intersection, complement, concatenation, Kleene star and projection.

While some of the characterizations we give are for arbitrary arity relations, we were only able to show decidability for binary arity. Indeed, the decidability of these characterizations relies, crucially, on the decidability of testing for inclusion between synchronized classes, which has only been shown for binary relations [14].

We do not include closure under union since it can be easily seen that all classes defined in this way are closed under union. The most involved result is closure under intersection. The main property we will prove is that $\text{Rel}(C)$ is closed under intersection if, and only if, $\text{Rel}(C) \subseteq \text{Rel}(D)$ for some D whose Parikh-image is *injective* (*i.e.*, there are no two distinct words of D with the same Parikh-image). Further, we show that this can be tested, and such a language D can be effectively constructed, whenever possible. In the same vein, we obtain that $\text{Rel}(C)$ is closed under complement if, and only if, $\text{Rel}(C) = \text{Rel}(D)$ for some D with a *bijective* Parikh-image. (Observe that closure under complement implies closure under intersection in view of the fact that all classes are closed under union.)

Related work

The formalization of the framework to describe synchronized classes of relations has been introduced only recently [19]. As mentioned, the problem of containment between classes of relations has been addressed in [14] for the binary case. The formalism of synchronizations has been also extended beyond rational relations by means of semi-linear constraints [18] in the context of querying graph databases.

The paper [9] studies relations with origin information, as induced by non-deterministic (one-way) finite state transducers. Origin information can be seen as a way to describe a synchronization between input and output words — somehow in the same spirit of our synchronization languages — and was exploited to recover decidability of the equivalence problem for transducers. The paper [20] pursues further this principle by studying “distortions” of the origin information, called resynchronizations. The paper [29] studies the *uniformization problem* for synchronized relations.

Organization

After a preliminary Section 2, we show the main result characterizing closure under intersection in Section 3. In Section 4 we study closure under complement and another variant that we call “relativized complement”. In Section 5 we give characterizations for closure under concatenation, Kleene star and projection. We conclude with Section 6. Detailed proofs of all statements not included in the body can be found in the Appendix.

2 Preliminaries

We denote by $\mathbb{N}$ the set of non-negative integers. $\mathbb{A}, \mathbb{B}$ denote arbitrary finite alphabets and for $k \in \mathbb{N}$, $k \geq 1$, $\Bbbk$ denotes the k-letter alphabet $\{1, \dots, k\}$. For a word $w \in \mathbb{A}^*$, $|w|$ is its length, and $|w|_a$ is the number of occurrences of symbol a in w.

Regular languages

We use standard notation for regular expressions without complement, namely, for expressions built up from the empty set, the empty word ε and the symbols $a \in \mathbb{A}$, using the operations $\cdot$, $\cup$, and $(\,)^*$. For economy of space and clarity we use the abbreviated notation $(\,)^n$, $(\,)^{<n}$, $(\,)^{\geq n}$, $(\,)^{n*}$, and $(\,)^{*n}$ —the last two being shorthands for $((\,)^n)^*$ and $((\,)^*)^n$ respectively. We abuse the notation $(\,)^k$ to also denote the cartesian product of k copies of the same set (typically $(\mathbb{A}^*)^k$) when there is no risk of confusion. We also identify regular expressions with the defined languages; for example, we may write $abbc \in a \cdot b^{\geq 2} \cdot (c \cup d)^*$, $b(ab)^* = (ba)^*b$ and $\{a, b\}^* \cdot c = (a \cup b)^* \cdot c$. The *star-height* of a regular expression is the maximum number of nested Kleene stars $(\,)^*$. Given $u = a_1 \cdots a_n \in \mathbb{A}^*$ and $v = b_1 \cdots b_n \in \mathbb{B}^*$, we write $u \otimes v$ for the word $(a_1, b_1) \cdots (a_n, b_n) \in (\mathbb{A} \times \mathbb{B})^*$. Similarly, given $U \subseteq \mathbb{A}^*$, $V \subseteq \mathbb{B}^*$, we write $U \otimes V \subseteq (\mathbb{A} \times \mathbb{B})^*$ for the set $\{u \otimes v \colon u \in U, v \in V, |u| = |v|\}$. Given two languages L, L' over $\mathbb{A}$, we write $L \subseteq_{reg} L'$ to denote that L is a regular subset of L'.

A regular expression $C \subseteq \mathbb{2}^*$ is *concat-star*, if it is of the form

$$C = C_1^* u_1 C_2^* u_2 \cdots C_n^* u_n, \tag{$\star$}$$

for $n \in \mathbb{N}$, words $u_1, \dots, u_n$, and regular expressions $C_1, \dots, C_n$ where none of the C_i's describes the empty language. The C_i^*'s from $(\star)$ are called *components* of the concat-star. A concat-star expression like $(\star)$ is *smooth* if either $n \leq 2$ or there are no $\ell, s \in \mathbb{2}$ and $1 \leq i < n$ such that $C_i \subseteq \ell^*, C_{i+1} \subseteq s^*$. We say that a regular language L is concat-star (resp. smooth) if it admits a concat-star (resp. smooth) expression.

Parikh-images and linear sets

The *Parikh-image* of $w \in \mathbb{2}^*$ is the pair associating each symbol of $\mathbb{2}$ to its number of occurrences in w, *i.e.* $\pi(w) = (|w|_1, |w|_2)$. We naturally extend this to languages $L \subseteq \mathbb{2}^*$ by letting $\pi(L) \stackrel{\text{def}}{=} \{\pi(w) \colon w \in L\}$ $(\subseteq \mathbb{N}^2)$. A language $C \subseteq \mathbb{2}^*$ is *Parikh-injective* if for every $u, v \in C$, if $\pi(u) = \pi(v)$ then $u = v$; it is *Parikh-surjective* if $\pi(C) = \mathbb{N}^2$; and it is *Parikh-bijective* if it is both Parikh-injective and -surjective. We will use the product order $(\leq, \mathbb{N}^2)$, defined by $(n, m) \leq (n', m')$ iff $n \leq n'$ and $m \leq m'$. Given a vector $\bar{x} \in \mathbb{N}^2$ and a set $X = \{\bar{x}_1, \dots, \bar{x}_n\} \subseteq \mathbb{N}^2$ (in our case, the Parikh-image of words from $\mathbb{2}^*$), we define the *linear set generated by X and $\bar{x}$* as $\langle \bar{x}, X \rangle = \{\bar{x} + \alpha_1 \cdot \bar{x}_1 + \cdots + \alpha_n \bar{x}_n \colon \alpha_i \in \mathbb{N}\}$. For economy of space we write $\langle X \rangle$ as short for $\langle \bar{0}, X \rangle$, where $\bar{0} = (0, 0)$. Note that, in particular, $\langle \emptyset \rangle = \{\bar{0}\}$. A *semi-linear set* is a finite union of linear sets. The following fact will be useful in the next section.

▶ **Lemma 1.** *For every semi-linear set $V \subseteq \mathbb{N}^2$ there exists a Parikh-injective language $C \subseteq_{reg} \mathbb{2}^*$ such that $\pi(C) = V$.*

Two sets of vectors $X, Y \subseteq \mathbb{N}^2$ are *independent* if $\bar{0} \notin X \cup Y$ and $\langle X \rangle \cap \langle Y \rangle = \{\bar{0}\}$; otherwise they are *dependent*. We say that two languages over $\mathbb{2}$ are *Parikh-independent* (resp. Parikh-dependent) if their Parikh-images are. We abuse notation and say that $\bar{x}$ and $\bar{y}$ are (in)dependent whenever $\{\bar{x}\}$ and $\{\bar{y}\}$ are (in)dependent, and likewise for words. We will need the following simple observation later.

▷ Observation 2. If u and v are Parikh-independent, for every $s, t, s', t' \in \mathbb{N}$, if $\pi(u^s v^{s'}) = \pi(v^t u^{t'})$, then $s' = t$ and $t' = s$.

Indeed, we have that $s \cdot \pi(u) + s' \cdot \pi(v) = t' \cdot \pi(u) + t \cdot \pi(v)$. Let us assume that $s' \leq t$ (the case in which is $\geq$ is similar). Then $t' \leq s$ and so we have $(s - t') \cdot \pi(u) = (t - s') \cdot \pi(v)$ which implies $s - t' = 0 = t - s'$ since u and v are Parikh-independent. Then $s' = t$ and $t' = s$.

2.1 Synchronized relations

A *synchronization* of a tuple $(w_1, \ldots, w_k)$ of words over $\mathbb{A}$ is a word over $\mathbb{k} \times \mathbb{A}$ such that the projection onto $\mathbb{A}$ of positions labeled by i is exactly w_i, for $i = 1, \ldots, k$. For example, the words $(1, a)(1, b)(2, a)$ and $(1, a)(2, a)(1, b)$ are two possible synchronizations of the same pair (ab, a). Every word $w \in (\mathbb{k} \times \mathbb{A})^*$ is a synchronization of a unique tuple $(w_1, \ldots, w_k)$ of words over $\mathbb{A}$, where for all $i \in \{1, \ldots, k\}$, $i^{|w_i|} \otimes w_i$ is the projection of w onto the alphabet $\{i\} \times \mathbb{A}$. We denote such tuple $(w_1, \ldots, w_k)$ by $[\![w]\!]_k$ and extend the notation to languages $L \subseteq (\mathbb{k} \times \mathbb{A})^*$ by denoting the unique k-ary relation synchronized by L as $[\![L]\!]_k \stackrel{\text{def}}{=} \{[\![w]\!]_k : w \in L\}$. In our previous example, $[\![(1, a)(1, b)(2, a)]\!]_2 = [\![(1, a)(2, a)(1, b)]\!]_2 = (ab, a)$, and $[\![\{(1, a)(2, a), (1, a)(2, b), (1, b)(2, a), (1, b)(2, b)\}^*]\!]_2$ is the equal-length relation on the alphabet $\{a, b\}$.

In this setup, we define classes of relations by restricting the set of admitted synchronizations. One way of doing so is to fix a language $C \subseteq_{reg} \mathbb{k}^*$, called *control language*, and let L vary over all regular languages over $\mathbb{k} \times \mathbb{A}$ whose projections onto $\mathbb{k}$ are in C. Thus, given $k \in \mathbb{N}$ and $C \subseteq_{reg} \mathbb{k}^*$, we define the class of k-ary C*-controlled relations* as

$$\text{REL}_k(C) \stackrel{\text{def}}{=} \big\{([\![L]\!]_k, \mathbb{A}) : L \subseteq_{reg} C \otimes \mathbb{A}^*, \mathbb{A} \text{ is a finite alphabet}\big\}.$$

Whenever k is clear from the context, we write $[\![w]\!]$, $[\![L]\!]$ and $\text{REL}(C)$. For economy of space, we write $C =_{\text{REL}} D$ as short for $\text{REL}(C) = \text{REL}(D)$, and we say that C is REL-equivalent to D. Similarly, we write $C \subseteq_{\text{REL}} D$ as short for $\text{REL}(C) \subseteq \text{REL}(D)$ and we say that C is REL-contained in D. The definition makes explicit the alphabet used for each relation, in contrast to previous definitions of synchronized classes [19, 14]. The reason for this is that in particular we study closure under complement, which requires the alphabet to be specified. However, we observe that synchronized classes are closed under taking super-alphabets, and thus the alphabet can be often disregarded. We then write $R \in \text{REL}(C)$ to denote $(R, \mathbb{A}) \in \text{REL}(C)$ for some $\mathbb{A}$.

▷ Observation 3. If $(R, \mathbb{A}) \in \text{REL}(C)$ then $(R, \mathbb{A}') \in \text{REL}(C)$ for every $\mathbb{A} \subseteq \mathbb{A}'$. If $(R, \mathbb{A}) \in \text{REL}(C)$ then $(R, \mathbb{A}_R) \in \text{REL}(C)$, where $\mathbb{A}_R \subseteq \mathbb{A}$ is the set of symbols present in R.

Clearly, $C \subseteq_{reg} D \subseteq_{reg} \mathbb{k}^*$ implies $C \subseteq_{\text{REL}} D$, but the converse does not hold: $\text{REL}_2(C_{\text{Rec}}) =$ *Recognizable* $\subsetneq$ *Automatic* $= \text{REL}_2(C_{\text{Aut}})$, but $C_{\text{Rec}} \not\subseteq C_{\text{Aut}}$. Moreover, different control languages may induce the same class of synchronized relations. For any two regular $C, D \subseteq_{reg} \mathbb{k}^*$ it is decidable to test whether $C \subseteq_{\text{REL}} D$ in the case $k = 2$ [14], but for arbitrary k-ary relations the decidability of the class containment problem is open. Henceforward, *Rational* will denote the class $\text{REL}(\mathbb{2}^*)$ of rational relations.

We restate some properties from [14] that we will use throughout (the proofs in [14] are for the case $k = 2$ but they can be easily generalized to arbitrary k). We will use the notation $R \cdot S$ to denote the usual concatenation of relations, more specifically, given $R, S \subseteq (\mathbb{A}^*)^k$, $R \cdot S = \{(u \cdot u', v \cdot v') \colon (u, v) \in R \text{ and } (u', v') \in S\}$.

▶ **Lemma 4** (Lemma 2 of [14]). *For every $C, D, C', D' \subseteq_{reg} \Bbbk^*$,*

1. *if $R \in \textsc{Rel}(C \cdot D)$, there are $R_1, \dots, R_n \in \textsc{Rel}(C)$, $R'_1, \dots, R'_n \in \textsc{Rel}(D)$ such that $R = \bigcup_i R_i \cdot R'_i$;*
2. *if $R \in \textsc{Rel}(C^*)$, there are $R_1, \dots, R_n \in \textsc{Rel}(C)$ and $I \subseteq_{reg} \{1, \dots, n\}^*$ such that $R = \bigcup_{w \in I} R_{w[1]} \cdots R_{w[|w|]}$;*
3. *For every $R \in \textsc{Rel}(C \cup D)$, there are $R_1 \in \textsc{Rel}(C)$, $R_2 \in \textsc{Rel}(D)$ such that $R = R_1 \cup R_2$.*
4. *if $C \subseteq D$, then $C \subseteq_{\textsc{Rel}} D$;*
5. *if $C \subseteq_{\textsc{Rel}} D$ and $C' \subseteq_{\textsc{Rel}} D'$, then $C \cdot C' \subseteq_{\textsc{Rel}} D \cdot D'$ and $C \cup C' \subseteq_{\textsc{Rel}} D \cup D'$;*
6. *if $C \subseteq_{\textsc{Rel}} D$, then $C^* \subseteq_{\textsc{Rel}} D^*$;*
7. *for every partition I, J of $\{1, \dots, k\}$ such that $C \subseteq I^*$ and $D \subseteq J^*$, we have $C \cdot D =_{\textsc{Rel}} D \cdot C$;*
8. *if C is finite, then $C \cdot D =_{\textsc{Rel}} D \cdot C$;*
9. *if $C \subseteq_{\textsc{Rel}} D$ then $\pi(C) \subseteq \pi(D)$; moreover, if C is finite, the converse also holds.*

The following decomposition lemma, which is an immediate consequence of [14, Proposition 3 plus Lemma 2 P7] and basic properties from Lemma 4, will be used throughout.

▶ **Lemma 5.** *Every $C \subseteq_{reg} \mathbb{2}^*$ is effectively* $\textsc{Rel}$*-equivalent to a finite union of smooth languages, i.e. given $C \subseteq_{reg} \mathbb{2}^*$, one can compute a finite set of smooth languages such that C is* $\textsc{Rel}$*-equivalent to their union.*

In addition to these, our characterization results make use of the following easy properties of relations controlled by Parikh-injective and Parikh-bijective languages.

▶ **Lemma 6.** *For any $C \subseteq_{reg} \Bbbk^*$ and $L, M \subseteq_{reg} C \otimes \mathbb{A}^*$,*

1. *if C is Parikh-injective, and $w, w' \in C \otimes \mathbb{A}^*$, then $[\![w]\!] = [\![w']\!]$ implies $w = w'$;*
2. *$[\![L]\!] \cup [\![M]\!] = [\![L \cup M]\!]$;*
3. *if C is Parikh-injective then $[\![L]\!] \cap [\![M]\!] = [\![L \cap M]\!]$ and $[\![L]\!] \setminus [\![M]\!] = [\![L \setminus M]\!]$;*
4. *if C is Parikh-bijective then $(\mathbb{A}^*)^k \setminus [\![L]\!] = [\![(C \otimes \mathbb{A}^*) \setminus L]\!]$;*
5. *if C is Parikh-surjective then $1^* \cdots k^* \subseteq_{\textsc{Rel}} C$.*

Proof. The first two items follow immediately from definitions.

3. $[\![L \cap M]\!] \subseteq [\![L]\!] \cap [\![M]\!]$ is always true. For the other containment, let $(w_1, \dots, w_k) \in [\![L]\!] \cap [\![M]\!]$, then there exist $w \in L, w' \in M$ such that $[\![w]\!] = [\![w']\!] = (w_1, \dots, w_k)$. Since C is Parikh-injective, by item 1, $w = w' \in L \cap M$ synchronizes $(w_1, \dots, w_k)$ which concludes the proof.
   $[\![L]\!] \setminus [\![M]\!] \subseteq [\![L \setminus M]\!]$ is always true. For the other containment, let $w \in L \setminus M$. Then $[\![w]\!] \in [\![L]\!]$. By way of contradiction, suppose that $[\![w]\!] \in [\![M]\!]$. In this case, there exists $w' \in M$ such that $[\![w]\!] = [\![w']\!]$. Since C is Parikh-injective, by item 1, $M \not\ni w = w' \in M$ which is a contradiction.
4. For $\subseteq$, note that, since C is Parikh-surjective, $(\mathbb{A}^*)^k = [\![C \otimes \mathbb{A}^*]\!]$, and so the result follows from the previous item.
5. We make use of closure under componentwise letter-to-letter relations (*cf.* Lemma 8 of Section 2.2). Suppose $C \subseteq_{reg} \Bbbk^*$ is Parikh-surjective, and let $R \in \textsc{Rel}(1^* \cdots k^*)$. As an immediate consequence of Mezei's theorem, we have the following:

▷ Claim 7. For every k, $\textsc{Rel}(1^* \cdots k^*) = \{\bigcup_{i \in I} L_{i,1} \times \cdots \times L_{i,k} \colon I$ is finite and $L_{i,j} \subseteq_{reg} \mathbb{A}^*$ for some finite alphabet $\mathbb{A}\}$.

Then $R = \bigcup_{i \in I} L_{i,1} \times \cdots \times L_{i,k}$ for a finite I and regular languages $L_{i,j}$. For any $i \in I$ and $j \in \mathbb{k}$ consider $T_{i,j}$ as the letter-to-letter relation $T_{i,j} = \{(u,v) : |u| = |v|$ and $v \in L_{i,j}\} \in \textsc{Rel}((12)^*)$. Note that, by Parikh-surjectivity, $U = (\mathbb{A}^*)^k = [\![C \otimes \mathbb{A}^*]\!] \in \textsc{Rel}(C)$ and therefore $U \circ (T_{i,1}, \ldots, T_{i,k}) = L_{i,1} \times \cdots \times L_{i,k}$. Then, by closure under union and componentwise letter-to-letter relations (Lemma 8), it follows that $R = \bigcup_{i \in I} U \circ (T_{i,1}, \ldots, T_{i,k}) \in \textsc{Rel}(C)$. ◀

2.2 Universal closure properties

There are some closure properties which are shared by *all* classes of synchronized relations, that is, by every $\textsc{Rel}(C)$ with $C \subseteq_{reg} \mathbb{k}^*$. We highlight the most salient ones.

An *alphabetic morphism* between two finite alphabets $\mathbb{A}, \mathbb{B}$ is a morphism $h : \mathbb{A}^* \to \mathbb{B}^*$ between the free monoids such that $h(a) \in \mathbb{B}$ for every $a \in \mathbb{A}$. Its application is extended to any relation $R \subseteq (\mathbb{A}^*)^k$ as follows $h(R) = \{(h(u_1), \ldots, h(u_k)) \colon (u_1, \ldots, u_k) \in R\} \subseteq (\mathbb{B}^*)^k$; and its *inverse* is applied to $S \subseteq (\mathbb{B}^*)^k$ as $h^{-1}(S) = \{(u_1, \ldots, u_k) \colon (h(u_1), \ldots, h(u_k)) \in S\} \subseteq (\mathbb{A}^*)^k$. A *letter-to-letter* relation is one from $\textsc{Rel}((12)^*)$.

We define the following closure properties over classes $\mathcal{C}$ of k-ary relations.

- $\mathcal{C}$ is closed under union if for all $(R, \mathbb{A}), (S, \mathbb{A}) \in \mathcal{C}$, $(R \cup S, \mathbb{A}) \in \mathcal{C}$;
- $\mathcal{C}$ is closed under (inverse) alphabetic morphisms if for all $(R, \mathbb{A}) \in \mathcal{C}$ and $h : \mathbb{A}^* \to \mathbb{B}^*$ (resp. $g : \mathbb{B}^* \to \mathbb{A}^*$) an alphabetic morphism, $(h(R), \mathbb{B}) \in \mathcal{C}$ (resp. $(g^{-1}(R), \mathbb{B}) \in \mathcal{C}$);
- $\mathcal{C}$ is closed under componentwise letter-to-letter relations if for every $(R, \mathbb{A}) \in \mathcal{C}$ and $(T_1, \mathbb{A}), \ldots, (T_k, \mathbb{A}) \in \textsc{Rel}((12)^*)$ the following relation over the alphabet $\mathbb{A}$ is also in $\mathcal{C}$: $R \circ (T_1, \ldots, T_k) \stackrel{\text{def}}{=} \{(u_1, \ldots, u_k) \colon$ there is $(v_1, \ldots, v_k) \in R$ s.t. $(v_i, u_i) \in T_i$ for every $i\}$.
- $\mathcal{C}$ is closed under recognizable projections if for all $(R, \mathbb{A}) \in \mathcal{C}$ and $(S, \mathbb{A}) \in \textsc{Rel}(1^* \cdots k^*)$, $(R \cap S, \mathbb{A}) \in \mathcal{C}$.

▶ **Lemma 8.** *For every $k \in \mathbb{N}$ and $C \subseteq_{reg} \mathbb{k}^*$, $\textsc{Rel}(C)$ is closed under union, alphabetic morphisms, inverse alphabetic morphisms, componentwise letter-to-letter relations, and recognizable projections.*

Proof. Closure under union follows from the fact that if $L, L' \subseteq_{reg} C \otimes \mathbb{A}^*$, then $L \cup L' \subseteq_{reg} C \otimes \mathbb{A}^*$ and $[\![L]\!] \cup [\![L']\!] = [\![L \cup L']\!]$ (Lemma 6). Closure under letter-to-letter relations follows from the fact that, given $L \subseteq_{reg} C \otimes \mathbb{A}^*$ and k letter-to-letter relations $T_1, \ldots, T_k$ over $\mathbb{A}$, there exists $L' \subseteq_{reg} C \otimes \mathbb{A}^*$ such that $[\![L']\!] = [\![L]\!] \circ (T_1, \ldots, T_k)$ (one can build an automaton recognizing such language from the automata for $L, T_1, \ldots, T_k$). Since any (inverse) alphabetic morphism can be implemented as a letter-to-letter relation, it follows that $\textsc{Rel}(C)$ is closed under (inverse) alphabetic morphisms. Finally, closure under recognizable projections follows from closure under letter-to-letter relations and closure under union, since for every $R \in \textsc{Rel}(C)$ and $S = \bigcup_{i \in I} L_{i,1} \times \cdots \times L_{i,k} \in \textsc{Rel}(1^* \cdots k^*)$ (recall Claim 7) we have that $R \cap S = \bigcup_{i \in I} R \circ (T_{i,1}, \ldots, T_{i,k})$ for $T_{i,j} = \{(w,w) \colon w \in L_{i,j}\}$. ◀

3 Closure under intersection

We say that a class $\mathcal{C}$ of k-ary relations is *closed under intersection* if for all $(R, \mathbb{A}), (S, \mathbb{A}) \in \mathcal{C}$, $(R \cap S, \mathbb{A}) \in \mathcal{C}$. In this section we show a decidable characterization of the languages $C \subseteq_{reg} \mathbb{2}^*$ for which $\textsc{Rel}(C)$ is closed under intersection. Further, for $C \subseteq_{reg} \mathbb{2}^*$, if $\textsc{Rel}(C)$ is closed under intersection, it is *effectively* closed, that is, for every $R, S \in \textsc{Rel}(C)$ over an alphabet

$\mathbb{A}$, one can compute $R \cap S$ as a synchronized relation, that is, as some $L \subseteq_{reg} (\mathbb{2} \times \mathbb{A})^*$ so that $[\![L]\!] = R \cap S$. The main result is the following.

▶ **Theorem 9.** *For every $C \subseteq_{reg} \mathbb{2}^*$, $\textsc{Rel}(C)$ is closed under intersection if, and only if, $C \subseteq_{\textsc{Rel}} D$ for some Parikh-injective $D \subseteq_{reg} \mathbb{2}^*$.*

At the end of this section we give an effective procedure to decide, given $C \subseteq_{reg} \mathbb{2}^*$, whether $\textsc{Rel}(C)$ is closed under intersection. Decidability can be seen as the fact that the set of languages $C \subseteq_{reg} \mathbb{2}^*$ for which there is a Parikh-injective language $D \subseteq_{reg} \mathbb{2}^*$ such that $C \subseteq_{\textsc{Rel}} D$ is both computably enumerable and co-computably enumerable. While showing that it is c.e. is straightforward, proving co-c.e. involves all the developments of this section. Concretely, we define some *bad conditions* that characterize all languages C such that $\textsc{Rel}(C)$ is *not* closed under intersection, and in this way we obtain that the set of languages $C \subseteq_{reg} \mathbb{2}^*$ which satisfy any of the bad conditions is c.e.

We will start by giving a sufficient condition for $\textsc{Rel}(C)$ to be closed under intersection. The following simple lemma (which was already proved in [19]) follows from Lemma 6.

▶ **Lemma 10.** *If $C \subseteq_{reg} \mathbb{2}^*$ is Parikh-injective, then $\textsc{Rel}(C)$ is closed under intersection.*

This lemma implies that any language which is $\textsc{Rel}$-equivalent to a Parikh-injective one gives rise to a closed under intersection class. A natural question is whether the converse holds but it doesn't seem to. For instance, if $C = 1^*2^* \cup (12)^*$, $\textsc{Rel}(C)$ is closed under intersection but it seems unlikely that C is $\textsc{Rel}$-equivalent to a Parikh-injective language.

Another sufficient condition for $\textsc{Rel}(C)$ to be closed under intersection is that $C =_{\textsc{Rel}} D \cup X$ for some Parikh-injective $D, X \subseteq_{reg} \mathbb{2}^*$ such that $X \subseteq_{\textsc{Rel}} 1^*2^*$ (in fact, it can be seen that injectivity of X is not really necessary). We will prove that this condition is also necessary, and thus we will have another characterization of closure under intersection. This is not obvious and we will prove a stronger statement, which we present below (Theorem 12). Also, in particular, we will show that if $\textsc{Rel}(C)$ is closed under intersection, we can compute a Parikh-injective $D \subseteq_{reg} \mathbb{2}^*$ such that $C \subseteq_{\textsc{Rel}} D$, which allows us in turn to compute the intersection of two relations in $\textsc{Rel}(C)$ as a synchronized relation.

For $C \subseteq_{reg} \mathbb{2}^*$, we denote by $\textsc{Rel}(C)^\cap$ the closure under intersection of $\textsc{Rel}(C)$, *i.e.*, the smallest class of relations containing $\textsc{Rel}(C)$ and being closed under intersection. We present three properties on $C \subseteq_{reg} \mathbb{2}^*$ that we call the *bad conditions*, which will characterize the languages such that $\textsc{Rel}(C)$ is *not* closed under intersection.

Bad conditions

For $C \subseteq_{reg} \mathbb{2}^*$, consider the following properties:

(A) There exist $u_1, u_2, v, z \in \mathbb{2}^*$ such that

1. u_i and v are Parikh-independent for $i = 1, 2$,
2. $\pi(u_i) \geq (1, 1)$ for some i,
3. $\{u_1, u_2\}$ and $\{v\}$ are Parikh-dependent,
4. $u_1^* u_2^* z \subseteq_{\textsc{Rel}} C$ and $v^* z \subseteq_{\textsc{Rel}} C$.

(B) There exist $u, v, z \in \mathbb{2}^*$ such that

1. u and v are Parikh-independent,
2. $\pi(u) \geq (1, 1)$ or $\pi(v) \geq (1, 1)$,
3. $u^* v^* z \subseteq_{\textsc{Rel}} C$ and $v^* u^* z \subseteq_{\textsc{Rel}} C$.

(C) There exist $u, v, w, z \in \mathbb{2}^*$ such that

1. $u \in 1^* \setminus \{\varepsilon\}$, $w \in 2^* \setminus \{\varepsilon\}$,
2. $\pi(v) \geq (1, 1)$,
3. $u^* v^* w^* z \subseteq_{\text{REL}} C$ or $w^* v^* u^* z \subseteq_{\text{REL}} C$.

For example, $1^*(12)^*(122)^*$ satisfies A for $u_1 = 1$, $u_2 = 122$, $v = 12$, $z = \varepsilon$; $1^*(12)^*1^*$ satisfies B for $u = 1$, $v = 12$, $z = \varepsilon$; and $1^*(12)^*2^*$ satisfies C for $u = 1$, $v = 12$, $w = 2$, $z = \varepsilon$.

▷ Observation 11. The bad conditions are $\subseteq_{\text{REL}}$-upward closed: If $C \subseteq_{\text{REL}} D$ and C satisfies property A (resp. B, C), then D also satisfies property A (resp. B, C).

We can now present the characterization theorem.

▶ **Theorem 12.** *For $C \subseteq_{reg} \mathbb{2}^*$, the following are equivalent:*

1. $\text{REL}(C)$ *is closed under intersection (i.e.* $\text{REL}(C)^{\cap} = \text{REL}(C)$*);*
2. $\text{REL}(C)^{\cap}$ *is definable (i.e. there exists $D \subseteq_{reg} \mathbb{2}^*$ such that* $\text{REL}(C)^{\cap} = \text{REL}(D)$*);*
3. $\text{REL}(C)^{\cap} \subseteq$ *Rational;*
4. *for all $R, S \in \text{REL}(C)$, $R \cap S \in$ Rational;*
5. *C does not satisfy any of the bad conditions;*
6. *there exist $D, X \subseteq_{reg} \mathbb{2}^*$ Parikh-injective such that $C =_{\text{REL}} D \cup X$ and $X \subseteq_{\text{REL}} 1^*2^*$;*
7. *there exists $D \subseteq_{reg} \mathbb{2}^*$ Parikh-injective such that $C \subseteq_{\text{REL}} D$.*

From $1 \Leftrightarrow 7$ and transitivity of $\subseteq_{\text{REL}}$, closure under intersection is $\subseteq_{\text{REL}}$-downward closed:

▶ **Corollary 13.** *For $C, D \subseteq_{reg} \mathbb{2}^*$, if $C \subseteq_{\text{REL}} D$ and $\text{REL}(D)$ is closed under intersection, then $\text{REL}(C)$ is closed under intersection.*

We first explain the main proof strategy for obtaining Theorem 12, and present the three key technical results we will need to prove (Propositions 14, 16 and 17).

Proof idea of Theorem 12

The proof strategy is by showing $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 6 \Rightarrow 1$ on the one hand, and $6 \Rightarrow 7 \Rightarrow 3$ on the other hand. First observe that $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$ are trivial. We next prove $6 \Rightarrow 1$, $7 \Rightarrow 3$ and $6 \Rightarrow 7$.

For $6 \Rightarrow 1$, suppose that $C =_{\text{REL}} D \cup X$ for some Parikh-injective languages D, X such that $X \subseteq_{\text{REL}} 1^*2^*$. Let $R, S \in \text{REL}(C)$. Then, by item 3 of Lemma 4, there exist $R_1, S_1 \in \text{REL}(D)$, $R_2, S_2 \in \text{REL}(X)$ such that $R = R_1 \cup R_2$ and $S = S_1 \cup S_2$. Note that:

- $R_1 \cap S_1 \in \text{REL}(D) \subseteq \text{REL}(C)$ by Lemma 10 applied to D;
- $R_2 \cap S_2 \in \text{REL}(X) \subseteq \text{REL}(C)$ by Lemma 10 applied to X; and
- $R_1 \cap S_2, R_2 \cap S_1 \in \text{REL}(D)$ by closure under recognizable projections (Lemma 8).

It only remains to observe that $R \cap S = (R_1 \cap S_1) \cup (R_1 \cap S_2) \cup (R_2 \cap S_1) \cup (R_2 \cap S_2)$ and obtain that $R \cap S \in \text{REL}(C)$ due to closure under union (Lemma 8).

On the other hand, $7 \Rightarrow 3$ can be derived from $1 \Rightarrow 3$. Indeed, suppose that $C \subseteq_{\text{REL}} D$ for some Parikh-injective language D. By Lemma 10, $\text{REL}(D)$ is closed under intersection and so, by $1 \Rightarrow 3$, $\text{REL}(C)^{\cap} \subseteq \text{REL}(D)^{\cap} \subseteq$ *Rational*.

For $6 \Rightarrow 7$, suppose that $C =_{\text{REL}} D \cup X$ for some Parikh-injective languages D, X such that $X \subseteq_{\text{REL}} 1^*2^*$. By Lemma 1, closure under complement of semi-linear sets and Parikh's Theorem [27], it follows that there exists $\hat{D} \subseteq_{reg} \mathbb{2}^*$ Parikh-injective such that $\pi(\hat{D}) = \mathbb{N}^2 \setminus \pi(D)$. Note that $D \cup \hat{D}$ is Parikh-bijective. Since $D \cup \hat{D}$ is Parikh-surjective, by

Lemma 6, item 5, $X \subseteq_{\mathrm{REL}} 1^*2^* \subseteq_{\mathrm{REL}} D \cup \hat{D}$ and so, by Lemma 4, item 3 plus closure under union of $D \cup \hat{D}$, we have $C =_{\mathrm{REL}} D \cup X \subseteq_{\mathrm{REL}} D \cup \hat{D}$.

The main difficulty will lie on the proofs of $4 \Rightarrow 5$ and $5 \Rightarrow 6$. For $4 \Rightarrow 5$, we will prove the contrapositive statement:

▶ **Proposition 14.** *If $C \subseteq_{reg} \mathbb{2}^*$ satisfies any of the bad conditions, then there exist $R, S \in$ $\mathrm{REL}(C)$ such that $R \cap S \notin$ Rational.*

To prove $5 \Rightarrow 6$, we define some basic regular languages over $\mathbb{2}$ that we call *basic injective*. A language $C \subseteq \mathbb{2}^*$ is *basic injective* if it can be expressed as u^*v^*z for $u, v, z \in \mathbb{2}^*$ such that if $u, v \neq \varepsilon$, then u and v are Parikh-independent. In particular this implies the following.

▶ **Lemma 15.** *Every basic injective language is Parikh-injective.*

Proof. Let $C = u^*v^*z$ be basic injective. The cases in which u and/or v are empty are straightforward. We will then assume that u and v are Parikh-independent. Suppose then that $\pi(u^r v^s z) = \pi(u^{r'} v^{s'} z)$ for some $r, s, r', s' \in \mathbb{N}$. By Observation 2, $r = r'$ or $s = s'$ which concludes the proof. ◀

Note that singleton sets and languages of the form u^*z for u an arbitrary word are basic injective. The interest of basic injective languages stems from the fact that we can prove the following two results, from which it is not hard to get $5 \Rightarrow 6$.

▶ **Proposition 16.** *If $C \subseteq_{reg} \mathbb{2}^*$ does not satisfy any of the bad conditions, then C is REL-equivalent to a finite union of basic injective languages.*

▶ **Proposition 17.** *If C is a finite union of basic injective languages that are not REL-contained in 1^*2^* and C does not satisfy any of the bad conditions, then C is REL-equivalent to a Parikh-injective regular language.*

To show $5 \Rightarrow 6$ from the two statements above, suppose that C does not satisfy any of the bad conditions. By Proposition 16, $C =_{\mathrm{REL}} X' \cup D'$, for $X' = \bigcup_{i \in I} X_i$ and $D' = \bigcup_{j \in J} D_j$, where I, J are finite, for every $i \in I$, X_i is basic injective and $X_i \subseteq_{\mathrm{REL}} 1^*2^*$, and for every $j \in J$, D_j is basic injective and $D_j \not\subseteq_{\mathrm{REL}} 1^*2^*$. Note that, from the definition of basic injective plus [14, Proposition 7] plus basic properties from Lemma 4, it follows readily that for each $i \in I$, there exist $\ell, s, \ell', s' \in \mathbb{N}$ such that X_i is REL-equivalent to $1^{\ell *} 1^{\ell'} 2^{s *} 2^{s'}$. Therefore X' is REL-equivalent to a Parikh-injective language X such that $X \subseteq_{\mathrm{REL}} 1^*2^*$. On the other hand, by Observation 11, since $D' \subseteq_{\mathrm{REL}} C$ and C does not satisfy any of the bad condition, neither does D'. Hence, by Proposition 17, D' is REL-equivalent to a Parikh-injective regular language D. Thus $C =_{\mathrm{REL}} X \cup D$ which concludes the proof.

We dedicate the rest of the section to prove Propositions 14, 16 and 17.

Proof idea of Proposition 14

We show the proof idea for condition A. The other two conditions follow a similar proof strategy. Suppose that condition A holds, and consider the 3-letter alphabet $\mathbb{A} = \{a_1, a_2, c\}$. Let R, S be the following relations in $(\mathbb{A}^*)^2$:

$$R = [\![(u_1^* \otimes a_1^*) \cdot (u_2^* \otimes a_2^*) \cdot z \otimes c^*]\!], \qquad S = [\![(v^* \otimes \{a_1, a_2\}^*) \cdot z \otimes c^*]\!],$$

note that $R, S \in \mathrm{REL}(C)$ by condition A.4. It is not hard to show that $|R \cap S| = \infty$ due to condition A.3. We show that $R \cap S \notin$ *Rational*. By means of contradiction, suppose there is an automaton over the alphabet $\mathbb{2} \times \mathbb{A}$ such that the language recognized by this automaton synchronizes $R \cap S$. Since the language is infinite, there is a non-trivial cycle $q_0 \xrightarrow{w_1} q \xrightarrow{w_2} q \xrightarrow{w_3} q_f$ inside some accepting run. By a pumping argument, it can be seen that: 1) $[\![w_2]\!]$ is necessarily of the form (a_i^s, a_i^t) for some i, s, t partly due to A.2; 2) $(s, t) \in \langle\{\pi(u_j)\}\rangle$ for some j; and 3) $(s, t) \in \langle\{\pi(v)\}\rangle$. Since 2) plus 3) are in contradiction with A.1, it follows that $R \cap S \notin$ *Rational*. ◀

Proof idea of Proposition 16

It can be seen that one can reduce to the case in which C is of the form $w_1^* \cdots w_n^* z$ with w_i and w_{i+1} Parikh-independent for all $i = 1, \ldots, n-1$. For this kind of languages, if $n \leq 2$ the result follows trivially since they are already basic injective. A straightforward case inspection shows that if $n \geq 3$ then at least one of the bad conditions holds. ◀

Proof idea of Proposition 17

In order to prove Proposition 17 we show the following stronger statement, which gives a characterization of closure under intersection based on the decomposition into basic injective languages. We denote the *commutative closure* of a language $C \subseteq_{reg} \mathbb{2}^*$ by $[C]_\pi = \{w \in \mathbb{2}^* : \pi(w) \in \pi(C)\}$.

▶ **Lemma 18.** *Given a finite set of basic injective languages $\{B_i\}$ that are not* $\textsc{Rel}$*-contained in 1^*2^*, the following are equivalent:*

1. $\textsc{Rel}(\bigcup_i B_i)$ *is closed under intersection;*
2. *for all $R, S \in \textsc{Rel}(\bigcup_i B_i)$, $R \cap S \in$ Rational;*
3. $\bigcup_i B_i$ *does not satisfy any of the bad conditions;*
4. *for every i, j $B_i \cup B_j$ does not satisfy any of the bad conditions;*
5. *for every i, j, $B_i \cap [B_j]_\pi$ is regular and $B_i \cap [B_j]_\pi \subseteq_{\textsc{Rel}} B_j$;*
6. $\bigcup_i B_i =_{\textsc{Rel}} C$ *for some Parikh-injective $C \subseteq_{reg} \mathbb{2}^*$.*

Proposition 17 follows from Lemma 18 since it is its implication $3 \Rightarrow 6$. In order to give a proof for Lemma 18, we first define the following property, which is at the core of the next lemmas. A pair of languages B_1, B_2, is said to verify the *dichotomy property* if either

- $B_1 \cup B_2$ satisfies one of the bad conditions; or
- $B_1 \cap [B_2]_\pi$ is regular and $B_1 \cap [B_2]_\pi \subseteq_{\textsc{Rel}} B_2$.

Note that $B_1 \cap [B_2]_\pi$ may not be regular in general, for example if $B_1 = 1^*2^*$ and $B_2 = (12)^*$. The main ingredient to prove Lemma 18 is given by the following statement.

▶ **Lemma 19.** *Every pair of basic injective languages B_1, B_2 such that $B_1, B_2 \not\subseteq_{\textsc{Rel}} 1^*2^*$ satisfies the dichotomy property.*

Proof of Lemma 18. $1 \Rightarrow 2$ is trivial; $2 \Rightarrow 3$ follows from the contrapositive of Proposition 14; $3 \Rightarrow 4$ holds by Observation 11; and $4 \Rightarrow 5$ follows from Lemma 19. For $5 \Rightarrow 6$, we proceed by induction on the number of basic injective languages in $\{B_i\}$. The base case is the empty language, which is (vacuously) Parikh-injective. For the inductive step, consider a union $B \cup \bigcup_i B_i$. First observe that, by Lemma 15, B is Parikh-injective. By inductive hypothesis, there exists $D \subseteq_{reg} \mathbb{2}^*$ Parikh-injective such that $\bigcup_i B_i =_{\textsc{Rel}} D$. Also, since $B \cap [\bigcup_i B_i]_\pi = \bigcup_i B \cap [B_i]_\pi$, by hypothesis both $B \cap [\bigcup_i B_i]_\pi$ and $B \setminus [\bigcup_i B_i]_\pi$ are regular, and $B \cap [\bigcup_i B_i]_\pi \subseteq_{\textsc{Rel}} \bigcup_i B_i$. Now it only remains to observe that $(B \setminus [\bigcup_i B_i]_\pi) \cup D$ is Parikh-injective and $\textsc{Rel}$-equivalent to $B \cup \bigcup_i B_i$. Finally, $6 \Rightarrow 1$ follows from Lemma 10. ◀

Decidability

We finally discuss briefly the decidability procedure to test whether a class $\textsc{Rel}(C)$ is closed under intersection.

▶ **Proposition 20.** *It is decidable wether a given $C \subseteq_{reg} \mathbb{2}^*$ is such that $\textsc{Rel}(C)$ is closed under intersection.*

Proof idea. It follows by the equivalence $1 \Leftrightarrow 5 \Leftrightarrow 7$ of Theorem 12, together with the fact that the set of languages $C \subseteq_{reg} \mathbb{2}^*$ for which there is a Parikh-injective language $D \subseteq_{reg} \mathbb{2}^*$ such that $C \subseteq_{\textsc{Rel}} D$ is computably enumerable; and the fact that the set of languages $C \subseteq_{reg} \mathbb{2}^*$ which satisfy any of the bad conditions is computably enumerable. ◀

Note that whenever $\textsc{Rel}(C)$ is closed under intersection, it is effectively so: given $L_1, L_2 \subseteq_{reg} C \otimes \mathbb{A}^*$ it is possible to compute $L \subseteq_{reg} (\mathbb{2} \times \mathbb{A})^*$ with $[\![L]\!] = [\![L_1]\!] \cap [\![L_2]\!]$. Indeed, by the previous proposition we can compute some Parikh-injective D such that $C \subseteq_{\textsc{Rel}} D$. By the results of [14], one can compute $L'_1, L'_2 \subseteq_{reg} D \otimes \mathbb{A}^*$ such that $[\![L'_1]\!] = [\![L_1]\!]$ and $[\![L'_2]\!] = [\![L_2]\!]$; and thus $L = L'_1 \cap L'_2$ is such that $[\![L]\!] = [\![L_1]\!] \cap [\![L_2]\!]$ due to injectivity of D and Lemma 6.

4 Closure under complement

We say that a class $\mathcal{C}$ of k-ary relations is *closed under complement* if for every $(R, \mathbb{A}) \in \mathcal{C}$, $((\mathbb{A}^*)^k \setminus R, \mathbb{A}) \in \mathcal{C}$. For every $\textsc{Rel}_k(C)$ and alphabet $\mathbb{A}$, note that there is a unique largest relation $(U, \mathbb{A}) \in \textsc{Rel}_k(C)$ that contains all relations $(R, \mathbb{A}) \in \textsc{Rel}_k(C)$; this is $U = [\![C \otimes \mathbb{A}^*]\!]_k$. Thus, a natural alternative definition for complement could take U, instead of $(\mathbb{A}^*)^k$, as the universe. We say that $\textsc{Rel}_k(C)$ is *closed under relativized complement* if for all $(R, \mathbb{A}) \in \textsc{Rel}_k(C)$ we have $([\![C \otimes \mathbb{A}^*]\!]_k \setminus R, \mathbb{A}) \in \textsc{Rel}_k(C)$. In this section, we give effective characterizations of the languages $C \subseteq_{reg} \mathbb{2}^*$ for which $\textsc{Rel}(C)$ is closed under complement and relativized complement.

Relativized complement

We show that closure under relativized complement, perhaps surprisingly, is equivalent to closure under intersection, and therefore it is decidable whether $\textsc{Rel}(C)$ is closed under relativized complement for a given $C \subseteq_{reg} \mathbb{2}^*$.

▶ **Proposition 21.** *For $C \subseteq_{reg} \mathbb{2}^*$, $\textsc{Rel}(C)$ is closed under relativized complement if, and only if, $\textsc{Rel}(C)$ is closed under intersection.*

Proof. For the left-to-right direction, let $(R, \mathbb{A}), (S, \mathbb{A}) \in \textsc{Rel}(C)$. Recall that $\textsc{Rel}(C)$ is always closed under union and note that $R \cap S = [\![C \otimes \mathbb{A}^*]\!] \setminus (([\![C \otimes \mathbb{A}^*]\!] \setminus R) \cup ([\![C \otimes \mathbb{A}^*]\!] \setminus S))$, and therefore $(R \cap S, \mathbb{A}) \in \textsc{Rel}(C)$. For the right-to-left direction, let $L \subseteq_{reg} C \otimes \mathbb{A}^*$. We want to check that $[\![C \otimes \mathbb{A}^*]\!] \setminus [\![L]\!] \in \textsc{Rel}(C)$. By the characterization of the previous section (Theorem 12, implication $1 \Rightarrow 6$) we can assume that $C = D \cup X$, for $X \subseteq_{\textsc{Rel}} 1^*2^*$ and X, D Parikh-injective. Then,

$$\begin{aligned}
[\![C \otimes \mathbb{A}^*]\!] \setminus [\![L]\!] &= [\![(D \cup X) \otimes \mathbb{A}^*]\!] \setminus [\![L]\!] = ([\![(D \otimes \mathbb{A}^*) \cup (X \otimes \mathbb{A}^*)]\!]) \setminus [\![L]\!] \\
&= ([\![D \otimes \mathbb{A}^*]\!] \cup [\![X \otimes \mathbb{A}^*]\!]) \setminus [\![L]\!] = ([\![D \otimes \mathbb{A}^*]\!] \setminus [\![L]\!]) \cup ([\![X \otimes \mathbb{A}^*]\!] \setminus [\![L]\!]) \\
&= \underbrace{[\![(D \otimes \mathbb{A}^*) \setminus L]\!]}_{R} \cup \underbrace{[\![X \otimes \mathbb{A}^* \setminus L]\!]}_{S}. && \text{(by Lemma 6, item 3)}
\end{aligned}$$

Since $R, S \in \textsc{Rel}(C)$, by Lemma 8, $R \cup S \in \textsc{Rel}(C)$, and thus $[\![C \otimes \mathbb{A}^*]\!] \setminus [\![L]\!] \in \textsc{Rel}(C)$. ◀

Note that if $C \subseteq \mathbb{2}^*$ is Parikh-surjective, then $[\![C \otimes \mathbb{A}^*]\!] = (\mathbb{A}^*)^2$, and hence closure under relativized complement and closure under complement coincide. Thus, by Proposition 21:

▷ Observation 22. If $C \subseteq \mathbb{2}^*$ is Parikh-surjective, then $\text{Rel}(C)$ is closed under complement if, and only if, $\text{Rel}(C)$ is closed under intersection.

Complement

Let $\text{Rel}(C)^c$ be the closure under complement of $\text{Rel}(C)$, *i.e.*, the smallest class closed under complement containing $\text{Rel}(C)$. The following lemma gives sufficient conditions for our characterization.

▶ **Lemma 23.** *For any $C \subseteq_{reg} \mathbb{2}^*$,*
1. *if C is Parikh-bijective, then $\text{Rel}(C)$ is closed under complement;*
2. *if $\text{Rel}(C)$ is closed under complement, then C is Parikh-surjective.*

Proof. For item 1, let $L \subseteq_{reg} C \otimes \mathbb{A}^*$. By item 4 of Lemma 6, $(\mathbb{A}^*)^2 \setminus [\![L]\!] = [\![C \otimes \mathbb{A}^* \setminus L]\!] \in \text{Rel}(C)$ which concludes the proof.

For item 2, let $L = C \otimes \{a\}^*$. Then $((\{a\}^*)^2 \setminus [\![L]\!], \{a\}) \in \text{Rel}(C)$ and so there exists $L' \subseteq_{reg} C \otimes \{a\}^*$ such that $[\![L']\!] = (\{a\}^*)^2 \setminus [\![L]\!]$. Then $[\![L \cup L']\!] = [\![L]\!] \cup [\![L']\!] = (\{a\}^*)^2$. Therefore, the Parikh-image of the projection of $L \cup L'$ onto the first component must be $\mathbb{N}^k$ and so C is Parikh-surjective since both L and L' (and hence $L \cup L'$) are $\subseteq_{reg} C \otimes \{a\}^*$. ◀

From Lemma 23 plus Observation 22, we have that $\text{Rel}(C)$ is closed under complement if, and only if, $\text{Rel}(C)$ is closed under intersection and C is Parikh-surjective. At the end of this section, we will use this to prove that closure under complement is a decidable property.

We now give a characterization for closure under complement without referring to closure under intersection.

▶ **Theorem 24.** *For $C \subseteq_{reg} \mathbb{2}^*$, the following are equivalent:*
1. *there exists $D \subseteq_{reg} \mathbb{2}^*$ Parikh-bijective such that $C =_{\text{Rel}} D$;*
2. *$\text{Rel}(C)$ is closed under complement (i.e. $\text{Rel}(C)^c = \text{Rel}(C)$);*
3. *$\text{Rel}(C)^c$ is definable (i.e. there is $D \subseteq_{reg} \mathbb{2}^*$ such that $\text{Rel}(C)^c = \text{Rel}(D)$).*

Before proving the above theorem, we observe that we cannot obtain the third and fourth equivalent statements that we have in Theorem 12.

▶ **Lemma 25.** *There is $C \subseteq_{reg} \mathbb{2}^*$ with $\text{Rel}(C)^c \subseteq$ Rational but $\text{Rel}(C)^c$ not definable.*

Proof. Consider any language which is Parikh-injective but not Parikh-surjective, *e.g.* $C = (12)^*$. Then, by item 2 of Lemma 23, plus Theorem 24, we have that $\text{Rel}(C)^c$ is not definable. The result is then an immediate consequence of the following:

▷ Claim. If $C \subseteq_{reg} \mathbb{2}^*$ is Parikh-injective, then $\text{Rel}(C)^c \subseteq$ *Rational*.

Indeed, by Parikh's Theorem [27], $\pi(C)$ is a semi-linear set and then so is $\mathbb{N}^2 \setminus \pi(C)$ (see for example [22]). By Lemma 1, $\mathbb{N}^2 \setminus \pi(C) = \pi(D)$ for some Parikh-injective language D. It follows then that $C \cup D$ is Parikh-bijective and so, by Lemma 23, item 1, $\text{Rel}(C \cup D)$ is closed under complement. Then $\text{Rel}(C)^c \subseteq \text{Rel}(C \cup D) \subseteq$ *Rational*. ◀

Proof of Theorem 24. 1 ⇒ 2 follows from item 1 of Lemma 23; and 2 ⇒ 3 is trivial.

2 ⇒ 1: Suppose that $\text{Rel}(C)$ is closed under complement. By Lemma 8, $\text{Rel}(C)$ is also closed under union and so under intersection. Therefore, by Theorem 12, there

exist Parikh-injective languages $D, X \subseteq_{reg} \mathbb{2}^*$ such that $X \subseteq_{\text{REL}} 1^*2^*$ and $C =_{\text{REL}} D \cup X$. It follows then that $[\![D \otimes \mathbb{A}^*]\!] \in \text{REL}(C)$ and so $R = (\mathbb{A}^*)^2 \setminus [\![D \otimes \mathbb{A}^*]\!] \in \text{REL}(C)$. Let $L \subseteq_{reg} (D \cup X) \otimes \mathbb{A}^*$ be such that $[\![L]\!] = R$. By definition of R, we get that $L \subseteq_{reg} X \otimes \mathbb{A}^*$ and so $X' = \{u \colon \exists v \text{ such that } u \otimes v \in L\} \subseteq_{reg} X$. Besides, also by definition of R, $\pi(X') = \mathbb{N}^2 \setminus \pi(D)$ and so $D \cup X'$ is Parikh-bijective. It only remains to observe that $C =_{\text{REL}} D \cup X'$: $\supseteq$ is trivial and $\subseteq$ follows from the fact that 1^*2^* is REL-contained in any Parikh-surjective language (Lemma 6, item 5) and so $X \subseteq_{\text{REL}} D \cup X'$.

$3 \Rightarrow 2$: Let $D \subseteq_{reg} \mathbb{2}^*$ such that $\text{REL}(C)^c = \text{REL}(D)$. Since $\text{REL}(D)$ is closed under complement, by $2 \Rightarrow 1$, we can assume wlog that D is Parikh-bijective. By means of contradiction, suppose that $\text{REL}(C)$ is not closed under complement. Therefore, by Observation 22, either C is not Parikh-surjective or $\text{REL}(C)$ is not closed under intersection. We show that in both cases we arrive to a contradiction. If $\text{REL}(C)$ is not closed under intersection, by Theorem 12 (implication $\neg 1 \Rightarrow \neg 4$), there are $R, S \in \text{REL}(C)$ such that $R \cap S \notin \mathit{Rational}$; but since $R \cap S = (\mathbb{A}^*)^2 \setminus ((\mathbb{A}^*)^2 \setminus R \cup (\mathbb{A}^*)^2 \setminus S) \in \text{REL}(C)^c = \text{REL}(D) \subseteq \mathit{Rational}$ (recall that $\text{REL}(D)$ is closed under union by Lemma 8), we have a contradiction. On the other hand, if C is not Parikh-surjective, there exists $\bar{x} \in \pi(D) \setminus \pi(C)$. Let $u \in D$ be such that $\pi(u) = \bar{x}$ and let us consider the singleton relation $R = \{[\![u \otimes a^{|u|}]\!]\}$. It is clear that $(R, \{a, b\}) \in \text{REL}(D) = \text{REL}(C)^c$. Then, either $(R, \{a, b\})$ or its complement $((\{a, b\}^*)^2 \setminus R, \{a, b\})$ should be in $\text{REL}(C)$. But it is easy to see that both relations contain a tuple with Parikh-image $\bar{x}$: $[\![u \otimes a^{|u|}]\!] \in R$ and $[\![u \otimes b^{|u|}]\!] \in (\{a, b\}^*)^2 \setminus R$. Since $\bar{x} \notin \pi(C)$, none of the relations is in $\text{REL}(C)$, which is a contradiction. ◀

Decidability

From Observation 22 and item 2 of Lemma 23, decidability of closure under complement follows immediately: $\text{REL}(C)$ is closed under complement if, and only if, $\text{REL}(C)$ is closed under intersection and C is Parikh-surjective. The former is decidable due to Proposition 20, and the latter is decidable through Parikh's Theorem, since universality for semi-linear sets is decidable (see, *e.g.*, [22]).

▶ **Proposition 26.** *Given $C \subseteq_{reg} \mathbb{2}^*$, testing whether $\text{REL}(C)$ is closed under complement is decidable.*

5 Closure under concatenation, Kleene star, and projection

In this section, we characterize languages $C \subseteq_{reg} \mathbb{k}^*$ such that $\text{REL}(C)$ is closed under concatenation, Kleene star, and projection.

- $\mathcal{C}$ is *closed under concatenation* if for all $R, S \in \mathcal{C}$, $R \cdot S \in \mathcal{C}$, where $\cdot$ is the component-wise concatenation operation (*e.g.*, $\{(a, ab), (b, a)\} \cdot \{(b, c)\} = \{(ab, abc), (bb, ac)\}$);
- $\mathcal{C}$ is *closed under Kleene star* if for all $R \in \mathcal{C}$, $R^* \in \mathcal{C}$ for $R^* = \bigcup_{i \in \mathbb{N}} R^{(i)}$, where $R^{(0)} = \{(\varepsilon, \ldots, \varepsilon)\}$, and $R^{(i+1)} = R \cdot R^{(i)}$;
- $\mathcal{C}$ is *closed under projection* if for all $(R, \mathbb{A}) \in \mathcal{C}$ and $K \subseteq \mathbb{k}$, $(R|_K, \mathbb{A}) \in \mathcal{C}$, where $R|_K \subseteq \mathbb{A}^{*k}$ is the projection of R onto the components in K (with ε in the other components). For example, for $R = \{(aa, ab, b), (a, bbb, aab), (aa, ab, ba)\}$ and $K = \{1, 2\}$ we have $R|_K = \{(aa, ab, \varepsilon), (a, bbb, \varepsilon)\}$.

We now give characterizations for closure under concatenation and Kleene star. As we show, closure under concatenation is in fact a necessary condition for closure under Kleene star.

▶ **Proposition 27.** *For every $C, C_1, C_2, C_3 \subseteq_{reg} \mathbb{k}^*$,*

1. *$C_1 \cdot C_2 \subseteq_{\textsc{Rel}} C_3$ iff for every $R_1 \in \textsc{Rel}(C_1)$, $R_2 \in \textsc{Rel}(C_2)$ we have $R_1 \cdot R_2 \in \textsc{Rel}(C_3)$;*
2. *$\textsc{Rel}(C)$ is closed under concatenation iff $C \cdot C \subseteq_{\textsc{Rel}} C$;*
3. *if $\textsc{Rel}(C)$ is closed under Kleene star, then it is closed under concatenation; and*
4. *$\textsc{Rel}(C)$ is closed under Kleene star iff $C^* \subseteq_{\textsc{Rel}} C$.*

Proof sketch. For the left-to-right direction of item 1, let $L_1 \subseteq_{reg} C_1 \otimes \mathbb{A}^*$ and $L_2 \subseteq_{reg} C_2 \otimes \mathbb{A}^*$. Then we only have to observe that $[\![L_1]\!] \cdot [\![L_2]\!] = [\![L_1 \cdot L_2]\!] \in \textsc{Rel}(C_1 \cdot C_2) \subseteq \textsc{Rel}(C_3)$ as we wanted. The right-to-left direction follows from Lemma 8 together with property 1 of Lemma 4. Note that item 2 is a particular case of item 1.

We now turn to item 3. For simplicity assume $k = 2$. Suppose $\textsc{Rel}(C)$ is closed under Kleene star, and take arbitrary $R_1, R_2 \in \textsc{Rel}(C)$ over an alphabet $\mathbb{A}$. Define R'_i over the alphabet $\mathbb{A} \times \{\overline{lst_i}, lst_i\}$ as the result of replacing every pair $(a_1 \cdots a_n, b_1 \cdots b_m) \in R_i$ with $((a_1, \overline{lst_i}) \cdots (a_{n-1}, \overline{lst_i})(a_n, lst_i), (b_1, \overline{lst_i}) \cdots (b_{m-1}, \overline{lst_i})(b_m, lst_i))$. Intuitively, lst_i marks the last symbols of tuples from R_i. It is easy to see that $R'_1, R'_2 \in \textsc{Rel}(C)$ using closure under componentwise letter-to-letter relations. Observe that $R'_1 \cdot R'_2 \subseteq (R'_1 \cup R'_2)^*$ and, by closure under union and Kleene star, that $(R'_1 \cup R'_2)^* \in \textsc{Rel}(C)$. Let $L \subseteq_{reg} C \otimes (\mathbb{A} \times \{\overline{lst}_1, lst_1, \overline{lst}_2, lst_2\})^*$ such that $[\![L]\!] = (R'_1 \cup R'_2)^*$. It is easy to see that there is $L' \subseteq_{reg} L$ such that $[\![L']\!] = R'_1 \cdot R'_2$, and thus that $R'_1 \cdot R'_2 \in \textsc{Rel}(C)$. Again by closure under component-wise letter-to-letter relations we obtain that $R_1 \cdot R_2 \in \textsc{Rel}(C)$, this time using the relation that projects onto the first component.

Finally, we prove item 4. For the right-to-left direction, let $R \in \textsc{Rel}(C)$ and take $L \subseteq_{reg} C \otimes \mathbb{A}^*$ such that $[\![L]\!] = R$. Therefore $R^* = [\![L]\!]^* = [\![L^*]\!] \in \textsc{Rel}(C^*) \subseteq \textsc{Rel}(C)$ as wanted. For the converse, first observe that $\textsc{Rel}(C)$ is also closed under concatenation due to item 3. Let $R \in \textsc{Rel}(C^*)$. By item 2 of Lemma 4, we have the following:

there are $R_1, \dots, R_n \in \textsc{Rel}(C)$ and $I \subseteq_{reg} \{1, \dots, n\}^*$ such that $R = \bigcup_{w \in I} R_{w[1]} \cdots R_{w[|w|]}$.

Consider any regular expression E denoting the language I above, and replace each occurence of $i \in \{1, \dots, n\}$ with R_i, in such a way that the resulting expression E' denotes R. Then, by finite application of closure under Kleene star, concatenation and union as given by E', we obtain that $R \in \textsc{Rel}(C)$. ◀

For $C \subseteq_{reg} \mathbb{k}^*$ and $K \subseteq \mathbb{k}$, let $C|_K$ be the projection of C onto the alphabet K (which is also regular). We give the following characterization of closure under projection.

▶ **Lemma 28.** *For every $k \in \mathbb{N}$ and $C \subseteq_{reg} \mathbb{k}^*$, $\textsc{Rel}_k(C)$ is closed under projection iff $\textsc{Rel}_k(C|_K) \subseteq \textsc{Rel}_k(C)$ for every $K \subseteq \mathbb{k}$.*

Decidability

For the binary case, by previous results [14], it is decidable to test whether a synchronized class is included in another, and thus the characterizations for Kleene star and concatenation are decidable. We leave the general case as an open question.

6 Concluding remarks and future work

We discuss the decidability of paradigmatic problems within $\textsc{Rel}(C)$. First, note that the emptiness problem for relations reduces to the emptiness problem for automata: $[\![L]\!] = \emptyset$ if, and only if, $L = \emptyset$ —and thus the emptiness problem is always decidable. Further, by the results we have shown together with Lemma 6 we obtain the following.

▶ **Lemma 29.** *For $C \subseteq_{reg} \mathbb{2}^*$, if $\textsc{Rel}(C)$ is closed under...*

- *...intersection, then equivalence and containment problems within* $\textsc{Rel}(C)$ *are decidable;*
- *...complement, then the universality problem within* $\textsc{Rel}(C)$ *is decidable;*
- *...*Op*, then the* Op *operation within* $\textsc{Rel}(C)$ *is computable, for* Op $\in$ {*intersection, complement, concatenation, Kleene star, projection*}.

Proof of Lemma 29. Given $L, M \subseteq_{reg} C \otimes \mathbb{A}^*$, the containment problem between $[\![L]\!]$ and $[\![M]\!]$ amounts to checking if $[\![L]\!] \setminus [\![M]\!]$ is empty. Since $\textsc{Rel}(C)$ is closed under intersection, by Theorem 12, there exists a Parikh-injective language D such that $C \subseteq_{\textsc{Rel}} D$. Moreover, our decidability proof, shows that we can effectively compute such language D. Therefore, by the results on [14], we can effectively construct $L', M' \subseteq_{reg} D \otimes \mathbb{A}^*$ such that $[\![L]\!] = [\![L']\!]$, and $[\![M]\!] = [\![M']\!]$. Then, by Lemma 6, item 3, $[\![L]\!] \setminus [\![M]\!] = [\![L']\!] \setminus [\![M']\!] = [\![L' \setminus M']\!]$ and so the containment problem within $\textsc{Rel}(C)$ reduces to the emptiness problem within $\textsc{Rel}(D)$. The equivalence problem obviously reduces to the containment problem.

The universality problem for $([\![L]\!], \mathbb{A})$ amounts to checking whether $(\mathbb{A}^*)^k \setminus [\![L]\!]$ is empty. Since $\textsc{Rel}(C)$ is closed under complement, by Theorem 24, there exists a Parikh-bijective language D such that $C =_{\textsc{Rel}} D$. As before, we can effectively compute such language D, and therefore, by the results on [14], we can effectively construct $L' \subseteq_{reg} D \otimes \mathbb{A}^*$ such that $[\![L']\!] = [\![L]\!]$. By Lemma 6, item 4, we thus obtain $(\mathbb{A}^*)^k \setminus [\![L]\!] = (\mathbb{A}^*)^k \setminus [\![L']\!] = [\![(D \otimes \mathbb{A}^*) \setminus L']\!]$ and so the containment problem within $\textsc{Rel}(C)$ reduces to the emptiness problem within $\textsc{Rel}(D)$.

We prove the last item only for intersection; similar (or simpler) arguments can be used for all the other operations. Given $L, M \subseteq_{reg} C \otimes \mathbb{A}^*$, with a similar argument than the one used in the previous item, we can effectively construct a Parikh-injective language D and $L', M' \subseteq_{reg} D \otimes \mathbb{A}^*$ such that $[\![L]\!] = [\![L']\!]$, and $[\![M]\!] = [\![M']\!]$. Then, by Lemma 6, item 3, $[\![L]\!] \cap [\![M]\!] = [\![L']\!] \cap [\![M']\!] = [\![L' \cap M']\!]$ and the result follows. ◀

One can then conclude that classes of synchronized binary relations are generally "well-behaved": *a*) it is decidable to test whether a class is closed under Boolean connectives; *b*) every synchronized class closed under intersection (resp. complement, etc.), is *effectively* closed under intersection (resp. complement, etc.); *c*) every synchronized class which is closed under Boolean connectives has decidable paradigmatic problems (in the sense of Lemma 29); *d*) at least for the binary case, the characterizations for Kleene star and concatenation are decidable.

We leave as future work the question of whether it is decidable to test if $\textsc{Rel}(C)$ is closed under Kleene star, concatenation and projection when $C \subseteq_{reg} \mathbb{k}$. We also leave open the characterization for closure under complement and intersection for k-ary relations. Although it is conceivable that the same characterization for closure under intersection holds for arbitrary arity relations, we were not able to show it — the main issue is that it is not clear how to generalize the bad conditions to a k-ary alphabet, nor what would be the analog of item 6 in Theorem 12.

▷ Conjecture 30. For every $k \in \mathbb{N}$ and $C \subseteq_{reg} \mathbb{k}^*$, $\textsc{Rel}(C)$ is closed under intersection if, and only if, $C \subseteq_{\textsc{Rel}} D$ for some Parikh-injective $D \subseteq_{reg} \mathbb{k}^*$.

References

1. Parosh Aziz Abdulla, Bengt Jonnson, Marcus Nilsson, and Mayank Saksena. A survey of regular model checking. In *International Conference on Concurrency Theory (CONCUR)*, pages 35–48, 2003.
2. Renzo Angles and Claudio Gutiérrez. Survey of graph database models. *ACM Comput. Surv.*, 40(1), 2008.
3. Kemafor Anyanwu and Amit Sheth. ρ-queries: enabling querying for semantic associations on the semantic web. In *12th International World Wide Web Conference (WWW'03)*, pages 690–699, 2003.
4. Pablo Barceló, Diego Figueira, and Leonid Libkin. Graph logics with rational relations. *Logical Methods in Computer Science (LMCS)*, 9(3:01), 2013. `doi:10.2168/LMCS-9(3:1)2013`.
5. Pablo Barceló, Leonid Libkin, Anthony Widjaja Lin, and Peter T. Wood. Expressive languages for path queries over graph-structured data. *ACM Trans. Database Syst.*, 37(4):31, 2012.
6. Pablo Barceló and Pablo Muñoz. Graph logics with rational relations: The role of word combinatorics. *ACM Trans. Comput. Log.*, 18(2):10:1–10:41, 2017. URL: `https://doi.org/10.1145/3070822`, `doi:10.1145/3070822`.
7. Jean Berstel. *Transductions and Context-Free Languages*. B. G. Teubner, 1979.
8. Achim Blumensath and Erich Grädel. Automatic structures. In *Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 51–62. IEEE Computer Society Press, 2000. `doi:10.1109/LICS.2000.855755`.
9. Mikołaj Bojańczyk. Transducers with origin information. In *International Colloquium on Automata, Languages and Programming (ICALP)*, volume 8573 of *Lecture Notes in Computer Science*, pages 26–37. Springer, 2014. `doi:10.1007/978-3-662-43951-7`.
10. Ahmed Bouajjani, Bengt Jonsson, Marcus Nilsson, and Tayssir Touili. Regular model checking. In *International Conference on Computer Aided Verification (CAV)*, pages 403–418, London, UK, 2000. Springer-Verlag.
11. Julius Richard Büchi. Weak second-order arithmetic and finite automata. *Mathematical Logic Quarterly*, 6(1-6):66–92, 1960.
12. Olivier Carton. The growth ratio of synchronous rational relations is unique. *Theor. Comput. Sci.*, 376(1-2):52–59, 2007. URL: `https://doi.org/10.1016/j.tcs.2007.01.012`, `doi:10.1016/j.tcs.2007.01.012`.
13. Christian Choffrut. Relations over words and logic: A chronology. *Bulletin of the EATCS*, 89:159–163, 2006.
14. María Emilia Descotte, Diego Figueira, and Gabriele Puppis. Resynchronizing classes of word relations. In *International Colloquium on Automata, Languages and Programming (ICALP)*, Leibniz International Proceedings in Informatics (LIPIcs), pages 381:1–381:13. Leibniz-Zentrum für Informatik, 2018. `doi:10.4230/LIPIcs.ICALP.2018.381`.
15. Samuel Eilenberg and Marcel-Paul Schützenberger. Rational sets in commutative monoids. *Journal of Algebra*, 13(2):173–191, 1969.
16. Calvin C. Elgot and Jorge E. Mezei. On relations defined by generalized finite automata. *IBM Journal of Research and Development*, 9(1):47–68, 1965. `doi:10.1147/rd.91.0047`.
17. Ronald Fagin, Benny Kimelfeld, Frederick Reiss, and Stijn Vansummeren. Document Spanners: A formal approach to information extraction. *Journal of the ACM*, 62(2):12:1–12:51, 2015. `doi:10.1145/2699442`.
18. Diego Figueira and Leonid Libkin. Path logics for querying graphs: Combining expressiveness and efficiency. In *Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 329–340. IEEE Computer Society Press, 2015. `doi:10.1109/LICS.2015.39`.
19. Diego Figueira and Leonid Libkin. Synchronizing relations on words. *Theory of Computing Systems*, 57(2):287–318, 2015. `doi:10.1007/s00224-014-9584-2`.
20. Emmanuel Filiot, Ismaël Jecker, Christof Löding, and Sarah Winter. On equivalence and uniformisation problems for finite transducers. In *International Colloquium on Automata, Languages and Programming (ICALP)*, volume 55 of *Leibniz International Proceedings in*

Informatics (LIPIcs), pages 125:1–125:14. Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.ICALP.2016.125.

21 Christiane Frougny and Jacques Sakarovitch. Synchronized rational relations of finite and infinite words. *Theoretical Computer Science*, 108(1):45–82, 1993. doi:10.1016/0304-3975(93)90230-Q.

22 Seymour Ginsburg and Edwin Spanier. Semigroups, Presburger formulas, and languages. *Pacific Journal of Mathematics*, 16(2):285–296, 1966.

23 Ryuichi Ito. Every semilinear set is a finite union of disjoint linear sets. *J. Comput. Syst. Sci.*, 3(2):221–231, 1969. doi:10.1016/S0022-0000(69)80014-0.

24 Bengt Jonsson and Marcus Nilsson. Transitive closures of regular relations for verifying infinite-state systems. In *International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, pages 220–234. Springer, 2000.

25 R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

26 Maurice Nivat. Transduction des langages de Chomsky. *Annales de l'Institut Fourier*, 18:339–455, 1968.

27 Rohit Parikh. On context-free languages. *Journal of the ACM*, 13(4):570–581, 1966. doi:10.1145/321356.321364.

28 Anthony Widjaja To and Leonid Libkin. Algorithmic metatheorems for decidable LTL model checking over infinite systems. In *International Conference on Foundations of Software Science and Computational Structures (FOSSACS)*, pages 221–236, 2010.

29 Sarah Winter. Uniformization problems for synchronizations of automatic relations on words. In *International Colloquium on Automata, Languages and Programming (ICALP)*, Leibniz International Proceedings in Informatics (LIPIcs). Leibniz-Zentrum für Informatik, 2018. URL: http://arxiv.org/abs/1805.02444.

A Missing proofs to Section 2

Before proving Lemma 1, we begin with a known normal form for semilinear sets.

▶ **Lemma 31** ([15, 23]). *For every semilinear set $V \subseteq \mathbb{N}^k$ there exist a finite I, and $\bar{x}_i \in \mathbb{N}^k$, $P_i \subseteq \mathbb{N}^k$ for every $i \in I$ such that $V = \bigcup_{i \in I} \bar{x}_i + \mathbb{N} \cdot P_i$, all P_i's are linearly independent sets, and $(\bar{x}_i + \mathbb{N} \cdot P_i) \cap (\bar{x}_j + \mathbb{N} \cdot P_j) = \emptyset$ for all $i \neq j$.*

▷ Lemma 1. For every semi-linear set $V \subseteq \mathbb{N}^2$ there exists a Parikh-injective language $C \subseteq_{reg} \mathbb{2}^*$ such that $\pi(C) = V$.

Proof. Let us assume that V is in the normal form given by Lemma 31. Note that, since we are in dimension 2, every P_i has at most 2 vectors, namely $P_i = \{\bar{y}_i, \bar{z}_i\}$ (where perhaps $\bar{y}_i = \bar{z}_i$). For every linear set $\bar{x}_i + \mathbb{N} \cdot P_i$, consider words u_i, v_i, w_i such that $\pi(u_i) = \bar{x}_i$, $\pi(v_i) = \bar{y}_i$ and $\pi(w_i) = \bar{z}_i$ (let us assume that $v_i = w_i$ if $\bar{y}_i = \bar{z}_i$). Then it follows that $C = \bigcup_{i \in I} v_i^* w_i^* u_i$ is Parikh-injective and $\pi(C) = V$. ◀

B Missing proofs to Section 3

▷ Lemma 10. If $C \subseteq_{reg} \mathbb{2}^*$ is Parikh-injective, then $\mathrm{REL}(C)$ is closed under intersection.

Proof. Given $L, M \subseteq_{reg} C \otimes \mathbb{A}^*$, note that $L \cap M \subseteq_{reg} C \otimes \mathbb{A}^*$, hence $(\llbracket L \cap M \rrbracket, \mathbb{A}) \in \mathrm{REL}(C)$. Further, by Lemma 6, $\llbracket L \cap M \rrbracket = \llbracket L \rrbracket \cap \llbracket M \rrbracket$. Thus, $(\llbracket L \rrbracket \cap \llbracket M \rrbracket, \mathbb{A}) \in \mathrm{REL}(C)$. ◀

▷ Proposition 14. If $C \subseteq_{reg} \mathbb{2}^*$ satisfies any of the bad conditions, then there exist $R, S \in \mathrm{REL}(C)$ such that $R \cap S \notin \mathit{Rational}$.

Proof. *Case* A. Suppose that condition A is satisfied. Consider the 3-letter alphabet $\mathbb{A} = \{a_1, a_2, c\}$. Let R, S be the following relations in $(\mathbb{A}^*)^2$:

$$R = \llbracket (u_1^* \otimes a_1^*) \cdot (u_2^* \otimes a_2^*) \cdot z \otimes c^* \rrbracket, \qquad S = \llbracket (v^* \otimes \{a_1, a_2\}^*) \cdot z \otimes c^* \rrbracket,$$

note that $R, S \in \mathrm{REL}(C)$ by condition A.4. We show that $|R \cap S| = \infty$. Since by condition A.3 we have $\{u_1, u_2\}$ and $\{v\}$ are Parikh-dependent, let $s_1, s_2, s_3 \in \mathbb{N}$ be such that $\pi(u_1^{s_1} \cdot u_2^{s_2}) = \pi(v^{s_3}) \neq \bar{0}$. For $x_1, x_2, x_3 \in \mathbb{N}$, let $R_{x_1, x_2} \subseteq R$ (resp. $S_{x_3} \subseteq S$) be the subrelation obtained by replacing each u_1^*, u_2^* with $u_1^{x_1}$, $u_2^{x_2}$ in the definition (resp. replacing v^* with v^{x_3}). It then follows that $R_{s_1, s_2} \cap S_{s_3} \neq \emptyset$ and further $R_{\ell \cdot s_1, \ell \cdot s_2} \cap S_{\ell \cdot s_3} \neq \emptyset$ for every $\ell \in \mathbb{N}$. Since $\pi(u_1^{s_1} \cdot u_2^{s_2}) \neq \bar{0}$, this implies that there is an infinite sequence of tuples in $R \cap S$ of increasing size, and thus $|R \cap S| = \infty$.

We show that $R \cap S \notin \mathit{Rational}$. By definition we have that

$$\text{for every } (w_1, w_2) \in R \text{ and } \ell \in \mathbb{2},\ (|w_1|_{a_\ell}, |w_2|_{a_\ell}) \in \langle \{\pi(u_\ell)\} \rangle, \text{ and} \tag{1}$$

$$\text{for every } (w_1, w_2) \in S,\ \sum_{b \in \{a_1, a_2\}} (|w_1|_b, |w_2|_b) \in \langle \{\pi(v)\} \rangle. \tag{2}$$

By means of contradiction, suppose $R \cap S \in \mathit{Rational}$ or, equivalently, that there is an automaton over the alphabet $\mathbb{2} \times \mathbb{A}$ such that the language recognized by this automaton synchronizes $R \cap S$. Since we have shown that $R \cap S$ is infinite, the automaton must have a non-trivial cycle $q_0 \xrightarrow{w_1} q \xrightarrow{w_2} q \xrightarrow{w_3} q_f$ inside some accepting run. In particular, $\llbracket w_1 w_3 \rrbracket \in R \cap S$. Let $\llbracket w_2 \rrbracket = (a_{i_1}^{s_1}, a_{i_2}^{s_2})$ with $i_1, i_2 \in \mathbb{2}$, $s_1, s_2 \in \mathbb{N}$ (it is easy to see that there cannot be more than one symbol of $\mathbb{A}$ in any given component). First observe that $i_1 = i_2$, since otherwise, due to item A.2, for some $t \in \mathbb{2}$ we have that $\pi(u_t) \geq (1, 1)$, and in this case, we would obtain that $\llbracket w_1 w_3 \rrbracket \in R$ would contradict (1) for $\ell = t$.

Since $w_2 \neq \varepsilon$, there is some r such that $s_r > 0$. Therefore we have

- $(s_1, s_2) \in \langle\{\pi(u_{i_r})\}\rangle$, as otherwise we would obtain that $[\![w_1 w_3]\!] \in R$ would contradict (1) for $\ell = i_1 = i_2$; and
- $(s_1, s_2) \in \langle\{\pi(v)\}\rangle$, as otherwise $[\![w_1 w_3]\!] \in S$ would contradict (2).

We therefore have that u_{i_r} and v are Parikh-dependent, but this is not possible because of condition A.1. The contradiction originates from assuming that $R \cap S \in$ *Rational*. Hence, $R \cap S \notin$ *Rational*. Note that the proof above can be replicated using only two letters, by replacing c with the letter a_1 (at the price of making the proof less clean). We can therefore conclude that condition A implies that $\text{REL}(C)$ is not closed under intersection, even when restricted to two-letter alphabets.

Case B. For $\alpha_1, \alpha_2 \in \{a, b, \star\}$ and $w \in \mathbb{2}^*$, let $w_{(\alpha_1,\alpha_2)} \in (\mathbb{2} \times \{a, b\})^*$ denote the set of words resulting from replacing in w each $i \in \mathbb{2}$ with (i, a) if $\alpha_i \in \{a, \star\}$ and each $i \in \mathbb{2}$ with (i, b) if $\alpha_i \in \{b, \star\}$ —for example $112_{(a,\star)} = \{(1, a)(1, a)(2, a), (1, a)(1, a)(2, b)\}$.

Now suppose that B holds. We will assume that $|u|_1, |v|_1 > 0$ (the other cases being similar). Consider the following relations in $\text{REL}(C)$:

$$R = [\![u^*_{(a,a)} \cdot v^*_{(b,b)} \cdot z_{(a,a)}]\!] \in \text{REL}(u^* v^* z),$$
$$S = [\![v^*_{(a,\star)} \cdot u^*_{(b,\star)} \cdot z_{(a,a)}]\!] \in \text{REL}(v^* u^* z).$$

We show that the projection onto the first component of $R \cap S$ contains only words of the form $a^n b^n a^{|z|_1}$ (Claim 32 below), and that it forms an infinite language (Claim 33 below), implying that it is not regular and thus that $R \cap S \not\subseteq$ *Rational*. (If, in turn, we had assumed $|u|_2, |v|_2 > 0$, we would produce similar relations and obtain such a result but this time on the second component.)

▷ Claim 32. $\{w_1 : \exists w_2, \dots, w_k\, (w_1, \dots, w_k) \in R \cap S\} \subseteq \{a^n b^n a^{|z|_1} : n \geq 0\}$.

Proof. By definition of R, S, any tuple in $R \cap S$ is of the form

$$[\![u^s_{(a,a)} \cdot v^{s'}_{(b,b)} \cdot z_{(a,a)}]\!] = [\![w_t \cdot w'_{t'} \cdot z_{(a,a)}]\!]$$

for some $w_t \in v^t_{(a,\star)}$, $w'_{t'} \in u^{t'}_{(b,\star)}$, and $s, s', t, t' \geq 0$, which imples $[\![u^s_{(a,a)} \cdot v^{s'}_{(b,b)}]\!] = [\![w_t w'_{t'}]\!]$. Since by hypothesis u and v are Parikh-independent, applying Observation 2 we obtain $t = s'$ and $s = t'$. Therefore, every tuple from $R \cap S$ is of the form $[\![u^s_{(a,a)} \cdot v^t_{(b,b)} \cdot z_{(a,a)}]\!] = [\![w_t \cdot w'_s \cdot z_{(a,a)}]\!]$. Let us now focus on the first component of any such tuple. Since in particular the number of a's in the first component of $[\![u^s_{(a,a)} \cdot v^t_{(b,b)} \cdot z_{(a,a)}]\!]$ is equal to the number of a's in the first component of $[\![w_t \cdot w'_s \cdot z_{(a,a)}]\!]$, we have $|u^s|_1 = |v^t|_1$. This means that the first component of $[\![u^s_{(a,)} \cdot v^t_{(b,)} \cdot z_{(a,)}]\!]$ is of the form $a^n b^n a^{|z|_1}$ for $n = |u^s|_1 = |v^t|_1$. ◀

▷ Claim 33. $|R \cap S| = \infty$, and further $\{w_1 : \exists w_2 \text{ s.t. } (w_1, w_2) \in R \cap S\}$ is infinite.

Proof. Since $|u|_1 > 0$ and $|v|_1 > 0$, there exist $s, t > 0$ such that $|u^s|_1 = |v^t|_1$. Therefore, $[\![u^s_{(a,a)} \cdot v^t_{(b,b)}]\!] = [\![w_t \cdot w'_s]\!]$ for some $w_t \in v^t_{(a,\star)}$, $w'_s \in u^s_{(b,\star)}$. It then follows that $R \cap S$ contains $[\![u^{n \cdot s}_{(a,a)} \cdot v^{n \cdot t}_{(b,b)}]\!] = [\![w^n_t \cdot w'^n_s]\!]$ for every $n \geq 0$. Since $|u|_1 > 0$, the projection onto the first component yields an infinite language. ◀

By the last two claims, it follows that $\{w_1 : \exists w_2 \text{ s.t. } (w_1, w_2) \in R \cap S\}$ is not regular. Indeed, it is easy to see, by pumping arguments, that there cannot exist an automaton accepting any infinite set inside $\{a^n b^n a^{|z|_1} \mid n \geq 0\}$. Then $R \cap S \notin$ *Rational* – if there would be $L \subseteq_{reg} (\mathbb{2} \times \mathbb{A})^*$ such that $[\![L]\!] = R \cap S$, by projecting L onto the 1-labeled components, we would have a regular language inside $(\{1\} \times \mathbb{A})^*$ synchronizing $\{w_1 : \exists w_2 \text{ s.t. } (w_1, w_2) \in R \cap S\}$ and thus it would be regular.

Case C. Let us assume that $u \in 1^*$ and $w \in 2^*$, the other case is symmetric. Consider the following relations over the alphabet $\mathbb{A} = \{a, b, c\}$, using the same terminology as before.

$$R = [\![(u^* \otimes a^*) \cdot (v^* \otimes b^*) \cdot (w^* \otimes a^*) \cdot z \otimes c^*]\!], \qquad S = [\![v^*_{(a,b)} \cdot v^*_{(b,a)} \cdot z_{(c,c)}]\!]$$

Note that $R, S \in \mathrm{REL}(C)$, and, that R consists of all pairs

$$(a^{|u|_1 \cdot n} b^{|v|_1 \cdot m} c^{|z|_1}, b^{|v|_2 \cdot m} a^{|w|_2 \cdot \ell} c^{|z|_2})$$

for $n, m, \ell \in \mathbb{N}$. By C.2 $|v|_1 > 0$ and $|v|_2 > 0$ and hence R has a linear dependence on the number of b's in the first and second component, and S has a linear dependence on the number of as' in the first component and the nuber of b's in the second component. Thus,

$$R \cap S = \{(a^{|u|_1 \cdot n} b^{|v|_1 \cdot m} c^{|z|_1}, b^{|v|_2 \cdot m} a^{|w|_2 \cdot \ell} c^{|z|_2}) : \\ n, m, \ell \in \mathbb{N}, (|u|_1 \cdot n, |v|_2 \cdot m), (|v|_1 \cdot m, |w|_2 \cdot \ell) \in \langle\{\pi(v)\}\rangle\}$$

creates a linear dependence between the number of a's and the number of b's in the first component. Since $\{w\colon (w, w') \in R \cap S\}$ is infinite due to C.1 and sits inside $a^*b^*c^*$ by definition, this implies that $R \cap S \notin \mathit{Rational}$. ◀

To prove Proposition 16, we need to prove first the following technical lemmas (Lemmas 34, 35 and 36):

▶ **Lemma 34.** *For every $C \subseteq_{reg} \mathbb{2}^*$ having two Parikh-independent words and $w \in \mathbb{2}^*$, we have that C^*w satisfies property* A.

Proof. Let $u, v \in C$ such that u and v are Parikh-independent. From this, it follows easily that both uuv, uv and uvv, uv are Parikh-independent. It is also immediate that $\pi(uuv) \geq (1, 1)$ (also $\pi(uvv) \geq (1, 1)$), and that $\{uuv, uvv\}$ and $\{uv\}$ are Parikh-dependent. It only remains to observe that $(uuv)^*(uvv)^*w \subseteq C^*w$ and $(uv)^*w \subseteq C^*w$. ◀

▶ **Lemma 35.** *For every two Parikh-dependent words $u, v \in \mathbb{2}^*$ there exist $n \in \mathbb{N}$ and $z_1, \dots, z_n \in \mathbb{2}^*$ such that $u^*v^* =_{\mathrm{REL}} \bigcup_{i=1}^n v^* z_i$. Moreover, $\pi(z_i) \in \langle\{\pi(u)\}\rangle$ for every $i \in \{1, \cdots, n\}$.*

Proof. Let $l, m > 0$ be such that $l \cdot \pi(u) = m \cdot \pi(v)$. Then, by using the properties in Lemma 4, we have

$$u^*v^* = u^{l*}u^{<l}v^* =_{\mathrm{REL}} v^{m*}v^*u^{<l} = v^*u^{<l} = \bigcup_{i=0}^{l-1} v^*u^i$$

from where the result follows. ◀

▶ **Lemma 36.** *For every $C \subseteq_{reg} \mathbb{2}^*$ and $\bar{x} \in \mathbb{N}^2$ such that $\pi(C) \subseteq \langle\{\bar{x}\}\rangle$, there exist $n \in \mathbb{N}$ and words $u, z_1, \dots, z_n \in \mathbb{2}^*$ with Parikh-image in $\langle\{\bar{x}\}\rangle$ such that $C^* =_{\mathrm{REL}} \bigcup_{i=1}^n u^* z_i$.*

Proof. Let $u \in \mathbb{2}^*$ such that $\pi(u) = \bar{x}$. We proceed by induction on the star-height s of C^* (the maximum number of nested Kleene-stars), for any given regular expression representing C.

- Base case $s = 1$: In this case, $C = \{u_1, \dots, u_m\}$ is finite. Then there exist $r_1, \dots, r_m \in \mathbb{N}$ such that $\pi(u_i) = r_i \cdot \pi(u)$ for all i. Then, by using the properties in Lemma 4, we have

  $$C^* =_{\mathrm{REL}} \{u^{r_1}, \dots, u^{r_m}\}^* = u^{r_1 *} \cdots u^{r_m *}.$$

  We conclude by applying Lemma 35 repeatedly and again properties from Lemma 4.

- Inductive step: Suppose $s > 1$, and then without loss in generality (see Lemma 5 plus properties in Lemma 4) that $C^* = (\bigcup_i C^*_{i,1} \cdots C^*_{i,n} u_i)^*$ where each $C_{i,j}$ has star-height strictly smaller than s Since $\pi(C^*) \subseteq \langle\{\bar{x}\}\rangle$ we have, for each i, that $\pi(u_i) \in \langle\{\bar{x}\}\rangle$ and thus, by property 9 of Lemma 4, $u_i =_{\text{REL}} u^{\ell_i}$ for some $\ell_i \geq 0$. Then, by properties in Lemma 4, we have that $C^* =_{\text{REL}} (\bigcup_i C^*_{i,1} \cdots C^*_{i,n} u^{\ell_i})^*$. We further have that $\pi(C_{i,j}) \subseteq \langle\{\bar{x}\}\rangle$; indeed, if there were some word $v \in C_{i,j}$ with $\pi(v) \notin \langle\{\bar{x}\}\rangle$, we would obtain a word vu^{ℓ_i} with $\pi(vu^{\ell_i}) \in \pi(C^*)$ and $\pi(vu^{\ell_i}) \notin \langle\{\bar{x}\}\rangle$, contradicting the hypothesis. We can therefore apply the inductive hypothesis to each $C_{i,j}$ which implies, together with the application of some properties from Lemma 4, that $C^* =_{\text{REL}} D^*$ where D is a finite union of languages of the form $w_1^* \cdots w_s^* z$ in which all words involved are powers of u. Then, applying Lemma 35 repeatedly (and, again properties from Lemma 4), we have that for some $r_j, t_j > 0$, $C^* =_{\text{REL}} (\bigcup_{j=1}^m u^{r_j *} u^{t_j})^* = \bigcup_{1 \leq s \leq m} \bigcup_{1 \leq j_1 < \cdots < j_s \leq m} u^{r_{j_1} *} u^{t_{j_1} *} \cdots u^{r_{j_s} *} u^{t_{j_s} *} u^{t_{j_1}} \cdots u^{t_{j_s}}$ (where $\{j_1, \ldots, j_s\}$ represents the disjuncts that are iterated at least once). Finally, by Lemma 35 and properties from Lemma 4 again, we have the desired result. ◀

▷ Proposition 16. If $C \subseteq_{reg} \mathbb{2}^*$ does not satisfy any of the bad conditions, then C is REL-equivalent to a finite union of basic injective languages.

Proof. We show the contrapositive, *i.e.*, that every $C \subseteq_{reg} \mathbb{2}^*$ that is not REL-equivalent to a finite union of basic injective languages satisfies at least one of the bad conditions. First observe that, by Lemma 5 plus Observation 11, without loss in generality we can assume that C is smooth. Also, by Lemma 34 plus Observation 11, if C has some component with two Parikh-independent words, then it satisfies property A. Then we can assume without loss in generality that the Parikh-image of every component of C lies inside the linear set generated by just one vector. We can then apply Lemma 36 to each component, which together with properties from Lemma 4, implies that C is REL-equivalent to a finite union of languages of the form $w_1^* \cdots w_n^* z$. Then, since C is not REL-equivalent to a finite union of basic injective languages, there must exist at least one of those conjuncts $w_1^* \cdots w_n^* z$ that is not REL-equivalent to a finite union of basic injective languages. Therefore, by Observation 11, without loss in generality we can assume that C is of the form $w_1^* \cdots w_n^* z$. Moreover, using also Lemma 35, we can assume that w_i and w_{i+1} are Parikh-independent for all $i = 1, \ldots, n-1$. Also, by inspecting the proofs of Lemmas 35 and 36, one can check that the "smoothness" is not lost, *i.e.* either $n = 2$, $w_1 = 1^p$ and $w_2 = 2^q$ for some $p, q \in \mathbb{N}$, or for $r = 1, 2$, there is no $i \in \{1, \ldots, n-1\}$ such that $w_i \in r^*$ and $w_{i+1} \in (3-r)^*$. For this kind of languages, if $n \leq 2$, the result follows immediately, since they are already basic injective. For $n \geq 3$, we proceed by case distinction according to the letters from $\mathbb{2}$ that are used in each of the w_i's. Since we are going to prove that in every case at least one of the bad conditions is satisfied, by Observation 11, we can assume without loss in generality that our language is $w_1^* w_2^* w_3^* z$, and w_2 is Parikh-independent with both w_1 and w_3.

1. There exists $r \in \mathbb{2}$ such that $w_1 \in r^*$, $w_3 \in (3-r)^*$ and $\pi(w_2) \geq (1, 1)$. In this case, condition C is satisfied.
2. There exists $r \in \mathbb{2}$ such that $w_1, w_3 \in r^*$ and $\pi(w_2) \geq (1, 1)$. In this case, condition B is satisfied.
3. We are neither in case 1 nor 2 and w_1, w_2, w_3 are pairwise Parikh-independent. In this case, condition A is satisfied: one only has to observe that, by basic linear algebra, for any three pairwise independent vectors en $\mathbb{N}^2$, there is one of them such that its singleton set is dependent with the set formed by the other two.
4. We are in none of the previous cases. Since w_1 and w_2 are Parikh-independent, we necessarily have that w_1 and w_3 are Parikh-dependent. In this case, we will prove

that condition B is satisfied. Indeed, let w be a word of shortest possible length which is Parikh-dependent with w_1. Then there exist $p, q \in \mathbb{N}$ such that $\pi(w_1) = p \cdot \pi(w)$ and $\pi(w_3) = q \cdot \pi(w)$. For $r = lcm(p, q)$, by properties from Lemma 4, we have that $w^{r*}w_2^*z \subseteq_{\text{REL}} w_1^*w_2^*z \subseteq w_1^*w_2^*w_3^*z$ and $w_2^*w^{r*}z \subseteq_{\text{REL}} w_2^*w_3^*z \subseteq w_1^*w_2^*w_3^*z$. It only remains to observe that

- w^r is Parikh-independent with w_2, which follows from the fact that it is Parikh-dependent with w_1, and w_1 is Parikh-independent with w_2, and
- $\pi(w) \geq (1, 1)$, which holds since otherwise we would be either in case 1 or 2.

Note that, by definition of smoothness plus the fact that there were no two consecutive Parikh-dependent w_i's, we covered all the possible cases. ◀

The only missing link in the proof of Lemma 18 is the proof of Lemma 19, which requires some effort. Before giving its proof, we show a reduction of the general problem into simpler cases.

▶ **Lemma 37.** *Let $B_1 =_{\text{REL}} \bigcup_{i \in I} B_{1,i}$ and $B_2 =_{\text{REL}} \bigcup_{j \in J} B_{2,j}$. If the pair $B_{1,i}, B_{2,j}$ satisfies the dichotomy property for all $i \in I, j \in J$, then B_1, B_2 satisfies the dichotomy property.*

Proof. If there exist i, j such that $B_{1,i} \cup B_{2,j}$ satisfies any of the bad conditions, then it is immediate that $B_1 \cup B_2$ satisfies the same condition by Observation 11. Otherwise, for every i, j we have $B_{1,i} \cap [B_{2,j}]_\pi$ regular and $B_{1,i} \cap [B_{2,j}]_\pi \subseteq_{\text{REL}} B_{2,j}$. Then $B_1 \cap [B_2]_\pi = \bigcup_{i,j} B_{1,i} \cap [B_{2,j}]_\pi$ is regular. Moreover, $B_1 \cap [B_2]_\pi = \bigcup_{i,j} \underbrace{B_{1,i} \cap [B_{2,j}]_\pi}_{\subseteq_{\text{REL}} B_{2,j}} \subseteq_{\text{REL}} \bigcup_j B_{2,j} =_{\text{REL}} B_2$. ◀

Using the previous lemma, we will prove Lemma 19 gradually, starting from some restricted cases.

Let $B_1 = u_1^*u_2^*u$, $B_2 = v_1^*v_2^*v$ be two basic injective languages such that $B_1, B_2 \not\subseteq_{\text{REL}} 1^*2^*$. We say that they *agree* if for all $i, j \in \mathbb{2}$, if u_i and v_j are Parikh-dependent then $u_i = v_j$. We say that they *have the same tail* if $u = v$. We first prove the case in which B_1 and B_2 agree and have the same tail.

▶ **Lemma 38.** *Every pair of basic injective languages B_1, B_2 that agree, have the same tail, and are such that $\text{REL}(B_1), \text{REL}(B_2) \not\subseteq$ Recognizable satisfies the dichotomy property.*

Proof. Let $B_1 = u_1^*u_2^*u$, $B_2 = v_1^*v_2^*u$ and assume that $B_1 \cup B_2$ does not satisfy any of the bad conditions. Let $I = \{i \in \mathbb{2} : \exists j \in \mathbb{2} \text{ s.t. } u_i = v_j\} = \{i_1, \ldots, i_r\}$ for $i_1 < \cdots < i_r$. For each $\ell \in I$, let $j_\ell \in \mathbb{2}$ be the index such that $u_{i_\ell} = v_{j_\ell}$ (there is only one such index since B_2 is basic injective), and let $J = \{j_\ell : \ell \in I\}$. We first prove

$$B_1 \cap [B_2]_\pi = u_{i_1}^* \cdots u_{i_r}^* u. \tag{3}$$

Note that the $\supseteq$-containment of (3) holds trivially. For the $\subseteq$-containment, by means of contradiction, we show that if $B_1 \cap [B_2]_\pi \supsetneq u_{i_1}^* \cdots u_{i_r}^* u$ then $B_1 \cup B_2$ would satisfy condition A, which is in contradiction with our hypothesis. If this happens, in particular there is a word $u_1^{s_1}u_2^{s_2}u \in B_1 \cap [B_2]_\pi$ such that $s_i > 0$ for some $i \notin I$. Let $\bar{I} = \mathbb{2} \setminus I$ (note that $\bar{I} \neq \emptyset$) and $\bar{J} = \mathbb{2} \setminus J$ (since $|I| = |J|$, we also have $\bar{J} \neq \emptyset$). Then there are $t_1, t_2 \in \mathbb{N}$ such that

$$\sum_{i \in I} s_i \cdot \pi(u_i) + \sum_{i \in \bar{I}} s_i \cdot \pi(u_i) = \sum_{i \in J} t_i \cdot \pi(u_i) + \sum_{i \in \bar{J}} t_i \cdot \pi(v_i).$$

Let $I_1 = \{i \in I \colon s_i \geq t_i\}$ and $I_2 = I \setminus I_1$. Then we have the following

$$\sum_{i \in I_1} (s_i - t_i) \cdot \pi(u_i) + \sum_{i \in \bar{I}} s_i \cdot \pi(u_i) = \sum_{i \in I_2} (t_i - s_i) \cdot \pi(u_i) + \sum_{i \in \bar{J}} t_i \cdot \pi(v_i).$$

Note that at least one of the sets I_1, I_2 or $\bar{I}$ has to be empty (they are pairwise disjoint inside $\mathbb{2}$). But we assumed that $\bar{I}$ was not empty. Then either I_1 or I_2 is empty from which it follows easily that condition A is satisfied by $B_1 \cup B_2$ (recall that $\bar{J} \neq \emptyset$). For example, if $I_2 = \emptyset$, $\bar{I} \neq \emptyset$, $I_1 \neq \emptyset$ and $\bar{J} = \{j_0\}$, we have that u_1, u_2 is Parikh-dependent with v_{j_0}; for some i $\pi(u_i) \geq (1,1)$ since $B_1 \not\subseteq_{\text{REL}} 1^*2^*$; u_1 (resp. u_2) is Parikh-independent with v_{j_0} because otherwise we would have $j_0 \in J$; and both $u_1^* u_2^* u \subseteq_{\text{REL}} B_1 \cup B_2$, $v_{j_0}^* u \subseteq_{\text{REL}} B_1 \cup B_2$.

We finish the proof by showing that if (3) holds, then $B_1 \cap [B_2]_\pi \subseteq_{\text{REL}} B_2$. Note that the cases in which $|I| = 0$ or $|I| = 1$ are trivial. Suppose then that $I = \mathbb{2}$. If $j_1 < j_2$, the result is trivial, and, if $j_1 > j_2$, we have a contradiction because $B_1 \cup B_2$ would satisfy condition B. ◀

We then allow tails to be different.

▶ **Lemma 39.** *Every pair of basic injective languages B_1, B_2 that agree and such that $\text{REL}(B_1), \text{REL}(B_2) \not\subseteq$ Recognizable satisfies the dichotomy property.*

Proof. Let $B_1 = u_1^* u_2^* u$, $B_2 = v_1^* v_2^* v$. We are going to split the proof in four cases according to the value of $n \in \{0, \ldots, 4\}$, which is the number of non-empty words in $\{u_1, v_1, u_2, v_2\}$.

- if $n = 0$, $B_1 = u, B_2 = v$. If $\pi(u) = \pi(v)$ we have $B_1 \cap [B_2]_\pi = \{u\} \subseteq_{\text{REL}} B_2$; and if $\pi(u) \neq \pi(v)$ we have $B_1 \cap [B_2]_\pi = \emptyset \subseteq_{\text{REL}} B_2$.
- if $n = 1$, we have two cases to consider: $B_1 = u_1^* u$ with $u_1 \neq \varepsilon$ and $B_2 = v$ or $B_1 = u$ and $B_2 = v_1^* v$ with $v_1 \neq \varepsilon$. In both cases, it is immediate that $B_1 \cap [B_2]_\pi$ is finite and so, by property 9 of Lemma 4, $B_1 \cap [B_2]_\pi \subseteq_{\text{REL}} B_2$.
- if $n = 2$, in case that $B_1 = u$ or $B_2 = v$, $B_1 \cap [B_2]_\pi$ is finite and we can use the previous argument. Then we can assume that $B_1 = u_1^* u$ and $B_2 = v_1^* v$ for $u_1, v_1 \neq \varepsilon$. If $B_1 \cap [B_2]_\pi = \emptyset$, then the result is trivial. Otherwise, there exist $s, t \in \mathbb{N}$ such that

$$s \cdot \pi(u_1) + \pi(u) = t \cdot \pi(v_1) + \pi(v). \tag{4}$$

  Therefore, by properties from Lemma 4, one can check that

$$B_1 = u_1^* u_1^s u \cup \bigcup_{\ell < s} u_1^\ell u \quad \text{and} \quad B_2 = v_1^* v_1^t v \cup \bigcup_{\ell < t} v_1^\ell v =_{\text{REL}} v_1^* u_1^s u \cup \bigcup_{\ell < t} v_1^\ell v,$$

  the latter equality holding due to (4). By Lemma 37, it is enough to check that the result holds for any two disjuncts from B_1 and B_2 respectively. Observe that each such pair consists of two basic injective languages that satisfy the hypothesis of the lemma. For the pair of disjuncts $u_1^* u_1^s u$ and $v_1^* u_1^s u$, we can apply Lemma 38 since they have equal tail. For all the other cases, we obtain pairs of basic injective languages with strictly less amount of non trivial components in total, and thus we can use previous cases. In this way we obtain that every pair of disjuncts satisfies the dichotomy property, and thus B_1 and B_2 too.
- if $n = 3$, in case $B_1 \cap [B_2]_\pi = \emptyset$, the result is trivial. Otherwise, we have two different cases to consider: $B_1 = u_1^* u$ and $B_2 = v_1^* v_2^* v$ for u_1, v_1, v_2 non-empty or $B_1 = u_1^* u_2^* u$ and $B_2 = v_1^* v$ for u_1, u_2, v_1 non-empty. Suppose that $B_1 = u_1^* u$ and $B_2 = v_1^* v_2^* v$ with $u_1, v_1, v_2 \neq \varepsilon$. Then, there exist $s, t_1, t_2 \in \mathbb{N}$ such that

$$s \cdot \pi(u_1) + \pi(u) = t_1 \cdot \pi(v_1) + t_2 \cdot \pi(v_2) + \pi(v). \tag{5}$$

  Therefore, by properties from Lemma 4, one can check that

$$B_1 = u_1^* u_1^{s_1} u \cup \bigcup_{\ell < s} u_1^\ell u \quad \text{and} \quad B_2 =_{\text{REL}} v_1^* v_2^* u_1^s u \cup \bigcup_{i=1}^{2} \bigcup_{\ell < t_i} v_{3-i}^* v_i^\ell v,$$

the latter equality holding due to (5). By Lemma 37, it is enough to check that the result holds for any two disjuncts from B_1 and B_2 respectively. Observe that each such pair consists of two basic injective languages that satisfy the hypothesis of the lemma. For the pair of disjuncts $u_1^* u_1^s u$ and $v_1^* v_2^* u_1^s u$, we can apply Lemma 38 since they have equal tail. For all the other cases, we obtain pairs of basic injective languages with strictly less amount of non trivial components in total, and thus we can use previous cases. In this way we obtain that every pair of disjuncts satisfies the dichotomy property, and thus B_1 and B_2 too.

- if $n = 4$, in case $B_1 \cap [B_2]_\pi = \emptyset$, the result is trivial. Otherwise, there exist $s_1, s_2, t_1, t_2 \in \mathbb{N}$ such that

$$s_1 \cdot \pi(u_1) + s_2 \cdot \pi(u_2) + \pi(u) = t_1 \cdot \pi(v_1) + t_2 \cdot \pi(v_2) + \pi(v) \tag{6}$$

Therefore, by properties from Lemma 4, one can check that

$$B_1 = u_1^* u_2^* u_1^{s_1} u_2^{s_2} u \cup \bigcup_{i=1}^{2} \bigcup_{\ell < s_i} u_{3-i}^* u_i^\ell u \quad \text{and} \quad B_2 =_{\text{REL}} v_1^* v_2^* u_1^{s_1} u_2^{s_2} u \cup \bigcup_{i=1}^{2} \bigcup_{\ell < t_i} v_{3-i}^* v_i^\ell v,$$

the latter equality holding due to (6). By Lemma 37, it is enough to check that the result holds for any two disjuncts from B_1 and B_2 respectively. Observe that each such pair consists of two basic injective languages that satisfy the hypothesis of the lemma. For the pair of disjuncts $u_1^* u_2^* u_1^{s_1} u_2^{s_2} u$ and $v_1^* v_2^* u_1^s u_2^{s_2} u$, we can apply Lemma 38 since they have equal tail. For all the other cases, we obtain pairs of basic injective languages with strictly less amount of non trivial components in total, and thus we can use previous cases. In this way we obtain that every pair of disjuncts satisfies the dichotomy property, and thus B_1 and B_2 too.

◀

Finally, we prove Lemma 19 by lifting the unnecessary 'agree' condition.

▷ **Lemma 19.** Every pair of basic injective languages B_1, B_2 such that $B_1, B_2 \not\subseteq_{\text{REL}} 1^* 2^*$ satisfies the dichotomy property.

Proof. Let $B_1 = u_1^* u_2^* u$, $B_2 = v_1^* v_2^* v$. We proceed by induction on the cardinality n of the set

$$\{i \in \mathbb{2} : \text{ there is } j \in \mathbb{2} \text{ s.t. } u_i \neq v_j \text{ and } u_i \text{ and } v_j \text{ Parikh-dependent}\}$$

(*i.e.*, the number of 'disagreements'). The base case $n = 0$ is already proved in Lemma 39. For the inductive step, let $i, j \in \mathbb{2}$ be such that u_i and v_j are Parikh-dependent but $u_i \neq v_j$. Then there is a word $w \in \mathbb{2}^*$ and $p, q \in \mathbb{N}$ such that $\pi(u_i) = \pi(w^p)$ and $\pi(v_j) = \pi(w^q)$. Let $r = lcm(p, q)$. Assume that $i = j = 1$ (other cases are analogous). Then, it follows that

$$B_1 =_{\text{REL}} \bigcup_{\ell < \frac{r}{p}} w^{r*} u_2^* u_1^\ell u \quad \text{and} \quad B_2 =_{\text{REL}} \bigcup_{\ell < \frac{r}{q}} w^{r*} v_2^* v_1^\ell v.$$

Notice that each disjunct is a basic injective language that satisfies the hypothesis of the lemma. By Lemma 37, it is enough to check that the dichotomy property holds for any pair of disjuncts from B_1 and B_2 respectively. Any such pair of languages has one less disagreement because now the first word of the first language of the pair agrees with the first word of the second language. Then, we can use the inductive hypothesis and the dichotomy property for B_1 and B_2 follows. ◀

Before proving Proposition 20, we show some preliminary results.

▶ **Lemma 40.** *The problem of whether a language $L \subseteq_{reg} \mathbb{A}^*$ is Parikh-injective is decidable.*

Proof. Suppose $\mathbb{A} = \{a_1, \ldots, a_n\}$ and let $\hat{\mathbb{A}} = \{\hat{a}_1, \ldots, \hat{a}_n\}$ be the alphabet where we replace every symbol with its 'marked' version. For every $w \in \mathbb{A}^*$ let $\hat{w} \in \hat{\mathbb{A}}^*$ be the result of replacing every a_i with $\hat{a}_i$ in w. Given a pair of words $u = b_1 \cdots b_m \in \mathbb{A}^*$ and $u' = b'_1 \cdots b'_{m'} \in \hat{\mathbb{A}}^*$ such that $m = m'$, let $u \oplus u'$ denote the word $b_1 b'_1 \cdots b_m b'_m \in (\mathbb{A} \cup \hat{\mathbb{A}})^*$. It is not hard to see that $S = \{u \oplus \hat{v} : u, v \in L, |u| = |v|, u \neq v\}$ is a regular language over $\mathbb{A} \cup \hat{\mathbb{A}}$. Let $Z \subseteq \mathbb{N}^{\mathbb{A} \cup \hat{\mathbb{A}}}$ be the semi-linear set $Z = \{\bar{x} : \bar{x}(a_i) = \bar{x}(\hat{a}_i) \text{ for every } i\}$. If follows that L is Parikh-injective if, and only if, $\pi(S) \cap Z = \emptyset$, and that the latter is decidable by effectiveness of intersection of semi-linear sets. ◀

▶ **Lemma 41.** *The set of languages $C \subseteq_{reg} \mathbb{2}^*$ for which there is a Parikh-injective language $D \subseteq_{reg} \mathbb{2}$ such that $C \subseteq_{\text{REL}} D$ is computably enumerable.*

Proof. This proposition follows from Lemma 40 plus the fact that checking REL-containment is decidable [14]. ◀

▶ **Lemma 42.** *The set of languages $C \subseteq_{reg} \mathbb{2}^*$ which satisfies any of the bad conditions is computably enumerable.*

Proof. This proposition also follows easily from the fact that REL-containment is decidable [14]. ◀

▷ Proposition 20. It is decidable wether a given $C \subseteq_{reg} \mathbb{2}^*$ is such that $\text{REL}(C)$ is closed under intersection.

Proof. By the equivalence $1 \Leftrightarrow 7$ of Theorem 12, we obtain that the problem is c.e. (Lemma 41) and, by the equivalence $1 \Leftrightarrow 5$ of the same theorem, we obtain that it is co-c.e. (Lemma 42). We conclude then that it is decidable. ◀

C Missing proofs to Section 5

▷ Proposition 27. For every $C, C_1, C_2, C_3 \subseteq_{reg} \mathbb{k}^*$,

1. $C_1 \cdot C_2 \subseteq_{\text{REL}} C_3$ iff for every $R_1 \in \text{REL}(C_1)$, $R_2 \in \text{REL}(C_2)$ we have $R_1 \cdot R_2 \in \text{REL}(C_3)$;
2. $\text{REL}(C)$ is closed under concatenation iff $C \cdot C \subseteq_{\text{REL}} C$;
3. if $\text{REL}(C)$ is closed under Kleene star, then it is closed under concatenation; and
4. $\text{REL}(C)$ is closed under Kleene star iff $C^* \subseteq_{\text{REL}} C$.

Proof. Items 1 and 2. For the left-to-right direction, let $L_1 \subseteq_{reg} C_1 \otimes \mathbb{A}^*$ and $L_2 \subseteq_{reg} C_2 \otimes \mathbb{A}^*$. Then we only have to observe that $[\![L_1]\!] \cdot [\![L_2]\!] = [\![L_1 \cdot L_2]\!] \in \text{REL}(C_1 \cdot C_2) \subseteq \text{REL}(C_3)$ as we wanted. The right-to-left direction follows from Lemma 8 together with property 1 of Lemma 4. Note that item 2 is a particular case of item 1.

Item 3. We now turn to item 3, since we will use this item to prove item 4. Suppose $\text{REL}(C)$ is closed under Kleene star, and take arbitrary $R_1, R_2 \in \text{REL}(C)$ over an alphabet $\mathbb{A}$. We define relations R'_1, R'_2 as the result of the component-wise application of some letter-to-letter relations T_1, T_2 to R_1, R_2, respectively. For each $i = 1, 2$ let Z_i be the regular language $(\overline{lst_i})^* \cdot lst_i \cup \{\varepsilon\}$ over the binary alphabet $\{\overline{lst_i}, lst_i\}$, where $\overline{lst_i}$ will intuitively stand for "not the last letter of a pair from R_i" and symbol lst_i for "the last letter of a pair from R_i". For each $i \in \{1,2\}$, we define $T_i \subseteq \mathbb{A}^* \times (\mathbb{A} \times \{\overline{lst}_i, lst_i\})^*$ as

$$T_i = \{(u,v) \in \mathbb{A}^* \times (\mathbb{A} \times \{\overline{lst}_i, lst_i\})^* : v \in u \otimes Z_i\} \in \textsc{Rel}((12)^*).$$

For example, T_1 associates aab with $(a, \overline{lst}_1)(a, \overline{lst}_1)(b, lst_1)$. Consider the component-wise application of T_1 and T_2 to R_1 and R_2 respectively, that is: $R'_1 = R_1 \circ (T_1, T_1)$ and $R'_2 = R_2 \circ (T_2, T_2)$. For example, if $(aab, b, ba) \in R_1$ then

$$((a, \overline{lst}_1)(a, \overline{lst}_1)(b, lst_1), (b, lst_1), (b, \overline{lst}_1)(a, lst_1)) \in R'_1.$$

By closure under component-wise letter-to-letter relations we have that $R'_1, R'_2 \in \textsc{Rel}(C)$ over the alphabet $\mathbb{A} \times \{\overline{lst}_1, lst_1, \overline{lst}_2, lst_2\}$.

Observe that $R'_1 \cdot R'_2 \subseteq (R'_1 \cup R'_2)^*$ and, by closure under union and Kleene star, that $(R'_1 \cup R'_2)^* \in \textsc{Rel}(C)$. Let $L \subseteq_{reg} C \otimes (\mathbb{A} \times \{\overline{lst}_1, lst_1, \overline{lst}_2, lst_2\})^*$ such that $[\![L]\!] = (R'_1 \cup R'_2)^*$. We show that there is $L' \subseteq_{reg} L$ such that $[\![L']\!] = R'_1 \cdot R'_2$, and thus that $R'_1 \cdot R'_2 \in \textsc{Rel}(C)$. Consider the language

$$S = \{w \in (\Bbbk \times (\mathbb{A} \times \{\overline{lst}_1, lst_1, \overline{lst}_2, lst_2\}))^* : [\![w]\!] \in (\mathbb{A}^* \otimes (Z_1 \cdot Z_2))^k\}.$$

Note that S is regular, and that $[\![L \cap S]\!] = R'_1 \cdot R'_2$. Since $L \cap S \subseteq_{reg} C \otimes (\mathbb{A} \times \{\overline{lst}_1, lst_1, \overline{lst}_2, lst_2\})^*$, it follows that $R'_1 \cdot R'_2 \in \textsc{Rel}(C)$. Again by closure under component-wise letter-to-letter relations we obtain that $R_1 \cdot R_2 \in \textsc{Rel}(C)$, this time using the relation that projects onto the first component: $R_1 \cdot R_2 = (R'_1 \cdot R'_2) \circ (T, T)$ for

$$T = \{(v,u) \in (\mathbb{A} \times \{\overline{lst}_1, lst_1, \overline{lst}_2, lst_2\})^* \times \mathbb{A}^* : v \in u \otimes (Z_1 \cdot Z_2)\} \in \textsc{Rel}((12)^*).$$

Item 4. For the right-to-left direction of 4, let $R \in \textsc{Rel}(C)$ and take $L \subseteq_{reg} C \otimes \mathbb{A}^*$ such that $[\![L]\!] = R$. Therefore $R^* = [\![L]\!]^* = [\![L^*]\!] \in \textsc{Rel}(C^*) \subseteq \textsc{Rel}(C)$ as wanted. For the left-to-right direction, first observe that $\textsc{Rel}(C)$ is also closed under concatenation due to item 3. Let $R \in \textsc{Rel}(C^*)$. By property 2 of Lemma 4, we have the following:

there are $R_1, \dots, R_n \in \textsc{Rel}(C)$ *and* $I \subseteq_{reg} \{1, \dots, n\}^*$ *such that* $R = \bigcup_{w \in I} R_{w[1]} \cdots R_{w[|w|]}$.

Consider any regular expression E denoting the language I above, and replace each occurence of $i \in \{1, \dots, n\}$ with R_i, in such a way that the resulting expression E' denotes R. Then, by finite application of closure under Kleene star, concatenation and union as given by E', we obtain that $R \in \textsc{Rel}(C)$. ◀

▷ Lemma 28. For every $k \in \mathbb{N}$ and $C \subseteq_{reg} \Bbbk^*$, $\textsc{Rel}_k(C)$ is closed under projection iff $\textsc{Rel}_k(C|_K) \subseteq \textsc{Rel}_k(C)$ for every $K \subseteq \Bbbk$.

Proof. The left-to-right direction follows from the fact that every relation $([\![L]\!], \mathbb{A}) \in \textsc{Rel}_k(C|_K)$ is the projection of a relation in $\textsc{Rel}_k(C)$ onto the components K. Indeed, let $M \subseteq_{reg} (\Bbbk \times \mathbb{A})^*$ the greatest language such that $M|_K = L$. Now consider $L' = M \cap (C \otimes \mathbb{A}^*)$. It follows that L' is regular and $[\![L']\!]|_K = [\![L]\!]$. The right-to-left direction follows from the fact that the projection onto K of any relation $R \in \textsc{Rel}_k(C)$ is a relation of $\textsc{Rel}_k(C|_K)$. Indeed, one can take an automaton recognizing a language that synchronizes R and replace all transitions labeled by an element from $(\Bbbk \setminus K) \times \mathbb{A}$ with an ε-transition. ◀