



HAL
open science

Topologies for Error-Detecting Variable-Length Codes

Jean Néraud

► **To cite this version:**

| Jean Néraud. Topologies for Error-Detecting Variable-Length Codes. 2023. <hal-04195106v2>

HAL Id: hal-04195106

<https://hal.science/hal-04195106v2>

Preprint submitted on 6 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Topologies for Error-Detecting Variable-Length Codes

Jean Néraud

Univ Rouen Normandie, LITIS UR 4108, F-76000 Rouen, France

Abstract

Given a finite alphabet A and a quasi-metric d over A^* , we introduce the relation $\tau_{d,k} \subseteq A^* \times A^*$ such that $(x, y) \in \tau_{d,k}$ holds whenever $d(x, y) \leq k$. The error detection capability of variable-length codes is expressed in term of conditions over $\tau_{d,k}$. With respect to the prefix metric, the factor one, and any quasi-metric associated with some free monoid (anti-)automorphism, we prove that one can decide whether a given regular variable-length code satisfies any of those error detection constraints.

1 Introduction

In Computer Science, the transmission of finite sequences of symbols (the so-called *words*) via some channel constitutes one of the most challenging research fields. With the notation of the free monoid, some classical models may be informally described as indicated in the following:

Two finite *alphabets*, say A and B , are required, every information being modeled by a unique word, say u , in B^* (the *free monoid* generated by B). Usually, in order to facilitate the transmission, beforehand u is transformed in $w \in A^*$, the so-called *input word*: this is done by applying some fixed one-to-one *coding* mapping $\phi : B^* \rightarrow A^*$. In numerous cases, ϕ is an injective *monoid homomorphism*, whence $X = \phi(B)$ is a *variable-length code* (for short, a *code*): equivalently every equation among the words of X is necessarily trivial. Such a translation is particularly illustrated by the well-known examples of the Morse and Huffman codes. Next, w is transmitted via a fixed *channel* into $w' \in A^*$, the so-called *output word*: should w' be altered by some *noise* and then the resulting word $\phi^{-1}(w') \in B^*$ could be different from the initial word u . In the most general model of transmission, the channel is represented by some *probabilistic transducer*. However, in the framework of error detection, most of the models only require that highly likely errors need to be taken into account: in the present paper, we assume the transmission channel modeled by some *binary word relation*, namely $\tau \subseteq A^* \times A^*$. In order to retrieve u , the homomorphism ϕ , and thus the code X , must satisfy specific constraints, which of course depend on the channel τ : in view of some formalization, we denote

by $\widehat{\tau}$ the *reflexive closure* of τ , and by $\underline{\tau}$ its *anti-reflexive restriction* that is, $\tau \setminus \{(w, w) : w \in A^*\}$.

About the channel itself, the so-called *synchronization constraint* appears mandatory: it sets that, for each input word factorized $w = x_1 \cdots x_n$, where x_1, \dots, x_n are *codewords* in X , every output word has to be factorized $w' = x'_1 \cdots x'_n$, with $(x_1, x'_1), \dots, (x_n, x'_n) \in \widehat{\tau}$. In order to ensure such a constraint, as for the Morse code, some pause symbol could be inserted after each codeword x_i .

Regarding the code X , in order to minimize the number of errors, in most cases some close neighborhood constraint is applied over $\widehat{\tau}(X)$, the set of the images of the codewords under $\widehat{\tau}$. In the most frequent use, such a constraint consists of some minimal distance condition: the smaller the distance between the input codeword $x \in X$ and any of its corresponding output words $x' \in \widehat{\tau}(X)$, the more optimal is error detection. In view of that, we fix over A^* a *quasi-metric* d , in the meaning of [36] (the difference with a *metric* is that d needs not to satisfy the symmetry axiom).

As outlined in [5], given an error tolerance level $k \geq 0$, a corresponding binary word relation, denoted in the present paper by $\tau_{d,k}$, can be associated in such a way that we set $(w, w') \in \tau_{d,k}$ (or equivalently, $w' \in \tau_{d,k}(w)$), whenever $d(w, w') \leq k$ holds.

Below, in the spirit of [14, 26], we draw some specification regarding error detection capability. Recall that a subset X of A^* is *independent* with respect to $\tau \subseteq A^* \times A^*$ (for short, *τ -independent*) iff. $\tau(X) \cap X = \emptyset$ holds. This notion, which appears dual with the one of *closed set* [26], relies on the famous *dependence systems* [6, 14]. In addition, given a family of codes, say \mathcal{F} , a code $X \in \mathcal{F}$ is *maximal in \mathcal{F}* whenever $X \subseteq Y$, with $Y \in \mathcal{F}$, implies $Y = X$.

Given a code $X \subseteq A^*$, we introduce the four following conditions:

- (c1) *Error detection: X is $\underline{\tau}_{d,k}$ -independent.*
- (c2) *Error correction: $x, y \in X$ and $\tau_{d,k}(x) \cap \tau_{d,k}(y) \neq \emptyset$ implies $x = y$.*
- (c3) *X is maximal in the family of $\underline{\tau}_{d,k}$ -independent codes.*
- (c4) *$\widehat{\tau}_{d,k}(X)$ is a code.*

A few comments on Conds. (c1)–(c4):

– By definition, Cond. (c1) is satisfied iff. the quasi-distance between pairs of different elements of X is greater than k that is, X is able to detect at most k errors in the transmission of any codeword.

– Cond. (c2) sets a classical definition: it expresses that if some transmission error has been detected in an output word, necessarily such a word comes from a unique input codeword.

– With Cond. (c3), in the family of $\underline{\tau}_{d,k}$ -independent codes, X cannot be improved. From this point of view, fruitful investigations have been done in several famous classes determined by code properties [13, 15, 18, 19].

– At last, Cond. (c4) expresses that the factorization of any output message over the set $\widehat{\tau}_{d,k}(X)$ is done in a unique way. Actually, since d is a quasi-metric,

the corresponding relation $\tau_{d,k}$ is reflexive, therefore Cond. **(c4)** is equivalent to $\tau_{d,k}(X)$ itself being a code.

Actually, in most of the cases it could be very difficult, even impossible, to satisfy all together Conds. **(c1)**–**(c4)**: for instance, as shown in [14, 26], there are regular codes satisfying **(c1)** that cannot satisfy **(c2)**. Furthermore some compromise has to be adopted: in view of this, given a regular code X , a natural question consists in examining whether each of those conds. is satisfied in the frameworks of classical free monoid quasi-metrics. From this point of view, in [26], we considered the so-called *edit relations*, some peculiar compositions of one-character *deletion*, *insertion*, and *substitution*: such relations involve the famous Levenshtein and Hamming metrics [11, 17, 21], which are priority related to *subsequences* in words. In the present paper, we focus on the following quasi-metrics, the two first ones involving *factors*:

- The *prefix* metric is defined by $d_P(w, w') = |w| + |w'| - 2|w \wedge w'|$, where $|w|$ stands for the length of the word w , and $w \wedge w'$ denotes the maximum length common *prefix* of w and w' : we set $\mathcal{P}_k = \tau_{d_P, k}$.

- The *factor* metric, for its part, is defined by $d_F(w, w') = |w| + |w'| - 2|f|$, where f is a maximum length common factor of w, w' : we set $\mathcal{F}_k = \tau_{d_F, k}$.

- A third type of topology can be introduced in connection with *monoid automorphisms* or *anti-automorphisms* (for short, we write *(anti-)automorphisms*): such a topology particularly involves the domain of DNA sequence comparison. By anti-automorphisms of the free monoid, we mean any one-to-one mapping onto A^* , say θ , st. the equation $\theta(uv) = \theta(v)\theta(u)$ holds for any $u, v \in A^*$: as in the case of automorphisms, each of those mappings actually extends to A^* some permutation of A (for involvements in the framework of closed codes, see [28]). With every (anti-)automorphism θ we associate the quasi-metric d_θ , defined as follows:

- (1) $d_\theta(w, w') = 0$ is equivalent to $w = w'$;
- (2) we set $d_\theta(w, w') = 1$ whenever $w' = \theta(w)$ holds, with $w \neq w'$;
- (3) in all other cases we set $d_\theta(w, w') = 2$.

It can be easily verified that, $k \geq 2$ implies $\tau_{d,k} = A^* \times A^*$. In other words, wrt. error detection constraints only the cond. $k = 1$ takes sense: by definition we have $\tau_{d_\theta, 1} = \widehat{\theta}$ and $\tau_{\theta, 1} = \underline{\theta}$.

This paper relates an extended and augmented version of the study we presented in [27]. In particular, answers are provided to some of the open questions that were asked: they concern the behavior of regular codes wrt. \mathcal{F}_k . We prove the following result:

Theorem. *With the preceding notation, given a regular code $X \subseteq A^*$, for every $k \geq 1$, it can be decided whether X satisfies any of Conds. **(c1)**–**(c4)** wrt. \mathcal{P}_k , \mathcal{F}_k , and $\widehat{\theta}$.*

Some comments about the proof:

- Regarding Cond. **(c1)**, we establish that, for each of the mentioned quasi-metrics, $\tau_{d,k}(X)$ is a *regular* subset of A^* . When d is the prefix metric, this is

done by proving that $\overline{\tau_{d,k}}$ itself is a regular relation that is, a regular subset of the monoid $A^* \times A^*$ in the sense of [9].

In the case where d corresponds to the factor metric, although the question of the regularity of $\overline{\tau_{d,k}}$ still remains open, the result is obtained thanks to the construction of a peculiar finite set covering for $\overline{\mathcal{F}_k} \subseteq A^* \times A^*$. Actually $\overline{\mathcal{F}_k}$ is precisely the union of the sets in that finite family. Moreover the so-called *conjugacy*, some concept from combinatorics on words [22] allows to prove that each of those sets is regular. Regarding $\widehat{\theta} = \tau_{d_\theta,1}$, we prove that in any case X satisfies Conds. **(c1)**, **(c2)**.

– In the case of the relation \mathcal{P}_k (resp., \mathcal{F}_k), we prove that X satisfies Cond. **(c2)** iff. it satisfies Cond. **(c1)** wrt. \mathcal{P}_{2k} (resp., \mathcal{F}_{2k}).

– Wrt. each of the quasi-metrics raised in the paper we established that, given a regular code $X \subseteq A^*$, X is maximal in the family of the codes independent wrt. $\overline{\tau_{d,k}}$ iff. it is *complete* that is, every word of A^* is a factor of some word in X^* , the free submonoid of A^* generated by X . Actually this is done by proving that any non-complete $\overline{\tau_{d,k}}$ -independent code can be embedded into some complete one: in other words it cannot be maximal. In order to establish such a property, in the spirit of [2, 20, 24, 25, 28, 37], we provide specific regularity-preserving embedding formulas:

their schemes are based upon the methodology from [7]. Notice that, in [13, 16, 18, 19, 35], wrt. peculiar families of sets, algorithmic methods for embedding a set into some maximal (but not necessarily complete) one were also provided.

– Regarding Cond. **(c4)**, for each of the preceding relations, the set $\widehat{\tau_{d,k}}(X) = \overline{\tau_{d,k}}(X)$ is regular, therefore in any case, by applying the famous Sardinas and Patterson algorithm [34], one can decide whether that cond. is satisfied.

We now shortly describe the contents of the paper:

– Section 2 is devoted to the preliminaries: we recall fundamental notions about words, word binary relations, regular sets, and codes. Taking account that d is a quasi-metric, wrt. $\overline{\tau_{d,k}}$ some equivalent formulations of the error correction cond. **(c2)** are provided.

– The aim of Sect. 3 is to study the relation \mathcal{P}_k . In addition, corresponding results wrt. the so-called *suffix* metric are set.

– Sect. 4 is devoted to a preliminary study about \mathcal{F}_k and $\overline{\mathcal{F}_k}$.

We construct the above-mentioned finite covering for $\overline{\mathcal{F}_k} \subseteq A^* \times A^*$.

Furthermore we prove that $\overline{\mathcal{F}_k}$ is regularity-preserving.

– The decidability results involving the factor metric are established in Sect. 5.

– Sect. 6 is devoted to quasi-metrics associated to (anti-)automorphisms.

– The paper concludes with some possible directions for further research (Sect. 7).

– In order to make the decidability results clearer, if needed, an appendix is added at the end of the paper. It provides some basic support in order to further implementing corresponding algorithms. Regarding the main study, in no way that appendix can constitute any prerequisite.

2 Preliminaries

Several definitions and notations have already been settled. In what follows, we bring precision about concepts such as words, automata, regular relations and variable-length codes. If necessary, we suggest the reader that he (she) report to classical books such as [1, 8, 12, 32]. Some classical decidability results are also set (for corresponding schemes of implementation see the appendix).

2.1 Words

In the whole paper, we fix a finite alphabet A , with $|A| \geq 2$. We denote by ε the *empty word* that is, the word with *length* 0, and we set $A^+ = A^* \setminus \{\varepsilon\}$. Given two words $v, w \in A^*$, v is a *prefix* (resp., *suffix*, *factor*) of w if words u, u' exist st. $w = vu$ (resp., $w = u'v$, $w = u'vu$). In the case where the equation $w = uv$ holds, we set $u = wv^{-1}$ and $v = u^{-1}w$. We denote by $P(w)$ (resp., $S(w)$, $F(w)$) the set of the words that are prefixes (resp., suffixes, factors) of w . In the case where we have $v \neq w$, with $v \in P(w)$ (resp., $v \in S(w)$), we say that v is a *proper prefix* (resp., *proper suffix*) of w . More generally, given $X \subseteq A^*$, we denote by $P(X)$ the union of the sets $P(x)$, for all the words $x \in X$ (the sets $S(X)$ and $F(X)$ are defined in a similar way). Given a word $w \in A^*$, we denote by w^R its *reversal* that is, for $a_1, \dots, a_n \in A$, we have $w^R = a_n \dots a_1$ whenever $w = a_1 \dots a_n$ holds.

A word $w \in A^*$ is *overlapping* whenever some $v \in A^*$ exists st. $wv \in A^*w$, with $1 \leq |v| \leq |w| - 1$; otherwise, w is *overlapping-free* that is, $wv \in A^*w$ with $|v| \leq |w| - 1$ implies $v = \varepsilon$. For instance, $w = ababa$ is overlapping: taking $v = ba$, we have $wv = (ababa)(ba) = (ab)(ababa) \in A^*w$; on the contrary, $w = ababb$ is overlapping-free. Note that w is overlapping-free iff. w^R itself is overlapping free. Classically, the following property holds (see e.g. [1, Proposition 1.3.6]):

Proposition 2.1. *Given $z_0 \in A^+$, let a be the initial letter of z_0 and let $b \in A \setminus \{a\}$. Then the word $z_0ab^{|z_0|}$ is overlapping-free.*

Two words $w, w' \in A^+$ are *conjugate* iff. a pair of words α, β exist st. $w = \alpha\beta$ and $w' = \beta\alpha$, with $\beta \neq \varepsilon$. The following result brings additional information:

Proposition 2.2. [22, Proposition 1.3.4] *Given a pair of non-empty words w, w' , the two following conds. are equivalent:*

- (i) w and w' are conjugate.
- (ii) $t \in A^*$ exists st. $wt = tw'$. More precisely $\alpha \in A^*$, $\beta \in A^+$, and $n \in \mathbb{N}$ exist st. $w = \alpha\beta$, $w' = \beta\alpha$, and $t \in (\alpha\beta)^n\alpha$.

2.2 Words binary relations

Let M be an arbitrary monoid. The following basic concepts are involved by our study:

Operations among subsets of M Given two sets $X, Y \subseteq M$, their *concatenation product* is $XY = \{xy : x \in X, y \in Y\}$ and the *Kleene star* of X is $X^* = \{x_1, \dots, x_n : x_1, \dots, x_n \in X, n \geq 1\} \cup \{\varepsilon\}$. Union, concatenation product, and Kleene star constitute the so-called *regular operations* into 2^M . In addition, the *left quotient* (resp., *right quotient*) of X by Y is $Y^{-1}X = \{z \in M : (\exists x \in X), (\exists y \in Y), x = yz\}$ (resp., $XY^{-1} = \{z \in M : (\exists x \in X), (\exists y \in Y), x = zy\}$).

M -automata A M -automaton, say \mathcal{A} , is defined on the basis of a finite labeled graph. The vertices are the so-called *states* and the M -labeled edges are the *transitions*: let Q and $E \subseteq Q \times M \times Q$ be the corresponding sets. The automaton is *finite* whenever E is a finite set. Two subsets of Q are actually distinguished namely I , the initial states, and T , the terminal ones. In the paper, the transition $(q, m, q') \in E$ is commonly denoted by $q \xrightarrow{m} q'$. Denoting by 1_M the identity element in the monoid M , for any $q \in Q$, $q \xrightarrow{1_M} q$ is a transition: accordingly we left it out in any graphical representation of \mathcal{A} . A *successful path* is a chain $(i \xrightarrow{m_0} q_1, q_1 \xrightarrow{m_1} q_2, \dots, q_n \xrightarrow{m_n} t)$, with $i \in I$ and $t \in T$. For convenience, we denote it by $i \xrightarrow{m_0} q_1 \xrightarrow{m_1} q_2 \dots q_n \xrightarrow{m_n} t$. The so-called *behavior* of \mathcal{A} , which we denote by $|\mathcal{A}|$, is the subset with elements all the labels $m_0 m_1 \dots m_n \in M$ of the corresponding successful paths. Classically, for any finite A^* -automaton \mathcal{A} there is another finite automaton \mathcal{A}' , with transitions labeled by characters of A , st. $|\mathcal{A}| = |\mathcal{A}'|$.

Regular sets A set $X \subseteq M$ is *regular* (or equivalently *rational*) if it belongs to the *regular closure* of the finite subsets of M that is, the smallest (wrt. the inclusion) subset of 2^M that contains the finite subsets and which is closed under the regular operations (see [32, Sect. II.1]). By definition, the family of regular sets is closed under the regular operations. The following property is attributed to Elgot and Mezei [9]:

Theorem 2.3. *Let M be a monoid and $X \subseteq M$. The following conds. are equivalent:*

- (i) X is regular.
- (ii) X is the behavior of some finite M -automaton.

Example 2.4. Let $M = \{a\}^* \times \{b\}^*$, and $X = \{(a^{n+1}, b^{2n}) : n \geq 0\}$. We have $X = YZ$, with $Y = \{(a, b^2)\}^*$ and $Z = \{(a, \varepsilon)\}$, whence X is regular. Actually, X is the behavior of the finite M -automaton represented in Figure 1.

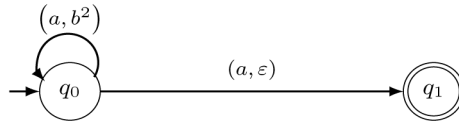


Figure 1: A M -automaton with behavior X : by convention, the initial state is identified with an input arrow and the terminal one with a double circle; transitions with label 1_M are left out.

In what follows, we recall some classical closure properties of regular sets:

Proposition 2.5. *Given a pair of monoids M, N , and a monoid homomorphism $h : M \rightarrow N$, the image of every regular subset of M under h is a regular subset of N .*

In the most general case the family of regular subsets of M is not closed neither under intersection, nor complementation, nor inverse monoid homomorphism. However, in the case where $M = A^*$, the following noticeable result holds:

Proposition 2.6. *The family of regular subsets of A^* is closed under boolean operations, regular operations, left-quotient (resp. right-quotient), and direct (resp. inverse) monoid endomorphism.*

In particular the finite union (resp., intersection, concatenation) of regular sets is itself regular.

Regular relations Let M, N be two monoids. A binary relation from M into N consists in any subset τ of $M \times N$. Since $M \times N$ itself is a monoid, those relations are directly involved by the preceding concept of regularity. The *composition* in this order of τ by τ' is defined by $\tau \cdot \tau'(x) = \tau'(\tau(x))$ (the notation τ^k refers to that operation). The *inverse* of τ is the relation $\tau^{-1} \subseteq N \times M$ defined by $(w, w') \in \tau^{-1}$ whenever $(w', w) \in \tau$; in addition its *complement* is $\bar{\tau} = M \times N \setminus \tau$. Composition and inverse preserve regularity among relations; however in the most general case a regular relation is not preserved under complementation. In the paper the following result (see [8, Sect. IX.3] or [32, Sect. IV.1.3]) will be frequently applied:

Proposition 2.7. *Given a regular relation $\tau \subseteq A^* \times A^*$, and a regular set $X \subseteq A^*$, the set $\tau(X)$ is regular.*

At last, $id_{A^*} = \{(w, w) | w \in A^*\}$ and its complement $\overline{id_{A^*}}$ are regular relations (see Fig. 2).

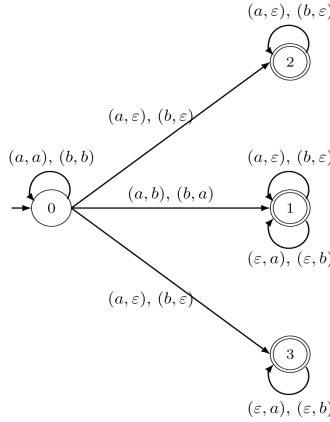


Figure 2: An automaton with behavior $\overline{id_{A^*}}$ in the case where $A = \{a, b\}$ (arrows are multi-labeled).

2.3 Variable-length codes

Given a subset X of A^* , and $w \in X^*$, let $x_1, \dots, x_n \in X$ st. w is the result of the concatenation of the words x_1, x_2, \dots, x_n , in this order. In view of specifying the factorization of w over X , we use the notation $w = (x_1)(x_2) \cdots (x_n)$, or equivalently: $w = x_1 \cdot x_2 \cdots x_n$. For instance, over the set $X = \{a, ab, ba\}$, the word $bab \in X^*$ can be factorized as $(ba)(b)$ or $(b)(ab)$ (equivalently denoted by $ba \cdot b$ or $b \cdot ab$).

A set X is a *variable-length code* (a *code* for short) if for any pair of finite sequences of words in X , say $(x_i)_{1 \leq i \leq n}$, $(y_j)_{1 \leq j \leq p}$, the equation $x_1 \cdots x_n = y_1 \cdots y_p$ implies $n = p$, and $x_i = y_i$ for each integer $i \in [1, n]$. In other words, every element of X^* has a unique factorization over X , equivalently the submonoid X^* is *free*. In particular, $X \neq \{\varepsilon\}$ is a *prefix* (resp., *suffix*) *code* whenever, for every pair of words $x, y \in X$, the cond. $x \in P(y)$ (resp., $x \in S(y)$) implies $x = y$. In addition, X is a *bifix* code if it is both a prefix code and a suffix one. Any set $X \neq \{\varepsilon\}$ whose elements have a common length is a complete bifix code, we say that it is a *uniform code*.

A positive *Bernoulli distribution* consists in a total mapping μ from the alphabet A into $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ (the set of the non-negative real numbers) st. $\sum_{a \in A} \mu(a) = 1$. Such a mapping is extended into a unique monoid homomorphism from A^* into (\mathbb{R}_+, \times) , which is itself extended into a unique positive *Bernoulli measure* $\mu : 2^{A^*} \rightarrow \mathbb{R}_+$. In order to do so, for each word $w \in A^*$, we set $\mu(\{w\}) = \mu(w)$; in addition given two disjoint subsets X, Y of A^* , we set $\mu(X \cup Y) = \mu(X) + \mu(Y)$. According to the famous *Kraft inequality*, for every code X we have $\mu(X) \leq 1$. In the whole paper, we take for μ the so-called *uniform* Bernoulli measure: it is determined by $\mu(a) = 1/|A|$, for each $a \in A$.

The two following results are classical: the first one is due to Schützenberger

(see [1, Theorem 2.5.16]) and the second provides some answer to a question actually set in [31].

Theorem 2.8. *Given a regular code $X \subseteq A^*$, the following conds. are equivalent:*

- (i) X is complete.
- (ii) X is a maximal code.
- (iii) $\mu(X) = 1$ holds.

Theorem 2.9. [7]. *Let $X \subseteq A^*$ be a non-complete code and let $z \notin F(X^*)$ be an overlapping-free word. Set $U = A^* \setminus (X^* \cup A^*zA^*)$ and $Y = (zU)^*z$. Then $Z = X \cup Y$ is a complete code.*

With the cond. of Theorem 2.9 if X is regular then the same holds for the resulting code Z . In addition, since z is an overlapping-free word in $A^* \setminus F(U)$, the following property holds: it will be applied further in the paper (see Prop. 3.11).

Lemma 2.10. *With the preceding notation, the set Uz is a prefix code.*

Finally, the following result is the basis of the decidability properties we establish in the paper:

Proposition 2.11. *Given a monoid M and a regular set $X \subseteq M$, what follows holds:*

- (i) In any case it can be decided whether $X = \emptyset$.
- (ii) If $M = A^*$ then it can be decided whether X is a code.
- (iii) If $M = A^*$ then one can decide whether $\mu(X) = 1$ holds.

2.4 Some equivalent formulation of the error correction condition.

Let A be an alphabet and $d : A^* \times A^* \rightarrow \mathbb{R}^+$ be a quasi-metric over A^* . Given a positive integer k , by definition we have $\tau_{d,k} \subseteq \tau_{d,k+1}$. In addition, if d is a metric then $\tau_{d,k}$ is a symmetric relation that is, the equation $\tau_{d,k}^{-1} = \tau_{d,k}$ holds. We close Sect. 2 by proving the following result, which will be applied several times in the rest of the paper:

Lemma 2.12. *Given a quasi-metric d over A^* and $X \subseteq A^*$, the following conds. are equivalent:*

- (i) X satisfies Cond. **(c2)** wrt. $\tau_{d,k}$.
- (ii) For every $x \in X$ we have $\tau_{d,k}^{-1}(\tau_{d,k}(x)) \cap X = \{x\}$.
- (iii) X satisfies Cond. **(c1)** wrt. $\tau_{d,k} \cdot \tau_{d,k}^{-1}$.

Proof. – Beforehand, we prove that any singleton satisfies each of Conds. (i)–(iii) of the lemma. Let $X = \{x\}$. By definition, trivially X satisfies Cond. **(c2)**. In addition, since $\tau_{d,k}$ is reflexive, we have $x \in \tau_{d,k}^{-1}(x)$, thus $x \in \tau_{d,k} \cdot \tau_{d,k}^{-1}(x)$. We obtain $\tau_{d,k} \cdot \tau_{d,k}^{-1}(X) \cap X = \{x\}$, whence X satisfies Cond. (ii). Finally, $\tau_{d,k} \cdot \tau_{d,k}^{-1}(x) \cap X = \{x\}$ implies $\tau_{d,k} \cdot \tau_{d,k}^{-1}(X) \cap X = \emptyset$, thus X satisfies Cond. **(c1)** wrt. $\tau_{d,k} \cdot \tau_{d,k}^{-1}$. In the rest of the proof we assume $|X| \geq 2$.

– In what follows we prove that Conds. (i) and (ii) are equivalent. Firstly, assuming that X satisfies Cond. (i), consider $x \in X$ and $y \in \tau_{d,k}^{-1}(\tau_{d,k}(x)) \cap X$. By construction $\tau_{d,k}(x) \cap \tau_{d,k}(y) \neq \emptyset$ holds, therefore we have $x = y$, thus $\tau_{d,k}^{-1}(\tau_{d,k}(x)) \cap X = \{x\}$. Conversely, assuming that Cond. (ii) holds, let $x, y \in X$ st. $\tau_{d,k}(x) \cap \tau_{d,k}(y) \neq \emptyset$. Some $z \in \tau_{d,k}(x) \cap \tau_{d,k}(y)$ exists, moreover we have $y \in \tau_{d,k}^{-1}(z) \subseteq \tau_{d,k}^{-1}(\tau_{d,k}(x))$. It follows from $y \in X$ and $\tau_{d,k}^{-1}(\tau_{d,k}(x)) \cap X = \{x\}$ that $y = x$.

– We prove that Cond. (i) implies Cond. (iii) in arguing by contrapositive. Assuming that X does not satisfy Cond. (iii), by definition we have $\tau_{d,k} \cdot \tau_{d,k}^{-1}(X) \cap X \neq \emptyset$, whence $x, y \in X$ exist st. $y \in \tau_{d,k} \cdot \tau_{d,k}^{-1}(x)$. By definition, we have $y \in \tau_{d,k} \cdot \tau_{d,k}^{-1}(x)$, with $y \neq x$. In other words $z \in A^*$ exists st. we have $z \in \tau_{d,k}(x)$ and $y \in \tau_{d,k}^{-1}(z)$ that is, $z \in \tau_{d,k}(x) \cap \tau_{d,k}(y)$, thus $\tau_{d,k}(x) \cap \tau_{d,k}(y) \neq \emptyset$. It follows from $y \neq x$ that X cannot satisfy Cond. **(c2)** wrt. $\tau_{d,k}$.

– Once more arguing by contrapositive, we prove that Cond. (iii) implies Cond. (ii). Assume that $x \in X$ exists st. $\tau_{d,k}^{-1}(\tau_{d,k}(x)) \cap X \neq \{x\}$. There is some $y \neq x$ st. $y \in \tau_{d,k} \cdot \tau_{d,k}^{-1}(X) \cap X$: by definition X cannot be $\tau_{d,k} \cdot \tau_{d,k}^{-1}$ -independent. \square

3 Error detection and the prefix metric

We start with a few examples:

Example 3.1. Over $A = \{a, b\}$, consider the finite prefix code $X = \{a, ba, b^2\}$.

– X satisfies Cond. **(c1)** wrt. \mathcal{P}_1 (that is, X is 1-error-detecting). Indeed, it follows from $a \wedge ba = a \wedge b^2 = \varepsilon$ and $ba \wedge b^2 = b$ that $d_P(a, ba) = |a| + |ba| = 3$, $d_P(a, b^2) = |a| + |b^2| = 3$, and $d_P(ba, b^2) = |ba| + |b^2| - 2|b| = 2$. Consequently, for each pair of different words $x, y \in X$, we have $(x, y) \notin \mathcal{P}_1$ that is, $\mathcal{P}_1(X) \cap X = \emptyset$.

– Regarding Cond. **(c2)**, in view of Lemma 2.12, we compute $\mathcal{P}_1 \cdot \mathcal{P}_1^{-1}(ba) \cap X = \mathcal{P}_1^2(ba) \cap X$. It follows from $\mathcal{P}_1(ba) = \{b, ba, ba^2, bab\}$ that:

$$\begin{aligned} \mathcal{P}_1^2(ba) &= \{\varepsilon, b, ba, b^2, ba^2, bab, ba^3, ba^2b, baba, bab^2\}, \text{ therefore:} \\ \mathcal{P}_1^2(ba) &= \{\varepsilon, b, b^2, ba^2, bab, ba^3, ba^2b, baba, bab^2\}. \end{aligned}$$

This implies $\mathcal{P}_1^2(ba) \cap X = \{b^2\}$, whence X cannot satisfy Cond. **(c1)** wrt. $\mathcal{P}_1 \cdot \mathcal{P}_1^{-1}$. According to Lemma 2.12, X cannot satisfy Cond. **(c2)** wrt. \mathcal{P}_1 .

– We have $\mu(X) = 1/2 + 1/4 + 1/4 = 1$ therefore, according to Theorem 2.8, X is a maximal code. Since it is \mathcal{P}_1 -independent, X is maximal in the family of \mathcal{P}_1 -independent codes that is, it satisfies Cond. **(c3)**.

– Since we have $X \subsetneq \mathcal{P}_1(X)$, and since X is a maximal code, the set $\widehat{\mathcal{P}}_1(X) = \mathcal{P}_1(X)$ cannot be a code that is, X cannot satisfy Cond. **(c4)** (we verify that we have $\varepsilon \in \mathcal{P}_1(X)$).

Example 3.2. Let $n \geq 2$ and $k \in [1, n-1]$. Consider the uniform code $X = A^n$.

– For any $x \in X$ we have $\mathcal{P}_k(x) \subseteq (A^{n-k} \cup \dots \cup A^{n-1}) \cup (A^{n+1} \cup \dots \cup A^{n+k})$: this implies $\mathcal{P}_k(X) \cap X = \emptyset$, thus X satisfies Cond. **(c1)** wrt. \mathcal{P}_k .

- However, Cond. **(c2)** is not satisfied by X : indeed, given two different characters a, b , we have $a^{n-1} \in \mathcal{P}_k(a^n) \cap \mathcal{P}_k(a^{n-1}b)$.
- As mentioned in the preliminaries X is complete. According to Theorem 2.8 X is a maximal (bifix) code, hence it is maximal in the family of $\underline{\mathcal{P}}_k$ -independent codes that is, X satisfies Cond. **(c3)** wrt. \mathcal{P}_k .
- We have $X \subsetneq \mathcal{P}_k(X)$: since X is a maximal code, it cannot satisfy Cond. **(c4)**.

Example 3.3. Over the alphabet $A = \{a, b\}$, consider the regular bifix code $X = \{ab^n a : n \geq 0\} \cup \{ba^n b : n \geq 0\}$ and the relation \mathcal{P}_1 .

– It follows from $\mathcal{P}_1(X) = \bigcup_{n \geq 0} \{ab^n, ab^n a^2, ab^n ab, ba^n, ba^n ba, ba^n b^2\}$, that $\underline{\mathcal{P}}_1(X) \cap X = \emptyset$, whence X satisfies Cond. **(c1)**.

– For $n \neq 0$ we have: $\underline{\mathcal{P}}_1 \cdot \underline{\mathcal{P}}_1^{-1}(ab^n a) = \underline{\mathcal{P}}_1^2(ab^n a) \setminus \{ab^n a\}$, thus:

$\underline{\mathcal{P}}_1 \cdot \underline{\mathcal{P}}_1^{-1}(ab^n a) = \{ab^n, ab^{n-1}, ab^{n+1}, ab^n a^2, ab^n a^3, ab^n a^2 b, ab^n ab, ab^n aba, ab^n ab^2\}$. Similarly, for $n = 0$ we have $\underline{\mathcal{P}}_1 \cdot \underline{\mathcal{P}}_1^{-1}(ab^n a) = \underline{\mathcal{P}}_1^2(a^2) = \mathcal{P}_1(\{a, a^2, a^3, a^2 b\}) \setminus \{a^2\} = \{\varepsilon, a, ab, a^3, a^2 b, a^4, a^3 b, a^2 b, a^2 ba, a^2 b^2\}$. In any case we obtain $X \cap \underline{\mathcal{P}}_1 \cdot \underline{\mathcal{P}}_1^{-1}(ab^n a) = \emptyset$. Similarly, we have $X \cap \underline{\mathcal{P}}_1 \cdot \underline{\mathcal{P}}_1^{-1}(ba^n b) = \emptyset$. Consequently, X is $\underline{\mathcal{P}}_1 \cdot \underline{\mathcal{P}}_1^{-1}$ -independent therefore, according to Lemma 2.12, X satisfies Cond. **(c2)** wrt. \mathcal{P}_1 .

– Regarding Cond. **(c3)**, we have $\mu(X) = 2 \cdot 1/4 \sum_{n \geq 0} (1/2)^n = 1$: according to Theorem 2.8, X is a maximal code, whence it is maximal in the family of $\underline{\mathcal{P}}_1$ -independent codes.

– Since we have $X \subsetneq \mathcal{P}_1(X)$, X cannot satisfy Cond. **(c4)** (we verify that $a, a^2 \in \mathcal{P}_1(X)$).

3.1 A preliminary study of the relation \mathcal{P}_k

Given a pair of words w, w' , let $p = w \wedge w'$ and let u, u' be the unique pair of words st. $w = pu$ and $w' = pu'$. By definition we have $u \wedge u' = \varepsilon$ and $d_P(w, w') = |w| + |w'| - 2|p| = |u| + |u'|$, therefore the following property comes from the definition of \mathcal{P}_k (see Fig. 3):

Claim 1. *With the preceding notation, each of the following properties holds:*

- (i) $(w, w') \in \mathcal{P}_k$ is equivalent to $0 \leq |u| + |u'| \leq k$.
- (ii) $(w, w') \in \underline{\mathcal{P}}_k$ is equivalent to $1 \leq |u| + |u'| \leq k$.

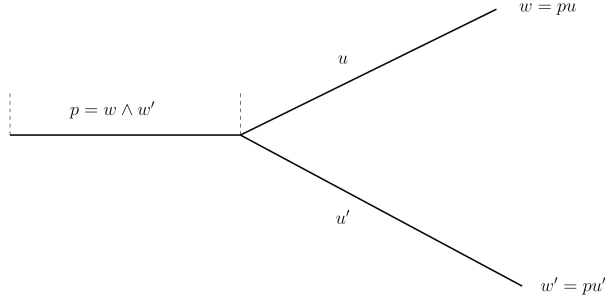


Figure 3: We have $(w, w') \in \mathcal{P}_k$ iff. $|u| + |u'| \leq k$.

In order to prove the further lemma 3.5, it is convenient to temporarily move into the more general framework of the factor metric:

Lemma 3.4. *Given a positive integer pair k, k' we have $\mathcal{F}_k \cdot \mathcal{F}_{k'} = \mathcal{F}_{k+k'}$.*

Proof. We assume wlog. $k \leq k'$. Firstly, we consider a pair of words $(w, w') \in \mathcal{F}_k \cdot \mathcal{F}_{k'}$. By definition, some word $w'' \in A^*$ exists st. we have $(w, w'') \in \mathcal{F}_k$, $(w'', w') \in \mathcal{F}_{k'}$ that is, $d_{\mathbb{F}}(w, w'') \leq k$ and $d_{\mathbb{F}}(w'', w') \leq k'$. This implies $d_{\mathbb{F}}(w, w') \leq d_{\mathbb{F}}(w, w'') + d_{\mathbb{F}}(w'', w') \leq k + k'$, thus $(w, w') \in \mathcal{F}_{k+k'}$. Conversely let $(w, w') \in \mathcal{F}_{k+k'}$, and let f be a word with maximum length in $\mathbb{F}(w) \cap \mathbb{F}(w')$. Regarding $|w| - |f|$ and $|w'| - |f|$, exactly one of the two following cases occurs:

(a) *The case where at least one of the integers $|w| - |f|$, $|w'| - |f|$ belongs to $[k, k+k']$.* Since $\mathcal{F}_{k+k'}$ is a symmetric relation, wlog. we assume $k \leq |w| - |f| \leq k + k'$. With this condition a pair of words u, v exist st. $w \in A^h u f v A^{h'}$, with $h + h' = k$. By construction we have $|u f v| = |w| - k$ and $d_{\mathbb{F}}(w, u f v) = k$, thus $(w, u f v) \in \mathcal{F}_k$. We observe that $t_1 f t_2 \in \mathbb{F}(u f v) \cap \mathbb{F}(w')$ implies $t_1 f t_2 \in \mathbb{F}(w) \cap \mathbb{F}(w')$. Since f is a maximum length word in $\mathbb{F}(w) \cap \mathbb{F}(w')$ we obtain $t_1 = t_2 = \varepsilon$. In other words, f remains a maximum length word in $\mathbb{F}(u f v) \cap \mathbb{F}(w')$. This implies $d_{\mathbb{F}}(u f v, w') = |u f v| + |w'| - 2|f| = (|w| - k) + |w'| - 2|f| = (|w| + |w'| - 2|f|) - k = d_{\mathbb{F}}(w, w') - k$. It follows from $d_{\mathbb{F}}(w, w') \leq k + k'$ that $d_{\mathbb{F}}(u f v, w') \leq k'$, thus $(u f v, w') \in \mathcal{F}_{k'}$. Since we have $(w, u f v) \in \mathcal{F}_k$, this implies $(w, w') \in \mathcal{F}_k \cdot \mathcal{F}_{k'}$.

(b) *The case where we have $0 \leq |w| - |f| < k$ and $0 \leq |w'| - |f| < k$.* Since we assume $k \leq k'$, by definition we have $d_{\mathbb{F}}(w, f) < k$, $d_{\mathbb{F}}(f, w') < k \leq k'$ that is, $(w, f) \in \mathcal{F}_k$, $(f, w') \in \mathcal{F}_{k'}$, thus $(w, w') \in \mathcal{F}_k \cdot \mathcal{F}_{k'}$. \square

Lemma 3.4 will be further applied in Sect. 4; regarding \mathcal{P}_k , it leads to the following statement:

Lemma 3.5. *Given a positive integer pair k, k' we have $\mathcal{P}_k \cdot \mathcal{P}_{k'} = \mathcal{P}_{k+k'}$.*

Proof. Firstly, in the proof of Lemma 3.4, by merely substitute $d_{\mathbb{P}}$ to $d_{\mathbb{F}}$, the condition $(w, w') \in \mathcal{P}_k \cdot \mathcal{P}_{k'}$ implies $(w, w') \in \mathcal{P}_{k+k'}$. Conversely, take for f a

maximum length word in $P(w) \cap P(w')$. Once more by substituting d_P to d_F exactly one of the two following conds. can occur:

(a) In the case where $|w| - |f| \in [k, k + k']$ holds, we have $w \in A^h u f v A^{h'}$, with $h = 0$, $h' = k$, and $u = \varepsilon$. We obtain $(w, f v) \in \mathcal{P}_k$ and $(f v, w') \in \mathcal{P}_{k'}$, thus $(w, w') \in \mathcal{P}_k \cdot \mathcal{P}_{k'}$.

(b) In the case where both the conditions $0 \leq |w| - |f| < k$ and $0 \leq |w'| - |f| < k$ hold, we directly obtain $(w, f) \in \mathcal{P}_k$, $(f, w') \in \mathcal{P}_{k'}$, thus $(w, w') \in \mathcal{P}_k \cdot \mathcal{P}_{k'}$. \square

Consequently the equations $\mathcal{F}_k \cdot \mathcal{F}_{k'} = \mathcal{F}_{k'} \cdot \mathcal{F}_k$ and $\mathcal{P}_k \cdot \mathcal{P}_{k'} = \mathcal{P}_{k'} \cdot \mathcal{P}_k$ hold. In addition, the following property holds:

Corollary 3.6. *Given a pair of positive integer k, n we have $\mathcal{P}_{nk} = \mathcal{P}_k^n$.*

Proof. We argue by induction on $n \geq 1$. Trivially, the equation holds for $n = 1$. Assuming that $\mathcal{P}_{nk} = \mathcal{P}_k^n$ holds, according to Lemma 3.5 we obtain $\mathcal{P}_{(n+1)k} = \mathcal{P}_{nk} \cdot \mathcal{P}_k = \mathcal{P}_k^n \cdot \mathcal{P}_k = \mathcal{P}_k^{n+1}$. \square

As another consequence of Lemma 3.5, in the framework of the prefix metric, Lemma 2.12 leads to the following result:

Lemma 3.7. *Given a set $X \subseteq A^*$ and $k \geq 1$ the three following conds. are equivalent:*

- (i) X satisfies Cond. **(c2)** wrt. \mathcal{P}_k .
- (ii) For every $x \in X$ we have $\mathcal{P}_{2k}(x) \cap X = \{x\}$.
- (iii) X satisfies Cond. **(c1)** wrt. \mathcal{P}_{2k} .

Proof. Since the relation \mathcal{P}_k is symmetric we have $\mathcal{P}_k \cdot \mathcal{P}_k^{-1} = \mathcal{P}_k^2$. By taking $n = 2$ in the statement of Corollary 3.6, we obtain $\mathcal{P}_k \cdot \mathcal{P}_k^{-1} = \mathcal{P}_{2k}$. The rest of the proof merely consists in substituting \mathcal{P}_k to $\tau_{d,k}$ in the proof of Lemma 2.12. \square

3.2 On the regularity of $\underline{\mathcal{P}}_k$

In view of Claim 1 we introduce the three following sets: E stands for the set of all the pairs of non-empty words (u, u') , with different initial characters, and $\text{st. } |u| + |u'| \leq k$. In addition F (resp., G) stands for the set of all the pairs (u, ε) (resp., (ε, u')), with $1 \leq |u| \leq k$ (resp., $1 \leq |u'| \leq k$). By construction, E , F , and G are finite sets. Regarding Conds. **(c1)**, **(c2)**, the following property will have noticeable involvement:

Proposition 3.8. *For every $k \geq 1$, both the relations \mathcal{P}_k and $\underline{\mathcal{P}}_k$ are regular.*

Proof. – In what follows we indicate the construction of a finite $A^* \times A^*$ -automaton with behavior $\underline{\mathcal{P}}_k$, namely $\mathcal{R}_{P,k}$. The states are 0, 1, 2, 3, the unique initial one being 0, and the terminal being 1, 2, 3. The transitions are listed hereunder (see Fig 4):

- $0 \xrightarrow{(a,a)} 0$, for every $a \in A$;
- $0 \xrightarrow{(u,u')} 2$ for every $(u, u') \in E$;

$$0 \xrightarrow{(u,\varepsilon)} 1 \text{ for every } (u,\varepsilon) \in F;$$

$$0 \xrightarrow{(\varepsilon,u')} 3 \text{ for every } (\varepsilon,u') \in G.$$

– Let $(w, w') \in |\mathcal{R}_{\mathcal{P},k}|$, and let $p = w \wedge w'$. By construction, there are $t \in \{1, 2, 3\}$ and $(u, u') \in E \cup F \cup G$, st. $w = pu$, $w' = pu'$, with $0 \xrightarrow{(p,p)} 0 \xrightarrow{(u,u')} t$ being a successful path. Since $(u, u') \in E \cup F \cup G$ implies $1 \leq |u| + |u'| \leq k$, we are in the cond. (ii) of Claim 1, hence $(w, w') \in \underline{\mathcal{P}}_k$ holds.

– Conversely, assume $(w, w') \in \underline{\mathcal{P}}_k$. According to Claim 1, words p , u , and u' exist st. $p = w \wedge w'$, $w = pu$, $w' = pu'$, $u \wedge u' = \varepsilon$, and $1 \leq |u| + |u'| \leq k$. The last cond. implies that at least one of the two words u, u' is non-empty. The cond. $u \neq \varepsilon$ with $u' \neq \varepsilon$ implies $(u, u') \in E$. Similarly, $u \neq \varepsilon$ and $u' = \varepsilon$ (resp., $u = \varepsilon$ and $u' \neq \varepsilon$), implies $(u, u') \in F$ (resp., $(u, u') \in G$). In any case we obtain $(u, u') \in E \cup F \cup G$, therefore by construction some $t \in \{1, 2, 3\}$ exists st. $0 \xrightarrow{(p,p)} 0 \xrightarrow{(u,u')} t$ is a successful path in $\mathcal{R}_{\mathcal{P},k}$. In other words we have $(w, w') = (pu, pu') \in |\mathcal{R}_{\mathcal{P},k}|$.

– As a consequence, we have $\underline{\mathcal{P}}_k = |\mathcal{R}_{\mathcal{P},k}|$: according to Theorem 2.3, the relation $\underline{\mathcal{P}}_k$ is regular. In addition, according to Prop. 2.6 the relation $\mathcal{P}_k = \underline{\mathcal{P}}_k \cup id_{A^*}$ itself is regular. \square

Example 3.9. Let $k = 2$. We have $E = \{(a, b) : a, b \in A, a \neq b\}$, $F = \{(a, \varepsilon) : a \in A\} \cup \{(ab, \varepsilon) : a, b \in A\}$, and $G = \{(\varepsilon, a) : a \in A\} \cup \{(\varepsilon, ab) : a, b \in A\}$ (see Fig. 4).

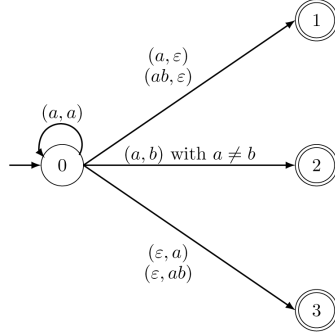


Figure 4: The case where we have $k = 2$: in the automaton $\mathcal{R}_{\mathcal{P},k}$, the edges are multi-labeled (a, b stand for every pair of characters in A).

We note that in [29] the author introduces an interesting $(A^* \times A^*) \times \mathbb{N}$ -automaton (equivalently transducer with input in $A^* \times A^*$ and output in \mathbb{N}) (see Fig. 5). This automaton allows to compute $d_{\mathcal{P}}$ as follows: for every $(w, w') \in A^* \times A^*$, the distance $d_{\mathcal{P}}(w, w')$ is the least $d \in \mathbb{N}$ for which $((w, w'), d)$ is the label of some successful path. Furthermore, an alternative proof of

the regularity of \mathcal{P}_k can be obtained. Indeed, by construction, denoting by $\mathcal{D} \subseteq (A^* \times A^*) \times \mathbb{N}$ the behavior of such an automaton, we have $\mathcal{P}_k = \mathcal{D}^{-1}([1, k])$. Since $[1, k]$ is a finite subset of the one-generator monoid \mathbb{N} , it is regular. In addition, since regular relations are closed under inverse \mathcal{P}_k itself is regular. However, we note that such a construction cannot involve the relation $\underline{\mathcal{P}}_k$ itself that is, it does not affect Cond. **(c1)**.

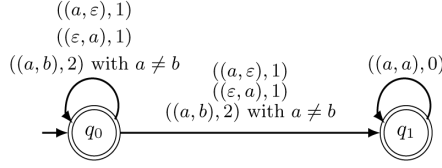


Figure 5: The automaton from [29] in the case where we have $A = \{a, b\}$

As a consequence of Prop. 3.8, we obtain the following noticeable result:

Proposition 3.10. *Given a regular code $X \subseteq A^*$ and $k \geq 1$, it can be decided whether X satisfies any of Conds. **(c1)**, **(c2)**, and **(c4)** wrt. \mathcal{P}_k .*

Proof. Let X be a regular code. We consider one by one our conditions:

– *Cond. (c1)* According to Prop. 3.8, the relation $\underline{\mathcal{P}}_k$ is regular; according to Prop. 2.7, $\underline{\mathcal{P}}_k(X)$ is a regular subset of A^* therefore, according to Prop. 2.6, the set $\underline{\mathcal{P}}_k(X) \cap X$ itself is regular. Consequently, according to Prop. 2.11, one can decide whether $\underline{\mathcal{P}}_k(X) \cap X \neq \emptyset$.

– *Cond. (c2)*: According to Lemma 3.7 the set X satisfies that cond. wrt. \mathcal{P}_k iff. it satisfies Cond. **(c1)** wrt. \mathcal{P}_{2k} : in view of the above, this can be decided.

– *Cond. (c4)*: Since X is regular, according to Props. 2.7, 3.8, the set $\widehat{\mathcal{P}}_k(X) = \underline{\mathcal{P}}_k(X)$ is regular. According to Prop. 2.11, one can decide whether this set is a code. \square

3.3 Maximal \mathcal{P}_k -independent codes

Regarding Cond. **(c3)**, we start with the following result:

Proposition 3.11. *Every regular $\underline{\mathcal{P}}_k$ -independent code can be embedded into some complete one.*

Proof. Let X be a regular $\underline{\mathcal{P}}_k$ -independent code. The result is trivial if X is complete: in the sequel we assume X being non-complete. By definition, a word z_0 exists in $A^* \setminus F(X^*)$. Without loss of generality, we assume $|z_0| \geq k$: otherwise, we substitute to z_0 some word z_0u , with $|u| = k - |z_0|$ (it follows from $z_0 \in F(z_0u)$ that $z_0u \notin F(X^*)$). Let a be the initial character of z_0 , let b

be a character different of a , and let $z = z_0ab^{|z_0|}$. According to Lemma 2.1, z is overlapping-free. We introduce the three following sets: $U = A^* \setminus (X^* \cup A^*zA^*)$, $Y = z(Uz)^*$, and $Z = X \cup Y$. According to Theorem 2.9, Z is a regular complete code. We will prove that Z is $\overline{\mathcal{P}_k}$ -independent that is, $\overline{\mathcal{P}_k}(X \cup Y) \cap (X \cup Y) = \emptyset$, thus $\overline{\mathcal{P}_k}(X) \cap X = \overline{\mathcal{P}_k}(X) \cap Y = \overline{\mathcal{P}_k}(Y) \cap X = \overline{\mathcal{P}_k}(Y) \cap Y = \emptyset$. Actually, since X is $\overline{\mathcal{P}_k}$ -independent, we already have $\overline{\mathcal{P}_k}(X) \cap X = \emptyset$.

(a) Firstly, we prove that $\overline{\mathcal{P}_k}(X) \cap Y = \emptyset$. By contradiction assume that a pair of words $x \in X$ and $y \in Y$ exist st. $(x, y) \in \overline{\mathcal{P}_k}$. By construction we have $|z_0| \geq k$. According to Claim 1 we obtain $1 \leq |(x \wedge y)^{-1}x| + |(x \wedge y)^{-1}y| \leq k \leq |z_0|$. This implies $|(x \wedge y)^{-1}y| \leq |z_0|$, thus $|x \wedge y| \geq |y| - |z_0|$. By construction, $|y| \geq |z|$ holds: we obtain $|x \wedge y| \geq |z_0ab^{|z_0|}| - |z_0| = |ab^{|z_0|}|$, thus $|x \wedge y| \geq |z_0| + 1$. Since both the words $x \wedge y$ and z_0 are prefixes of y , this implies $z_0 \in P(x \wedge y)$, thus $z_0 \in P(x)$: a contradiction with $z_0 \notin F(X^*)$.

(b) Now, by contradiction we assume $\overline{\mathcal{P}_k}(Y) \cap X \neq \emptyset$. Let $y \in Y$ and $x \in X$ st. $(y, x) \in \overline{\mathcal{P}_k}$. Since \mathcal{P}_k and $\overline{id_{A^*}}$ are symmetrical relations, $\overline{\mathcal{P}_k} = \mathcal{P}_k \cap \overline{id_{A^*}}$ itself is symmetrical. We obtain $(x, y) \in \overline{\mathcal{P}_k}$, thus $\overline{\mathcal{P}_k}(X) \cap Y \neq \emptyset$: this contradicts the conclusion of the preceding case (a).

(c) It remains to prove that $\overline{\mathcal{P}_k}(Y) \cap Y = \emptyset$. Once more arguing by contradiction, we assume that $y, y' \in Y$ exist st. $(y, y') \in \overline{\mathcal{P}_k}$. Let $p = y \wedge y'$, $u = p^{-1}y$, and $u' = p^{-1}y'$. We compare the words p , y , and y' :

(c1) At first, we assume $p \in \{y, y'\}$ that is, wlog. $y' = p$, thus $y = y'u$. More precisely, it follows from $(y, y') \in \overline{\mathcal{P}_k}$ that $y \neq y'$, hence y' is a proper prefix of y . By construction we have $y, y' \in z(Uz)^*$ that is, two sequences of words in U , namely u_1, \dots, u_m and u'_1, \dots, u'_n , exist st. $y = zu_1zu_2 \dots u_mz$ and $y' = zu'_1z \dots u'_nz$. It follows from $y = y'u$ that $zu_1zu_2 \dots u_mz = zu'_1z \dots u'_nz$, thus $u_1zu_2 \dots u_mz = u'_1z \dots u'_nz$. With this condition, at least one of the two words u_1z and u'_1z is a prefix of the other one. Since z is overlapping free, according to Lemma 2.10 this implies $u_1z = u'_1z$, thus $u_1 = u'_1$. By induction each of the equations $u_2 = u'_2, \dots$, and $u_n = u'_n$ also holds. From the fact that y' is a proper prefix of y , we obtain $m \geq n + 1$ and $u = u_{n+1}z \dots u_mz$. This implies $|u| \geq |z| \geq |z_0| + 1$, thus $|u| \geq k + 1$: a contradiction with $|u| = d_P(y, y') \leq k$.

(c2) Consequently we have $p \notin \{y, y'\}$, thus $u \neq \varepsilon$ and $u' \neq \varepsilon$. By construction, $b^{|z_0|}$ is a suffix of z , which itself is a suffix of both the words $y, y' \in Y$. Consequently, at least one of the words u and $b^{|z_0|}$ is a suffix of the other one; similarly at least one of the two words u' and $b^{|z_0|}$ is a suffix of the other one. It follows from $1 \leq |u| \leq k \leq |z_0|$ and $1 \leq |u'| \leq k \leq |z_0|$ that u, u' are non-empty suffixes of $b^{|z_0|}$ that is, $u, u' \in bb^*$. But this cond. implies that the word pb remains a prefix of both the words y and y' : a contradiction with $p = y \wedge y'$ (see Fig. 6). \square

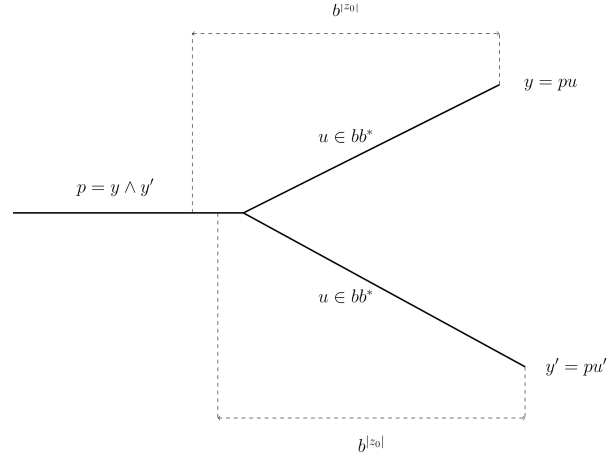


Figure 6: Proof of Prop. 3.11: the case where we have $y, y' \in Y$ and $(y, y') \in \mathcal{P}_k$, with $p \notin \{y, y'\}$.

As a consequence, we obtain the following result:

Proposition 3.12. *Given a regular $\underline{\mathcal{P}}_k$ -independent code $X \subseteq A^*$, the four following conds. are equivalent:*

- (i) X is maximal in the family of $\underline{\mathcal{P}}_k$ -independent codes.
- (ii) X is a maximal code.
- (iii) X is complete.
- (iv) We have $\mu(X) = 1$.

Proof. According to Theorem 2.8 for every regular code, Conds. (ii), (iii), and (iv) are equivalent. Trivially Cond. (ii) implies Cond. (i). We prove that Cond. (i) implies Cond. (iii) in arguing by contrapositive. Assuming X non-complete, according to Prop. 3.11, a regular $\underline{\mathcal{P}}_k$ -independent code strictly containing X exists, hence X is not maximal as a $\underline{\mathcal{P}}_k$ -independent code. \square

Corollary 3.13. *Every non-maximal regular $\underline{\mathcal{P}}_k$ -independent code can be embedded into some maximal one.*

Proof. Let X be a regular non-maximal regular $\underline{\mathcal{P}}_k$ -independent code. According to Prop. 3.12, X is non-complete. According to Prop. 3.11, a complete regular $\underline{\mathcal{P}}_k$ -independent code Y exists st. $X \subsetneq Y$. Once more according to Prop. 3.12, Y is maximal as a regular $\underline{\mathcal{P}}_k$ -independent code. \square

Finally, we obtain the following result:

Proposition 3.14. *One can decide whether a given regular code $X \subseteq A^*$ satisfies Cond. (c3) wrt. \mathcal{P}_k .*

Proof. Once more according to Prop. 3.12, the code X satisfies Cond. (c3) iff. $\mu(X) = 1$ holds. According to Prop. 2.11, one can decide whether X satisfies the last condition. \square

3.4 The suffix metric and the relation \mathcal{S}_k

Given a pair of words w, w' , their *suffix* distance is $d_S = |w| + |w'| - 2|s|$, where s denotes the longest word in $S(w) \cap S(w')$: set $\mathcal{S}_k = \tau_{d_S, k}$. For every pair $w, w' \in A^*$, we have $d_S(w, w') = d_P(w^R, w'^R)$, hence $(w, w') \in \mathcal{S}_k$ is equivalent to $(w^R, w'^R) \in \mathcal{P}_k$. In particular, starting from the preceding automaton $\mathcal{R}_{\mathcal{P}, k}$ (see proof of Prop. 3.8), an automaton with behavior $|\underline{\mathcal{S}_k}|$ can be constructed, whence $\underline{\mathcal{S}_k}$ is a regular relation (see Fig. 7).

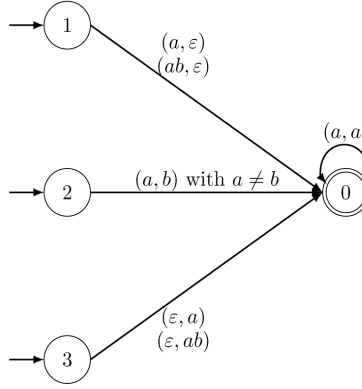


Figure 7: An automaton with behavior $\underline{\mathcal{S}_2}$.

By applying the preceding study to the reversal words, we obtain the following result:

Proposition 3.15. *Given a regular code, one can decide whether it satisfies any of Conds. (c1)–(c4) wrt. \mathcal{S}_k .*

4 On the combinatorial structure of $\underline{\mathcal{F}_k}$

As in the preceding section, we start with some examples:

Example 4.1. Over $A = \{a, b\}$, consider the code $X = \{a, ba, bb\}$

- X cannot satisfy Cond. (c1) wrt. \mathcal{F}_1 : indeed, we have $ba \in \underline{\mathcal{F}_1}(a) \cap X$.
- Since we have $\mathcal{F}_1 \cdot \mathcal{F}_1^{-1}(a) \cap X = \{a, ba\}$, X cannot satisfy the cond. (ii) in Lemma 2.12, therefore it cannot satisfy Cond. (c2).
- We have $\mu(X) = \mu(a) + \mu(b)\mu(a) + \mu(b)^2 = 1/2 + 1/4 + 1/4 = 1$, therefore X is a maximal code. However, since it is not a $\underline{\mathcal{F}_1}$ -independent codes, X cannot satisfy Cond. (c3) wrt. \mathcal{F}_1 .
- Finally, it follows from $\varepsilon \in \mathcal{F}_1(X)$ that $\widehat{\mathcal{F}_1}(X) = X \cup \mathcal{F}_1(X)$ is not a code, whence it cannot satisfy Cond. (c4).

Example 4.2. Take $A = \{a, b\}$ and consider the context-free bifix code $X = \{a^n b^n : n \geq 1\}$.

– It follows from $\overline{\mathcal{F}_1}(X) = \bigcup_{n \geq 1} \{a^{n-1}b^n, a^{n+1}b^n, ba^n b^n, a^n b^{n-1}, a^n b^n a, a^n b^{n+1}\}$ that $\overline{\mathcal{F}_1}(X) \cap X = \emptyset$, thus X is 1-error-detecting wrt. \mathcal{F}_1 (Cond. **(c1)**).

– Regarding error correction, we have $a^{n+1}b^{n+1} \in \mathcal{F}_1(a^n b^{n+1}) \subseteq \mathcal{F}_1^2(a^n b^n)$, thus $a^{n+1}b^{n+1} \in \mathcal{F}_1^2(a^n b^n)$. This implies $\overline{\mathcal{F}_1} \cdot \overline{\mathcal{F}_1}^{-1}(a^n b^n) \cap X \neq \{a^n b^n\}$, whence X cannot satisfy the cond. (ii) in Lemma 2.12 that is, X cannot satisfy Cond. **(c2)** wrt. \mathcal{F}_1 .

– We have $\mu(X) = \sum_{n \geq 1} (\frac{1}{4})^n = 1/3 < 1$, whence X does not satisfy Cond. **(c3)**.

– Finally, since we have $(a^n b^{n-1})(ba^n b^n) = (a^n b^n)(a^n b^n)$, the set $\widehat{\mathcal{F}_1}(X) = \mathcal{F}_1(X)$ no more satisfies Cond. **(c4)**.

4.1 A few generalities about $\overline{\mathcal{F}_k}$

Given $w, w' \in A^*$, let f be a maximum length word in $F(w) \cap F(w')$ and let (u, v, u', v') be a tuple of words st. $w = ufv = u'f v'$: we have $d_F(w, w') = |w| + |w'| - 2|f| = |u| + |v| + |u'| + |v'|$. The following statement, which comes from the definition, provides an extension of Claim 1 in the framework of the factor metric (see Fig. 8):

Claim 2. *With the preceding notation, each of the following properties holds:*

- (i) $(w, w') \in \mathcal{F}_k$ is equivalent to $0 \leq |u| + |v| + |u'| + |v'| \leq k$.
- (ii) $(w, w') \in \underline{\mathcal{F}_k}$ is equivalent to $1 \leq |u| + |v| + |u'| + |v'| \leq k$.

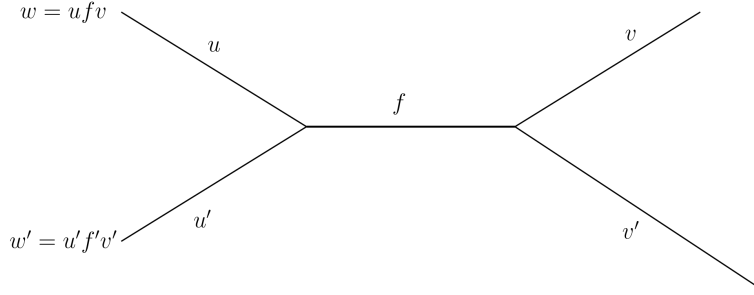


Figure 8: We have $(w, w') \in \mathcal{F}_k$ iff. $|u| + |v| + |u'| + |v'| \leq k$.

As attested by the following example, a maximum length word in $F(w) \cap F(w')$ needs not to be unique. Because of this, the proof of Prop. 3.8 cannot be extended to the framework of $\underline{\mathcal{F}_k}$.

Example 4.3. $w = cccbabababbab$, $w' = bbabbaababaacca$. There are two words of maximum length in $F(w) \cap F(w')$, namely $f_1 = ababa$ and $f_2 = babba$.

Since d_F is a metric, the relation \mathcal{F}_k is reflexive and symmetric. In addition, according to Lemma 3.4, Corollary 3.6, directly translates into the following statement:

Corollary 4.4. *Given a pair of positive integer k, n we have $\mathcal{F}_{nk} = \mathcal{F}_k^n$.*

As a consequence of Corollary 4.4, we obtain the following result:

Lemma 4.5. *Given a positive integer k we have $\mathcal{F}_k = \mathcal{F}_1^k = (\mathcal{P}_1 \cup \mathcal{S}_1)^k$.*

Proof. – With the condition of Corollary 4.4, by taking $n = 1$: we obtain $\mathcal{F}_1^k = \mathcal{F}_k$.

– Trivially, we have $\mathcal{P}_1 \cup \mathcal{S}_1 \subseteq \mathcal{F}_1$. For the converse, let $(w, w') \in \mathcal{F}_1$, let f be a maximum length word in $F(w) \cap F(w')$, and let (u, v, u', v') st. $w = uf v$, $w' = u' f v'$. According to Claim 2, we have $0 \leq |u| + |v| + |u'| + |v'| \leq 1$, thus $|u| + |v| + |u'| + |v'| \in \{0, 1\}$. More precisely, at most one of the integers $|u|, |v|, |u'|$ and $|v'|$ is non-zero: this implies $(w, w') \in \mathcal{P}_1 \cup \mathcal{S}_1$. As a consequence we have $\mathcal{F}_1 = \mathcal{P}_1 \cup \mathcal{S}_1$, thus $\mathcal{F}_1^k = (\mathcal{P}_1 \cup \mathcal{S}_1)^k$. \square

Regarding regular relations, the following result is a consequence of Lemma 4.5:

Corollary 4.6. *Each of the following properties holds:*

- (i) *For every integer $k \geq 1$, the relation \mathcal{F}_k is regular.*
- (ii) *The relation $\underline{\mathcal{F}}_1$ is regular.*

Proof. According to Prop. 3.8, \mathcal{P}_1 and \mathcal{S}_1 are regular relations. According to Prop. 2.6, the relations $\mathcal{F}_1 = \mathcal{P}_1 \cup \mathcal{S}_1$ and $\mathcal{F}_k = \mathcal{F}_1^k$ are regular: this establishes the property (i). Regarding the property (ii), we have $\underline{\mathcal{F}}_1 = (\mathcal{P}_1 \cup \mathcal{S}_1) \cap \overline{id_{A^*}} = (\mathcal{P}_1 \cap \overline{id_{A^*}}) \cup (\mathcal{S}_1 \cap \overline{id_{A^*}}) = \underline{\mathcal{P}}_1 \cup \underline{\mathcal{S}}_1$. Once more according to Prop. 3.8, the relations $\underline{\mathcal{P}}_1$ and $\underline{\mathcal{S}}_1$ are regular, whence $\underline{\mathcal{F}}_1$ itself is regular. \square

Regarding Conds. **(c1)**, **(c2)**, the property (ii) in Corollary 4.6 may appear promising. Unfortunately, we do not know whether it could be extended to the relation \mathcal{F}_k , for any $k \geq 1$. For instance, taking $k = 2$ we have $\underline{\mathcal{F}}_1 = (\mathcal{P}_1 \cup \mathcal{S}_1)^2 \cap \overline{id_{A^*}} = \mathcal{P}_1^2 \cup \mathcal{P}_1 \mathcal{S}_1 \cup \mathcal{S}_1 \mathcal{P}_1 \cup \mathcal{S}_1^2$ thus, according to Corollary 3.6, $\underline{\mathcal{F}}_1 = \underline{\mathcal{P}}_1^2 \cup \underline{\mathcal{P}}_1 \underline{\mathcal{S}}_1 \cup \underline{\mathcal{S}}_1 \underline{\mathcal{P}}_1 \cup \underline{\mathcal{S}}_1^2$. Although the relations $\underline{\mathcal{P}}_1$ and $\underline{\mathcal{S}}_1$ are regular (see Prop. 3.8), we do not know whether the same holds for $\underline{\mathcal{P}}_1 \underline{\mathcal{S}}_1$ and $\underline{\mathcal{S}}_1 \underline{\mathcal{P}}_1$. However, a clever examination of the structure of $\mathcal{F}_k \subseteq A^* \times A^*$ will allow to overcome this obstacle: to be more precise, afterwards we will prove that $\underline{\mathcal{F}}_k$ preserves the regularity of sets.

4.2 A set covering for $\underline{\mathcal{F}}_k$

We start by the following property: regarding Claim 2, it actually allows to get away of the maximum length constraint over the words in $F(w) \cap F(w')$.

Lemma 4.7. *Given a pair of words $(w, w') \in A^* \times A^*$, we have $(w, w') \in \mathcal{F}_k$ iff. a tuple of words (g, u, u', v, v') exists st. each of the following conds. holds:*

- (i) $0 \leq |u| + |u'| + |v| + |v'| \leq k$ holds.
- (ii) We have $w = ugv$ and $w' = u'gv'$.

Proof. Let $(w, w') \in \mathcal{F}_k$, let f be a maximum length word in $F(w) \cap F(w')$, and let (u, u', v, v') st. $w = ufv$, $w' = u'fv'$. Taking $g = f$, trivially we obtain the cond. (ii) of Lemma 4.7. In addition, according to Claim 2, we obtain the cond. (i). Conversely, assume that there is a tuple of words (g, u, u', v, v') satisfying both the conds. (i) and (ii). With the cond. (ii), the word g belongs to $F(w) \cap F(w')$: by the maximality of $|f|$ we have $|g| \leq |f|$. This implies $d_F(w, w') = |w| + |w'| - 2|f| \leq |w| + |w'| - 2|g|$, thus $d_F(w, w') \leq |u| + |u'| + |v| + |v'|$. In view of the cond. (i) we obtain $d_F(w, w') \leq k$, thus $(w, w') \in \mathcal{F}_k$. \square

Notice that, in the statement of Lemma 4.7, the word g needs not to be a factor of any maximum length word in $F(w) \cap F(w')$, as attested by what follows:

Example 4.8. (*Example 4.3 continued*)

Let $w = cccbababbab$, $w' = bbabbaababaacca$, and $k = 23$. Recall that there are two words of maximum length in $F(w) \cap F(w')$, namely $f_1 = ababa$ and $f_2 = babba$: we have $d_F(w, w') = |w| + |w'| - 2|f_1| = 13 + 16 - 10 = 19$, thus $(w, w') \in \mathcal{F}_k$. Taking $g = ccc$, we have $w = ugv$, $w' = u'gv'$, with $u = \varepsilon$, $v = babababbab$, $u' = bbabbaababaa$, and $v' = a$. Although we have neither $g \in F(f_1)$ nor $g \in F(f_2)$, it follows from $|u| + |u'| + |v| + |v'| = 23 \leq k$ that (g, u, u', v, v') satisfies both Conds. (i), (ii) of Lemma 4.7.

In view of Lemma 4.7, in what follows, we indicate the construction of a finite family of subsets of $A^* \times A^*$, namely $(S_\omega)_{\omega \in \Omega_k}$. Firstly, we denote by Ω_k the set of the tuple (u, u', v, v') satisfying the cond. (i) in Lemma 4.7. By construction, Ω_k is finite. Secondly, for each $\omega = (u, u', v, v') \in \Omega_k$ we denote by S_ω the set of the pairs $(ugv, u'gv')$, for all $g \in A^*$. The following result emphasizes a connection between the family $(S_\omega)_{\omega \in \Omega_k}$ and the relation \mathcal{F}_k :

Lemma 4.9. *Each of the following properties holds:*

- (i) *For any $\omega \in \Omega_k$, the relation $S_\omega \subseteq A^* \times A^*$ is regular.*
- (ii) *We have $\mathcal{F}_k = \bigcup_{\omega \in \Omega_k} S_\omega$.*

Proof. – For establishing the property (i), we observe that, given $\omega = (u, u', v, v') \in \Omega_k$, the relation S_ω is the behavior of a finite automaton, namely \mathcal{A}_ω . The states are 0, 1, 2, the initial state being 0 and the terminal one being 2. The transitions are $0 \xrightarrow{(u, u')} 1$, $1 \xrightarrow{(v, v')} 2$, and $1 \xrightarrow{(a, a)} 1$, for every $a \in A$ (see Fig. 9). By construction the successful paths are $0 \xrightarrow{(u, u')} 1 \xrightarrow{(a_1, a_1)} 1 \dots 1 \xrightarrow{(a_n, a_n)} 1 \xrightarrow{(v, v')} 2$, with $n \geq 0$ and $a_1, \dots, a_n \in A$. Since the corresponding labels are $(ugv, u'gv')$, for all $g = a_1 \dots a_n \in A^*$, the behavior of \mathcal{A}_ω is S_ω .

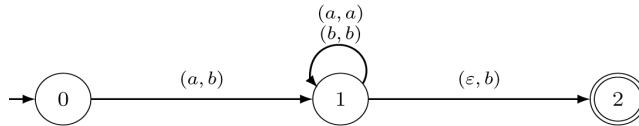


Figure 9: Automaton \mathcal{A}_ω , with $\omega = (a, b, \varepsilon, b)$.

– Now we proceed to establish the property (ii). Firstly, we prove that $\mathcal{F}_k \subseteq \bigcup_{\omega \in \Omega_k} S_\omega$. Assuming that $(w, w') \in \mathcal{F}_k$, according to Lemma 4.7, a tuple of words (g, u, u', v, v') exists st. $0 \leq |u| + |u'| + |v| + |v'| \leq k$, $w = ugv$, and $w' = u'gv'$: set $\omega = (u, u', v, v')$. By construction the cond. $0 \leq |u| + |u'| + |v| + |v'| \leq k$ implies $\omega \in \Omega_k$; in addition $w = ugv$, $w' = u'gv'$ implies $(w, w') \in S_\omega$. For proving that $\bigcup_{\omega \in \Omega_k} S_\omega \subseteq \mathcal{F}_k$, we consider $\omega = (u, u', v, v') \in \Omega_k$, and $(w, w') \in S_\omega$. By definition, $g \in A^*$ exists st. $w = ugv$ and $w' = u'gv'$. On the one hand, by construction $\omega \in \Omega_k$ implies $|u| + |u'| + |v| + |v'| \leq k$: we obtain the cond. (i) of Lemma 4.7. On the other hand, from the fact that the equations $w = ugv$ and $w' = u'gv'$ hold, we are further in the cond. (ii) of the same lemma: this implies $(w, w') \in \mathcal{F}_k$. \square

Lemma 4.9 attests that the family $(S_\omega)_{\omega \in \Omega_k}$ constitutes a set covering for $\mathcal{F}_k \subseteq A^* \times A^*$. In that family, the sets need not to be pairwise disjoint, even pairwise different. For instance, let $k \geq 2$, and let $u, v \in A^*$ such that $|u| + |v| \leq k/2$. With this cond. both the tuples $\omega_0 = (\varepsilon, \varepsilon, \varepsilon, \varepsilon)$ and $\omega_1 = (u, u, v, v)$ belong to Ω_k . From the fact that $(\varepsilon \cdot uv \cdot \varepsilon, \varepsilon \cdot uv \cdot \varepsilon) = (u \cdot \varepsilon \cdot v, u \cdot \varepsilon \cdot v)$, we obtain $(uv, uv) \in S_{\omega_0} \cap S_{\omega_1}$. As another example, taking $\omega_2 = (u, u', pt, pt') \in \Omega_k$ and $\omega_3 = (u, u', t, t') \in \Omega_k$, we have $S_{\omega_2} = S_{\omega_3}$. Indeed, $(w, w') \in \omega_2$ holds iff. some $g \in A^*$ exists st. $w = u \cdot g \cdot pt$ and $w' = u' \cdot g \cdot pt'$ that is, $w = u \cdot gp \cdot t$ and $w' = u' \cdot gp \cdot t'$. Nevertheless, such redundancies have no incidence on the rest of our study. In particular they do not involve the properties of the set R_ω , which is further constructed.

In the sequel, starting from $(S_\omega)_{\omega \in \Omega_k}$, we construct a covering for $\underline{\mathcal{F}}_k$. Beforehand, we note that the following property holds:

Claim 3. *Let $\omega = (u, u', v, v') \in \Omega_k$. Then $u = u'$ with $v = v'$ implies $S_\omega \subseteq id_{A^*}$.*

Proof. $u = u'$ with $v = v'$ implies $S_\omega = \{(ugv, ugv) : g \in A^*\}$. Furthermore $(w, w') \in S_\omega$ implies $w' = w$. \square

In view of Claim 3, we introduce the set Ω'_k of all the tuple $(u, u', v, v') \in \Omega_k$ st. at least one of the conds. $u \neq u'$ or $v \neq v'$ holds. With this notation, the following property holds:

Lemma 4.10. *We have $\underline{\mathcal{F}}_k = \bigcup_{\omega \in \Omega'_k} (S_\omega \setminus id_{A^*})$.*

Proof. According to Lemma 4.9, we have $\underline{\mathcal{F}}_k = \overline{id_{A^*}} \cap \bigcup_{\omega \in \Omega_k} S_\omega$, whence we have:

$$\underline{\mathcal{F}}_k = \bigcup_{\omega \in \Omega'_k} (\overline{id_{A^*}} \cap S_\omega) = \bigcup_{\omega \in \Omega'_k} (S_\omega \setminus id_{A^*}) \cup \bigcup_{(u, u', v, v') \in \Omega_k} (S_{(u, u', v, v')} \setminus id_{A^*}).$$

Moreover, in view of Claim 3 the identity $\bigcup_{(u, u', v, v') \in \Omega_k} (S_{(u, u', v, v')} \setminus id_{A^*}) = \emptyset$ holds. \square

4.3 On the combinatorial structure of $S_\omega \cap id_{A^*}$

The following result brings a noticeable combinatorial characterization of those tuples (g, u, u', v, v') satisfying the equation $ugv = ugv'$:

Lemma 4.11. *Let $\omega = (u, u', v, v') \in \Omega'_k$ and $g \in A^*$ st. $(ugv, u'gv') \in S_\omega$. We have $ugv = u'gv'$ iff. $\alpha \in A^*$, $\beta \in A^+$, and $n \in \mathbb{N}$ exist st. exactly one of the following conds. holds:*

- (i) $u = u'\alpha\beta$, $v' = \beta\alpha v$, and $g \in (\alpha\beta)^n\alpha$.
- (ii) $u' = u\alpha\beta$, $v = \beta\alpha v'$, and $g \in (\alpha\beta)^n\alpha$.

Proof. – With the cond. (i) we have $ugv = (u'\alpha\beta)(\alpha\beta)^n\alpha v = u'(\alpha\beta)^n\alpha(\beta\alpha v) = u'(\alpha\beta)^n\alpha v' = u'gv'$. Similarly, the cond. (ii) implies $u'gv' = (u\alpha\beta)(\alpha\beta)^n\alpha v' = u(\alpha\beta)^n\alpha(\beta\alpha v') = ugv$.

– Conversely, assume that the equation $ugv = u'gv'$ holds. It follows from $|u| + |v| = |u'| + |v'|$ that the conds. $u = u'$ and $v = v'$ are equivalent. Furthermore, $(u, u', v, v') \in \Omega'_k$ implies $u \neq u'$ and $v \neq v'$. Consequently, either u' is a proper prefix of u (and v is a proper suffix of v'), or u is a proper prefix of u' (and v' is a proper suffix of v).

(a) In the case where u' is a proper prefix of u , let $t, t' \in A^+$ st. $u = u't$ and $v' = t'v$. The equation $ugv = u'gv'$ implies $u'tgv = u'gt'v$, thus $tg = gt'$. According to Prop. 2.2, t, t' are conjugate words, furthermore $\alpha \in A^*$, $\beta \in A^+$, and $n \in \mathbb{N}$ exist st. $t = \alpha\beta$, $t' = \beta\alpha$, and $g \in (\alpha\beta)^n\alpha$. moreover we have $u = u't = u'\alpha\beta$, $v' = t'v = \beta\alpha v$: we are in the cond. (i).

(b) In the case where u is a proper prefix of u' and v' is a proper suffix of v , we note that, by construction, the conds. $\omega = (u, u', v, v') \in \Omega'_k$ and $\omega' = (u', u, v', v) \in \Omega'_k$ are equivalent. By substituting (u', u, v', v) to (u, u', v, v') in the arguments we applied in the preceding case (a), we obtain the cond. (ii) of our lemma. \square

Let $\omega = (u, u', v, v') \in \Omega'_k$. Note that Lemma 4.11 does not guarantee the unicity of the pair of words (α, β) . Accordingly, we introduce the following sets:

- $R_\omega^{(i)}$ is the union of the sets $R_{\omega, \alpha, \beta}^{(i)} = \{u(\alpha\beta)^n\alpha v : n \geq 0\}$, for all word pairs $\alpha \in A^*$, $\beta \in A^+$ st. $u = u'\alpha\beta$ and $v' = \beta\alpha v$.
- $R_\omega^{(ii)}$ is the union of the sets $R_{\omega, \alpha, \beta}^{(ii)} = \{u'(\alpha\beta)^n\alpha v' : n \geq 0\}$, for all word pairs $\alpha \in A^*$, $\beta \in A^+$ st. $u' = u\alpha\beta$ and $v = \beta\alpha v'$.
- $R_\omega = R_\omega^{(i)} \cup R_\omega^{(ii)}$.

As indicated above, either u' is a proper prefix of u or u is a proper prefix of u' . Accordingly at most one of the sets $R_\omega^{(i)}$, $R_\omega^{(ii)}$ may be non-empty.

Example 4.12. Let $A = \{a, b\}$, $k = 10$.

– Firstly, consider the tuple $\omega = (aba, a, b, b^3) \in \Omega'_k$. Since u' is a proper prefix of u , only the cond. (i) of Lemma 4.11 may hold, whence we have $R_\omega^{(ii)} = \emptyset$. With the preceding notation, $u = u'\alpha\beta$ implies $\alpha\beta = ba$; similarly $v' = \beta\alpha v$ implies $\beta\alpha = b^2$. It is straightforward to verify that no pair of words (α, β) may satisfy such constraints, hence we have $R_\omega^{(i)} = \emptyset$ and $R_\omega = \emptyset$.

– Now, consider the tuple $\omega = (aba, a, b^2, ab^3) \in \Omega'_k$. Once more u' is a proper prefix of u , thus $R_\omega^{(ii)} = \emptyset$ holds. Regarding the set $R_\omega^{(i)}$ we have $\alpha\beta = ba$, $\beta\alpha = ab$. The set of the pairs (α, β) st. $\alpha\beta = ab$ is $\{(\varepsilon, ab), (a, b)\}$ (recall that we set $\beta \neq \varepsilon$). Similarly, the equation $\beta\alpha = ba$ is satisfied by the pairs $(\alpha, \beta) \in \{(\varepsilon, ba), (a, b)\}$. Consequently only the pair (a, b) may satisfy both the preceding constraints. We verify that the equations $ugv = aba \cdot (ba)^n b \cdot b^2 = a \cdot (ba)^n b \cdot ab^3 = u'gv'$ hold. This corresponds to $R_\omega = R_{\omega, a, b}^{(i)} = \{(ab)^{n+2}b^2 : n \geq 0\}$.

As direct consequences of Lemma 4.11 the two following properties hold:

Corollary 4.13. *Given $\omega \in \Omega'_k$ and $(w, w') \in S_\omega$, the conds. $w = w'$ and $w \in R_\omega$ are equivalent.*

Proof. By construction $g \in A^*$ exists st. $w = ugv$, $w' = u'gv'$. We have $ugv = u'gv'$ iff. exactly one of the conds. (i) or (ii) of Lemma 4.11 is satisfied that is, $w \in R_\omega^{(i)} \cup R_\omega^{(ii)} = R_\omega$. \square

Corollary 4.14. *Given $\omega \in \Omega'_k$ and $X \subseteq A^*$ we have $\underline{S}_\omega(X) = S_\omega(X \setminus R_\omega)$.*

Proof. We have $y \in \underline{S}_\omega(X)$ iff. $x \in X$ exists st. $(x, y) \in S_\omega$ with $x \neq y$. According to Corollary 4.13, $x \neq y$ is equivalent to $x \notin R_\omega$, hence $y \in \underline{S}_\omega(X)$ is equivalent to $y \in \underline{S}_\omega(X \setminus R_\omega)$. \square

We conclude Sect. 4 by proving that \mathcal{F}_k is a regularity-preserving relation:

Proposition 4.15. *If $X \subseteq A^*$ is regular, then $\underline{\mathcal{F}}_k(X)$ is a regular subset of A^* .*

Proof. We proceed through the three following steps:

– Let $\omega = (u, v, u', v') \in \Omega'_k$, and let α, β be words satisfying the cond. of Lemma 4.11. The set $R_{\omega, \alpha, \beta}^{(i)} = \{u(\alpha\beta)^n \alpha v : n \geq 0\}$ is regular: indeed, classically it is the behavior of the finite A^* -automaton represented in Figure 10.

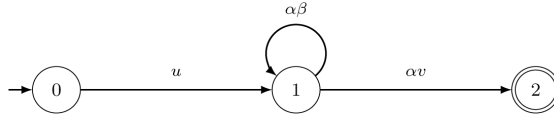


Figure 10: Proof of Prop. 4.15. A finite A^* -automaton with behavior $\{u(\alpha\beta)^n \alpha v : n \geq 0\}$.

The set $R_\omega^{(i)}$ is the union of the sets $R_{\omega, \alpha, \beta}^{(i)}$, for all the pairs of words (α, β) satisfying the cond. of Lemma 4.11. Since in any case we have $|\alpha|, |\beta| \leq \max\{|u|, |u'|, |v|, |v'|\} \leq k$, there are only a finite number of such pairs (α, β) , hence according to Prop. 2.6, $R_\omega^{(i)}$ is a regular set. Similarly, $R_\omega^{(ii)}$ is regular, hence $R_\omega = R_\omega^{(i)} \cup R_\omega^{(ii)}$ itself is regular.

– Let $X \subseteq A^*$ be a regular set. According to Corollary 4.14, for every $\omega \in \Omega'_k$ we have $\underline{S}_\omega(X) = S_\omega(X \cap \overline{R_\omega})$. According to Prop. 2.6, since R_ω is a regular subset of A^* , the same holds for the set $\overline{R_\omega}$. Since the relation $S_\omega \subseteq A^* \times A^*$ is regular (see Lemma 4.9), according to Prop. 2.7, the set $\underline{S}_\omega(X) = S_\omega(X \cap \overline{R_\omega})$ itself is regular.

– According to Lemma 4.10 we have $\underline{\mathcal{F}}_k(X) = \bigcup_{\omega \in \Omega'_k} \underline{S}_\omega(X)$. As established above, for each $\omega \in \Omega'_k$ the set $\underline{S}_\omega(X)$ is regular. Since Ω'_k is a finite set, $\underline{\mathcal{F}}_k(X)$ itself is regular. \square

5 Error detection conds. wrt. \mathcal{F}_k

We now have enough properties to establish decidability results. First and foremost, according to Corollary 4.4, and since \mathcal{F}_k is a symmetric relation, Lemma 2.12 translates as follows:

Lemma 5.1. *Given a set $X \subseteq A^*$ and $k \geq 1$ the following conds. are equivalent:*

- (i) X satisfies Cond. **(c2)** wrt. \mathcal{F}_k .
- (ii) For every $x \in X$ we have $\mathcal{F}_{2k}(x) \cap X = \{x\}$.
- (iii) X satisfies Cond. **(c1)** wrt. \mathcal{F}_{2k} .

In view of Prop. 4.15, the following decidability result holds:

Proposition 5.2. *It can be decided whether a given regular set satisfies any of Conds. **(c1)**, **(c2)**, or **(c4)** wrt. \mathcal{F}_k .*

Proof. Let $X \subseteq A^*$ be a regular set. We examine one by one our conditions:

– *Cond. **(c1)**:* According to Prop. 4.15, $\underline{\mathcal{F}}_k(X)$ is a regular subset of A^* , therefore $\underline{\mathcal{F}}_k(X) \cap X$ itself is regular. According to Prop. 2.11, one can decide whether it is the empty set or not.

– *Cond. **(c2)**:* Since \mathcal{F}_k is a symmetric relation, according to Lemma 5.1, X satisfies Cond. **(c2)** iff. it satisfies Cond. **(c1)** wrt. \mathcal{F}_{2k} : as indicated above, this cond. can be decided.

– *Cond. **(c4)**:* According to Corollary 4.6 \mathcal{F}_k is a regular relation. By Prop. 2.7, the set $\widehat{\mathcal{F}}_k(X) = \mathcal{F}_k(X)$ is regular. According to Prop. 2.6 $X \cup \mathcal{F}_k(X)$ is a regular set furthermore, according to Prop. 2.11, one can decide whether it is a code or not. \square

Regarding Cond. **(c3)**, we start by proving the following result:

Proposition 5.3. *Every regular $\underline{\mathcal{F}}_k$ -independent code can be embedded into some complete one.*

Proof. Let X be a regular $\underline{\mathcal{F}}_k$ -independent code: wlog. we assume X non-complete. At first, we will construct a special word $z_1 \in A^* \setminus F(X^*)$. Let $z_0 \notin F(X^*)$, with $|z_0| \geq k$, let a be its initial character, and let b be a character

different of a . Consider the word $z = z_0ab^{|z_0|}$ as constructed in the proof of Prop. 3.11 and set $z_1 = a^{|z|}bz = a^{2|z_0|+1}bz_0ab^{|z_0|}$. According to Prop. 2.1, z_1 is overlapping-free indeed, as indicated in the preliminaries, $z^Rba^{|z|}$, the reversal of z_1 , is overlapping-free. Now, we set $U_1 = A^* \setminus (X^* \cup A^*z_1A^*)$, $Y_1 = z_1(U_1z_1)^*$, and $Z_1 = X \cup Y_1$. According to Theorem 2.9, the set Z_1 is a regular complete code. In what follows we prove that Z_1 is \mathcal{F}_k -independent that is, $\overline{\mathcal{F}_k}(X \cup Y_1) \cap (X \cup Y_1) = \emptyset$. Since X itself is \mathcal{F}_k -independent, this amounts to prove that each of the three equations $\overline{\mathcal{F}_k}(X) \cap Y_1 = \emptyset$, $\overline{\mathcal{F}_k}(Y_1) \cap X = \emptyset$, and $\overline{\mathcal{F}_k}(Y_1) \cap Y_1 = \emptyset$ holds.

(a) For proving that $\overline{\mathcal{F}_k}(X) \cap Y_1 = \emptyset$ holds, we argue by contradiction: let $x \in X$, $y \in Y_1$ st. $(x, y) \in \mathcal{F}_k$. Let f be a word with maximum length in $F(x) \cap F(y)$ and let $u, v \in A^*$ st. $y = ufv$. According to Claim 2, we have $|u| + |v| \leq d_F(x, y) \leq k$. By construction the word $a^{|z_0|}$ (resp., $b^{|z_0|}$) is a prefix (resp., suffix) of $y \in Y_1$, therefore at least one of the words u and $a^{|z_0|}$ (resp., v and $b^{|z_0|}$) is a prefix (resp., suffix) of the other one. More precisely, it follows from $|u| \leq k \leq |z_0|$ and $|v| \leq k \leq |z_0|$ that we have $u \in P(a^{|z_0|})$ and $v \in S(b^{|z_0|})$. On the other hand, by construction, $y \in Y_1$ implies either $y = z_1$, or $y = z_1wz_1$ for some $w \in A^+$. With the first condition, the equation $f = u^{-1}z_1v^{-1} = a^{2|z_0|+1-|u|}bz_0b^{|z_0|-|v|}$ holds. With the second cond. we have $f = u^{-1}wv^{-1} = a^{2|z_0|+1-|u|}wbz_0b^{|z_0|-|v|}$. In each case z_0 is a factor of $f \in F(x)$: a contradiction with $z_0 \notin F(X^*)$. Consequently we have $\overline{\mathcal{F}_k}(X) \cap Y_1 = \emptyset$.

(b) By contradiction we assume $\overline{\mathcal{F}_k}(Y_1) \cap X \neq \emptyset$. Let $y \in Y_1$ and $x \in X$ st. $(y, x) \in \mathcal{F}_k$. Since the relations \mathcal{F}_k and $\overline{id_{A^*}}$ are symmetrical, $\mathcal{F}_k = \mathcal{F}_k \cap \overline{id_{A^*}}$ itself is symmetrical, therefore, we have $(x, y) \in \mathcal{F}_k$, thus $\overline{\mathcal{F}_k}(X) \cap Y_1 \neq \emptyset$: this contradicts the conclusion of Case (a). (c) It remains to prove that $\overline{\mathcal{F}_k}(Y_1) \cap Y_1 = \emptyset$. By contradiction, we assume that a pair of different words $y, y' \in Y_1$ exist st. $(y, y') \in \mathcal{F}_k$. Let f be a word with maximum length in $F(y) \cap F(y')$. Once more according to Claim 2, words u, u', v, v' exist st. $y = ufv$, $y' = u'fv'$, with $|u| + |u'| + |v| + |v'| = d_F(w, w') \leq k$.

(c1) At first, we compare the words v, v' with ε . Firstly, by contradiction assume that both the conds. $v \neq \varepsilon, v' \neq \varepsilon$ hold. Necessarily we have $2 \leq |v| + |v'| \leq k \leq |z_0|$. By construction $v, v' \in S(Y_1)$ holds: this implies $v, v' \in S(b^{|z_0|})$, whence $i, j \in [1, |z_0|]$ exist st. $v = b^i, v' = b^j$ (see Fig. 11). We obtain $fb \in F(y) \cap F(y')$, a contradiction with $|f|$ being maximum. Consequently at least one of the conds. $v = \varepsilon$ or $v' = \varepsilon$ holds: wlog. we assume $v' = \varepsilon$, thus $f \in S(y')$. On the one hand, according to the maximality of $|f|$, it follows from $z_1 \in F(y) \cap F(y')$ that $|f| \geq |z_1|$. Since we have $f, z_1 \in S(y')$ we obtain $f \in A^*z_1$, that is, $fv \in A^*z_1v$: in particular we have $|fv| \geq |z_1v| \geq |z_1|$. On the other hand, by construction both the words z_1 and fv are suffixes of y : it follows from $|fv| \geq |z_1|$ that $fv \in A^*z_1$. Accordingly both the words z_1v and z_1 are suffixes of f : we obtain $z_1v \in A^*z_1$. Since z_1 is an overlapping-free word, by definition the cond. $|v| = |v| + |v'| \leq |z_0| \leq |z_1| - 1$ implies $v = \varepsilon = v'$.

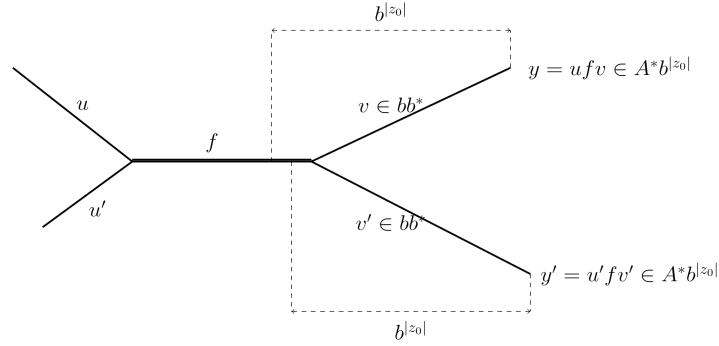


Figure 11: Proof of Prop. 5.3. Case (c1): we have $(y, y') \in \underline{\mathcal{F}}_k$, with $v \neq \varepsilon$ and $y' = v' \neq \varepsilon$.

(c2) Now, we compare u, u' with ε . In order to do so we apply to the prefixes of y and y' arguments similar to the preceding ones. Firstly, by contradiction, we assume $u \neq \varepsilon$ and $u' \neq \varepsilon$. It follows from $2 \leq |u| + |u'| \leq k \leq |z_0|$ and $u, u' \in P(Y_1)$ that $u, u' \in P(a^{|z_0|}) \setminus \varepsilon$. Accordingly $i, j \in [1, |z_0|]$ exist st. $u = a^i, u' = a^j$, thus $af \in F(y) \cap F(y')$: a contradiction with f being of maximal length in $F(y) \cap F(y')$. Consequently, at least one of the two words u, u' is the empty word: wlog. we assume $u' = \varepsilon$, thus $y' = f v' = f$. As observed in the case (c1), according to the maximality of $|f|$ we have $|f| \geq |z_1|$. Since both the words f and z_1 are prefixes of y' , this implies $f \in z_1 A^*$, thus $uf \in uz_1 A^*$. In addition, by construction both the words z_1 and uf are prefixes of y : this implies $uf \in uz_1 A^* \cap z_1 A^*$ (see Fig. 12) that is, $uz_1 \in P(f)$ and $z_1 \in P(f)$, thus $uz_1 \in z_1 A^*$. Since z_1 is overlapping-free, we obtain $u = \varepsilon = u' = v = v'$, thus $y = y'$: a contradiction with $(y, y') \in \underline{\mathcal{F}}_k$. As a consequence, we have $\underline{\mathcal{F}}_k(Y_1) \cap Y_1 = \emptyset$: this completes the proof. \square

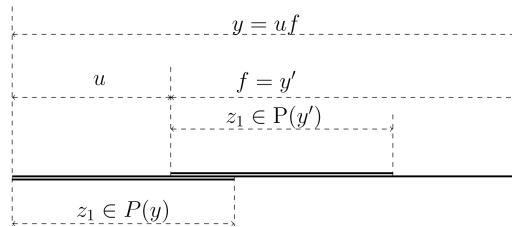


Figure 12: Proof of Prop. 5.3: Case (c2) $(y, y') \in \underline{\mathcal{F}}_k$, with $v = v' = \varepsilon$, and $u' = \varepsilon$.

According to Prop. 5.3, by merely substituting \mathcal{F}_k to \mathcal{P}_k in the proofs of the statements 3.12–3.14, we obtain the following result:

Proposition 5.4. *Given a regular code X , each of the following properties holds:*

- (i) *The three following conds. are equivalent:*
 - X is maximal in the family of $\underline{\mathcal{F}_k}$ -independent codes.
 - X is complete.
 - $\mu(X) = 1$.
- (ii) *One can decide whether X satisfies Cond. (c3) wrt. \mathcal{F}_k .*
- (iii) *If X is $\underline{\mathcal{F}_k}$ -independent, it can be embedded into a maximal $\underline{\mathcal{F}_k}$ -independent code.*

6 Error detection wrt. topologies associated with (anti-)automorphisms

Let θ be (anti-)automorphism of A^* . As justified in the introduction, we focus on the relation $\tau_{d_\theta,1} = \widehat{\theta} = \theta \cup id_{A^*}$. Recall that we have $\underline{(\widehat{\theta})} = \widehat{\theta} \setminus id_{A^*} = \underline{\theta}$.

Example 6.1. Let $A = \{a, b\}$ and θ be the automorphism defined by $\theta(a) = b$, $\theta(b) = a$. It follows from $\theta^{-1}(a) = b = \theta(a)$ that $\theta^{-1} = \theta = \underline{\theta}$. Consider the regular prefix code $X = \{a^n b : n \geq 0\}$.

– X satisfies Cond. (c1) wrt. $\tau_{d_\theta,1} = \widehat{\theta}$. Indeed, we have $\underline{(\widehat{\theta})}(X) = \underline{\theta}(X) = \theta(X) = \{b^n a : n \geq 0\}$, thus $\underline{\theta}(X) \cap X = \emptyset$.

– We have $\widehat{(\widehat{\theta})}^{-1} = (\theta \cup id_{A^*})^{-1} = \theta^{-1} \cup id_{A^*}^{-1} = \theta \cup id_{A^*} = \widehat{\theta}$. Furthermore we obtain $\widehat{\theta} \cdot \widehat{(\widehat{\theta})}^{-1} = (\theta \cup id_{A^*})^2 = \theta^2 \cup \theta \cup id_{A^*} = \theta \cup id_{A^*} = \widehat{\theta}$. As indicated above, X satisfies Cond. (c1) wrt. $\widehat{\theta} = \widehat{\theta} \cdot \widehat{(\widehat{\theta})}^{-1}$: according to Prop. 2.12, X also satisfies Cond. (c2) wrt. $\widehat{\theta}$.

– According to Theorem 2.8, it follows from $\mu(X) = \frac{1}{2} \sum_{n \geq 0} (\frac{1}{2})^n = 1$ that X is a maximal prefix code, therefore X is maximal in the family of $\underline{(\widehat{\theta})}$ -independent codes (Cond. (c3)).

– Finally, it follows from $ba \in \theta(X) \setminus X$ that $X \subsetneq \widehat{\theta}(X)$. Consequently, since X is a maximal code, $\widehat{\theta}(X)$ cannot be a code that is, X cannot satisfy Cond. (c4) (we verify that $a, b, ab \in \widehat{\theta}(X)$).

Example 6.2. Over the alphabet $A = \{a, b\}$, take for θ the anti-automorphism defined by $\theta(a) = b$, and $\theta(b) = a$ and, once more, consider the code $X = \{a^n b : n \geq 0\}$.

– It follows from $\theta(X) \cap X = \{ab\}$, that the equation $x = \theta(x)$ is equivalent to $x = ab$. This implies $\underline{\theta}(X) = \{ab^n : n \neq 1\}$, thus $\underline{\theta}(X) \cap X = \emptyset$, whence X satisfies Cond. (c1).

– As in Example 6.1, we have $\theta^{-1} = \theta$. Once more this implies $\widehat{\theta} \cdot \widehat{(\widehat{\theta})}^{-1} = \widehat{\theta}$, thus X satisfies Cond. (c2) wrt. $\widehat{\theta}$.

– Similarly, we have $\mu(X) = 1$, whence X satisfies Cond. (c3).

– Lastly, it follows from $ab^2 \in \theta(X) \setminus X$ that $X \subsetneq \widehat{\theta}(X)$: since X is a maximal code, $\widehat{\theta}(X)$ cannot be a code that is, X cannot satisfy Cond. **(c4)**.

Example 6.3. Over the alphabet $\{A, C, G, T\}$, let θ denote the Watson-Crick anti-automorphism (see e.g. [15, 17]), which is defined by $\theta(A) = T$, $\theta(T) = A$, $\theta(C) = G$, and $\theta(G) = C$. We have $\theta^{-1} = \theta = \underline{\theta} = \widehat{\theta}$. Consider the prefix code $X = \{A, C, GA, G^2, GT, GCA, GC^2, GCG, GCT\}$.

– It follows from $\theta(X) = \{T, G, TC, C^2, AC, TGC, G^2C, CGC, AGC\}$ that we have $\widehat{\theta}(X) \cap X = \emptyset$, whence satisfies Cond. **(c1)**.

– As in the examples 6.1, 6.2, it follows from $\theta^{-1} = \theta$ that $\widehat{\theta} \cdot \left(\widehat{\theta}\right)^{-1} = \widehat{\theta}$, therefore, X satisfies Cond. **(c2)** wrt. $\tau_{d_\theta, 1} = \widehat{\theta}$.

– We have $\mu(X) = 2/4 + 3/4^2 + 4/4^3 = 3/4 < 1$, hence X cannot satisfy Cond. **(c3)**.

– At last, it follows from $G, G^2 \in \widehat{\theta}(X) = \theta(X) \cup X$ that Cond. **(c4)** is not satisfied.

Example 6.4. In each of the preceding examples, since the mapping θ satisfies $\theta^{-1} = \theta$, the quasi-metric d_θ is actually a metric. Of course, d_θ could be only a quasi-metric. For instance over $A = \{a, b, c\}$, taking for θ the automorphism generated by the cycle (a, b, c) , we obtain $d_\theta(a, b) = 1$ and $d_\theta(b, a) = 2$ (we have $b = \theta(a)$ and $a \neq \theta(b)$).

6.1 Questions involving regular sets

We start with the following result:

Proposition 6.5. *With the preceding notation, each of the following properties holds:*

(i) *If θ is an automorphism, then the relations $\tau_{d_{\theta, 1}} = \widehat{\theta}$ and $\tau_{\underline{d_{\theta, 1}}} = \underline{\theta}$ are regular.*

(ii) *If θ is an anti-automorphism, then $\widehat{\theta}$ is a non-regular relation.*

(iii) *In any case, $\widehat{\theta}$ is regularity-preserving.*

Proof. – In the case where θ is an automorphism of A^* , it is a regular relation: indeed, θ is the behavior of the one-state automaton with transitions $q_0 \xrightarrow{(a, \theta(a))} q_0$ for all $a \in A$. Set $B = \{a \in A : \theta(a) \neq a\}$. Starting with the preceding automaton, we obtain an automaton with behavior $\widehat{\theta}$ by merely adding the transitions $(q_0, (a, a), q_0)$, for all $a \in B$ (see Fig. 13): in other words, $\widehat{\theta}$ is a regular relation.

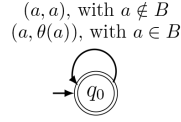


Figure 13: An automaton with behavior $\tau_{d_{\theta,1}} = \widehat{\theta}$ in the case where θ is an automorphism.

The relation $\widehat{(\theta)} = \underline{\theta}$, for its part, is the set of all the pairs (w, w') with $w \neq w'$ and st. $w' = \theta(w)$. By construction we have $|w| = |w'|$. Let $u = w \wedge w'$. Necessarily, we have $u \in (A \setminus B)^*$, moreover $a, b \in B$ and $s, s' \in A^*$ exist such that $w = uas$, $w' = ubs'$. Since θ is a free monoid automorphism, necessarily we have $b = \theta(a)$ and $s' = \theta(s)$. Accordingly, $\underline{\theta}$ is the behavior of the two-state automaton provided in Fig. 14, whence it is a regular relation.

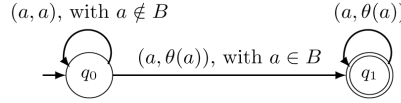


Figure 14: An automaton with behavior $\underline{\theta}$, in the case where θ is an automorphism.

– Let θ be an anti-automorphism of A^* . Classically, an automorphism of A^* , say h , exists st. $\theta = t \cdot h$, where $t : w \rightarrow w^R$, is the so-called *transposition* onto A^* . In addition t cannot be a regular relation (see e.g. [32, Example IV.1.10]). By contradiction, assume θ being regular. Since we have $t = \theta \cdot h^{-1}$, and since h^{-1} itself is a monoid automorphism, according to Prop. 2.5 the transposition t should be a regular relation: a contradiction with the above.

– For proving the property (iii), we consider a regular set $X \subseteq A^*$. In the case where θ is an automorphism, the relation $\widehat{\theta} = \theta \cup id_{A^*}$ is regular: according to Prop. 2.5, $\widehat{\theta}(X)$ is regular. In the case where θ is an anti-automorphism, with the preceding notation although the transposition t is not a regular relation, the set $t(X)$ is classically known to be regular (see e.g. [32, Proposition I.1.1]). According to Prop. 2.5, the set $\widehat{\theta}(X) = h(t(X))$ itself is regular. \square

6.2 Decidability results for Conds. (c1)–(c4)

Regarding Conds. (c1), (c2), (anti-)automorphisms satisfy the following noticeable property:

Lemma 6.6. *A regu $X \subseteq A^*$ satisfies Cond. (c1) wrt. $\widehat{\theta}$, iff. it satisfies Cond. (c2).*

Proof. Firstly, assume that X is $\underline{\theta}$ -independent, and let $x, y \in X$ st. $\tau_{d_{\theta,1}}(x) \cap \tau_{d_{\theta,1}}(y) = \underline{\theta}(x) \cap \underline{\theta}(y) \neq \emptyset$: necessarily we have $\theta(x) \cap \theta(y) \neq \emptyset$. Since θ is

a one-to-one mapping, this implies $x = y$ therefore, by definition X satisfies Cond. **(c2)**. For proving the converse, we argue by contrapositive. Assuming that Cond. **(c1)** does not hold that is, $X \cap \widehat{\theta}(X) = X \cap \theta(X) \neq \emptyset$, a pair of words $x, y \in X$ exist st. $y = \theta(x)$, with $x \neq y$. It follows from $\widehat{\theta}(x) = \{x\} \cup \{\theta(x)\} = \{x, y\}$ and $\widehat{\theta}(y) = \{y\} \cup \{\theta(y)\}$ that $\widehat{\theta}(x) \cap \widehat{\theta}(y) \neq \emptyset$, whence Cond. **(c2)** cannot hold. \square

As a consequence of Propositions 6.5, 6.6, we obtain the following result:

Proposition 6.7. *Given $X \subseteq A^*$, wrt. $\widehat{\theta}$ each of the following properties holds:*

- (i) *In any case, X satisfies both Conds. **(c1)**, **(c2)**.*
- (ii) *If $X \subseteq A^*$ is a regular code, it can be decided whether it satisfies Cond. **(c4)**.*

Proof. – For proving that X satisfies Cond. **(c2)**, we prove that it satisfies the cond. (ii) of Lemma 2.12. Actually, it follows from $\theta \subseteq \widehat{\theta}$ that $\theta \cdot \theta^{-1} \subseteq \widehat{\theta} \cdot \widehat{\theta}^{-1}$. Since θ is a one-to-one mapping, we have $\theta \cdot \theta^{-1} = id_{A^*}$: this implies $\widehat{\theta} \cdot \widehat{\theta}^{-1} = id_{A^*}$. As a consequence, for every $x \in X$ we have $\widehat{\theta} \cdot \widehat{\theta}^{-1}(x) \cap X = \{x\} \cap X = \{x\}$, thus $\widehat{\theta} \cdot \widehat{\theta}^{-1}(x) = \emptyset$. According to Lemma 2.12 X satisfies Cond. **(c1)** wrt. $\widehat{\theta}$ that is, according to Lemma 6.6, X also satisfies Cond. **(c1)**.

– According to Prop. 6.5, in any case the set $\theta(X)$ is regular therefore, according to Prop. 2.11, one can decide whether $\widehat{\theta}(X)$ is a code that is, whether X satisfies Cond. **(c4)**. \square

Before to study the behavior of Cond. **(c3)**, we note that the following property holds:

Claim 4. *If θ is an anti-automorphism, for all $w, w' \in A^*$, $w' \in F(w)$ implies $\theta(w') \in F(w)$.*

Proof. Let $u, v \in A^*$ st. $w = uw'v$. By definition, we have $\theta(w) = \theta(v)\theta(w')\theta(u)$. \square

As usual in the paper, we start by examining completeness from the point of view of embedding:

Proposition 6.8. *Every regular θ -independent code can be embedded into some complete one.*

Proof. The property trivially holds if X is complete. Assume X non-complete. According to Theorem 2.9, the result holds if θ is an automorphism: indeed the action of such a transformation merely consists in rewriting words by applying some permutation of A . In the sequel we assume that θ is an anti-automorphism. Classically, some positive integer n , the so-called *order* of the permutation θ , exists such $\theta^n = id_{A^*}$. As in the proofs of Prop. 3.11 and 5.3, in view of

Theorem 2.9, we start by constructing a convenient word in $A^* \setminus F(X^*)$. Let $z_0 \notin F(X^*)$, let a be its initial character, and let b be a character different of a . We assume wlog. $|z_0| \geq 2$ and $z_0 \notin aa^*$, for every $a \in A$ (otherwise, substitute z_0b to z_0). Since θ is a free monoid anti-automorphism, it is length-preserving that is, for every $i \geq 0$, the equation $|\theta^i(z_0)| = |\theta(z_0)|$ holds. Consequently, we have $|z_0\theta(z_0)\cdots\theta^{n-1}(z_0)| = n|z_0|$ therefore, according to Lemma 2.1 $z_2 = z_0\theta(z_0)\cdots\theta^{n-1}(z_0)ab^{n|z_0|}$ is an overlapping-free word. In addition, it follows from $z_0 \in A^* \setminus F(X^*)$ that $z_2 \in A^* \setminus F(X^*)$. Set $U_2 = A^* \setminus (X^* \cup A^*z_2A^*)$, $Y_2 = (z_2U_2)^*z_2$, and $Z_2 = X \cup Y_2$. According to Theorem 2.9, the set $Z_2 = X \cup Y_2$ is a complete regular code. Since we assume X being $\widehat{\theta}$ -independent, we have $\underline{\theta}(Z_2) \cap Z_2 = (\underline{\theta}(X) \cap Y_2) \cup (X \cap \underline{\theta}(Y_2)) \cup (\underline{\theta}(Y_2) \cap Y_2)$. In order to prove that Z_2 is $\underline{\theta}$ -independent, we argue by contradiction. Actually assuming that $\underline{\theta}(Z_2) \cap Z_2 \neq \emptyset$, exactly one of the three following conds. holds:

(a) *Cond.* $\underline{\theta}(X) \cap Y_2 \neq \emptyset$. With this condition, $x \in X$ exists st. $\theta(x) \in Y_2$. Since by construction z_2 is a prefix of any word in Y_2 , we have $z_2 \in F(\theta(x))$. In addition, it follows from $\theta(z_0) \in F(z_2)$ that $\theta(z_0) \in F(\theta(x))$. According to Claim 4, we obtain $\theta^n(z_0) \in F(\theta^n(x))$, thus $z_0 \in F(x)$, a contradiction with $z_0 \notin F(X^*)$.

(b) *Cond.* $X \cap \underline{\theta}(Y_2) \neq \emptyset$. Some pair of words $x \in X$, $y \in Y_2$ exist st. $x = \theta(y)$. By construction we have $\theta^{n-1}(z_0) \in F(z_2) \subseteq F(y)$. According to Claim 4 this implies $\theta(\theta^{n-1}(z_0)) \in F(\theta(y))$, thus $z_0 \in F(x)$: once more this contradicts $z_0 \notin F(X^*)$.

(c) *Cond.* $\underline{\theta}(Y_2) \cap Y_2 \neq \emptyset$. With this cond. there are different words $y, y' \in Y_2$ st. $y' = \theta(y)$. On the one hand, since θ is an anti-automorphism, $b^{n|z_0|} \in S(y')$ implies $\theta(b^{n|z_0|}) \in P(y')$. On the other hand, it follows from $Y_2 \subseteq z_0A^*$ and $y' \in Y_2$ that $z_0 \in P(y')$. More precisely, $|z_0| \leq |b^{n|z_0|}|$ implies $z_0 \in P(b^{n|z_0|})$, thus $z_0 = (\theta(b))^{n|z_0|}$. From the fact that we have $|z_0| \geq 2$ and $\theta(b) \in A$, this is incompatible with the construction of z_0 .

In each case we obtain a contradiction, whence Z_2 is $\underline{\theta}$ -independent: this completes the proof. \square

As a consequence, we obtain the following result-the proof is merely done by translating in term of (anti-)automorphism the one of Prop. 5.4 (recall that we have $\underline{(\widehat{\theta})} = \underline{\theta}$):

Proposition 6.9. *Given a regular code X , each of the following properties holds:*

- (i) *The three following conds. are equivalent:*
 - X is maximal as a $\underline{\theta}$ -independent code.
 - X is complete.
 - $\mu(X) = 1$ holds.
- (ii) *One can decide whether X satisfies Cond. (c3) wrt. $\widehat{\theta}$.*
- (iii) *If X is $\underline{\theta}$ -independent, then it can be embedded into a maximal $\underline{\theta}$ -independent code.*

We close the study with the following statement: it synthesizes the decidability results obtained in the whole paper:

Theorem 6.10. *With the preceding notation, each of the following properties holds:*

(i) *For every positive integer k , it can be decided whether X satisfies any of Conds. (c1)–(c4) wrt. \mathcal{P}_k , \mathcal{S}_k , or \mathcal{F}_k .*

(ii) *For every (anti-)automorphism θ of A^* , wrt. $\tau_{d_\theta,1} = \widehat{\theta}$, the code X satisfies Conds. (c1) and (c2). In addition one can decide whether it satisfies any of Conds. (c3), (c4).*

7 Concluding remark

From the point of view of decidability, we have now fully studied the behaviors of Conds. (c1)–(c4) wrt. \mathcal{P}_k , \mathcal{F}_k , and $\widehat{\theta}$. We also note that, in [16] the authors were interested in the question of embedding a non maximal τ -independent set $L \subseteq A^*$ into some maximal one. From this point of view, the results of Corollary 3.13 and Props. 5.4, 6.9 attest that, in the frameworks of \mathcal{P}_k , \mathcal{P}_k , and $\widehat{\theta}$, in any case such an embedding can be labeledy done for variable-length codes.

Regarding further research, several ways may be involved:

– At first, for $k \geq 2$ the question whether the relation $\underline{\mathcal{F}}_k$ is regular or not remains open.

– The so-called *subsequence metric* associates, with each pair of words (w, w') , the integer $\delta(w, w') = |w| + |w'| - 2|lcs(w, w')|$, where $lcs(w, w')$ stands for some maximum length subsequence common to w and w' . With such a definition that metric appears as a direct extension of the factor one. Equivalently, $\delta(w, w')$ is the minimum number of one character insertions and deletions that have to be applied for computing w' by starting from w . From this point of view, the frameworks of the Hamming and Levenshtein metrics are also involved.

We observe that, wrt. $\tau_{\delta,k}$, results very similar to Prop. 5.2, 5.4 have been established in [26], however, the question whether or not Conds. (c1), (c2) are decidable remains open.

– More generally, it appears natural to study, from the point of view of decidability, the behavior of each of the conds. (c1)–(c4) in the framework of other topologies in the free monoid. Without being exhaustive, we mention the so-called additivity preserving quasi-metrics [3], or metrics based on absent words [4].

– From another point of view, a quasi-metric being fixed over A^* , presenting families of codes satisfying all the best Conds. (c1)–(c4) would be desirable.

Appendix

In view of Theorem 6.10, in what follows we provide some outline of basic results in order to implement corresponding algorithms:

(i) *Regular operations, boolean operations*

The proof of Prop. 2.6 lays upon the fact that every regular (resp., boolean) operation among regular sets can be translated in term of corresponding ones among finite automata (see e.g. [12, Chapter 3]). Operations such as quotient, direct, or inverse image under monoid endomorphism are also involved.

(ii) *Deciding whether a regular set is empty*

In order to do so, starting with a finite automaton with behavior X , we will decide whether or not a successful path exists by applying some classical graph mining algorithm. More precisely, according to [12, Theorem 3.7], for a n -state automaton, we have $X = \emptyset$ iff. no path with length less than n can be successful.

(iii) *Regular expressions and automata for regular subsets of A^**

A description of a regular set $X \subseteq A^*$ by using only the operations union, product and Kleene star is called a *regular expression* [12, Sect. 2.5]. Several classical methods can be applied in order to switch between the representation of X by an automaton or by a corresponding regular expression [10, 33]. Furthermore, every regular subset of A^* can be described by a so-called *unambiguous* regular expression, which is exclusively built from unambiguous regular operations (see e.g. [8, Corollary VII.8.3]). In order to compute such an unambiguous regular expression, starting with a finite automaton with behavior X , several methods can be applied, the best-known certainly being McNaughton and Yamada algorithm (see [23] or [1, Proposition 4.1.8]). Regarding Bernoulli measure, unambiguous regular expressions allow to compute $\mu(X)$ by recursively applying the three following formulas: $\mu(R_1 + R_2) = \mu(R_1) + \mu(R_2)$, $\mu(R_1 R_2) = \mu(R_1)\mu(R_2)$, and $\mu(R^*) = (1 - \mu(R))^{-1}$.

(iv) *Image of a regular set under a regular word binary relation.*

The proof of Prop. 2.7, lays upon a result from [30], concerning a peculiar decomposition of regular relations. More precisely, given a regular relation $\tau \subseteq A^* \times A^*$, there are a finite alphabet Z , a regular set $K \subseteq Z^*$, and two monoid homomorphisms $\phi, \psi : Z^* \rightarrow A^*$ st. $\tau = \phi^{-1} \cdot \iota_K \cdot \psi$. In this equation, the relation $\iota_K \subseteq Z^* \times Z^*$ is defined by $\iota_K(z) = \{z\} \cap K$, for every $z \in Z^*$. In addition we have $Z = A' \cup A''$, where A' and A'' are disjoint copies of the alphabet A . From the point of view of implementation, some finite automata with behaviors ϕ , ψ , and ι_K can be explicitly computed. In particular, starting with a finite $A^* \times A^*$ -automaton with behavior τ , say \mathcal{A} , a finite automaton \mathcal{A}' with behavior ι_k can be constructed by associating, with each transition $p \xrightarrow{(a_1 \dots a_m, b_1 \dots b_n)} q$ (with $m, n \geq 0$) in \mathcal{A} , the transition $p \xrightarrow{a'_1 \dots a'_m b''_1 \dots b''_n} q$ in \mathcal{A}' . The mapping ϕ (resp., ψ), for its part, is constructed on the basis of the free monoid projection $\pi_{A'} : Z^* \rightarrow A'^*$ (resp., $\pi_{A''} : Z^* \rightarrow A''^*$). By the way, a finite A^* -automaton with behavior $\tau(X)$ can be effectively constructed. The reader could find more precise details in [32, Sect. IV.1.3.1].

(v) *Deciding whether a regular set is a code*

Given a regular set X , the question can be classically solved by applying Sardinas and Patterson algorithm [34]. Starting with the set $X^{-1}X \setminus \{\varepsilon\}$, an ul-

timately periodic sequence, say $(U_n)_{n \geq 0}$, is computed by applying the following induction formula: $U_{n+1} = U_n^{-1}X \cup X^{-1}U_n$. In view of Prop. 2.7, each term of $(U_n)_{n \geq 0}$ is a regular set. The algorithm necessarily stops: this corresponds to either $\varepsilon \in U_n$, or $U_n = U_p$, for some pair of different integers $p < n$: X is a code iff. the second cond. holds.

Example 7.1. Over $A = \{a, b\}$, the set $X = \{a, ab, baa\}$ is not a code. Indeed, we have $U_0 = \{b\}$, $U_1 = U_0^{-1}X \cup X^{-1}U_0 = \{aa\}$, $U_2 = U_1^{-1}X \cup X^{-1}U_1 = \{a\}$, and $U_3 = U_2^{-1}X \cup X^{-1}U_2 = \{\varepsilon, b\}$. We verify that the equation $ab \cdot a \cdot a = a \cdot baa$ holds among the words of X .

References

- [1] J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata*. Cambridge University Press, 2010.
- [2] V. Bruyère, Limin Wang, and Liang Zhang. On completion of codes with finite deciphering delay. *European J. Comb.*, 11(6):513–521, 1990.
- [3] C.S. Calude, A. Salomaa, and S. Yu. Additive distances and quasi-distances between words. *J. of Universal Comput. Sci.*, 8(2):141–122, 2002.
- [4] G. Castiglione, S. Mantaci, and A. Restivo. Some investigations on similarity measures based on absent words. *Fundam. Informaticae*, 171(1-4):97–112, 2020.
- [5] C. Choffrut and G. Pighizzini. Distances between languages and reflexivity of relations. *Theoret. Comp. Sci.*, 286(1-6):117–138, 2002.
- [6] P.M. Cohn. *Universal Algebra (Mathematics and Its Applications, 6)*. Springer, 1981.
- [7] A. Ehrenfeucht and S. Rozenberg. Each regular code is included in a maximal regular code. *RAIRO - Theoret. Inf. Appl.*, 20(1):89–96, 1985.
- [8] S. Eilenberg. *Automata, Languages and Machines*. Academic Press, 1974. eBook ISBN: 9780080873749.
- [9] C. C. Elgot and J. E. Mezei. On relations defined by generalized finite automata. *IBM J. Res. Develop.*, 9(1):47–68, 1965.
- [10] H. Gruber, M. Holzer, and M. Kutrib. Descriptive complexity of regular languages. In J-É. Pin, editor, *Handbook of Automata Theory*, volume [I], chapter 12, pages 411–458. EMS Press, Berlin, 2021.
- [11] R. W. Hamming. Error detecting and error correcting codes. *The Bell Syst. Techn. J.*, 29(2):147–160, 1950.

- [12] J. E. Hopcroft and D. Ullman. *Introduction to automata theory, languages and Computation*. Addison-Wesley publishing company Reading, MA, Menlo Park, CA, London, Amsterdam, Don Mills ONT, Sydney, 1979.
- [13] H. Jürgensen, Masashi Katsura, and S. Konstantinidis. Maximal solid codes. *J. of Automata, Languages and Combinatorics*, 6(1):25–50, 2001.
- [14] H. Jürgensen and S. Konstantinidis. Codes. In *Handbook of Formal Languages*, volume 1, chapter 8, pages 511–607. Springer-Verlag Berlin, 1997. ISBN 78-3-642-59136-5.
- [15] L. Kari, S. Konstantinidis, and S. Kopecki. On the maximality of languages with combined types of code properties. *Theoret. Comp Sci.*, 550:79–89, 2014.
- [16] S. Konstantinidis and M. Mastnak. Embedding rationally independent languages into maximal ones. *J. of Aut., Lang. and Comb.*, 21(4):311–338, 2016.
- [17] J. B. Kruskal. An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM Review*, 25:201–237, 1983.
- [18] Nguyen Huong Lam. Finite maximal infix codes. *Semigr. Forum*, 61:346–356, 2000.
- [19] Nguyen Huong Lam. Finite maximal solid codes. *Theoret. Comput. Sci.*, 262:333–347, 2001.
- [20] Nguyen Huong Lam. Completing comma-free codes. *Theot. Comput. Sci.*, 301(1):399–415, 2003.
- [21] V.I. Levenshtein. Binary codes capable of correcting deletions, insertion, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966. Translat. from Doklady Akademii Nauk SSSR 163(4):845-848, 1965.
- [22] M. Lothaire. *Combinatorics on Words*. Addison-Wesley Publishing Company (2nd edition in Cambridge: Cambridge University Press 1997), 1983.
- [23] R. McNaughton and H. Yamada. Regular expressions and state graphs for automata. *IRE Trans. Electronic Computers*, EC-9(1):39–47, 1960.
- [24] J. Néraud. On the completion of codes in submonoids with finite rank. *Fundam. Informaticae*, 74(4):549–562, 2006. ISSN 0169-2968.
- [25] J. Néraud. Completing circular codes in regular submonoids. *Theoret. Comput. Sci.*, 391(1-2):90–98, 2008.
- [26] J. Néraud. Variable-length codes independent or closed with respect to edit relations. *Inf. Comput.*, 288:104747, 2022.

- [27] J. Néraud. When variable-length codes meet the field of error detection. In D. Poulakis and G. Rahonis, editors, *Algebraic Informatics 9th International Conference on Algebraic Informatics, CAI 2022, Virtual Event, October 27-29, 2022, Proceedings*, volume 13706, pages 203–222. Lect. Notes in Comp. Sci., 2022.
- [28] J. Néraud and C. Selmi. Embedding a θ -invariant code into a complete one. *Theoret. Comput. Sci.*, 806:28–41, 2020.
- [29] Timothy Ng. Prefix distance between regular languages. In Yo-Sub Han and Kai Salomaa, editors, *Implementation and Applications of Automata, 21st International Conference, CIAA 2016, Seoul, South Korea, July 19-22, 2016, Proceedings*, volume 9705, pages 224–235. Lect. Notes in Comp. Sci., 2016.
- [30] M. Nivat. Transductions des langages de Chomsky. *Ann. Inst. Fourier (Grenoble)*, Tome 18(1):339–455, 1968.
- [31] A. Restivo. On codes having no finite completion. *Discr. Math.*, 17(1):309–316, 1977.
- [32] J. Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, 2009.
- [33] J. Sakarovitch. Automata and rational expressions. In J-É. Pin, editor, *Handbook of Automata Theory*, volume I, chapter 2, pages 39–78. EMS Press, 2021.
- [34] A. A. Sardinas and G. W. Patterson. A necessary and sufficient condition for the unique decomposition of coded messages. *IRE Internat. Conv. Rec.*, 8:104–108, 1953.
- [35] Do Long Van, Kieu Van Hung, and Phan Trung Huy. Codes and length-increasing transitive binary relations. In Dang Van Hung and M. Wirsing, editors, *Theoretical Aspects of Computing - ICTAC 2005 Second International Colloquium, Hanoi, Vietnam, October 17-21, 2005, Proceedings*, volume 3722, pages 29–48. Lect. Notes in Comp. Sci., 2005.
- [36] W. A. Wilson. On Quasi-Metric Spaces. *American J. of Math.*, 53:675–684, 1931.
- [37] Liang Zhang and Zhonghui Shen. Completion of recognizable bifix codes. *Theoret. Comput. Sci.*, 145(1-2):345–355, 1995.