HAL open science

When Variable-Length Codes Meet the Field of Error Detection

Jean Néraud

▶ To cite this version:

Jean Néraud. When Variable-Length Codes Meet the Field of Error Detection. 2022. hal-03764465v2

HAL Id: hal-03764465
https://hal.science/hal-03764465v2

Preprint submitted on 5 Oct 2022

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.

When Variable-Length Codes Meet the Field of Error Detection

Jean Néraud **

Université de Rouen, Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes, Avenue de l'Université, 76800 Saint-Étienne-du-Rouvray, France
neraud.jean@gmail.com
http:neraud.jean.free.fr;orcid:0000-0002-9630-461X

Abstract. Given a finite alphabet A and a binary relation $\tau \subseteq A^* \times A^*$, a set X is τ-*independent* if $\tau(X) \cap X = \emptyset$. Given a quasi-metric d over A^* (in the meaning of [27]) and $k \geq 1$, we associate the relation $\tau_{d,k}$ defined by $(x, y) \in \tau_{d,k}$ if, and only if, $d(x, y) \leq k$ [3]. In the spirit of [10, 20], the error detection-correction capability of variable-length codes can be expressed in term of conditions over $\tau_{d,k}$. With respect to the prefix metric, the factor one, and every quasi-metric associated to (anti-)automorphisms of the free monoid, we examine whether those conditions are decidable for a given regular code.

Keywords: Anti-reflexive, automaton, automorphism, anti-automorphism, Bernoulli measure, binary relation, channel, code, codeword, complete, distance, error correction, error detection, embedding, factor, free monoid, homomorphism, independent, input word, Kraft inequality, maximal, measure, metric, monoid, output word, prefix, quasi-metric, regular, subsequences, suffix, synchronization constraint, transducer, variable-length code, word

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 theory, 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 $\phi : B^* \longrightarrow A^*$. In numerous cases, ϕ is a *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

** Corresponding author

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 of 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 states that, for each input word factorized $w = x_1 \cdots x_n$, where $x_1, \cdots, x_n$ are *codewords* in X, every output word has to be factorized $w' = x'_1 \cdots x'_n$, with $(x_1, x'_1), \cdots, (x_n, x'_n) \in \widehat{\tau}$. In order to ensure such a constraint, as for the Morse code, some pause symbol could be inserted after each codeword x_i.

With regard to the code X, in order to minimize the number of errors, in most cases some close neighbourhood constraint over $\widehat{\tau}(X)$ is applied. 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 [27] (the difference with a *metric* is that d needs not to satisfy the symmetry axiom). As outlined in [3], given an error tolerance level $k \geq 0$, a corresponding binary word relation, that we denote by $\tau_{d,k}$, can be associated in such a way that $(w, w') \in \tau_{d,k}$ is equivalent to $d(w, w') \leq k$. Below, in the spirit of [10, 20], we draw some specification regarding error detection-correction capability. Recall that a subset X of A^* is *independent* with respect to $\tau \subseteq A^* \times A^*$ (for short, τ-*independent*) whenever $\tau(X) \cap X = \emptyset$: this notion, which appears dual with the one of *closed code* [20], relies to the famous *dependence systems* [4, 10]. 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 $Y = X$. 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 if, and only if, the distance between different elements of X is greater than k that is, the code X can detect at most k errors in the transmission of every codeword $x \in X$.

– Cond. (c2) states a classical definition: equivalently, for every codeword x we have $\tau_{d,k}^{-1}(\tau_{d,k}(x)) \cap X = \{x\}$ whenever $\tau_{d,k}(x)$ is non-empty [20].

– 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 classes determined by code properties [9, 11, 14, 15]. On the other hand, according to the famous Kraft inequality, given a *positive Bernoulli measure* over A^*, say μ, for every code X we have $\mu(X) \leq 1$. According to a famous result due to Schützenberger, given a *regular* code X, the condition $\mu(X) = 1$ itself corresponds to X being maximal in the whole family of codes, or equivalently X being *complete* that is, every word in A^* is a *factor* of some word in X^*. From this last point of view, no part of X^* appears spoiled.

– At last, Cond. (c4) expresses that the factorization of every output message over the set $\widehat{\tau_{d,k}}(X)$ is done in a unique way. Since d is a quasimetric, the corresponding relation $\tau_{d,k}$ is reflexive, therefore Cond. (c4) is equivalent to $\tau_{d,k}(X)$ being a code.

As shown in [10, 20], there are regular codes satisfying (c1) that cannot satisfy (c2). Actually, in most of the cases it could be very difficult, even impossible, to satisfy all together Conds. (c1)–(c4): necessarily 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 conditions is satisfied in the framework of some special quasi-metric. From this point of view, in [20], 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 [8, 13, 17], which are prioritary related to subsequences in words. In the present paper, we focuse on quasimetrics rather involving factors:

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

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

– 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 concerns the domain of DNA sequence comparison. By anti-automorphism of the free monoid, we mean a one-to-one mapping onto A^*, say θ, such that the equation $\theta(uv) = \theta(v)\theta(u)$ holds for every $u, v \in A^*$ (for involvements of such mappings in the dual notion of closed code, see [21]). 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 have $d_\theta(w, w') = 1$ whenever $w' = \theta(w)$ holds, with $w \neq w'$;
(3) in all other cases we set $d_\theta(w, w') = 2$.

By definition we have $\tau_{d_\theta,1} = \widehat{\theta}$ and $\underline{\tau_{\theta,1}} = \underline{\theta}$. In addition, a code X is $\underline{\theta}$-independent if, and only if, for every pair of different words $x, y \in X$, we have $d_\theta(x, y) = 2$ that is, X is capable to detect at most one error.

We will establish the following result:

Theorem. *With the preceding notation, given a regular code $X \subseteq A^*$, it can be decided whether X satisfies each of the following conditions:*

(i) *Conds.* (c1)–(c4) *wrt.* $\mathcal{P}_k$.
(ii) *Conds.* (c3), (c4) *wrt.* $\mathcal{F}_k$.
(iii) *If X is finite, Conds.* (c1), (c2) *wrt.* $\mathcal{F}_k$.

(iv) *Conds.* (c3), (c4) *wrt.* $\widehat{\theta}$*, for any (anti-)automorphism* θ *of* A^**. In that framework, in any case* X *satisfies Conds. (c1), (c2), which are actually equivalent.*

Some comments regarding the proof:

– For proving that X satisfies Cond. (c1) wrt. $\mathcal{P}_k$, the main argument consists in establishing that $\underline{\mathcal{P}_k}$ is a *regular* relation [24, Ch. IV] that is, $\underline{\mathcal{P}_k}$ can be simulated by a finite transducer.

– Once more wrt. $\mathcal{P}_k$, we actually prove that Cond. (c2) is equivalent to $(X \times X) \cap \underline{\mathcal{P}_{2k}} = \emptyset$. Since $X \times X$ is a *recognizable* relation [24, Ch. IV], it can be decided whether that equation holds.

– Regarding Cond. (c3), the critical step is reached by proving that, wrt. each of the quasi-metrics raised in the paper, for a regular code X, being maximal in the family of $\underline{\tau_{d,k}}$-independent codes is equivalent to being complete that is, $\mu(X) = 1$. This is done by proving that every non-complete $\underline{\tau_{d,k}}$-independent code, say X, can be embedded into some complete $\underline{\tau_{d,k}}$-independent one: in other words, X cannot be maximal in such a family of codes. In order to establish such a property, in the spirit of [2, 16, 18, 19, 21, 28], we provide specific regularity-preserving embedding formulas, whose schemes are based upon the method from [5]. Notice that, in [9, 12, 14, 15, 26], wrt. peculiar families of sets, algorithmic methods for embedding a set into some maximal one were also provided.

– With regard to Cond. (c4), for each of the preceding relations, the set $\widehat{\tau_{d,k}}(X) = \tau_{d,k}(X)$ is regular therefore, in any case, by applying the famous Sardinas and Patterson algorithm [25], one can decide whether that condition is satisfied.

We now shorty describe the contents of the paper:

– Section 2 is devoted to the preliminaries: we recall fundamental notions over codes, regular (resp., *recognizable*) relations, and automata.

– The aim of Sect. 3 is to study $\mathcal{P}_k$. We prove that, in any case, the corresponding relation $\underline{\mathcal{P}_k}$ is itself regular. Furthermore, given a regular code X, one can decide whether X satisfies any of Conds. (c1)–(c4). Some remarks are also formulated regarding the so-called *suffix* metric.

– Sect. 4 is concerned with the factor metric. We prove that, given a finite code, one can decide whether it satisfies any of Conds. (c1)–(c4). For a non-finite regular codes, we prove that one can decide whether it satisfies Conds. (c3), (c4), however, the question of the decidability of Conds. (c1), (c2) remains open.

– Sect. 5 is devoted to (anti-)automorphisms. We obtain results similar to those involving the relation $\mathcal{P}_k$.

– In Sect. 6, the paper concludes with some possible directions for further research.

2 Preliminaries

Several definitions and notation has already been stated in the introduction. In the whole paper, we fix a finite alphabet A, with $|A| \geq 2$, and we denote by ε the

word of length 0. Given two words $v, w \in A^*$, v is a *prefix* (resp., *suffix*, *factor*) of w if words u, u' exist such that $w = vu$ (resp., $w = u'v$, $w = u'vu$). We denote by $\mathrm{P}(w)$ (resp., $\mathrm{S}(w)$, $\mathrm{F}(w)$) the set of the words that are prefix (resp., suffix, factors) of w. More generally, given a set $X \subseteq A^*$, we set $\mathrm{P}(X) = \bigcup_{w \in X} \mathrm{P}(w)$; the sets $\mathrm{S}(X)$ and $\mathrm{F}(X)$ are defined in a similar way. A word $w \in A^*$, is *overlapping-free* if $wv \in A^*w$, with $|v| \leq |w| - 1$, implies $v = \varepsilon$.

Variable-length codes

We assume that the reader has a fundamental understanding with the main concepts of the theory of variable-length codes: we suggest, if necessary, that he (she) refers to [1]. Given a subset X of A, and $w \in X^*$, let $x_1, \cdots, x_n \in X$ such that w is the result of the concatenation of the words $x_1, x_2, \cdots, x_n$, 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 $i \in [1, n]$ (equivalently, the submonoid X^* is *free*). In other words, every element of X^* has a unique factorization over X. A set $X \neq \{\varepsilon\}$ is a *prefix* (resp., *suffix*) *code* if $x \in \mathrm{P}(y)$ (resp., $x \in \mathrm{S}(y)$) implies $x = y$, for every pair of words $x, y \in X$; X is a *bifix* code if it is both a prefix code and a suffix one. A set $X \subseteq A^*$ is *uniform* if all its elements have a common length. In the case where we have $X \neq \{\varepsilon\}$, the uniform set X is a bifix code.

Given a regular set X, the Sardinas and Patterson algorithm allows to decide whether or not X is a code. Since it will be used several times through the paper, it is convenient to shortly recall it. Actually, some ultimately periodic sequence of sets, namely $(U_n)_{n \geq 0}$, is computed, as indicated in the following:

$$U_0 = X^{-1}X \setminus \{\varepsilon\} \quad \text{and}: \quad (\forall n \geq 0) \quad U_{n+1} = U_n^{-1}X \cup X^{-1}U_n. \tag{1}$$

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 if, and only if, the second condition holds.

A positive *Bernoulli distribution* consists in some 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) such that $\sum_{a \in A} \mu(a) = 1$; that mapping is extended into a unique monoid homomorphism from A^* into $(\mathbb{R}_+, \times)$, which is itself extended into a unique positive measure $\mu : 2^{A^*} \longrightarrow \mathbb{R}_+$. In order to do this, for each word $w \in A^*$, we set $\mu(\{w\}) = \mu(w)$; in addition given two disjoint subsets X, Y of A^*, we set $\mu(X \cup Y) = \mu(X) + \mu(Y)$. 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 following results are classical: the first one is due to Schützenberger and the second provides some answer to a question actually stated in [23].

Theorem 1. [1, Theorem 2.5.16] *Given a regular code $X \subseteq A^*$, the following properties are equivalent:*

(i) *X is complete;*
(ii) *X is a maximal code;*
(iii) *we have $\mu(X) = 1$.*

Theorem 2. [5]. *Let $X \subseteq A^*$ be non-complete code, $z \notin \mathrm{F}(X^*)$ overlapping-free, $U = A^* \setminus (X^* \cup A^* z A^*)$, and $Y = (zU)^* z$. Then $Z = X \cup Y$ is a complete code.*

Clearly, if X is a regular set then the same holds for the resulting set Z.

Regular relations, recognizable relations

We also assume the reader to be familiar with the theory of regular relations and automata: if necessary, we suggest that he (she) refers to [6] or [24, Ch. IV].

Given two monoids, say M, N, a binary relation from M into N consists in any subset $\tau \subseteq M \times N$. For $(w, w') \in \tau$, we also set $w' \in \tau(w)$, and we set $\tau(X) = \{\tau(x) : x \in X\}$. 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); τ^{-1}, the *inverse* of τ, is defined by $(w, w') \in \tau^{-1}$ whenever $(w', w) \in \tau$. We denote by $\overline{\tau}$ the *complement* of τ, i.e. $(M \times N) \setminus \tau$.

A family of subsets of M, say $\mathcal{F} \subseteq 2^M$, is *regularly closed* if, and only if, the sets $X \cup Y$, XY, and X^* belong to $\mathcal{F}$, whenever we have $X, Y \in \mathcal{F}$. Given a family $\mathcal{F} \subseteq 2^M$, its *regular closure* is the smallest (wrt. the inclusion) subset of 2^M that contains $\mathcal{F}$ and which is regularly closed. A binary relation $\tau \subseteq M \times N$ is *regular* (or equivalently, *rational*) if, and only if, it belongs to the regular closure of the finite subsets of $M \times N$. Equivalently there is some finite $M \times N$-*automaton* (or equivalently, *transducer*), say $\mathcal{R}$, with *behavior* $|\mathcal{R}| = \tau$ [7, 24]. The family of regular relations is closed under inverse and composition.

The so-called *recognizable* relations constitute a noticeable subfamily in regular relations: a subset $R \subseteq M \times N$ is *recognizable* if, and only if, we have $R = R \cdot \phi \cdot \phi^{-1}$, for some monoid homomorphism $\phi : M \times N \longrightarrow P$, with P finite. Equivalently, finite families of recognizable subsets of M and N, namely $(T_i)_{\in I}$ and $(U_i)_{i \in I}$, exist such that $R = \bigcup_{i \in I} T_i \times U_i$ [24, Corollary II.2.20].

In the paper, we focus on $M = N = A^*$. With this condition, recognizable relations are closed under composition, complement and intersection, their intersection with a regular relation being itself regular [24, Sect. IV.1.4]. According to [24, Theorem IV.1.3], given a regular relation $\tau \subseteq A^* \times A^*$, and a regular (equivalently, recognizable) set $X \subseteq A^*$, the sets $\tau(X)$ and $\tau^{-1}(X)$ are regular. If X, Y are recognizable subsets of A^*, the same holds for $X \times Y$. At last the relation $id_{A^*} = \{(w, w) | w \in A^*\}$ and its complement $\overline{id_{A^*}}$, are regular but non-recognizable.

3 Error detection and the prefix metric

Given $w, w' \in A^*$, a unique pair of words u, u' exist such that $w = (w \wedge w')u$ and $w' = (w \wedge w')u'$: by definition, we have $d_{\mathrm{P}}(w, w') = |u| + |u'|$ (see Figure 1). In addition $\mathcal{P}_k$ is reflexive and symmetric that is, the equality $\mathcal{P}_k^{-1} = \mathcal{P}_k$ holds, and we have $\mathcal{P}_k \subseteq \mathcal{P}_{k+1}$. Below, we provide some example:

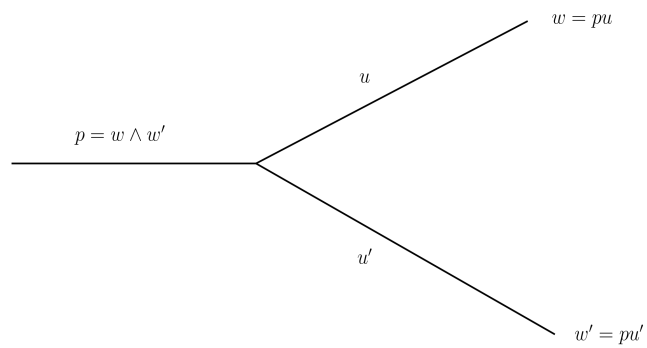

Fig. 1. We have $(w, w') \in \mathcal{P}_k$ iff. $|u| + |u'| \leq k$; similarly, $(w, w') \in \underline{\mathcal{P}_k}$ is equivalent to $1 \leq |u| + |u'| \leq k$.

Example 1. Let $A = \{a, b\}$. (1) The finite prefix code $X = \{a, ba, b^2\}$ satisfies Cond. (c1) wrt. $\mathcal{P}_1$ (in other words X is 1-error-detecting). Indeed, it follows from $d_{\mathrm{P}}(a, ba) = d_{\mathrm{P}}(a, bb) = 3$ and $d_{\mathrm{P}}(ba, bb) = 2$ that we have $(x, y) \notin \mathcal{P}_1$, for each pair of different words $x, y \in X$, that is $\underline{\mathcal{P}_1}(X) \cap X = \emptyset$. Cond. (c2) is not satisfied by X. Indeed, we have:

$\mathcal{P}_1^{-1}\mathcal{P}_1(ba) \cap X = \{ba, b, ba^2, bab, \varepsilon, b^2, ba^3, ba^2b, baba, bab^2\} \cap X = \{ba, b^2\}$. Classically, X is a maximal code, therefore, since it is $\underline{\mathcal{P}_1}$-independent, X is maximal in the family of $\underline{\mathcal{P}}_1$-independent codes that is, X satisfies Cond. (c3). Consequently, since we have $X \subsetneq \mathcal{P}_1(X)$, the code X cannot satisfy Cond. (c4) (we verify that we have $\varepsilon \in \mathcal{P}_1(X)$).

(2) For $n \geq k+1$, the complete uniform code $Y = A^n$ satisfies Conds. (c1), (c3). Since Y is a maximal code, it cannot satisfies Cond. (c4) (we have $Y \subsetneq \mathcal{P}_1(Y)$). Cond. (c2) is no more satisfied by Y: indeed, given two different characters a, b, we have $\mathcal{P}_k(a^n) \neq \emptyset$ and $a^{n-1}b \in \mathcal{P}_k^{-1}(\mathcal{P}_k(a^n)) \cap Y$ thus $\mathcal{P}_k^{-1}(\mathcal{P}_k(a^n)) \cap Y \neq \{a^n\}$.

(3) The regular bifix code $Z = \{ab^na : n \geq 0\} \cup \{ba^nb : n \geq 0\}$ satisfies Cond. (c1) wrt. $\mathcal{P}_1$. Indeed, we have:

$\underline{\mathcal{P}}_1(Z) = \bigcup_{n\geq 0}\{ab^n, ab^na^2, ab^nab, ba^n, ba^nba, ba^nb^2\}$, thus $\underline{\mathcal{P}}_1(Z) \cap Z = \emptyset$. For $n \neq 0$ we have $\mathcal{P}_1^{-1}\mathcal{P}_1(ab^na) = \{ab^n, ab^{n-1}, ab^na, ab^{n+1}, ab^na^2, ab^na^3, ab^nab, ab^na^2b, ab^naba, ab^nab^2\}$, moreover we have $\mathcal{P}_1^{-1}\mathcal{P}_1\left(a^2\right) = \mathcal{P}_1\left(\{a, a^3, a^2b\}\right)$, thus $\mathcal{P}_1^{-1}\mathcal{P}_1\left(a^2\right) = \{a, \varepsilon, a^2, ab, a^3, a^4, a^3b, a^2b, a^2ba, a^2b^2\}$: in any case we obtain $Z \cap \mathcal{P}_1^{-1}\mathcal{P}_1(ab^na) = \{ab^na\}$. Similarly, we have $Z \cap \mathcal{P}_1^{-1}\mathcal{P}_1(ba^nb) = \{ba^nb\}$, hence Z satisfies Cond. (c2). At last, we have $\mu(X) = 2 \cdot 1/4 \sum_{n\geq 0}(1/2)^n = 1$ therefore, according to Theorem 1, Z is a maximal code, whence it is maximal in the family of $\mathcal{P}_1$-independent codes (Cond. (c3)). Since we have $Z \subsetneq \mathcal{P}_1(Z)$, Z cannot satisfies Cond. (c4) (we verify that have $a, a^2 \in \mathcal{P}_1(Z) \subseteq \widehat{\mathcal{P}_1}(Z)$).

In the sequel, we will prove that, given a regular code X, one can decide whether any of Conds. (c1)-(c4) holds. Beforehand we establish the following property which, regarding Cond. (c1), plays a prominent part:

Proposition 1. *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 automaton with behavior $\underline{\mathcal{P}}_k$ (see Figure 2). This construction is based on the different underlying configurations in Figure 1. Firstly, we denote by E the finite set of all the pairs of non-empty words (u, u'), with different initial characters, and such that $|u| + |u'| \leq k$. In addition, F (resp., G) stands for the set of all the pairs (u, ε) (resp., (ε, u)), with $1 \leq |u| \leq k$. Secondly, we construct a finite tree-like $A^* \times A^*$-automaton with behavior $E \cup F \cup G$. Let 0 be the initial state, the other states being terminal. We complete the construction by adding the transitions $(0, (a, a), 0)$, for all $a \in A$. Let $\mathcal{R}_{\mathrm{P},k}$ be the resulting automaton.
By construction we have $|\mathcal{R}_{\mathrm{P},k}| = id_{A^*}(E \cup F \cup G)$. More precisely, $|\mathcal{R}_{\mathrm{P},k}|$ is the set of all the pairs (w, w') such that there are $v, u, u' \in A^*$ satisfying each of the three following conditions:

(1) $w = pu$, $w' = pu'$;
(2) if both the words u, u' are non-empty, their initial characters are different;
(3) $1 \leq |u| + |u'| \leq k$.

In other words, $|\mathcal{R}_{\mathrm{P},k}|$ is the sets of all the pairs (w, w') such that $1 \leq d_{\mathrm{P}}(w, w') = |u| + |u'| \leq k$, therefore we have $\underline{\mathcal{P}_k} = |\mathcal{R}_{\mathrm{P},k}|$. Consequently $\underline{\mathcal{P}_k}$ and $\mathcal{P}_k = \underline{\mathcal{P}_k} \cup id_{A^*}$ are regular relations. □

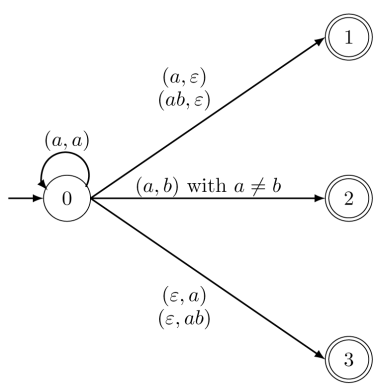

Fig. 2. The case where we have $k = 2$: in the automaton $\mathcal{R}_{\mathrm{P,k}}$, the arrows are muti-labelled (a, b stand for all pairs of characters in A) and terminal states are represented with double circles.

Remark 1. In [22] the author introduces a peculiar $(A^* \times A^*) \times \mathbb{N}$-automaton: in view of this, for every $(w, w') \in A^* \times A^*$, the distance $d_{\mathrm{P}}(w, w')$ is the least $d \in \mathbb{N}$ for which $((w, w'), d)$ is the label of some successful path. Furthermore, some alternative proof of the regularity of $\mathcal{P}_k$ can be obtained. However, we note

that such a construction cannot involve the relation $\underline{\mathcal{P}_k}$ itself that is, it does not affect Cond. (c1).

The following property is also used in the proof of Proposition 2:

Lemma 1. *Given a positive integer k we have $\mathcal{P}_{2k} = \mathcal{P}_k^2$.*

Proof. In order to prove that $\mathcal{P}_k^2 \subseteq \mathcal{P}_{2k}$, we consider a pair of words $(w, w') \in \mathcal{P}_k^2$. By definition, some word $w'' \in A^*$ exists such that we have $(w, w''), (w'', w') \in \mathcal{P}_k$ that is, $d_{\mathrm{P}}(w, w'') \leq k$, $d_{\mathrm{P}}(w'', w') \leq k$. This implies $d_{\mathrm{P}}(w, w') \leq d_{\mathrm{P}}(w, w'') + d_{\mathrm{P}}(w'', w') \leq 2k$ that is, $(w, w') \in \mathcal{P}_{2k}$.
Conversely, let $(w, w') \in \mathcal{P}_{2k}$, and let $p = w \wedge w'$. Regarding the integers $|w| - |p|$, $|w'| - |p|$, exactly one of the two following conditions occurs:

(a) Firstly, at least one of the integers $|w| - |p|$, $|w'| - |p|$ belongs to $[k + 1, 2k]$. Since $\mathcal{P}_{2k}$ is a symmetric relation, without loss of generality, we assume $k + 1 \leq |w| - |p| \leq 2k$. With this condition a non-empty word v exists such that $w = pvA^k$ that is, $pv \in \mathcal{P}_k(w)$ with $|pv| = |w| - k$. On the other hand, we have $p = pv \wedge w'$, thus $d_{\mathrm{P}}(pv, w') = (|w| - k) + |w'| - 2|p| = d_{\mathrm{P}}(w, w') - k$. It follows from $d_{\mathrm{P}}(w, w') \leq 2k$ that $d_{\mathrm{P}}(pv, w') \leq k$, thus $w' \in \mathcal{P}_k(pv)$: this implies $w' \in \mathcal{P}_k(\mathcal{P}_k(w))$, thus $(w, w') \in \mathcal{P}^2$.

(b) Secondly, in the case where we have $|w| - |p| \leq k$ and $|w'| - |p| \leq k$, by definition we have $p \in \mathcal{P}_k(w)$, $w' \in \mathcal{P}_k(w)$, thus $(w, w') \in \mathcal{P}_k^2$. □

We are now ready to establish the following result:

Proposition 2. *Given a regular code $X \subseteq A^*$, wrt. $\mathcal{P}_k$, it can be decided whether X satisfies any of Conds.* (c1)*,* (c2)*, and* (c4)*.*

Proof. Let X be a regular code. We consider one by one our Conds. (c1), (c2), and (c4):

– *Cond.* (c1) According to Proposition 1, $\underline{\mathcal{P}_k}(X)$ is a regular set, hence $\underline{\mathcal{P}_k}(X) \cap X$ itself is regular, therefore one can decide whether Cond. (c1) holds.

– *Cond.* (c2) Since $\mathcal{P}_k$ is a symmetric binary relation, and according to Lemma 1, we have: $\mathcal{P}_k^{-1}(\mathcal{P}_k(X)) \cap X = \mathcal{P}_k^2(X) \cap X = \mathcal{P}_{2k}(X) \cap X$. In addition $x \in X$ implies $\mathcal{P}_k(x) \neq \emptyset$, therefore Cond. (c2) is equivalent to $(X \times X) \cap \mathcal{P}_{2k} \subseteq id_{A^*}$. This last condition is equivalent to $(X \times X) \cap \left(\mathcal{P}_{2k} \cap \overline{id_{A^*}}\right) = \emptyset$, thus $(X \times X) \cap \underline{\mathcal{P}_{2k}} = \emptyset$. According to Proposition 1 and since $X \times X$ is a recognizable relation, the set $(X \times X) \cap \underline{\mathcal{P}_{2k}}$ is regular, therefore one can decide whether X satisfies Cond. (c2).

– *Cond.* (c4) According to Proposition 1, the set $\mathcal{P}_k(X)$ itself is regular. Consequently, by applying Sardinas and Patterson algorithm, it can be decided whether X satisfies Cond. (c4). □

It remains to study the bahaviour of X wrt. Cond. (c3). In order to do this, we proceed by establishing the two following results:

Proposition 3. *Let $X \subseteq A^*$ be a non-complete regular $\underline{\mathcal{P}_k}$-independent code. Then a complete regular $\underline{\mathcal{P}_k}$-independent code containing X exists.*

Proof. Beforehand, in view of Theorem 2, we indicate the construction of a convenient word $z \in A^* \setminus \mathrm{F}(X^*)$. Since X is a non-complete set, by definition some word z_0 exists in $A^* \setminus \mathrm{F}(X^*)$: without loss of generality, we assume $|z_0| \geq k$ (otherwise, we substitute to z_0 any word in $z_0 A^{k-|z_0|}$). Let a be the initial character of z_0, b be a character different of a, and $z = z_0 a b^{|z_0|}$. Classically, z is overlapping-free (see e.g. [1, Proposition 1.3.6]): set $U = A^* \setminus (X^* \cup A^* z A^*)$, $Y = z(Uz)^*$, and $Z = X \cup Y$.

According to Theorem 2, the set Z is a (regular) complete code. For proving that Z is $\underline{\mathcal{P}_k}$-independent, we argue by contradiction. By assuming that $\underline{\mathcal{P}_k}(Z) \cap Z \neq \emptyset$, according to the construction of Z, exactly one of the two following cases occurs:

(a) Firstly, $x \in X$, $y \in Y$ exist such that $(x, y) \in \underline{\mathcal{P}_k}$. With this condition we have $d_{\mathrm{P}}(x, y) = d_{\mathrm{P}}(y, x) = \left|(x \wedge y)^{-1} x\right| + \left|(x \wedge y)^{-1} y\right| \leq k$. According to the construction of Y, the word $z = z_0 a b^{|z_0|}$ is a suffix of y. It follows from $\left|(x \wedge y)^{-1} y\right| \leq k \leq |z_0|$ that $z_0 a b^{|z_0|-1} \in \mathrm{F}(x \wedge y)$, thus $z_0 \in \mathrm{F}(x)$: a contradiction with $z_0 \notin \mathrm{F}(X^*)$.

(b) Secondly, $y, y' \in Y$ exist such that $(y, y') \in \underline{\mathcal{P}_k}$. Let $p = y \wedge y'$, $u = p^{-1} y$, and $u' = p^{-1} y'$.

(b1) At first, assume $p \in \{y, y'\}$ that is, without loss of generality, $y' = p$. With such a condition, y' is a prefix of y. Since we have $y, y' \in z(Uz)^*$, and since z is an overlapping-free word, necessarily a word $v \in (Uz)^*$ exists such that $y = y'v$. It follows from $|v| = d_{\mathrm{P}}(y, y') \leq k \leq |z| - 1$ that $v = \varepsilon$, thus $y = y'$: a contradiction with $\underline{\mathcal{P}_k}$ being anti-reflexive.

(b2) Consequently we have $p \notin \{y, y'\}$ that is, $1 \leq |u| \leq k$ and $1 \leq |u'| \leq k$ (see Figure 3). By construction, we have $b^{|z_0|} \in \mathrm{S}(z) \subseteq \mathrm{S}(y) \cap \mathrm{S}(y')$: it follows from $1 \leq |u| \leq k \leq |z_0|$ and $1 \leq |u'| \leq k \leq |z_0|$ that $u, u' \in \mathrm{S}(b^{|z_0|}) \setminus \{\varepsilon\}$, thus $u, u' \in bb^*$: a contradiction with $p = y \wedge y'$.

In any case we obtain a contradiction, therefore Z is $\underline{\mathcal{P}_k}$-independent. □

Proposition 4. *Let $X \subseteq A^*$ be a regular code. Then X is maximal in the family of $\underline{\mathcal{P}_k}$-independent codes if, and only if, we have $\mu(X) = 1$.*

Proof. According to Theorem 1, $\mu(X) = 1$ implies X being a maximal code, thus X being maximal as a $\underline{\mathcal{P}_k}$-independent code. For the converse, we argue by contrapositive. Once more according to Theorem 1, $\mu(X) \neq 1$ implies X non-complete. According to Proposition 3, some $\underline{\mathcal{P}_k}$-independent code strictly containing X exists, hence X is not maximal as a $\underline{\mathcal{P}_k}$-independent code. □

If X is a regular set, $\mu(X)$ can be computed by making use of some rational series. As a consequence, we obtain the following result:

Proposition 5. *One can decide whether a given regular code X satisfies Cond.* (c3) *wrt. $\mathcal{P}_k$.*

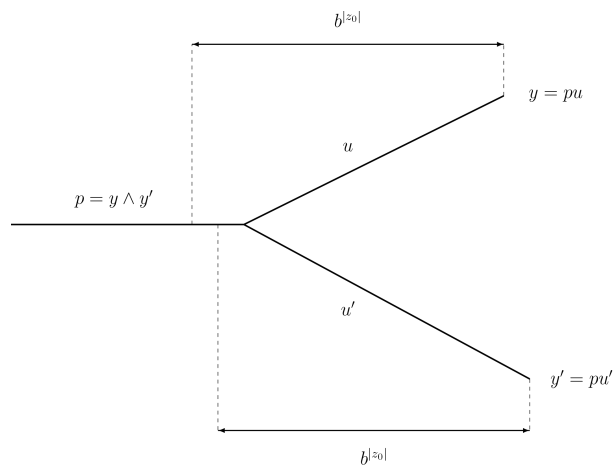

Fig. 3. Proof of Proposition 3: the case where we have $y, y' \in Y$ and $(y, y') \in \mathcal{P}_k$, with $p \notin \{y, y'\}$.

Remark 2. Given a pair of words w, w', their *suffix* distance is $d_{\mathrm{S}} = |w| + |w'| - 2|s|$, where s denotes the longest word in $\mathrm{S}(w) \cap \mathrm{S}(w')$. Let $\mathcal{S}_k$ be the binary relation defined by $(w, w') \in \mathcal{S}_k$ if, and only if, $d_{\mathrm{S}}(w, w') \leq k$. Given a word $w \in A^*$, denote by w^R its *reversal* that is, for $a_1, \cdots, a_n \in A$, we have $w^R = a_n \cdots a_1$ if, and only if, $w = a_1 \cdots a_n$ holds. For every pair $w, w' \in A^*$, we have $d_{\mathrm{S}}(w, w') = d_{\mathrm{P}}(w^R, w'^R)$, hence $(w, w') \in \mathcal{S}_k$ is equivalent to $(w^R, w'^R) \in \mathcal{P}_k$. As a consequence, given a regular code, one can decide whether it satisfies any of Conds. (c1)–(c4) wrt. $\mathcal{S}_k$.

4 Error detection and the factor metric

By definition, given a pair of words $w, w' \in A^*$, at least one tuple of words, say (u, v, u', v'), exists such that $d_{\mathrm{F}}(w, w') = |u| + |v| + |u'| + |v'|$. More precisely, we have $w = ufv$, $w' = u'fv'$, with f being of maximum length. Such a configuration is illustrated by Figure 4 which, in addition, can provide some support in view of examining the proof of Proposition 8.

Actually the word f, thus the tuple (u, v, u', v'), needs not to be unique (see Example 2(1)). Due to this fact, the construction in the proof of the preceding proposition 1 unfortunately cannot be extended in order to obtain a finite automaton with behaviour $\underline{\mathcal{F}}_k$.

For every positive integer k, the relation $\mathcal{F}_k$ is reflexive and symmetric that is, the equality $\mathcal{F}_k^{-1} = \mathcal{F}_k$ holds. In addition, with the preceding notation, we have $\mathcal{P}_k \cup \mathcal{S}_k \subseteq \mathcal{F}_k \subseteq \mathcal{F}_{k+1}$.

Example 2. (1) Let $w = bababababb$ and $w' = bbabbaababaa$. There are two words of maximum length in $F(w) \cap F(w')$, namely $f_1 = ababa$ and $f_2 = babba$. We

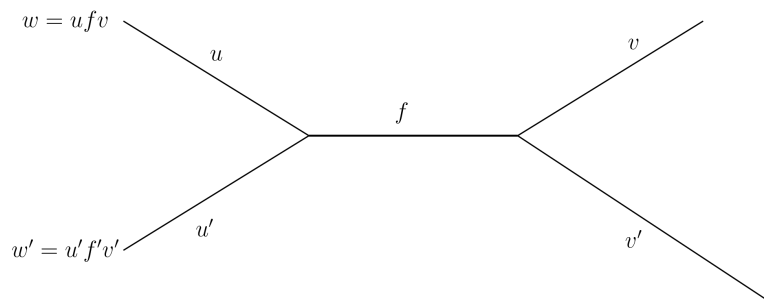

Fig. 4. We have $(w, w') \in \mathcal{F}_k$ iff. $|u| + |v| + |u'| + |v'| \leq k$.

have $d_{\mathrm{F}}(w, w') = |u_1| + |v_1| + |u'_1| + |v'_1| = |u_2| + |v_2| + |u'_2| + |v'_2| = 12$, where the tuples (u_1, v_1, u'_1, v'_1) and (u_2, v_2, u'_2, v'_2) satisfy the following equations:

$$w = u_1 f_1 v_1, w' = u'_1 f_1 v'_1, \quad \text{with} \quad u_1 = b, v_1 = bbab, u'_1 = bbabba, v'_1 = a,$$
$$w = u_2 f_2 v_2, w' = u'_2 f_2 v'_2, \quad \text{with} \quad u_2 = baba, v_2 = b, u'_2 = b, v'_2 = ababaa.$$

(2) Over the alphabet $\{a, b\}$, the code $X = \{a, ba, bb\}$ from Example 1 cannot satisfy Cond. (c1) wrt. $\mathcal{F}_1$. Indeed, we have $ba \in \underline{\mathcal{F}_1}(a) \cap X$. Since we have $\mathcal{F}_1^{-1}\mathcal{F}_1\,(a) \cap X = \{a, ba\}$, X cannot satisfy Cond. (c2). Althought X is complete, since it does not belong to the family of $\underline{\mathcal{F}_1}$-independent codes, X cannot satisfy Cond. (c3) wrt. $\mathcal{F}_1$. It follow from $\varepsilon \in \overline{\mathcal{P}_1}(X) \subseteq \mathcal{F}_1(X)$ that X can no more satisfy Cond. (c4).

(3) Take $A = \{a, b\}$ and consider the context-free bifix code $Y = \{a^n b^n : n \geq 1\}$. We have $\underline{\mathcal{F}_1}(Y) = \bigcup_{n \geq 1} \{a^{n-1}b^n, a^{n+1}b^n, ba^n b^n, a^n b^{n-1}, a^n b^n a, a^n b^{n+1}\}$. This implies $\underline{\mathcal{F}_1}(Y) \cap Y = \emptyset$, thus Y being 1-error-detecting wrt. $\mathcal{F}_1$ (Cond. (c1)). Regarding error correction, we have $a^{n+1}b^{n+1} \in \mathcal{F}_1^2(a^n b^n)$, therefore X cannot satisfy Cond. (c2) wrt. $\mathcal{F}_1$. We have $\mu(X) = \sum_{n \geq 1} \left(\frac{1}{4}\right)^n < 1$, whence Y cannot 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 $\widehat{\mathcal{F}_1}(Y) = \mathcal{F}_1(Y)$ cannot satisfy Cond. (c4).

The following property allows some noticeable connection between the frameworks of prefix, suffix, and factor metrics:

Lemma 2. *Given a positive integer k we have $\mathcal{F}_k = \mathcal{F}_1^k = (\mathcal{P}_1 \cup \mathcal{S}_1)^k$.*

Proof. – We start by proving that we have $\mathcal{F}_1^k = (\mathcal{P}_1 \cup \mathcal{S}_1)^k$. A indicated above, we have $\mathcal{P}_1 \cup \mathcal{S}_1 \subseteq \mathcal{F}_1$. Conversely, given $(w, w') \in \mathcal{F}_1$, some tuple of words (u, v, u', v') exists such that $w = ufv$, $w' = u'fv'$, with $0 \leq |u| + |v| + |u'| + |v'| \leq 1$, thus $|u| + |v| + |u'| + |v'| \in \{0, 1\}$. More precisely, at most one element of the set $\{|u|, |v|, |u'|, |v'|\}$ is a non-zero integer: this implies $(w, w') \in \mathcal{P}_1 \cup \mathcal{S}_1$. Consequently we have $\mathcal{F}_1 = \mathcal{P}_1 \cup \mathcal{S}_1$, thus $\mathcal{F}_1^k = (\mathcal{P}_1 \cup \mathcal{S}_1)^k$.

– Now, we prove we have $\mathcal{F}_1^k \subseteq \mathcal{F}_k$. Given a pair of words $(w, w') \in \mathcal{F}_1^k$, there is some sequence of words $(w_i)_{0 \leq i \leq k}$ such that $w = w_0$, $w' = w_k$, and $d_{\mathrm{F}}(w_i, w_{i+1}) \leq$

1, for each $i \in [0, k-1]$. We have $d_{\mathrm{F}}(w, w') \leq \sum_{0 \leq i \leq k-1} d_{\mathrm{F}}(w_i, w_{i+1}) \leq k$, thus $(w, w') \in \mathcal{F}_k$.

– For proving that the inclusion $\mathcal{F}_k \subseteq \mathcal{F}_1^k$ holds, we argue by induction over $k \geq 1$. The property trivially holds for $k = 1$. Assume that we have $\mathcal{F}_k \subseteq \mathcal{F}_1^k$, for some $k \geq 1$. Let $(w, w') \in \mathcal{F}_{k+1}$ and let $f \in \mathrm{F}(w) \cap \mathrm{F}(w')$ be a word with maximum length; set $w = ufv$, $w' = u'fv'$.

(a) Firstly, assume that at least one of the integers $|w| - |f|$, $|w'| - |f|$ belongs to $[2, k+1]$ that is, without loss of generality $2 \leq |w| - |f| = |u| + |v| \leq k+1$. With this condition, there are words $s \in \mathrm{S}(u)$, $p \in \mathrm{P}(v)$ such that $w \in A^h sfpA^{h'}$, with $sp \neq \varepsilon$ and $h + h' = 1$. On a first hand, it follows from $sfp \in \mathrm{F}(w)$ that $d_{\mathrm{F}}(w, sfp) = |w| - |sfp| = h + h' = 1$, thus $(w, sfp) \in \mathcal{F}_1$. On the other hand, f remains a word of maximum length in $\mathrm{F}(sfp) \cap \mathrm{F}(w')$ (otherwise words $s' \in \mathrm{S}(s)$, $p' \in \mathrm{P}(p)$ exist such that we have $s'fp' \in \mathrm{F}(w) \cap \mathrm{F}(w')$, with $|s'| + |p'| \geq 1$), whence we have $d_{\mathrm{F}}(sfp, w') = |sfp| + |w'| - 2|f|$. Since we have $|w| - |sfp| = 1$, we obtain $d_{\mathrm{F}}(sfp, w') = (|w| - 1) + |w'| - 2|f| = d_{\mathrm{F}}(w, w') - 1 \leq (k+1) - 1$, thus $(sfp, w') \in \mathcal{F}_k$ that is, by induction, $(sfp, w') \in \mathcal{F}_1^k$. Since we have $(w, sfp) \in \mathcal{F}_1$, this implies $(w, w') \in \mathcal{F}_1^{k+1}$.

(b) In the case where we have $|w| - |f| \leq 1$ and $|w'| - |f| \leq 1$ that is, $(w, f) \in \mathcal{F}_1$ and $(f, w') \in \mathcal{F}_k$, by induction we obtain $(f, w') \in \mathcal{F}_1^k$, therefore we have $(w, w') \in \mathcal{F}_1^{k+1}$.

In any case the condition $(w, w') \in \mathcal{F}_{k+1}$ implies $(w, w') \in \mathcal{F}_1^{k+1}$, hence we have $\mathcal{F}_{k+1} \subseteq \mathcal{F}_1^{k+1}$. As a consequence, for every $k \geq 1$ the inclusion $\mathcal{F}_k \subseteq \mathcal{F}_1^k$ holds: this completes the proof. □

As a direct consequence of Lemma 2, we obtain the following result:

Proposition 6. *Each of the following properties holds:*

(i) *The relation $\mathcal{F}_k$ is regular.*

(ii) *Given a regular code X, it can be decided whether X satisfies Cond.* (c4) *wrt. $\mathcal{F}_k$.*

(iii) *Given a finite code X, one can decide whether X satisfies any of Conds.* (c1)–(c4) *wrt. $\mathcal{F}_k$.*

Proof. In view of Sect. 3, the relations $\mathcal{P}_1$ and $\mathcal{S}_1$ are regular, therefore Property (i) comes from Lemma 2. The proof of Property (ii) is done by merely substituting $\mathcal{F}_k$ to $\mathcal{P}_k$ in the proof of the preceding proposition 2. In the case where X is finite, the same holds for $\mathcal{F}_k(X)$ and $\underline{\mathcal{F}_k}(X)$, furthermore Property (iii) holds. □

For non-finite regular sets, the question of the decidability of Conds. (c1), (c2) remains open. Indeed, presently no finite automaton with behavior $\underline{\mathcal{F}}_k = \mathcal{F}_k \cap \overline{id_{A^*}}$ is known. Actually, the following property holds:

Proposition 7. *For every $k \geq 1$, $\mathcal{F}_k$ is a non-recognizable regular relation.*

Proof. According to Proposition 6, the relation $\mathcal{F}_k$ is regular. By contradiction, assume $\mathcal{F}_k$ recognizable. As indicated in the preliminaries, with this condition a finite set I exists such that $\mathcal{F}_k = \bigcup_{i \in I}(T_i \times U_i)$. For every $n \geq 0$, we have

$(a^n b, a^n) \in \mathcal{F}_k$ therefore, since I is finite, there are $i \in I$ and $m, n \geq 1$, with $m - n \geq k$, such that $(a^n b, a^n), (a^m b, a^m) \in T_i \times U_i$. This implies $(a^n b, a^m) \in T_i \times U_i \subseteq \mathcal{F}_k$, thus $d_{\mathrm{F}}(a^n b, a^m) \leq k$: a contradiction with $d_{\mathrm{F}}(a^n b, a^m) = m - n + 1 \geq k + 1$. □

Regarding Cond. (c3), the following result holds:

Proposition 8. *Every non-complete regular $\underline{\mathcal{F}_k}$-independent code can be embedded into some complete one.*

Proof. In the family of $\underline{\mathcal{F}_{\mathrm{k}}}$-independent codes, we consider a non-complete regular set X. In view of Theorem 2, we will construct a convenient word $z_1 \in A^* \setminus F(X^*)$. Take a word $z_0 \notin \mathrm{F}(X^*)$, with $|z_0| \geq k$; let a be its initial character, and b be a character different of a. Consider the word z that was constructed in the proof of Proposition 3, that is $z = z_0 a b^{|z_0|}$. Set $z_1 = a^{|z|} b z = a^{2|z_0|+1} b z_0 a b^{|z_0|}$: since by construction, z_1^R, the reversal of z_1, is overlapping-free, the same holds for z_1. Set $U_1 = A^* \setminus (X^* \cup A^* z_1 A^*)$, $Y_1 = z_1 (U_1 z_1)^*$, and $Z_1 = X \cup Y_1$.

According to Theorem 2, the set Z_1 is a regular complete code. In order to prove that it is $\underline{\mathcal{F}_k}$-independent that is, $\underline{\mathcal{F}_k}(Z) \cap Z = \emptyset$, we argue by contradiction. Actually, exactly one of the following conditions occurs:

(a) The first condition sets that $x \in X$, $y \in Y_1$ exist such that $(x, y) \in \underline{\mathcal{F}_k}$. Let f be a word with maximum length in $\mathrm{F}(x) \cap \mathrm{F}(y)$: we have $y = ufv$, with $|u| + |v| \leq d_{\mathrm{F}}(x, y) \leq k$. It follows from $y \in z_1 (U_1 z_1)^*$ and $|z_1| \geq k + 1$ that we have $u \in \mathrm{P}(z_1)$ and $v \in \mathrm{S}(z_1)$. More precisely, according to the construction of y, we have $u \in \mathrm{P}\left(a^{|z_0|}\right)$, $v \in \mathrm{S}\left(b^{|z_0|}\right)$, and $u^{-1} y \in a^{|z_0|+1} A^*$, thus $f \in A^* z_0 A^*$. Since we have $f \in \mathrm{F}(x)$, we obtain $z_0 \in \mathrm{F}(x)$, a contradiction with $z_0 \notin \mathrm{F}(X^*)$.

(b) With the second condition, a pair of different words $y, y' \in Y_1$ exist such that $(y, y') \in \mathcal{F}_k$. Let f be a word with maximum length in $\mathrm{F}(y) \cap \mathrm{F}(y')$. As indicated above, words u, u', v, v' exist such that $w = ufv$, $w' = u'fv'$, with $|u| + |u'| + |v| + |v'| = d_{\mathrm{F}}(w, w') \leq k$.

(b1) At first, assume that both the words v, v' are different of ε. According to the construction of Y_1, since we have $v, v' \in \mathrm{S}(Y_1)$ with $|v| + |v'| \leq k$, a pair of positive integers i, j exist such that $v = b^i$, $v' = b^j$. This implies $fb \in \mathrm{F}(y) \cap \mathrm{F}(y')$, which contradicts the maximality of $|f|$.

(b2) As a consequence, at least one of the conditions $v = \varepsilon$, or $v' = \varepsilon$ holds. Without loss of generality, we assume $v' = \varepsilon$, thus $f \in \mathrm{S}(y')$. On a first hand, it follows from $z_1 \in \mathrm{F}(y) \cap \mathrm{F}(y')$ that $|f| \geq |z_1|$: since we have $f, z_1 \in \mathrm{S}(y')$, this implies $f \in A^* z_1$. On the other hand, since we have $z_1 \in \mathrm{S}(y)$, $fv \in \mathrm{S}(y)$, and $|f| \geq |z_1|$, we obtain $z_1 \in \mathrm{S}(fv)$. This implies $fv \in A^* z_1 v \cap A^* z_1$, thus $z_1 v \in A^* z_1$. It follows from $|v| \leq |z_1| - 1$ and z_1 being overlapping-free that $v = \varepsilon$. Similar arguments applied to the prefixes of y, y' lead to $u = u' = \varepsilon$: we obtain $y = y'$, a contradiction with $(y, y') \in \underline{\mathcal{F}_k}$.

In any case we obtain a contradiction, therefore Z_1 is $\underline{\mathcal{F}_k}$-independent. □

As a consequence, by merely substituting $\mathcal{F}_k$ to $\mathcal{P}_k$ in the proof of the propositions 4 and 5, we obtain the following statement:

Proposition 9. *Given a regular code X, each of the following properties holds:*
(i) *X is maximal as a $\underline{\mathcal{F}_k}$-independent code if, and only if, we have $\mu(X) = 1$.*
(ii) *One can decide whether X satisfies Cond.* (c3) *wrt. $\mathcal{F}_k$.*

5 Error detection in the topologies associated to (anti-)automorphisms

Given an (anti-)automorphism θ, we will examine Conds. (c1)–(c4) wrt. the quasi-metric d_θ that is, wrt. the relation $\tau_{d_\theta,1} = \widehat{\theta}$. Recall that we have $\underline{\left(\widehat{\theta}\right)} = \underline{\theta}$. Regarding error correction, the following noticeable property holds:

Proposition 10. *With respect to $\widehat{\theta}$, a regular code $X \subseteq A^*$ satisfies Cond.* (c1) *if, and only if, it satisfies Cond.* (c2).

Proof. – Firstly, assume that X is $\underline{\theta}$-independent, and let $x, y \in X$ such that $\tau_{d_\theta,1}(x) \cap \tau_{d_\theta,1}(y) = \underline{\theta}(x) \cap \underline{\theta}(y) \neq \emptyset$. It follows from $\underline{\theta}(x) \cap \underline{\theta}(y) = (\{\theta(x)\} \setminus \{x\}) \cap (\{\theta(y)\} \setminus \{y\})$ that $\theta(x) \neq x$, $\theta(y) \neq y$, and $\theta(x) = \theta(y)$: since θ is one-to-one, this implies $x = y$, therefore X satisfies Cond. (c2).

– Secondly, assume that Cond. (c1) does not hold wrt. $\hat{\theta}$ that is, $X \cap \underline{(\hat{\theta})}(X) = X \cap \underline{\theta}(X) \neq \emptyset$. Necessarily, a pair of different words $x, y \in X$ exist such that $y = \theta(x)$. It follows from $\widehat{\theta}(x) = \{x\} \cup \{\theta(x)\} = \{x, y\}$ and $\widehat{\theta}(y) = \{y\} \cup \{\theta(y)\}$ that $\widehat{\theta}(x) \cap \widehat{\theta}(y) \neq \emptyset$, hence Cond. (c2) cannot hold. □

Example 3. (1) Let $A = \{a, b\}$ and θ be the automorphism defined by $\theta(a) = b$, and $\theta(b) = a$. The regular prefix code $X = \{a^n b : n \geq 0\}$ satisfies Conds. (c1). Indeed, we have $\underline{\theta}(X) = \{b^n a : n \geq 0\}$, thus $\underline{\theta}(X) \cap X = \emptyset$. According to Proposition 10, it also satisfies Cond. (c2). We have $\mu(X) = \frac{1}{2} \sum_{n \geq 0} \left(\frac{1}{2}\right)^n = 1$, whence X is a maximal prefix code. Consequently X is maximal in the family of $\underline{\theta}$-independent codes (Cond. (c3)). Finally, we have $X \subsetneq \hat{\theta}(X) = \{a^n b : n \geq 0\} \cup \{b^n a : n \geq 0\}$, hence $\hat{\theta}(X)$ cannot be a code (we verify that $a, b, ab \in \hat{\theta}(X)$).

(2) Over the alphabet $A = \{a, b\}$, take for θ the anti-automorphism defined by $\theta(a) = b$, and $\theta(b) = a$. The regular prefix code $X = \{a^n b : n \geq 1\}$ satisfies Conds. (c1), (c2): indeed, we have $\underline{\theta}(X) = \{ab^n : n \neq 1\}$, thus $\underline{\theta}(X) \cap X = \emptyset$. As indicated above, X is a maximal prefix code, thus it satisfies Cond.(c3). At last, it follows from $X \subsetneq \widehat{\theta}(X)$ that X cannot satisfies Cond. (c4).

(3) With the condition above, consider $Y = X \setminus \{b, ab\} = \{a^n b : n \geq 2\}$. We have $\underline{\theta}(Y) \cap Y = \emptyset$, hence Y satisfies Conds. (c1), (c2). However, by construction, Y cannot satisfy Cond. (c3). Finally, we have $\hat{\theta}(Y) = \bigcup_{n \geq 2} \{a^n b, ab^n\}$, which remains a prefix code that is, Y satisfies Cond. (c4).

(4) Over the alphabet $\{A, C, G, T\}$, let θ denotes the Watson-Crick anti-automorphism (see eg. [11, 13]), which is defined by $\theta(A) = T$, $\theta(T) = A, \theta(C) = G$, and $\theta(G) = C$.
Consider the prefix code $Z = \{A, C, GA, G^2, GT, GCA, GC^2, GCG, GCT\}$. We have $\theta(Z) = \{T, G, TC, C^2, AC, TGC, G^2C, CGC, AGC\}$, hence Z satisfies Conds.

(c1), (c2). By making use of the uniform distribution, we have $\mu(Z) = 1/2 + 3/16 + 1/16 = 3/4$, hence Z cannot satisfy Cond. (c3). At last, it follows from $G, G^2 \in \widehat{\theta}(Z) = \theta(Z) \cup Z$ that Cond. (c4) is not satisfied.

(5) Notice that, in each of the preceding examples, since the (anti-)automorphism θ satisfies $\theta^2 = id_{A^*}$, actually the quasimetric d_θ is a metric.
Of course, (anti-)automorphisms exist in such a way that d_θ is 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 $a \neq \theta(b)$).

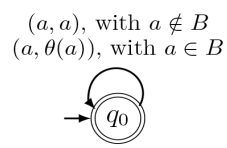

Fig. 5. An automaton with behavior $\tau_{d_{\theta,1}} = \hat{\theta}$, in the case where θ is an automorphism: a represents every character in A and we set $B = \{a \in A : \theta(a) \neq a\}$.

Regarding regulary and recognizability of relations we state:

Proposition 11. *With the preceding notation, each of the following properties holds:*

(i) If θ is an automorphism, then $\tau_{d_{\theta,1}} = \hat{\theta}$ and $\underline{\tau_{d_{\theta,1}}} = \underline{\theta}$ are non-recognizable regular relations.

(ii) If θ is an anti-automorphism, then it cannot be a regular relation.

(iii) Given an (anti-)automorphism θ, if X is a regular subset of A^*, then the same holds for $\hat{\theta}(X)$.

Proof. Let θ be an (anti-)automorphism onto A^*.

(i) In the case where θ is an automorphism of A^*, it is regular: indeed, trivially θ is the behavior of a one-state automaton with transitions $(0, (a, \theta(a)), 0)$, for all $a \in A$. Starting with this automaton we obtain a finite automaton with behaviour $\hat{\theta}$ by merely adding the transitions $(0, (a, a), 0)$, for all $a \in A$ (see Figure 6): consequently the relation $\hat{\theta}$ is regular.

By contradiction, assume θ recognizable. As in the proof of Proposition 7, a finite set I exists such that $\hat{\theta} = \bigcup_{i \in I}(T_i \times U_i)$. Since I is finite, there are $i \in I$, $a \in A$, and $m, n \geq 1$ such that $\left(a^n, \left(\hat{\theta}(a)\right)^n\right), \left(a^m, \left(\hat{\theta}(a)\right)^m\right) \in T_i \times U_i$, with $m \neq n$. This implies $\left(a^n, \left(\hat{\theta}(a)\right)^m\right) \in T_i \times U_i$ that is, $\hat{\theta}(a^n) = (\hat{\theta}(a))^m$. If we have $\hat{\theta}(a) = a$, we obtain $a^n = a^m$. Otherwise, we have $\hat{\theta}(a) = \theta(a)$, thus $\theta(a^n) = (\theta(a))^m$. In each case, this contradicts $m \neq n$.

Finally, the binary relation $\underline{(\hat{\theta})} = \underline{\theta}$ is the set of all the pairs (uas, ubs') that satisfy both the three following conditions:

(1) $u \in A^*$;

(2) $b = \theta(a)$, with $a, b \in A$ and $a \neq b$;

(3) $s' = \theta(s)$, with $s \in A^*$.

Consequently, $\underline{\theta}$ is the behaviour of the automaton in Figure 6, hence it is a regular relation. In addition, by merely substituting the relation $\underline{\theta}$ to $\hat{\theta}$ in the argument above, it can be easily prove that $\underline{\theta}$ recognizable implies $\theta(a^n) = (\theta(a))^m$, for some $a \in A$ and $m \neq n$: a contradiction with θ being a free monoid automorphism.

(ii) For every anti-automorphism θ, the relation $\hat{\theta}$ is the result of the composition of the so-called *transposition*, namely $t : w \to w^R$, by some automorphism of A^*, say h. As shown in [24, Example IV.1.10], the transposition is not a regular relation. Actually, the same argument can be applied for proving that the resulting relation $\hat{\theta}$ is non-regular.

(iii) Let X be a regular subset of A^*. If θ is an automorphism, the relation $\hat{\theta} = \theta \cup id_{A^*}$ is a regular relation, hence $\hat{\theta}(X)$ is regular. In the case where θ is an anti-automorphism, with the preceding notation, although the transposition is not a regular relation, the set $t(X)$ itself is regular (see eg. [24, Proposition I.1.1]). Consequently $\hat{\theta}(X) = h\,(t(X))$ is also regular. □

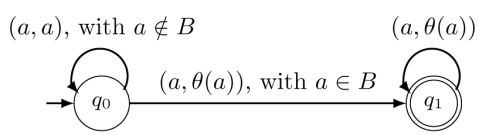

Fig. 6. With the notation in Example 6, an automaton with behavior $\underline{\theta}$, in the case where θ is an automorphism.

As a consequence of Proposition 11, we obtain the following result:

Proposition 12. *Given an (anti-)automorphism θ onto A^* and a regular code $X \subseteq A^*$, each of the two following properties holds:*

(i) In any case, X satisfies Cond. (c1), (c2).

(ii) It can be decided whether X satisfies Cond. (c4).

Proof. (i) Firstly, since θ is an (anti-)automorphism of A^*, it is a one-to-one mapping. For every $x \in X$, we have $\theta(x) \neq \emptyset$, moreover we have $\theta^{-1}(\theta(x)) \cap X = \{x\}$, consequently, in any case X satisfies Cond (c2). Secondly, according to Proposition 10, necessarily X also satisfies Cond. (c1).

(ii) According to Proposition 11, in any case the set $\theta(X)$ is regular: by applying Sardinas and Patterson algorithm, one can decide whether the code X satisfies Cond. (c4). □

It remains to study the behavior of regular codes with regard to Cond. (c3):

Proposition 13. *Every non-complete regular $\underline{\theta}$-independent code can be embedded into some complete one.*

Proof. According to Theorem 2, the result holds if θ is an automorphism: indeed the action of such a transformation merely consists of rewriting words by applying some permutation of A.

Now, we assume that θ is an anti-automorphism. Classically, some positive integer n, the *order* of the permutation θ, exists such $\theta^n = id_{A^*}$. As in the propositions 3 and 8, in view of Theorem 2, we construct a convenient word in $A^* \setminus \mathrm{F}(X^*)$.

Let $z_0 \notin \mathrm{F}(X^*)$, a be its initial character, and b be a character different of a. Without loss of generality, we assume $|z_0| \geq 2$ and $z_0 \notin aa^*$, for every $a \in A$ (otherwise, substitute $z_0 b$ to z_0). By definition, for every integer i, we have $\left|\theta^i(z_0)\right| = |\theta(z_0)|$, therefore it follows from $z_0\theta(z_0)\cdots\theta^{n-1}(z_0) \in A^* \setminus F(X^*)$ that $z_2 = z_0\theta(z_0)\cdots\theta^{n-1}(z_0)ab^{n|z_0|}$ is an overlapping-free word in $A^* \setminus \mathrm{F}(X^*)$. Set $U_2 = A^* \setminus (X^* \cup A^* z_2 A^*)$, $Y_2 = (z_2 U_2)^* z_2$, and $Z_2 = X \cup Y_2$.

According to Theorem 2, the set $Z_2 = X \cup Y_2$ is a complete regular code. In order to prove that it is $\underline{\theta}$-independent, we argue by contradiction. Actually, assuming that $\underline{\theta}(Z) \cap Z \neq \emptyset$, exactly one of the three following cases occurs:

(a) The first condition sets that $x \in X$ exists such that $\underline{\theta}(x) \in Y_2$. By construction, the sets X and Y_2 are disjoint (we have $z_2 \notin F(X^*)$), therefore we have $\theta(x) \neq x$: this implies $\underline{\theta}(x) = \theta(x)$. According to the definition of Y_2, we have $z_2 \in \mathrm{F}\left(\theta(x)\right)$, thus $\theta(z_0) \in \mathrm{F}\left(\theta(x)\right)$. It follows from $x = \theta^{n-1}\left(\theta(x)\right)$, that the word $z_0 = \theta^{n-1}\left(\theta(z_0)\right)$ is a factor of x, a contradiction with $z_0 \notin \mathrm{F}(X^*)$.

(b) With the second condition some pair of words $x \in X$, $y \in Y_2$ exist such that $x = \theta(y)$. It follows from $\theta^{n-1}(z_0) \in \mathrm{F}(z_2) \subseteq \mathrm{F}(y)$ that $z_0 = \theta\left(\theta^{n-1}(z_0)\right) \in \mathrm{F}\left(\theta(y)\right) = \mathrm{F}\left(x\right)$: once more this contradicts $z_0 \notin \mathrm{F}(X^*)$.

(c) The third condition sets that there are different words $y, y' \in Y_2$ such that $y' = \theta(y)$. Since θ is an anti-automorphism, $ab^{n|z_0|} \in \mathrm{S}(y)$ implies $\theta\left(b^{n|z_0|}\right) \in \mathrm{P}(y')$. Since we have $z_0 \in \mathrm{P}(Y_2)$, this implies $\left(\theta\left(b\right)\right)^{|z_0|} = z_0$. But we have $|z_0| \geq 2$, and $\theta(b) \in A$: this is incompatible with the construction of $z_0 \notin aa^*$.

In any case we obtain a contradiction, therefore Z_1 is $\underline{\theta}$-independent. □

As a consequence, given a regular code $X \subseteq A^*$, X is maximal in the family of $\underline{\theta}$-independent codes of A^* if, and only if, the equation $\mu(X) = 1$ holds. In other words, one can decide whether X satisfies Cond. (c3).

Finally, the following statement synthesizes the results of the whole study we relate in our paper:

Theorem 3. *With the preceding notation, given a regular code X one can decide whether it satisfies each of the following conditions:*

(i) *Conds.* (c1)–(c4) *wrt.* $\mathcal{P}_k$.

(ii) *Conds.* (c3), (c4) *wrt.* $\mathcal{F}_k$.

(iii) *In the case where X is finite, Conds.* (c1), (c2) *wrt.* $\mathcal{F}_k$.

(iv) *Conds.* (c3), (c4) *wrt.* $\widehat{\theta}$, *for any (anti-)automorphism* θ *of* A^*. *In that framework, in any case X satisfies Conds. (c1), (c2), which are actually equivalent.*

6 Concluding remark

With each pair of words (w, w'), the so-called *subsequence metric* associates the integer $\delta(w, w') = |w| + |w'| - 2s(w, w')$, where $s(w, w')$ stands for a maximum length common subsequence of w and w'. 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. We observe that, wrt. relation $\tau_{\delta,k}$, results very similar to the ones of the propositions 6 and 9 hold [20]. Moreover, in that framework, we still do not know whether Conds. (c1), (c2) can be decided, given a regular code X.

It is noticeable that, although the inclusion $\mathcal{F}_k \subseteq \tau_{\delta,k}$ holds (indeed factors are very special subsequences of words), we do not know any more whether the relation $\mathcal{F}_k$ is regular or not. In the case where the answer is no, can after all Conds. (c1), (c2) be decidable?

From another point of view, wrt. each of the relations we mentionned, presenting families of codes satisfying all the best Conds. (c1)–(c2) would be desirable.

References

1. Berstel, J., Perrin, D., Reutenauer, C.: Codes and Automata. Cambridge University Press (2010). https://doi.org/10.1017/CBO9781139195768
2. Bruyère, V., Wang, L., Zhang, L.: On completion of codes with finite deciphering delay. European J. Comb. **11**, 513–521 (1990). https://doi.org/10.1016/S0195-6698(13)80036-4
3. Choffrut, C., Pighizzini, G.: Distances between languages and reflexivity of relations. Theoret. Comp. Sci. **286**, 117–138 (2002). https://doi.org/10.1016/S0304-3975(01)00238-9
4. Cohn, P.M.: Universal Algebra (Mathematics and Its Applications, 6). Springer, Dordrecht (1981). https://doi.org/10.1007/978-94-009-8399-1
5. Ehrenfeucht, A., Rozenberg, S.: Each regular code is included in a regular maximal one. RAIRO - Theoret. Informatics and Appl. **20**, 89–96 (1986). https://doi.org/10.1051/ita/1986200100891
6. Eilenberg, S.: Automata, Languages and Machines. Academic Press (1974), eBook ISBN: 9780080873749
7. Elgot, C., Meizei, J.: On relations defined by generalized finite automata. IBM J. Res. Develop. **9**, 47–68 (1965). https://doi.org/10.1147/rd.91.0047
8. Hamming, R.: Error detecting and error correcting codes. The Bell Technical Journal **29**, 147–160 (1950). https://doi.org/10.1002/j.1538-7305.1950.tb00463.x
9. Jürgensen, H., Katsura, M., Konstantinidis, S.: Maximal solid codes. J. of Automata, Languages and Combinatorics **6**, 25–50 (2001). https://doi.org/10.25596/jalc-2001-025

10. Jürgensen, H., Konstantinidis, S.: Codes. In: Grzegorz Rozenberg, A.S. (ed.) Handbook of Formal Languages, vol. 1, chap. 8, pp. 511–607. Springer, Berlin, Heidelberg (1997). https://doi.org/10.1007/978-3-642-59136-5
11. Kari, L., Konstantinidis, S., Kopecki, S.: On the maximality of languages with combined types of code properties. Theoret. Comp Sci. **550**, 79–89 (2014). https://doi.org/10.1016/j.tcs.2014.07.015
12. Konstantinidis, S., Mastnak, M.: Embedding rationally independent languages into maximal ones. J. of Aut., Lang. and Comb. **21**, 311–338 (2016). https://doi.org/10.25596/jalc-2016-311
13. Kruskal, J.B.: An overview of sequence comparison: Time warps, string edits, and macromolecules. SIAM Review **25**, 201–234 (1983), https://www.jstor.org/stable/2030214
14. Lam, N.: Finite maximal infix codes. Semigr. Forum **61**, 346–356 (2000). https://doi.org/10.1007/PL00006Rank033
15. Lam, N.: Finite maximal solid codes. Theot. Comput. Sci. **262**, 333–347 (2001). https://doi.org/10.1016/S0304-3975(00)00277-2
16. Lam, N.: Completing comma-free codes. Theot. Comput. Sci. **301**, 399–415 (2003). https://doi.org/10.1016/S0304-3975(02)00595-9
17. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions and reversals. Cybernetics and Control Theory **10** (1966), Engl. transl. from Doklady Akademii Nauk SSSR 163(4): 845–848 (1965)
18. Néraud, J.: On the completion of codes in submonoids with finite rank. Fundam. Informaticae **74**, 549–562 (2006), ISSN 0169-2968
19. Néraud, J.: Completing circular codes in regular submonoids. Theoret. Comput. Sci. **391**, 90–98 (2008). https://doi.org/10.1016/j.tcs.2007.10.033
20. Néraud, J.: Variable-length codes independent or closed with respect to edit relations. Inf. Comput. (2021), in press. arXiv:2104.14185
21. Néraud, J., Selmi, C.: Embedding a θ-invariant code into a complete one. Theoret. Comput. Sci. **806**, 28–41 (2020). https://doi.org/10.1016/j.tcs.2018.08.022
22. Ng, T.: Prefix distance between regular languages. In: Han, Y.S., Salomaa, K. (eds.) Implementation and Applications of Automata – CIAA 2016. LNCS. vol. 9705, pp. 224–235. Springer Cham (2016), https://doi.org/10.1007/978-3-319-40946-7
23. Restivo, A.: On codes having no finite completion. Discr. Math. **17**, 309–316 (1977). https://doi.org/10.1016/0012-365X(77)90164-9
24. Sakarovitch, J.: Elements of Automata Theory. Cambridge University Press (2009). https://doi.org/10.1017/CBO9781139195218
25. Sardinas, A., Patterson, G.W.: A necessary and sufficient condition for the unique decomposition of coded messages. IRE Internat Con. Rec. **8**, 104–108 (1953)
26. Van, D.L., Van Hung, K., Huy, P.T.: Codes and length-increasing transitive binary relations. In: Van Hung, D., Wirsing, M. (eds.) Theoretical Aspects of Computing – ICTAC 2005. LNCS. vol. 3722, pp. 29–48. Springer, Berlin, Heidelberg, New York (2005), https://doi.org/10.1007/11560647_2
27. Wilson, W.A.: On Quasi-Metric Spaces. American J. of Math. **53**, 675–684 (1931). https://doi.org/10.2307/2371174
28. Zhang, L., Shen, Z.: Completion of recognizable bifix codes. Theoret. Comput. Sci. **145**, 345–355 (1995). https://doi.org/10.1016/0304-3975(94)00300-8