On some factorization problems

Marcella Anselmo Clelia De Felice* Antonio Restivo*

Abstract

We consider three notions of factorization arising in different frameworks: factorizing languages, factorization of the natural numbers, factorizing codes. A language $X \subseteq A^*$ is called *factorizing* if there exists a language $Y \subseteq A^*$ such that $XY = A^*$ and the product is unambiguous. This is a decidable property for recognizable languages X. If we consider the particular case of unary alphabets, we prove that finite factorizing languages can be constructed by using *Krasner factorizations*. Moreover, we extend Krasner's algorithm to factorizations of A^n. We introduce a class of languages, the *strong factorizing* languages, which are related to the *factorizing codes*, introduced by Schützenberger. We characterize strong factorizing languages having three words.

1 Introduction

In this paper we consider three notions of factorization arising in different frameworks: factorizing languages, factorization of the natural numbers, factorizing codes. The aim of this paper is to present some relations between these notions.

The notion of *factorizing language*, introduced in [1, 2], is related to some important questions in formal language theory (see e.g. [3, 5, 7]). We recall that a subset X of A^* is *factorizing* if there exists $Y \subseteq A^*$ such that $\underline{X}\ \underline{Y} = \underline{A}^*$. Here, the product of the languages X and Y is unambiguous [4]. Moreover $\underline{X}$ denotes the characteristic series of X. Several results about this class of languages have been

*Partially supported by ESPRIT-BRA Working Group 6317 *ASMICS* and Project 40% MURST *Algoritmi, Modelli di Calcolo e Strutture Informative*.

Received by the editors May 95.

Communicated by M. Boffa.

1991 *Mathematics Subject Classification :* 68Q45.

Key words and phrases : Factorizing codes, factorizing languages, Krasner factorizations.

proved in [1, 2]. In particular, the decidability of the existence of Y for finite (resp. recognizable) X (see Theorems 2.4, 2.7).

If we consider the particular case $A = \{a\}$, finding finite factorizing languages $X = a^T = \{a^t \mid t \in T\}$ is equivalent to characterize finite subsets T of $\mathbb{N}$ such that $T + R = \mathbb{N}$, where $R \subseteq \mathbb{N}$ and the sum is unambiguous. We will call (T, R) a *factorization* of $\mathbb{N}$. It is not too difficult to see that some of these pairs can be constructed starting from pairs (T', R') such that $T' + R' = \{0, \ldots, n-1\}$. One of the results of this paper proves that *any* factorization (T, R) of the natural numbers, with finite T, can be obtained in this way (Proposition 3.3). Notice that this result is no longer true without the hypothesis of finiteness on T, as it is well known that factorizations of $\mathbb{N}$ exist with both infinite sets as factors [13]. We recall that Krasner provided an algorithm constructing all pairs (T', R') such that $T' + R' = \{0, \ldots, n-1\}$ ([15], see [11, 16] for a recursive version of it). We will call them *Krasner factorizations* of $\mathbb{Z}_n$, as they are particular *factorizations* of the cyclic group $\mathbb{Z}_n$. In this framework, another result gives a slight generalization of Krasner's algorithm in several variables. Precisely, we prove that the same algorithm holds when we replace a letter a by a finite alphabet A (Proposition 3.11).

Finally, we stress some relations with the so-called *factorizing codes*. A *code* C is a subset of A^* such that any word in A^* has at most one factorization as a product of elements in C. A code C is *maximal* (over A) if it is not a proper subset of another code C' (over A), it is *factorizing* if there exist finite subsets P, S of A^* such that $\underline{S}\,\underline{C^*}\,\underline{P} = \underline{A^*}$. Maximal and factorizing codes are related by the following conjecture.

Conjecture 1.1 *(Schützenberger) [3] Every finite maximal code is factorizing.*

Thus, given a factorizing code C, there are two factorizing languages, S and SC^*, associated with it. We will call S a *strong factorizing language*. Thanks to a characterization of factorizing languages with three words (Theorem 2.8), we will characterize strong factorizing languages with the same cardinality (Propositions 4.7, 4.8, 4.12). In particular, we will prove that there exist factorizing languages with three words which are not strong factorizing. We recall that strong factorizing languages with two words have been considered in [10].

This paper is organized as follows. Section 2 contains general results about factorizing languages. Section 3 is dedicated to the relations with Krasner factorizations. Finally, in Section 4 we prove our results about strong factorizing languages.

2 Factorizing languages

In this section we will introduce the notion of *factorizing* language and recall the results in [1, 2] which will be subsequently referred to. Notice that we can decide whether a recognizable language is factorizing (Theorem 2.4). Moreover, some criteria exist for deciding it, whenever X is finite (Theorem 2.7).

Let us introduce some definitions and notations. Given a finite alphabet A, let $\langle A^*, \cdot, 1\rangle$ be the free monoid generated by it. Given a semiring K, the class $K\langle\langle A\rangle\rangle$ of *formal power series* in non-commuting variables A and coefficients in K is the set of functions $s : A^* \to K$. As usual, the value of s on $w \in A^*$ is denoted by (s, w) and

is referred to as the *coefficient* of the series. $K\langle\langle A\rangle\rangle$, equipped with the sum and the Cauchy product of formal power series, is a semiring. The power series is written as a formal sum $s = \Sigma_{w\in A^*}(s, w)w$. The set $supp(s) = \{w \in A^*|(s, w) \neq 0\}$ is the *support* of the series s. A series s having a finite support is called a *polynomial*. We denote by $K\langle A\rangle$ the semiring of the polynomial in non-commuting variables A and coefficients in K. The *characteristic series* of a language $X \subseteq A^*$, denoted $\underline{X}$, is the power series associating 1 with the words belonging to X and 0 with the words not belonging to X. In the following we will often identify X with its characteristic series without stating it explicitly. Some classical references to formal power series are [4, 9, 20].

Definition 2.1 A language $X \subseteq A^*$ factorizes $T \subseteq A^*$ if there exists $Y \subseteq A^*$ such that $\underline{X}\ \underline{Y} = \underline{T}$. If X factorizes A^*, then $X \subseteq A^*$ is called factorizing.

In other words, X factorizes A^* if there exists a language $Y \subseteq A^*$ such that any word $w \in A^*$ has a unique factorization $w = xy$, with $x \in X$ and $y \in Y$.

Remark 2.2 [1, 2] The notion of factorizing language does not depend on the alphabet A: if X factorizes A^*, then X factorizes any B^* with $B \supseteq A$. The hypothesis that the product is unambiguous is essential in order to have a non trivial definition: without this hypothesis, any language X would be factorizing, by taking $Y = A^*$.

Example 2.3 Bisections are pairs (X, Y), with $X, Y \subseteq A^*$ such that $\underline{X}^*\underline{Y}^* = \underline{A}^*$. A simple example is the pair (a^*b, a), for $A = \{a, b\}$. Hence, $(a^*b)^*$ is a factorizing language. Other examples of bisections and results about this problem can be found in [3, 14, 17].

Many examples of factorizing languages can be found in [1, 2]. One of the main results of these papers is the following theorem.

Theorem 2.4 *It is decidable whether a regular language $X \subseteq A^*$ factorizes A^* or not. For any $X \subseteq A^*$, there exists at most one language Y such that $\underline{X}\ \underline{Y} = \underline{A}^*$.*

Recall that a word $x \in A^*$ is a *prefix* (resp. *proper prefix*) of a word $w \in A^*$ if $w = xy$, with $y \in A^*$ (resp. $y \in A^+$). A subset C of A^+ is a *prefix code* if no word in C is a proper prefix of another word in C. The next proposition will be used later and its proof is in [2]. It gives a condition for stating if a language X is factorizing.

Proposition 2.5 *Let $X \subseteq A^*$, $C \subseteq A^*$ be a prefix code with $X \subseteq C^*$. Then, X factorizes A^* if and only if X factorizes C^*.*

Remark 2.6 The previous proposition can also be enunciated as follows. Let π be any bijection between C and an alphabet A_C and denote π the morphism between C^* and A_C^* naturally induced by it. Then $X \subseteq C^*$ factorizes A^* if and only if $\pi(X)$ factorizes A_C^*.

In order to decide whether a finite language X is factorizing, let us introduce some notations and definitions. One can find them in [2] with several examples and other results.

Let $w \in A^*$ be a word and $X \subseteq A^*$ be a language. An *even prefix sequence* of w with respect to X is a sequence $(x_1, x_2, \ldots, x_{2k})$ such that $x_i \in X \setminus 1$ for

$i \in \{1, \ldots, 2k\}$, $k \geq 0$ and $x = x_1 \cdots x_{2k}$ is a prefix of w. The set of the even prefix sequences of w with respect to X is denoted by $E_X(w)$. An *odd prefix sequence* of w with respect to X is a sequence $(x_1, x_2, \ldots, x_{2k+1})$ such that $x_i \in X \setminus 1$ for $i \in \{1, \ldots, 2k+1\}$, $k \geq 0$ and $x = x_1 \cdots x_{2k+1}$ is a prefix of w. The set of the odd prefix sequences of w with respect to X is denoted by $O_X(w)$. Set $d_X(w) = card(E_X(w)) - card(O_X(w))$.

Moreover, let us associate to X, the formal power series $s(X) = \underline{X}^{-1}\underline{A}^*$ and the subset $Z(X)$ of A^*

$$Z(X) = \{w = xz \mid x \in ((X-1)^2)^*,\ z = z_1 z_2 \in X - 1, z_2 \in A^+, xz_1 \in X - 1\}.$$

One can prove the following theorem.

Theorem 2.7 *The following conditions hold*

1. $d_X = s(X)$.
2. *A language X is factorizing if and only if for any word $w \in A^*$, $d_X(w) \in \{0, 1\}$.*
3. *A language X is factorizing if and only if $s(X) \in \mathbb{N}\langle\langle A \rangle\rangle$.*
4. *A language X is factorizing if and only if for any word $w \in X + Z(X)$, $d_X(w) \in \{0, 1\}$.*
5. *A language X is factorizing if and only if for any word $w \in X + Z(X)$, $(s, w) \in \mathbb{N}$.*

Proof. We just prove 2) $\Leftrightarrow$ 3) as one can find a proof of 1), 2), 4), 5) in [2]. Obviously, 2) $\Rightarrow$ 3). Conversely, suppose that $s(X) \in \mathbb{N}\langle\langle A \rangle\rangle$. Then, we have $\underline{X} \cdot s(X) = \underline{A}^*$. This relation implies that the formal power series $s(X)$ has coefficients $0, 1$, i.e., $d_X(w) \in \{0, 1\}$ for any $w \in A^*$. ∎

Finally, we recall the characterization of the factorizing languages having cardinality three given in [1, 2].

Theorem 2.8 *[1, 2] Let $X = 1 + v + w$, with $v \neq 1, w \neq 1$, $v, w \in A^*$. Then, X factorizes A^* if and only if one of the following cases holds:*

1. *$v + w$ is a prefix code.*
2. *$w = v^2$.*
3. *$w = v^{2k}u$, with $k \geq 1, u \neq 1$ and $v + u$ a prefix code.*

In case 1), $Y = ((v+w)^2)^*(A^* - (v+w)A^*)$; in case 2), $Y = (v^3)^*(A^* - vA^*)$ and in case 3) $Y = Y_0^*(uA^* + v^2uA^* + \ldots + v^{2k-2}uA^* + (v^2)^*)$, where $Y_0 = (v^2)^*((X-1)^2 - v^2)$.

3 Krasner factorizations

In this section we will investigate the relations between Krasner factorizations of $\mathbb{Z}_n$ and factorizations of A^*. We will see that any factorization (X, Y) of a^*, with finite X, can be constructed by using Krasner pairs (Proposition 3.2). Moreover we will extend Krasner algorithm to the factorizations of A^n (Proposition 3.11).

We recall that a *factorization* of the cyclic group $\mathbb{Z}_n$ is a pair (T, R) of subsets of $\mathbb{Z}_n$ such that any $z \in \mathbb{Z}_n$ can be written uniquely as a sum of an element of T and an element of R [12]. Then, $t + r = z \pmod n$. A particular class of factorizations of $\mathbb{Z}_n$ has been constructed by Krasner [15]. A *Krasner factorization* (I, J) of $\mathbb{Z}_n$ is a pair of subsets of $\{0, \ldots, n-1\}$ such that any $z \in \{0, \ldots, n-1\}$ can be written uniquely as $z = i + j$, with $i \in I$, $j \in J$. Krasner constructed such pairs (I, J) in [15]. We recall here a recursive version of this algorithm, since it is more useful for our aims.

Remark 3.1 [11, 16] Let (I, J) be a pair of subsets of $\{0, \ldots, n-1\}$, $I \neq \{0\}$, $J \neq \{0\}$. Then, (I, J) is a Krasner factorization of $\mathbb{Z}_n$ if and only if there exists a divisor h of n, $h \neq 1$, and a Krasner factorization (I_1, J_1) of $\mathbb{Z}_{n/h}$ such that

$$I = hJ_1, \qquad J = hI_1 + \{0, \ldots, h-1\}.$$

Moreover $1 \in J$ if and only if $1 \in J_1$.

Let us see how this notion is related to the factorizations of A^*. For a subset X of A^* and a subset I of $\mathbb{N}$, we set $X^I = \{x^i \mid x \in X, i \in I\}$. Let (I, J) be a Krasner factorization of $\mathbb{Z}_n$, $n \in \mathbb{N}$. As it has been pointed out in [1, 2] we have $\underline{A}^* = \underline{X} \cdot \underline{Y}$, with $\underline{X} = \underline{A}^I$, $\underline{Y} = \underline{A}^J(\underline{A}^n)^*$. In the next proposition we will state that *all* factorizations of a^*, with finite X, can be constructed in this way.

Proposition 3.2 *Let $X \subseteq a^*$ be a finite factorizing language. Then, there exists a Krasner factorization (I, J) of $\mathbb{Z}_n$ such that $\underline{X} = a^I$. Moreover, $\underline{X} \cdot \underline{Y} = a^*$ with $\underline{Y} = a^J(a^n)^*$.*

Notice that if (X, Y) is a factorization of a^*, then there exists a factorization (I, P) of the set $\mathbb{N}$ of natural numbers such that $\underline{X} = a^I$, $\underline{Y} = a^P$. Here, a pair (I, P) of subsets of $\mathbb{N}$ is a factorization of $\mathbb{N}$ if any element $n \in \mathbb{N}$ can be written uniquely as a sum of an element of I and an element of P. Then, Proposition 3.2 can also be stated as follows.

Proposition 3.3 *Let (I, P) be a factorization of $\mathbb{N}$ where I is a finite set. Then, there exists a Krasner factorization (I, J) of $\mathbb{Z}_n$ such that $P = J + \{mn \mid m \in \mathbb{N}\}$.*

Proof. We have $0 \in I \cap P$ and $1 \in I \cup P$. Suppose $1 \in I$ (in the other case, just replace I by P in the following proof). Let h be the greatest integer such that $\{0, \ldots, h-1\} \subseteq I$ and denote I as a union of intervals of consecutive integers:

$$I = \cup_{i=1}^{s}(k_i + \{0, \ldots, h_i\})$$

with $k_i + h_i + 1 < k_{i+1}$, for $i \in \{1, \ldots, s-1\}$.

Remark 3.4 $k_1 = 0, h_1 = h - 1$.

Remark 3.5 $h \in P$.

Lemma 3.6 $h_1 = h_2 = \ldots = h_s = h - 1$.

Proof. By contradiction, suppose that there exists $j \in \{1, \ldots, s\}$ such that $h_j \neq h - 1$. We have two cases

$$(*) \quad h_j > h - 1$$
$$(**) \quad h_j < h - 1.$$

(*) Suppose that there exists $j \in \{1, \ldots, s\}$ such that $h_j > h-1$. Then $k_j, k_j+h \in I$ and we have two different factorizations for $k_j + h$ in $I + P$

$$k_j + h = (k_j + h) + 0,$$

a contradiction.
(**) Suppose that there exists $j \in \{1, \ldots, s\}$ such that $h_j < h - 1$ and take j minimum with respect to this condition. Consider $k_j + h_j + 1$. By hypothesis, there exists $p \in P$ and $i \in I$ such that

$$i + p = k_j + h_j + 1.$$

We have $p \neq 0$ and then $p \geq h$. Consequently $i \leq k_j + (h_j - h + 1) < k_j$.
Then there exists $j' \in \{1, \ldots, s\}, j' < j$, such that $i \in k_{j'} + \{0, \ldots, h_{j'}\}$. By the minimality of j we have $h_{j'} = h - 1$, i.e.,

$$i \in k_{j'} + \{0, \ldots, h - 1\} \subseteq I.$$

Thus, $p + k_{j'} + \{0, \ldots, h - 1\}$ is a set of h consecutive integers containing $p + i = k_j + h_j + 1$; so $p + k_{j'} + \{0, \ldots, h - 1\}$ contains either $h_j + k_j$ or $h_j + k_j + h$. Consequently, there exists $h' \in \{0, \ldots, h-1\}$ such that $(k_{j'} + h') + p = (h_j + k_j) + 0$ or $(k_{j'} + h') + p = (k_j + h_j) + h$. This is a contradiction ($0, h \in P$, $k_{j'} + h' \leq k_{j'} + h_{j'} < k_j \leq k_j + h_j$). ∎

Lemma 3.7

$$\forall p \in P, \quad p = 0 \pmod{h}$$
$$\forall j \in \{1, \ldots, s\}, \quad k_j = 0 \pmod{h}.$$

Proof. By contradiction, suppose that there exists $j \in \{1, \ldots, s\}$ such that $k_j \neq 0 \pmod{h}$ or $p \in P$ such that $p \neq 0 \pmod{h}$. Take the minimum integer verifying this condition.
Then, there exist $q \in \mathbb{N}, r \in \{1, \ldots, h - 1\}$ such that

$$k_j = hq + r \quad (\text{resp. } p = hq + r).$$

In virtue of Remark 3.4 and Lemma 3.6, we have $hq \notin I$, $hq \notin P$, otherwise $hq + r$ would have two different factorizations in $I + P$ (either $hq \in P$, $r \in I$, or $hq \in I$ and so $hq + \{0, \ldots, h - 1\} \subseteq I$). Then, there exists $i \in I \setminus 0$, $p' \in P \setminus 0$ with $hq = i + p'$, i.e.,

$$i = hq_1 + r_1, \quad p' = hq_2 + r_2.$$

By using $p' < k_j$, $i < k_j$ (resp. $p' < p$, $i < p$), we get $r_1 = r_2 = 0$,

$$i = hq_1 = k_{j'}, \quad p' = hq_2, \quad q_1 + q_2 = q, \quad j' < j.$$

Thus, according to Lemma 3.6,

$$hq + r = k_j + 0 = (hq_1 + r) + hq_2$$

$$(\text{resp. } hq + r = 0 + p = (hq_1 + r) + hq_2),$$

would be two different factorizations of $hq + r$ in $I + P$, a contradiction. ∎

End of the proof. We prove the conclusion by induction over the cardinality of I. We have either

$$I = \cup_{i=1}^{s}(q_i h + \{0, \ldots, h-1\}), \quad P = \cup_{i=1}^{\infty} q'_i h \tag{1}$$

or

$$I = \cup_{i=1}^{s} q_i h, \quad P = \cup_{i=1}^{\infty}(q'_i h + \{0, \ldots, h-1\}). \tag{2}$$

If $I = \{0, \ldots, h-1\}$ (resp. $I = \{0\}$) then the conclusion holds by taking $n = h$ and $J = \{0\}$ (resp. $J = \{0, \ldots, h-1\}$).
Otherwise, consider (I_1, P_1) with $I_1 = \{q_1, \ldots, q_s\}, P_1 = \{q'_1, \ldots\}$. Obviously (I_1, P_1) is a factorization of $\mathbb{N}$.
Suppose that (1) holds. We have $card(I_1) < card(I)$. Then, by using the induction hypothesis, there exists J_1, with (I_1, J_1) a Krasner factorization of $\mathbb{Z}_n$ and $P_1 = J_1 + \{mn \mid m \in \mathbb{N}\}$. Thus, according to Remark 3.1, (I, hJ_1) is a Krasner factorization of $\mathbb{Z}_{nh}$ and $P = hP_1 = hJ_1 + \{mnh \mid m \in \mathbb{N}\}$. Now suppose that (2) holds. Notice that $1 \in I_1$, i.e., (I_1, P_1) is a factorization of $\mathbb{N}$ such that (1) holds and $card(I) = card(I_1)$. We can apply the argument above and get a Krasner factorization $(I_1, h'J_2)$ of $\mathbb{Z}_n$ with $P_1 = h'J_2 + \{mn \mid m \in \mathbb{N}\}$. Then, in virtue of Remark 3.1, $(hI_1, hh'J_2 + \{0, \ldots, h-1\}) = (I, hh'J_2 + \{0, \ldots, h-1\})$ is a Krasner factorization of $\mathbb{Z}_{nh}$ and it holds $P = hP_1 + \{0, \ldots, h-1\} = hh'J_2 + \{0, \ldots, h-1\} + \{mnh \mid m \in \mathbb{N}\}$. ∎

Remark 3.8 Notice that Proposition 3.3 is well known for the factorizations (I, P) of the cyclic group $\mathbb{Z}$. Indeed, let $I, P \subseteq \mathbb{Z}$ be such that (I, P) is a factorization of $\mathbb{Z}_n$. It is well known that if I is finite then P is periodic [12, 13]. Thus, P is the direct sum $P' + < n >$ of a finite subset P' of $\mathbb{Z}$ and a cyclic subgroup $< n >= \{mn \mid m \in \mathbb{Z}\}$ of $\mathbb{Z}$ [12, 13]. Moreover, up to a translation, we can suppose $I \subset \mathbb{N}$. Consequently $I + P' = [-n+1, n-1]$ (direct sum) and $P' = J \cup J'$ with (I, J) a Krasner factorization of $\mathbb{Z}_n$. Proposition 3.3 states that any factorization of $\mathbb{N}$ can be obtained starting from a factorization of $\mathbb{Z}$. We are not been able to give a direct proof of this result.

Remark 3.9 Notice that Lemma 3.6 has been proved for finite I, P in [15]. The proof is the same in the two cases. It has been reported here for the sake of the completeness.

Remark 3.10 Factorizations (I, P) of $\mathbb{N}$ exist, with I, P both infinite sets. So, Proposition 3.2 (resp. Proposition 3.3) does not hold without the hypothesis on X (resp. I). As a counterexample, consider the unambiguous factorization (X, Y) of a^* with $X = (1+a)(1+a^4)(1+a^{16})\ldots$ and $Y = (1+a^2)(1+a^8)(1+a^{32})\ldots$ (see [2, 13]).

The next proposition generalizes Krasner's algorithm to A^n.

Proposition 3.11 *Let $X, Y \subseteq A^*$ be two languages and $n \geq 1$ be a positive integer. Suppose that*

$$\underline{X} \cdot \underline{Y} = (\underline{A}^n - 1)/(\underline{A} - 1) = 1 + \underline{A} + \ldots + \underline{A}^{n-1}.$$

Then, there exists a Krasner factorization (I, J) of $\mathbb{Z}_n$ such that $\underline{X} = \underline{A}^I$, $\underline{Y} = \underline{A}^J$.

Proof. Let us prove that $X = A^I$, $Y = A^J$, for a pair (I, J) of subsets of $\mathbb{N}$. Then (I, J) will be a Krasner factorization and we have done. Indeed, let $x' \in X$ (resp. $y' \in Y$) be a word of maximal length in X (resp. Y). Then, $|x'| + |y'| = n - 1$ and so

$$A^{n-1} = \{x \in X \mid |x| = |x'|\}\{y \in Y \mid |y| = |y'|\} \subseteq A^{|x'|}A^{|y'|} = A^{n-1},$$

which implies

$$A^{|x'|} = \{x \in X \mid |x| = |x'|\} \subseteq X, \quad A^{|y'|} = \{y \in Y \mid |y| = |y'|\} \subseteq Y.$$

By contradiction, suppose that there exist X_1, $Y_1 \subseteq A^*$, $I', J' \subseteq \mathbb{N}$ such that

$$\underline{X} = \underline{A}^{I'} + \underline{X}_1, \quad \underline{Y} = \underline{A}^{J'} + \underline{Y}_1,$$

with $i = max\{|x| \mid x \in X_1\} < minI'$, $j = max\{|y| \mid y \in Y_1\} < minJ'$, $A^i \setminus X_1 \neq \emptyset$ or $A^j \setminus Y_1 \neq \emptyset$.
Thus, the product $\underline{X}\,\underline{Y}$ verifies the following equation

$$\underline{X}\,\underline{Y} = \underline{A}^{I'}\underline{A}^{J'} + \underline{X}_1\underline{A}^{J'} + \underline{A}^{I'}\underline{Y}_1 + \underline{X}_1\underline{Y}_1 = 1 + \underline{A} + \ldots + \underline{A}^{n-1}. \tag{3}$$

Suppose that $A^j \setminus Y_1 \neq \emptyset$ (a similar argument holds in the other case). Let $y_1, y_2 \in A^j$ be such that $y_1 \in Y_1, y_2 \notin Y_1$ and, as before, let x' be a word of maximal length in X. Obviously, we have

$$(\underline{A}^{I'}\underline{Y}_1, x'y_1) = 1, \tag{4}$$

and, by using $|x'y_2| = |x'y_1| \leq n - 1$, we also get

$$(\underline{X}\,\underline{Y}, x'y_2) = 1,$$

i.e., in virtue of (3),

$$(\underline{A}^{I'}\underline{A}^{J'} + \underline{X}_1\underline{A}^{J'} + \underline{A}^{I'}\underline{Y}_1 + \underline{X}_1\underline{Y}_1, x'y_2) = 1. \tag{5}$$

On the other hand, since x' (resp. y_2) has length equal to the maximal length in X (resp. Y_1) and $y_2 \notin Y_1$, we get

$$((\underline{A}^{I'} + \underline{X}_1)\underline{Y}_1, x'y_2) = (\underline{X}\,\underline{Y}_1, x'y_2) = 0.$$

Thus, according to (5),

$$(\underline{A}^{I'}\underline{A}^{J'} + \underline{X}_1\underline{A}^{J'}, x'y_2) = 1.$$

Recall that $|y_2| < minJ'$. Then, by using the previous equality, there exist $x_1, x_2 \in A^*$ with $x_1x_2 = x'$ and $x_2 \neq 1$ such that

$$(\underline{A}^{I'} + \underline{X}_1, x_1) = 1 = (\underline{A}^{J'}, x_2y_2). \tag{6}$$

Finally, as $|y_2| = |y_1|$,

$$(\underline{A}^{J'}, x_2y_1) = 1. \tag{7}$$

By using these relations, we have a contradiction. Indeed, according to (3),

$$1 \geq (\underline{X}\ \underline{Y}, x'y_1) \geq (\underline{A}^{I'}\underline{A}^{J'} + \underline{X}_1\underline{A}^{J'} + \underline{A}^{I'}\underline{Y}_1, x'y_1)$$

and, by using (4), (6) and (7),

$$1 \geq (\underline{X}\ \underline{Y}, x'y_1) \geq (\underline{A}^{I'}\underline{A}^{J'} + \underline{X}_1\underline{A}^{J'}, x'y_1) + (\underline{A}^{I'}\underline{Y}_1, x'y_1) \geq 2.$$

■

Let $C \subseteq A^*$ be a prefix code and (I, J) be a Krasner factorization of $\mathbb{Z}_n$. Then $\underline{C}^I$ will be referred to as a *Krasner polynomial*. As it has been pointed out, a Krasner polynomial is a factorizing language. This result can be seen as a generalization of cases 1) and 2) in Theorem 2.8. Analogously, the next proposition generalizes the third case in the same theorem. Moreover, it shows that a generalization of Proposition 3.2, for alphabets A with $card(A) > 1$, requires more than Krasner polynomials. Proof is essentially the same for proving case 3) of Theorem 2.8.

Proposition 3.12 *Let $C \subseteq A^*$, $u \in A^*$, with $C + u$ a prefix code. Then, for any $k \in \mathbb{N}, k \geq 1$, $1 + C + C^{2k}u$ is a factorizing language.*

Proof. Let A_C be any alphabet having the same cardinality than C. By Proposition 2.5 and Remark 2.6 it suffices to prove that $X = 1 + A_C + A_C^{2k}b$ factorizes $A^* = (A_C \cup b)^*$. By Theorem 2.7 this is equivalent to prove that the formal power series

$$s = ((\underline{X} - 1)^2)^*(1 - (\underline{X} - 1))\underline{A}^* = (\underline{A}_C^2 + (\underline{X} - 1)^2 - \underline{A}_C^2)^*(1 - (\underline{X} - 1))\underline{A}^*$$

has non-negative coefficients.
By using the equality of power series $(y + z)^* = (y^*z)^*y^*$ with $y = \underline{A}_C^2$, $z = (\underline{X} - 1)^2 - \underline{A}_C^2$, we get

$$s = ((\underline{A}_C^2)^*((\underline{X} - 1)^2 - \underline{A}_C^2))^*(\underline{A}_C^2)^*(1 - (\underline{X} - 1))\underline{A}^*. \tag{8}$$

Let us prove that $(\underline{A}_C^2)^*(1 - (\underline{X} - 1))\underline{A}^*$ has non-negative coefficients. Indeed,

$$\begin{aligned}
(\underline{A}_C^2)^*(1 - (\underline{X} - 1))\underline{A}^* &= (\underline{A}_C^2)^*(1 - \underline{A}_C - \underline{A}_C^{2k}\underline{b})\ \underline{A}^* \\
&= (\underline{A}_C^2)^* + (\underline{A}_C^2)^*\underline{A}_C\underline{A}^* + (\underline{A}_C^2)^*\underline{b}\ \underline{A}^* \\
&\quad -(\underline{A}_C^2)^*\underline{A}_C\underline{A}^* - (\underline{A}_C^2)^*\underline{A}_C^{2k}\underline{b}\ \underline{A}^* \quad (\text{as } A = A_C \cup b) \\
&= (\underline{A}_C^2)^* + (\underline{A}_C^2)^*\underline{b}\ \underline{A}^* - (\underline{A}_C^2)^*\underline{A}_C^{2k}\underline{b}\ \underline{A}^* \\
&= (\underline{A}_C^2)^* + (\underline{A}_C^2)^*(1 - \underline{A}_C^{2k})\underline{b}\ \underline{A}^* \\
&= (\underline{A}_C^2)^* + [(1 - \underline{A}_C^{2k})/(1 - \underline{A}_C^2)]\underline{b}\ \underline{A}^* \\
&= (\underline{A}_C^2)^* + (1 + \underline{A}_C^2 + \ldots + \underline{A}_C^{2k-2})\underline{b}\ \underline{A}^*.
\end{aligned}$$

By using this equation, (8) becomes

$$s = ((\underline{A}_C^2)^*((\underline{X}-1)^2 - \underline{A}_C^2))^*((\underline{A}_C^2)^* + (1 + \underline{A}_C^2 + \ldots + \underline{A}_C^{2k-2})\underline{b}\ \underline{A}^*) =$$

$$((\underline{A}_C^2)^*((\underline{A}_C^{2k}\ \underline{b})^2 + \underline{A}_C^{2k}\ \underline{b}\ \underline{A}_C + \underline{A}_C\underline{A}_C^{2k}\ \underline{b}))^*((\underline{A}_C^2)^* + (1 + \underline{A}_C^2 + \ldots + \underline{A}_C^{2k-2})\underline{b}\ \underline{A}^*),$$

which proves that s has non-negative coefficients. ∎

We end this section with the following open problem.

Problem 3.13 *Can we generalize Proposition 3.11 to a Krasner polynomial ?*

4 Strong factorizing languages

In this section we will introduce a class of factorizing languages, the *strong factorizing* languages, which are strictly related to some codes, the *factorizing* codes.

Thanks to the characterization of factorizing languages with three words (Theorem 2.8), we will characterize strong factorizing languages having the same cardinality (Propositions 4.7, 4.8 and 4.12). As a byproduct, we will prove the existence of factorizing languages which are not strong factorizing (Proposition 4.12).

First, let us recall some definitions which we need in the sequel. We recall that a subset C of A^* is a *code* if for any $c_1, \ldots, c_h, c'_1, \ldots, c'_k \in C$

$$c_1 \cdots c_h = c'_1 \cdots c'_k \quad \Rightarrow \quad h = k; \quad \forall i \in \{1, \ldots, h\} \quad c_i = c'_i.$$

An important class of codes is the class of maximal codes. A code C is *maximal* over A if for any code C' over A

$$C \subseteq C' \quad \Rightarrow \quad C = C'.$$

Many deep results about codes have been proved (see [3] for a complete survey on this topic and [8] for a list of open problems in this area). Nevertheless, the structure of these objects is not yet completely investigated. In particular, twenty years ago Schützenberger gave the following conjecture which is still open.

Conjecture 4.1 *[3, 8, 21] Any finite maximal code is factorizing.*

We recall that a code C is *factorizing* (over A) if there exist two finite subsets P, S of A^* such that

$$\underline{S}\ \underline{C}^*\underline{P} = \underline{A}^*.$$

As an example, maximal prefix codes C are factorizing, by taking $S = 1$ and P as the set of the proper prefixes of the elements in C. We will refer to [6, 8, 11, 18, 19, 21] for some partial results about this conjecture.

The notion of *strong factorizing* language is related to the notion of factorizing code.

Definition 4.2 A finite subset S of A^* is a *strong factorizing* language if there exist two finite subsets P, C of A^*, C being a code, such that:

$$\underline{S}\ \underline{C}^*\underline{P} = \underline{A}^*.$$

Remark 4.3 If S is a strong factorizing language, then S is a factorizing language and we have $\underline{S}\ \underline{Y} = \underline{A^*}$, where $Y = C^*P$. Moreover, C is a factorizing code.

Obviously, the construction of the factorizing codes is related to the construction of the strong factorizing languages which, in turn, could give information on the factorization conjecture. This observation naturally suggests the following questions.

Problem 4.4 *Is it decidable whether a language S is a strong factorizing language?*

Problem 4.5 *Can we characterize the structure of a strong factorizing language S?*

In [10] a strong factorizing language S is called a *polynomial having solutions*. This terminology is motivated by the following result.

Proposition 4.6 *[4] For finite subsets P, S, C of A^*, we have $\underline{S}\ \underline{C^*}\ \underline{P} = \underline{A^*}$ if and only if we have $\underline{P}(\underline{A} - 1)\underline{S} = \underline{C} - 1$. Moreover, if $\underline{P}(\underline{A} - 1)\underline{S} + 1$ has non-negative coefficients, then it is the characteristic series of a finite maximal code C.*

Let s be a formal power series with non-negative integer coefficients and denote it by $s \geq 0$. In the particular case of a strong factorizing language S, we have that the inequality $\underline{P}(\underline{A} - 1)\underline{S} + 1 \geq 0$ has at least one *solution* P.

Notice that there exist languages which are not strong factorizing. Take, as an example, $S = 1 + a + ab$. This language is not factorizing and so it is not strong factorizing [2]. A more interesting example of a language which is not strong factorizing is given in Proposition 4.12 below. Indeed, we will compare the class of factorizing languages with the one of strong factorizing languages and we show that they are different.

The following results are related to Theorem 2.8. Propositions 4.7, 4.8 show that in the first and in the second case of this theorem, X is also a strong factorizing language. Moreover, we will state that in the third case X is not a strong factorizing language (Proposition 4.12).

Proposition 4.7 *Let $S = 1 + C$ with C a prefix code. Then S is a strong factorizing language.*

Proof. Let P be the set of the proper prefixes of the elements of C. Notice that

$$\underline{P} - 1 \leq \underline{P} - 1 + \underline{C} \leq \underline{P}\ \underline{A} \quad \Rightarrow \quad (\underline{P} - 1)\underline{C} \leq \underline{P}\ \underline{A}\ \underline{C}.$$

By using the previous relations, we get

$$1 + \underline{P}(\underline{A} - 1)\underline{S} = 1 + \underline{P}(\underline{A} - 1)(1 + \underline{C}) =$$

$$\underline{P}\ \underline{A} - (\underline{P} - 1) - \underline{C} + \underline{P}\ \underline{A}\ \underline{C} - (\underline{P} - 1)\underline{C} \geq 0.$$

■

Proposition 4.8 *Let $S = 1 + v + v^2$, with $v \in A^*$. Then, S is a strong factorizing language.*

Proof. The set P of the proper prefixes of v verifies

$$1 + \underline{P}\ \underline{A} - \underline{P} - v \geq 0.$$

In virtue of this inequality, we get

$$0 \leq (1 + \underline{P}\ \underline{A} - \underline{P} - v) + (1 + \underline{P}\ \underline{A} - \underline{P} - v)v + (1 + \underline{P}\ \underline{A} - \underline{P} - v)v^2 =$$

$$1 + \underline{P}\ \underline{A} - \underline{P} + \underline{P}\ \underline{A}v - \underline{P}v + \underline{P}\ \underline{A}v^2 - \underline{P}v^2 - v^3 \leq 1 + \underline{P}(\underline{A} - 1)(1 + v + v^2).$$

■

Propositions 4.7 and 4.8 naturally suggest the following question.

Problem 4.9 *Can we generalize the previous results to a Krasner polynomial* $S = \underline{C}^I$ *?*

In the following examples we consider some strong factorizing languages and we compare the techniques introduced in Theorem 2.8 with the techniques introduced in Propositions 4.7, 4.8.

Example 4.10 Let $A = \{a, b\}$ and $S = 1 + a + ba$. We have $S = 1 + C$, with C a prefix code. Then, S is strong factorizing, in virtue of Proposition 4.7. By taking $P = 1$, one has $P(A - 1)S + 1 = a^2 + aba + b + bba = C'$. According to Proposition 4.6, we have that S is factorizing, by taking as Y the language $C'^*P = (a^2 + aba + b + bba)^*$. This is the same language $Y = ((a + ba)^2)^*(A^* - (a + ba)A^*)$ $= (aa + aba + baa + baba)^*(bbA^* + b + 1)$ that we obtain by using Theorem 2.8. In order to prove $Y \subseteq C'^*P = C'^*$, notice that $(a + ba)^2 \subseteq C'^*$. In addition, C' being a maximal suffix code, for any $w \in A^*$, we have $A^*bbbw \cap C'^* \neq \emptyset$ and so $bbA^* \subseteq C'^*$. Conversely, by using the equality of power series $(y + z)^* = (y^*z)^*y^*$ $= y^* + y^*z(y^*z)^*y^*$, we have $C'^* = (a^2 + aba + b + bba)^* = (a^2 + aba + b)^* +$ $(a^2 + aba + b)^*bba((a^2 + aba + b)^*bba)^*(a^2 + aba + b)^* \subseteq Y$.

Example 4.11 Let $A = \{a\}$ and $S = 1 + a^k + a^{2k}$, $k \geq 1$. In virtue of Proposition 4.8, S is strong factorizing. By taking $P = (1 + a + a^2 + \ldots + a^{k-1})$, we have $C = P(a - 1)(1 + a^k + a^{2k}) + 1 = (1 + a + a^2 + \ldots + a^{k-1})(a - 1)(1 + a^k + a^{2k}) + 1$ $= a^{3k}$. According to Proposition 4.6, we have that S is factorizing, by taking as Y the language $C^*P = (a^{3k})^*(1 + a + \ldots + a^{k-1})$. This is the same language that we obtain by using Theorem 2.8.

Let $S = 1 + v + v^{2k}u$, with $k \geq 1, u \neq 1$ and $\{v, u\}$ a prefix code. The next proposition states that S is not a strong factorizing language. The proof is decomposed into several intermediate results.

Proposition 4.12 *Let* $S = 1 + v + v^{2k}u$*, with* $k \geq 1, u \neq 1$ *and* $v + u$ *a prefix code. Then,* S *is not a strong factorizing language.*

In the sequel, we will use the following notations. We will denote by P_0 the set of the proper prefixes of v and by p_r the prefix of v of length r, with $r \in \{0, \ldots, n-1\}$, $n = |v|$. In the next lemma, we suppose that $v^h p_r$ can be written as a product of two words p and $v^{2k}u$. Roughly, we prove that a factorization $p_{r'}v'$ of v also exists, with $p_{r'} \neq 1$, $p_{r'}$ suffix of p and $v' \neq 1$, v' prefix of $v^{2k}u$.

Lemma 4.13 *Let P' be a finite subset of A^*. Suppose that*

$$(\underline{P'}v^{2k}u, v^h p_r) > 1,$$

with $h \in \mathbb{N}$ and $p_r \in P_0$. Then, there exist $h' \in \mathbb{N}$, $h' < h$ and $p_{r'} \in P_0 \setminus 1$ such that

$$v^{h'}p_{r'} \in P', \quad v^h p_r = v^{h'}p_{r'}v^{2k}u.$$

Proof. According to the hypothesis, there exists $p \in P'$ such that

$$pv^{2k}u = v^h p_r.$$

Then, there exists $h' \in \mathbb{N}$ and $p_{r'} \in P_0$ such that $p = v^{h'}p_{r'}$. We have $h' < h$, otherwise $p_{r'}v^{2k}u = p_r$, a contradiction. Moreover, we have $p_{r'} \neq 1$, otherwise

$$p = v^{h'} \quad \Rightarrow \quad v^{2k}u = v^{h-h'}p_r \quad \Rightarrow \quad u = v^{h-h'-2k}p_r,$$

a contradiction, since $\{v, u\}$ is a prefix code. ∎

The next corollary is obtained by taking $P' = PA$ in Lemma 4.13.

Corollary 4.14 *Let P be a finite subset of A^*. Let $p_r \in P_0$ and $h \in \mathbb{N}$. If $(\underline{P}\ \underline{A}v^{2k}u, v^h p_r) > 1$, then there exist $h' \in \mathbb{N}$, $h' < h$, $p_{r'} \in P_0$, $a \in A$ such that*

$$p_{r'}a \in P_0, \quad v^{h'}p_{r'} \in P, \quad v^h p_r = v^{h'}p_{r'}av^{2k}u.$$

The next lemma adds some more informations on the structure of P. It has been proved in [10]. We report here the proof for the sake of the completeness.

Lemma 4.15 *Let S be a strong factorizing language. Let S_1 be the set of the words in $S \setminus 1$ of minimal length and let P_1 be the set of the proper prefixes of the words in S_1. For a finite subset P of A^*such that $\underline{P}(\underline{A} - 1)\underline{S} + 1 \geq 0$ we have*

$$1 \in P \cap S, \quad P_1 \subseteq P.$$

Proof. It is obvious that $1 \in P \cap S$. Let us prove that $P_1 \subseteq P$. Let $v_i \in S_1$. As $(P, 1) = 1$, we get

$$1 \leq (\underline{P}\ \underline{S}, v_i) \leq (1 + \underline{P}\ \underline{A}\ \underline{S}, v_i).$$

According to this relation, there exist $p \in P, a \in A, v_j \in S$ such that

$$v_i = pav_j,$$

and, by using the definition of S_1, we get $v_j = 1$. Thus, P contains the proper prefix of maximal length of v_i. By contradiction, suppose that there exists a proper prefix p of v_i with $p \in P \setminus PA$, $p \neq 1$. Moreover, take p of maximal length with respect to this condition. Again, as $(S, 1) = 1$,

$$1 \leq (\underline{P}\ \underline{S}, p) \leq (1 + \underline{P}\ \underline{A}\ \underline{S}, p),$$

i.e., there exist $p' \in P, a \in A, v_j \in S$ such that

$$p = p'av_j.$$

As we supposed $p \in P \setminus PA$, then $v_j \neq 1$. Thus, $|v_j| < |p| < |v_i|$, a contradiction. ∎

Ab absurdo, we are supposing that $S = 1+v+v^{2k}u$ is a strong factorizing language, i.e., there exists a finite subset P of A^* such that $\underline{P}(\underline{A}-1)\underline{S}+1 \geq 0$. Under this hypothesis we prove two lemmata; the former is needed for proving the latter.

Lemma 4.16 *Let P be a finite subset of A^* such that*

$$\underline{P}(\underline{A}-1)(1+v+v^{2k}u)+1 \geq 0,$$

let $w \in A^+$ be such that

$$(\underline{P}\ \underline{A}\ v^{2k}u, w) = 1 \quad \Rightarrow \quad (\underline{P}\ v^{2k}u, w) = 1.$$

Then

$$(\underline{P}+\underline{P}\ v, w) \leq (\underline{P}\ \underline{A}+\underline{P}\ \underline{A}\ v, w).$$

Proof. Suppose $(\underline{P}\ \underline{A}\ v^{2k}u, w) = 0$. Then, according to the hypothesis,

$$(\underline{P}+\underline{P}\ v, w) \leq (\underline{P}+\underline{P}\ v+\underline{P}\ v^{2k}u, w) \leq$$

$$(\underline{P}\ \underline{A}+\underline{P}\ \underline{A}\ v+\underline{P}\ \underline{A}\ v^{2k}u, w) = (\underline{P}\ \underline{A}+\underline{P}\ \underline{A}\ v, w).$$

Otherwise, $(\underline{P}\ \underline{A}\ v^{2k}u, w) = 1$. Thus, again by using the hypothesis,

$$(\underline{P}+\underline{P}\ v, w) \leq (\underline{P}\ \underline{A}+\underline{P}\ \underline{A}\ v+\underline{P}\ \underline{A}\ v^{2k}u-\underline{P}\ v^{2k}u, w) = (\underline{P}\ \underline{A}+\underline{P}\ \underline{A}\ v, w).$$

■

The next lemma gives a precise description of the set $P \cap v^*P_0$.

Lemma 4.17 *Let P be a finite subset of A^* such that*

$$\underline{P}(\underline{A}-1)(1+v+v^{2k}u)+1 \geq 0.$$

Then, there exists $h \in \mathbb{N}$ such that $v^*P_0 \cap P = (1+v^2+\ldots+v^{2h})P_0$.

Proof. In virtue of Lemma 4.15, we have $P_0 \subseteq P$. By contradiction, suppose that the conclusion does not hold and let h' be the minimal counterexample. Precisely, we suppose that there exist $h, h' \in \mathbb{N}$, with $h' > 2h$, such that $(1+v^2+\ldots+v^{2h})P_0 \subseteq P$, $v^{h_1}P_0 \cap P = \emptyset$ for $2h < h_1 < h'$, $v^{h'}p_r \in P$ and one of the following cases holds:

$$\begin{array}{ll} (*) & h' \text{ odd} \\ (**) & h' \text{ even and } v^{h'}P_0 \not\subseteq P \\ (***) & h' \text{ even, } v^{h'}P_0 \subseteq P \text{ and } h' > 2h+2. \end{array}$$

First, notice that

$$(\underline{P}\ \underline{A}\ v^{2k}u, v^{h'}p_r) = 1 \quad \Rightarrow \quad (\underline{P}\ v^{2k}u, v^{h'}p_r) = 1. \tag{9}$$

Indeed, due to Corollary 4.14, if $(\underline{P}\ \underline{A}\ v^{2k}u, v^{h'}p_r) = 1$ then $v^{h'}p_r = v^{h''}p_{r'}av^{2k}u$, with $v^{h''}p_{r'} \in P$, $h'' < h'$ and $p_{r'}a \in P_0$. By using the hypothesis on h', we get $v^{h''}p_{r'}a \in P$ and $(\underline{P}\ v^{2k}u, v^{h'}p_r) = 1$. In virtue of Lemma 4.16, relation (9) implies

$$(\underline{P}+\underline{P}\ v, v^{h'}p_r) \leq (\underline{P}\ \underline{A}+\underline{P}\ \underline{A}\ v, v^{h'}p_r). \tag{10}$$

Now, take r minimal with respect to the condition $v^{h'}p_r \in P$.
$(*)$ Then, for $p_r \neq 1$, by using (10) we get

$$1 = (\underline{P}, v^{h'}p_r) \leq (\underline{P} + \underline{P}\ v, v^{h'}p_r) \leq (\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v, v^{h'}p_r) = (\underline{P}\ \underline{A}\ v, v^{h'}p_r) \leq 1.$$

Thus, $(\underline{P}\ \underline{A}\ v, v^{h'}p_r) = 1$ and $(\underline{P}\ v, v^{h'}p_r) = 0$ which implies $v^{h'-1}p_r \in PA$, $v^{h'-1}p_r \notin P$; a contradiction.
On the other hand, for $p_r = 1$, we get $(\underline{P}\ \underline{A}\ v, v^{h'}) = 0$ since, $h' - 2$ being odd, $v^{h'-2}p_{n-1} \notin P$. Thus, in virtue of (10),

$$1 = (\underline{P}, v^{h'}) \leq (\underline{P} + \underline{P}\ v, v^{h'}) \leq (\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v, v^{h'}) = (\underline{P}\ \underline{A}, v^{h'}) \leq 1, \qquad (11)$$

which implies $(\underline{P}\ \underline{A}, v^{h'}) = 1$, i.e., $v^{h'-1}p_{n-1} \in P$. By using the hypothesis, $v^{h'-1} \in P$ and (11) becomes

$$1 = (\underline{P}, v^{h'}) \leq (\underline{P} + \underline{P}\ v, v^{h'}) = 2 \leq (\underline{P}\ \underline{A}, v^{h'}) = 1,$$

i.e., a contradiction.
$(**)$ Notice that for no $p_r \in P_0 \setminus 1$ we have $v^{h'}p_r \in P \setminus PA$. Indeed, by using (10), we would have $(\underline{P}\ \underline{A}\ v, v^{h'}p_r) = 1$, i.e., $v^{h'-1}p_{r-1} \in P$ with $h' - 1$ odd, a contradiction. In particular, for the minimal r we have $p_r = 1$, i.e., $v^{h'} \in P$. Moreover, $v^{h'}p_{n-1} \notin P$, otherwise $v^{h'}P_0 \subseteq P$.
On the other hand, again

$$(\underline{P}\ \underline{A}\ v^{2k}u, v^{h'+1}) = 1 \quad \Rightarrow \quad (\underline{P}\ v^{2k}u, v^{h'+1}) = 1. \qquad (12)$$

Indeed, due to Corollary 4.14, if $(\underline{P}\ \underline{A}\ v^{2k}u, v^{h'+1}) = 1$ then $v^{h'+1} = v^{h''}p_{r'}av^{2k}u$, with $v^{h''}p_{r'} \in P$, $h'' < h' + 1$ and $p_{r'}a \in P_0$. Notice that $h'' < h'$ since $k \geq 1$. By using the hypothesis on h', we get $v^{h''}p_{r'}a \in P$ and $(\underline{P}\ v^{2k}u, v^{h'+1}) = 1$.
In virtue of Lemma 4.16, relation (12) implies

$$(\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v, v^{h'+1}) \geq (\underline{P} + \underline{P}\ v, v^{h'+1}). \qquad (13)$$

Now, $(\underline{P} + \underline{P}\ v, v^{h'+1}) \geq 1$, since $v^{h'} \in P$. Moreover, $(\underline{P}\ \underline{A}, v^{h'+1}) = 0$ since $v^{h'}p_{n-1} \notin P$ and $(\underline{P}\ \underline{A}\ v, v^{h'+1}) = 0$ since $v^{h'} \notin PA$, h' being even. This contradicts (13).
$(***)$ In this case, we have $(\underline{P}, v^{h'}) = 1$, $(\underline{P}\ \underline{A}\ v, v^{h'}) = 0$ and $(\underline{P}\ \underline{A}, v^{h'}) = 0$. Thus, by using (10), we get

$$0 = (\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v, v^{h'}) \geq (\underline{P} + \underline{P}\ v, v^{h'}) = 1,$$

i.e., a contradiction. ∎

We know that $(P, 1) = (P, v^0) = 1$. Let $h \in \mathbb{N}$ be the maximal integer such that $v^h \in P$. The next lemmata prove some relations on h we need for completing the proof.

Lemma 4.18 *For any proper prefix u_1 of u,*

$$(\underline{P}\ \underline{A}\ v^{2k}u, v^{2k+h}u_1) = 1 \quad \Rightarrow \quad (\underline{P}\ v^{2k}u, v^{2k+h}u_1) = 1.$$

Proof. Suppose that $(\underline{P}\ \underline{A}\ v^{2k}u, v^{2k+h}u_1) = 1$. Then, by definition, there exist $h' \in \mathbb{N}$, $p_{r'} \in P_0$, $a \in A$, with $h' < h$ and $p_{r'}a$ prefix of v, such that

$$v^{h'}p_{r'} \in P, \quad v^{h'}p_{r'}av^{2k}u = v^{2k+h}u_1. \tag{14}$$

If $r' \neq n-1$, then $p_{r'}a \in P_0$ and $(\underline{P}\ v^{2k}u, v^{2k+h}u_1) = 1$, according to Lemma 4.17. Otherwise, by using (14),

$$v^{h'+1}u = v^h u_1 \quad \Rightarrow \quad u = v^{h-h'-1}u_1,$$

a contradiction, since $\{u, v\}$ is a prefix code and $u_1 \neq u$. ∎

Lemma 4.19 *For any prefix u_1 of u, with $0 < |u_1| < |v|$, we have*

$$(\underline{P}\ \underline{A}\ v, v^h v^{2k}u_1) = 0.$$

Proof. By contradiction, suppose

$$(\underline{P}\ \underline{A}\ v, v^h v^{2k}u_1) = 1.$$

Then, we have $v^{2k+h-1}p_r \in PA$, with $p_r \in P_0 \setminus 1$ and $|p_r| = |u_1|$. By using Lemma 4.17, we obtain $v^{2k+h-1} \in P$ which implies $2k + h - 1 \leq h$; a contradiction. ∎

End of the proof. Recall that h is the maximal integer such that $v^h \in P$. According to Lemma 4.17, h is even and we have

$$(\underline{P}\ \underline{A}\ v^{2k}u, v^h v^{2k}u) = 0. \tag{15}$$

Moreover, as $v^h \in P$, we get

$$(\underline{P}\ v^{2k}u, v^h v^{2k}u) = 1. \tag{16}$$

Then, by using (15) and (16), we get

$$(\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v, v^h v^{2k}u) = (\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v + \underline{P}\ \underline{A}\ v^{2k}u, v^h v^{2k}u) \geq$$
$$(\underline{P} + \underline{P}\ v + \underline{P}\ v^{2k}u, v^h v^{2k}u) = (\underline{P} + \underline{P}\ v, v^h v^{2k}u) + 1.$$

Thus, $(\underline{P}\ \underline{A}, v^h v^{2k}u) = 1$ or $(\underline{P}\ \underline{A}\ v, v^h v^{2k}u) = 1$. Let us prove that, in both of cases, there exists a proper prefix u_1 of u, $u_1 \neq 1$, such that

$$(\underline{P}\ , v^h v^{2k}u_1) = 1. \tag{17}$$

This is obvious if $(\underline{P}\ \underline{A}, v^h v^{2k}u) = 1$, as the definition of h implies $|u| > 1$. So, suppose $(\underline{P}\ \underline{A}\ v, v^h v^{2k}u) = 1$. For proving (17), it suffices to notice that $|u| > |v|+1$, which holds in virtue of Lemmata 4.17, 4.19 and according to the definition of h.
Let u_1 be the proper prefix of u, $u_1 \neq 1$, of minimal length verifying (17). We have

$$(\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v + \underline{P}\ \underline{A}\ v^{2k}u, v^h v^{2k}u_1) \geq (\underline{P} + \underline{P}\ v + \underline{P}\ v^{2k}u, v^h v^{2k}u_1).$$

According to Lemmata 4.16, 4.18 and by using (17), we have

$$(\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v, v^{h+2k}u_1) \geq (\underline{P}\ v, v^{2k+h}u_1) + 1.$$

Then, $(\underline{P}\ \underline{A}, v^h v^{2k}u_1) = 1$ or $(\underline{P}\ \underline{A}\ v, v^h v^{2k}u_1) = 1$. Let us prove that, in both of cases, we have a contradiction.
This is obvious if $(\underline{P}\ \underline{A}, v^h v^{2k}u_1) = 1$, as the definition of h implies $|u_1| > 1$ and the definition of u_1 implies $|u_1| = 1$. So, suppose $(\underline{P}\ \underline{A}\ v, v^h v^{2k}u_1) = 1$. Thus, in virtue of Lemmata 4.17, 4.19 and according to the definition of h, we have $|u_1| > |v| + 1$, which contradicts the definition of u_1. ∎

We end this section with a lemma and a remark. Let S be a strong factorizing language and let P, C be finite subsets of A^*, with C a finite maximal code, such that

$$\underline{S}\ \underline{C^*}\ \underline{P} = \underline{A^*}.$$

Looking at this equation, it is trivial to notice that $P \cap S = \{1\}$; otherwise, for a non empty word $w \in P \cap S$, we get

$$(\underline{S}\ \underline{C^*}\ \underline{P}, w) \geq 2, \quad (\underline{A^*}, w) = 1.$$

Thus, we have proved the following lemma.

Lemma 4.20 *Let S be a strong factorizing language. For any finite subset P of A^* such that $\underline{P}(\underline{A}-1)\underline{S}+1 \geq 0$, we have $P \cap S = \{1\}$.*

Remark 4.21 We recall that $1+v$ is a strong factorizing language [10]. In the same paper the author also described the complete structure of all finite subsets P of A^* such that $\underline{P}(\underline{A}-1)(1+v)+1 \geq 0$. Lemma 4.17 gives a more precise description of some particular subsets P verifying the above relation. Precisely, we notice that for a finite subset P of v^*P_0, we have $\underline{P}(\underline{A}-1)(1+v)+1 \geq 0$ if and only if there exists $h \in \mathbb{N}$ such that $P = (1+v^2+\ldots+v^{2h})P_0$. Indeed, assume that $P = (1+v^2+\ldots+v^{2h})P_0$. It is a straightforward verification that $\underline{P}(\underline{A}-1)(1+v)+1 \geq 0$. Conversely, let $P \subseteq v^*P_0$ be such that $\underline{P}(\underline{A}-1)(1+v)+1 \geq 0$. For getting $P = (1+v^2+\ldots+v^{2h})P_0$, we have just to take the proof of Lemma 4.17, by skipping in it the relations (9) and (12). Moreover, denote by C the corresponding finite maximal code, verifying

$$(1+v)\ \underline{C^*}\ \underline{P} = \underline{A^*}.$$

Then, $v^{2h+2} \in C$. Indeed,

$$(\underline{P}\ v, v^{2h+2}) = 0 = (\underline{P}, v^{2h+2}) = (\underline{P}\ \underline{A}, v^{2h+2}), \quad (\underline{P}\ \underline{A}\ v, v^{2h+2}) = 1.$$

By using this relation, thanks to Proposition 4.6, we get

$$(\underline{C}-1, v^{2h+2}) = (\underline{P}\ \underline{A} + \underline{P}\ \underline{A}\ v, v^{2h+2}) - (\underline{P} + \underline{P}\ v, v^{2h+2}) = 1.$$

References

[1] M. Anselmo, A. Restivo, Factorizing languages, *Proc. IFIP 94* (B. Pehrson and I. Simon eds), Vol. **1**, Elsevier Sc. B. V. (1994) 445–450.

[2] M. Anselmo, A. Restivo, On languages factorizing the free monoid, *Intern. J. Algebra Comput.* **6** (1996), 15 p., to appear.

[3] J. Berstel, D. Perrin, *Theory of codes*, Academic Press, New York (1985).

[4] J. Berstel, C. Reutenauer, *Rational series and their languages*, EATCS Monographs **12**, Springer-Verlag (1988).

[5] A. Bertoni, P. Massazza, On the square root of regular languages, *Preprint* (1995).

[6] J. M. Boë, Factorisation par excès du monoïde libre, LIRMM *report* **94-005** (1994).

[7] V. Bruyère, Factorisation des ensembles préfixiels, *RAIRO Inform. Théor. et Appl.* **23** (1989) 295–315.

[8] V. Bruyère, Research topics in the theory of codes, *Bull. EATCS* **48** (1992) 412–424.

[9] S. Eilenberg, *Automata, languages and machines*, vol. A, Academic Press, New York (1974).

[10] C. De Felice, *Construction et complétion de codes finis*, Thèse de 3ème cycle, *Rapport LITP* **85-3** (1985).

[11] C. De Felice, Construction of a family of finite maximal codes, *Theoret. Comput. Sci.* **63** (1989) 157–184.

[12] L. Fuchs, *Abelian groups*, Pergamon Press, Oxford, London, New York and Paris (1960).

[13] G. Hajós, Sur la factorisation des groupes abéliens, *Casopis Pest. Mat. Fys.* **74** (1950) 157–162.

[14] D. Krob, Codes limités et factorisations finies du monoïde libre, *RAIRO Inform. Théor.* **21** (1987) 437–467.

[15] M. Krasner, B. Ranulac, Sur une propriété des polynômes de la division du cercle, *C. R. Acad. Sc. Paris* **240** (1937) 397–399.

[16] N. H. Lam, On codes having no finite completion, *Proc. STACS 94*, *Lecture Notes Comput. Sci.* **775** (1994) 691–698.

[17] M. Lothaire, *Combinatorics on words*, Encyclopedia of Mathematics and its Applications, **17**, Add. Wesley Pub. Comp. (1983).

[18] D. Perrin, M. P. Schützenberger, Un problème élémentaire de la théorie de l'information, *"Théorie de l'Information"*, Colloques Internat. CNRS **276**, Cachan (1977) 249–260.

[19] C. Reutenauer, Non commutative factorization of variable-length codes, *J. Pure Appl. Algebra* **36** (1985) 167–186.

[20] A. Salomaa and M. Soittola, *Automata-theoretic aspects of formal power series*, Springer, Berlin Heidelberg New York (1978).

[21] M. P. Schützenberger, Sur certains sous-monoïdes libres, *Bull. Soc. Math. France* **93** (1965) 209–223.

Marcella Anselmo, Clelia De Felice
Dipartimento di Informatica ed Applicazioni
Università di Salerno
I-84081 Baronissi (SA) (Italy)

Antonio Restivo
Dipartimento di Matematica e Applicazioni
Università di Palermo
via Archirafi 34
I-90123 Palermo (Italy)