FACTORIZING CODES AND KRASNER FACTORIZATIONS

CLELIA DE FELICE
Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (SA), Italy
E-mail: defelice@dia.unisa.it

ROSALBA ZIZZA
Dipartimento di Scienze dell'Informazione, Università di Milano, via Comelico 39, 20135 Milano, Italy
E-mail: rosalba@dsi.unimi.it

Motivated by Schützenberger's factorization conjecture, we investigate finite languages $P, S \subseteq A^*$ which can appear in a factorization $A^* = SC^*P$ of the free monoid, with the product being unambiguous and C being a factorizing code. In particular, we characterize all such P for $|S| = 3$ and all such S with $S \subseteq w^*, w \in A^*$. We propose conjectures on the structure of the left factors S in a factorization as above. The results which we obtain show a relationship with the Krasner factorizations.

1 Introduction

The algebraic theory of *variable length codes* (i.e., the *bases* of the free submonoids of a free monoid) was developed by Schützenberger and his school from the late 50's onwards[3,5,20]. Several classes of codes were introduced in connection with the original framework of the theory. In particular *maximal* codes (i.e., maximal objects in the class of codes for the order of set inclusion) and *factorizing* codes. The latter class of codes was introduced by Schützenberger, a finite code C being factorizing if it verifies the equation $\underline{C} - 1 = \underline{P}(\underline{A} - 1)\underline{S}$, where P, S are finite subsets of A^* and $\underline{X}$ denotes the characteristic polynomial of a finite $X \subseteq A^*$. Every factorizing code is a finite maximal code, the converse is the object of a famous conjecture by Schützenberger: *any finite maximal code is factorizing*[3,15,16].

Some partial results on this problem have been obtained, but this conjecture is still open. The result that was closest to a solution of the conjecture has been proved by Reutenauer. He shows that any finite maximal code C verifies the above-mentioned equation with P, S polynomials with integer coefficients [4,18]. Other results give a positive answer to the conjecture in the case of finite maximal codes over a two-letter alphabet $A = \{a, b\}$ having a constraint on the number of the occurrences of the letter b in each word[8,10,17,21].
Given a pair P, S of finite subsets of A^*, it is known that if $P(\underline{A} - 1)S + 1$ is a polynomial with nonnegative coefficients, then $\underline{C} = P(\underline{A} - 1)S + 1$

is the characteristic polynomial of a factorizing code [3,18]. Consequently the construction of all factorizing codes is equivalent to the construction of all pairs P, S of finite subsets of A^* such that $P(\underline{A}-1)S+1$ is a polynomial with nonnegative coefficients. We denote this by $P(\underline{A}-1)S+1 \geq 0$. For this reason in this case we will say that S *has solutions* and P is a *solution* for S. The problem of finding all the pairs (P, S) as above can be split into two different tasks:

1. The characterization of all S with solutions
2. Given S with solutions, the characterization of all the solutions P for S

Some partial results have been proved in [1] with respect to the first problem where polynomials S with solutions are called *strong factorizing* (see Section 2.3). In particular, in [1] polynomials S with solutions and $|S| = 3$ have been characterized. In [6], problem 2 has been solved for $S = 1 + X$, X being a finite prefix code.
Other results concerning the structure of the factorizing codes have been obtained using some factorizations of the cyclic groups [7,8,9].
In this paper, we give several examples of polynomials S with or without solutions (Section 3). In particular, we characterize polynomials S with solutions in the particular case $S \subseteq w^*$ (Proposition 3.2). We present conjectures on the structure of such polynomials S in the general case (Problem 3.1, Conjecture 3.1). Finally, we characterize all the solutions P for S with $|S| = 3$ (Propositions 4.1, 4.2). All the obtained results show a relationship with the so-called Krasner factorizations (Section 2.4). As a matter of fact, the main aim of our investigation is to find a generalization of these factorizations in several variables as suggested by Schützenberger.
This paper is organized as follows: Section 2 contains all the basic notions on words, codes and Krasner factorizations which we need in the sequel. We present our results on S with solutions and our characterization of the solutions P for S with $|S| = 3$ in Sections 3 and 4 respectively. An extended version of this paper contains all the missing proofs of the results presented in this abstract [11]. The manuscript is available upon request from the authors.

2 Known Results

2.1 Words

Let A^* be the *free monoid* generated by a finite alphabet A and let $A^+ = A^* \setminus 1$, where 1 is the empty word. In the next part of this paper, for two words $x, y \in A^*$ we set $x \leq y$ (resp. $x < y$) if x is a *prefix* (resp. *proper prefix*) of y, i.e., $y = xz$ for $z \in A^*$ (resp. $z \in A^+$). We denote $P(y)$ the set of the proper prefixes of y. Furthermore, $|w|$ will be the length of $w \in A^*$, $|X|$ the cardinality of any multiset X (the elements are counted by taking account of their multiplicity). We recall that a word $w \in A^+$ is *primitive* if for any $u \in A^*$, $w = u^h$ implies $h = 1$ [14]. We also known that for any word w a unique primitive word u exists such that $w = u^h$, $h \geq 1$ [14]. The following result is also used in the proof of Proposition 3.2.

Proposition 2.1 [14] *Let x, y be words. We have $xz = zx$ if and only if u, k, k' exist with $u \in A^*$, $k, k' \geq 0$ and such that $x = u^k$, $z = u^{k'}$.*

2.2 Codes and Formal Power Series

Let $\boldsymbol{Z}$ be the ring of the integer, let $\boldsymbol{Z}\langle\langle A\rangle\rangle$ be the ring of the *formal power series* with coefficients in $\boldsymbol{Z}$ and variables in A, let $\boldsymbol{Z}\langle A\rangle$ be the ring of the *polynomials* with integral coefficients and variables in A. For any $S \in \boldsymbol{Z}\langle\langle A\rangle\rangle$ we denote (S, w) the coefficient of $w \in A^*$ in S. We will write $S \geq 0$ or $S \in \boldsymbol{N}\langle\langle A\rangle\rangle$ (resp. $S \in \boldsymbol{N}\langle A\rangle$) when S is a formal power series (resp. polynomial) with coefficients in the set $\boldsymbol{N}$ of the nonnegative integers. For any subset X of A^*, $\underline{X} = \sum_{x \in X} x$ is the *characteristic* series of X. In the next part of this paper we will often identify X with $\underline{X}$ and we will write $x \in X$ instead of $(\underline{X}, x) = 1$ (see [4] or [19] as main references for this topic).

A *code* $C \subseteq A^*$ is a set of words such that any word in A^* has at most one factorization as a product of elements in C, i.e., for all $c_1, \ldots, c_h, c'_1, \ldots, c'_k \in C$ we have

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

A code C is *maximal* if, for any code C', $C \subseteq C'$ implies $C = C'$. $C \subseteq A^+$ is a *prefix* code if $C \cap CA^+ = \emptyset$ (see [3] for a complete treatment on codes).

2.3 Factorizing Codes

A finite code C is *factorizing* if there exist two finite subsets P, S of A^* such that $\underline{C} - 1 = \underline{P}(\underline{A} - 1)\underline{S}$. If C is a factorizing code then C is a finite maximal code (see [3]). The converse result was conjectured by Schützenberger. This question is still open and known as the *factorization conjecture*.

Conjecture 2.1 [3,20] *(Schützenberger). Any finite maximal code is factorizing.*

We have already recalled all known results on this conjecture in Section 1. In Section 1 we have also observed that constructing factorizing codes is the same as constructing all pairs P, S of finite subsets of A^* such that $P(\underline{A}-1)S+1 \geq 0$.

For such a pair, we will say that S *has solutions* and P is a *solution* for S. We recall that such polynomials S have been called *strong factorizing* in [1]. Indeed, in [2] any language X such that $\underline{X}\,\underline{Y} = \underline{A^*}$ is called a *factorizing language* (over A). Thus S is a factorizing language (over A) with $Y = C^*P$ when S verifies $P(\underline{A} - 1)S + 1 = \underline{C}$, with C a finite maximal code and P, S finite subsets of A^* [1].

All known results on the construction of the pairs (P, S) or of the factorizing codes C have been recalled in Section 1. We end this section with two known results. They will be used in the last two sections. Proposition 2.2 characterizes polynomials S with solutions when $|S| = 3$; Proposition 2.3 gives some properties of any solution P of a general polynomial S.

Proposition 2.2 [1] *Let $S = 1 + w_1 + w_2$ be a polynomial which has solutions. Then either $w_2 = w_1^2$ or $\{w_1, w_2\}$ is a prefix code.*

Proposition 2.3 [1] *Let S be a polynomial which has solutions and let P be a solution for S. Then $S \cap P = 1$ and $P_0 \subseteq P$, where P_0 is the set of the proper prefixes of the words of minimal length in $S \setminus 1$.*

2.4 Krasner Factorizations

The particular case of $A = \{a\}$ will be fundamental in Section 3. For any finite multiset $X \in \boldsymbol{N}_1 = \boldsymbol{N}\langle 1\rangle$ of nonnegative integers, let us denote $a^X = \sum_{n\in\boldsymbol{N}} (X, n)a^n$.

It is easy to see that if (P, S) is a pair of finite subsets of $A^* = a^*$ such that $P(a - 1)S + 1 \geq 0$, then we have:

$$P = a^I, \quad S = a^J, \quad a^I a^J = \frac{a^n - 1}{a - 1}$$

(I, J) pairs as above have been characterized by Krasner and Ranulac in [12]. We will call (I, J) a *Krasner factorization* of the cyclic group $\boldsymbol{Z}_n$ of order n. Here, we recall their recursive construction given in [7,13].

Proposition 2.4 [7,13] *(I, J) is a Krasner factorization of $\boldsymbol{Z}_n$ if and only if either $I = J = \{0\}$ or there exists a divisor k of n, $k > 1$, and a Krasner factorization (I_1, J_1) of $\boldsymbol{Z}_{n/k}$ such that*

$$I = kJ_1, \qquad J = kI_1 + \{0, \ldots, k - 1\}.$$

Furthermore, we have $I = \{0\}$ or $min(I \setminus 0) = k$.

The following proposition will be used in Section 3.

Proposition 2.5 *Let $I, J' \subseteq \boldsymbol{N}$, $I \neq \{0\}$, $h \in \boldsymbol{N}$, $h > 1$. Suppose that (hI, J') is a Krasner factorization of $\boldsymbol{Z}_n$. Then k exists, with $h|k|n$ and $J \subseteq \boldsymbol{N}$ exists such that (I, J) is a Krasner factorization of $\boldsymbol{Z}_{n/h}$, with $J = (k/h)I_1 + \{0, \ldots, (k/h) - 1\}$, $J' = kI_1 + \{0, \ldots, k-1\}$, $I_1 \subseteq \boldsymbol{N}$.*

3 Polynomials S with or without solutions

In this section we will investigate the problem of characterizing those polynomials S which have solutions. We will tackle this problem under three points of view. We first try to reduce this problem to the particular case of polynomials S with $S \not\subseteq X^*$, X being a code different from A (Conjecture 3.1, Proposition 3.1). This is a common reduction in the theory of codes where such a condition defines the operation of *composition* of codes [3]. Next, for a prefix code X we denote $P(X)$ the set of the proper prefixes of the elements of X and we set $X^I = \cup_{i \in I} X^i$.

Conjecture 3.1 Let S be a polynomial with solutions and $X = (S \setminus 1) \setminus (S \setminus 1)A^+$. Let X_1 be the finite maximal prefix code defined by $\underline{X}_1 - 1 = \underline{P}(X)(\underline{A} - 1)$. Then $S \subseteq X_1^*$.

In [1], the authors proved that $S = 1 + X$ has solutions for any finite prefix code X. We extend this result in the next proposition which is also related to Conjecture 3.1.

Proposition 3.1 *Let X be a finite prefix code, (I, J) a Krasner factorization of $\boldsymbol{Z}_n$, $n \in \boldsymbol{N}$. Let X_1 be the finite maximal prefix code defined by $\underline{X}_1 - 1 = \underline{P}(X)(\underline{A} - 1)$, and let L be such that $X^J(X-1)^L + X^J \geq 0$ with $L \subseteq \boldsymbol{N}$. Then $P = X^J P(X)$ is a solution for $S = X^I + X^L(X_1 \setminus X)X^I$. In particular, $S = X^I$ has solutions.*

Sketch of the proof:
Observe that $\underline{X^J P(X)} = \underline{X}^J \underline{P}(X)$ since X is a prefix code. Furthermore, for $L = \emptyset$, the two polynomials X^I and $X^I + X^L(X_1 \setminus X)X^I$ are equal. In order to obtain the conclusion we have to check that $\underline{P}(\underline{A} - 1)\underline{S} + 1 \geq 0$ when $P = X^J P(X)$ and $S = X^I + X^L(X_1 \setminus X)X^I$. $\square$

Next we consider the relations between polynomials with solutions and Krasner factorizations (Problem 3.1, Proposition 3.2). Proposition 3.1 will be used in the proof of the Proposition 3.2 which characterizes S with solutions in a particular case and that can be considered as a first step towards a positive answer to the following problem.

Problem 3.1 Let S be a polynomial with solutions and denote $S \cap w^* = w^I$ where $w \in A^*$ and $S \cap w^* \neq \emptyset$. Does $n \in \boldsymbol{N}$ and $J \subseteq \boldsymbol{N}$ always exist such that (I, J) is a Krasner factorization of $\boldsymbol{Z}_n$?

If Problem 3.1 had a positive answer, any polynomial S with solutions would have the form

$$S = w_1^{I_1} + w_2^{I_2} + \ldots + w_k^{I_k}$$

with (I_h, J_h) being a Krasner factorization of $\boldsymbol{Z}_{n_h}$, for $h \in \{1, \ldots, k\}$. We conjecture that such a polynomial always has a solution when $\{w_1, w_2, \ldots, w_k\}$ is a prefix code.

Proposition 3.2 *Let $S \subseteq w^*$, $w \in A^*$. S has solutions if and only if n, I, J exist with $S = w^I$, $n \in \boldsymbol{N}$ and (I, J) a Krasner factorization of $\boldsymbol{Z}_n$. Moreover for $S \neq 1$ and for any solution P for S, we have $P \cap w^* P(w) = w^J P(w)$ and thus $w^J P(w) \subseteq P$.*

PROOF :
Suppose that $S = w^I$, with (I, J) a Krasner factorization of $\boldsymbol{Z}_n$. Then, S has solutions in virtue of Proposition 3.1. Conversely, let $S \subseteq w^*$, i.e., $S = w^I$ with $I \subseteq \boldsymbol{N}$ and suppose that S has solutions with $S \neq 1$. First suppose that w is a primitive word. By hypothesis, $P \subseteq A^*$ exists such that $C = P(A-1)S + 1$ is a finite maximal code. Let us denote $P(w) = \{p_1, \ldots, p_k\}$ with $|p_i| = i$ and $w = p_k x$, $x \in A$, and let $J, H \subseteq \boldsymbol{N}$ such that

$$P \cap w^* = w^J, \qquad P \cap w^* p_k = w^H p_k$$

We have $J \neq \emptyset$ $(1 \in P)$ and $H \neq \emptyset$ (Proposition 2.3). Furthermore, it holds

$$C \cap w^* = (P(A-1)S + 1) \cap w^* = w^H p_k x w^I - w^J w^I + 1 \geq 0 \qquad (1)$$

We have $H \subseteq J$. Indeed, for any $h \in \boldsymbol{N}$, $0 < h \leq k$, and $s \in \boldsymbol{N}$ if $w^s p_h \in P$ then $w^s p_h \in PA$ (otherwise, looking at Proposition 2.3, we have $w^s p_h \in PAw^t$ with $s, t > 0$. Consequently we have a contradiction with Proposition 2.1). Thus, if $j \in H$ we have $w^j p_k \in P$ and so $w^j \in P$, i.e., $j \in J$. On the other hand, $J \setminus H = \emptyset$. Indeed, since C is a finite maximal code, then either $C \cap w^*$ is the empty set or it is w^n, for a unique $n \in \boldsymbol{N}$. Furthermore, by using a projection over a letter a in w and then the morphism $a \to 1$ in the inequality (1), we get

$$1 - |(J \setminus H) + I| = |C \cap w^*|$$

Thus, since $|I| \neq 1$, then $J \setminus H = \emptyset$, i.e., $J = H$, $C \cap w^* = w^n$ and (1) implies that (I, J) is a Krasner factorization of $\boldsymbol{Z}_n$. Furthermore, $w^s p_h \in P$ implies $w^s \in P$. Thus $s \in J = H$ and $w^s P(w) \subseteq P$. So, $P \cap w^* P(w) = w^J P(w)$.
Suppose now that $S = w^I$ with $w = u^h$, $h > 1$ and u a primitive word. Thus $S = u^{hI}$. Using the result which we proved above, (hI, J') is a Krasner factorization of $\boldsymbol{Z}_n$, for $J' \subseteq \boldsymbol{N}$, $n \in \boldsymbol{N}$ and $P \cap u^* P(u) = u^{J'} P(u)$. Using Proposition 2.5, $J \subseteq \boldsymbol{N}$ exists such that (I, J) is a Krasner factorization of $\boldsymbol{Z}_{n/h}$ and $J = (k/h) I_1 + \{0, \ldots, (k/h) - 1\}$, $\quad J' = k I_1 + \{0, \ldots, k-1\}$, with $I_1 \subseteq \boldsymbol{N}$ and $h | k | n$. Finally, it is easy to conclude that we have $P \cap w^* P(w) = w^J P(w)$. $\square$

We observe that the "if" part of the previous proposition ($S \subseteq w^*$ with solutions $\Rightarrow S = w^I$ with (I, J) being a Krasner factorization of $\boldsymbol{Z}_n$) could also be obtained in another way. Indeed, Lemma 2 in [2] states that $S = w^I$ is a factorizing language (over A) if and only if $S' = a^I$ is a factorizing language (over a). Furthermore, Proposition 3.3 in [1] states that a^I is a factorizing language over a if and only if (I, J) is a Krasner factorization of $\boldsymbol{Z}_n$. However, this argument does not give the property of P stated in the previous proposition.

The next examples show polynomials with no solution and enforce the existence of a relationship between Krasner factorizations and polynomials S with solutions.

Example 3.1 The polynomial $S_1 = 1 + ab + ba + abba + abbabb$ has no solution. Indeed, by contradiction, suppose that P is a solution for S_1. As $(PS_1, abbabb) \geq 1$, then $abbab \in P$ and $(PS_1, abbab) \geq 1$. Using Proposition 2.3, we get $(PA, abbab) = 1$, that is $abba \in P$, i.e., a contradiction with Proposition 2.3.

Example 3.2 For any prefix code $\{w_1, w_2\}$ with $|w_1| \leq |w_2|$ and $w_2 \notin A^* w_1 A^*$, the polynomial $S_2 = 1 + w_1 + w_2 + w_1 w_1 w_2$ has no solutions.

Remark 3.1 $S_2 \subseteq \{w_1, w_2\}^I$, $I = \{0, 1, 3\}$ and for such I no J exists with (I, J) being a Krasner factorization. Let $z \in \{w_1, w_2\}^2$. We can observe that $P(w_1) + P(w_2)$ is a solution for $1 + w_1 + w_2 + z$. We conjecture that any S with solutions containing S_2 must contain $w_1 w_2$.

Lastly, Proposition 2.2 shows that all polynomials S with solutions and $|S| = 3$ can be obtained from S' with solutions and $|S'| = 2$ by adding to S' a suitable word taken from a solution P' for S'. This observation suggests a generalization for the case of S of any cardinality and motivates a characterization of the solutions P for S, with $|S| = 3$ (Section 4).

4 Characterizing solutions P for S, with $|S| = 3$

In this section, we state the main result of the paper which characterizes the solutions P of the inequality $P(\underline{A} - 1)S + 1 \geq 0$ with $|S| = 3$. Looking at Proposition 2.2, we have to distinguish two cases. In both cases our characterization generalizes the analogous result for $|S| = 2$ given in [6]. The first case is when $S = 1 + w + w^2$, with $w \in A^*$. In the next proposition, once again we denote $P(w) = \{p_1, \ldots, p_k\}$, where $|p_i| = i$ and $k = |w| - 1$. Before giving Proposition 4.1 with all its technical details, let us give an intuitive idea of this result. In order to do so, we recall that P is a solution for $S = 1 + w$ if and only if P verifies Proposition 2.3 and for every $z \in P \setminus PA$, $z \neq 1$, we have $z = paw$, $paP(w) \cap P = \emptyset$, $zP(w) \subseteq P$, $zw \notin P$ [6]. This means that for $z \in P$, if a prefix t of z does not belong to P, then a gap of $|w|$ prefixes exists and $|w|$ words having t as a prefix are in P. It would be natural to imagine that an analogous gap of size $|w|$ or $|w^2|$ exists for the solutions P of $S = 1 + w + w^2$. As a matter of fact, condition 1) in Proposition 4.1 gives a supplementary constraint for z being in $P \setminus PA$. If this case occurs, then $z = paw$ or $z = paw^2$, $pa \notin P$ (condition 2)). Conditions $3) - 6)$ state that there is a gap in the set of the prefixes of z which are in P. The size of this gap is between $|w|$ and $|w^2|$ which are the same boundaries for the cardinality of the set of the words having $z = paw$ (resp. $z = paw^2$) as a prefix and which are a prefix of zw^2 (resp. zw). Conditions $3) - 6)$ give a very precise description of the above-mentioned sets of words. In the proof of the Proposition 4.1 we also need the next lemma, already proved in [6]:

Lemma 4.1 [6] *Let S be a polynomial with a solution P. Then, for any $w, t \in A^+$ such that $wt \in P$, $w \notin P$, a prefix y of wt exists such that*

$$y \in P \setminus PA, \quad w < y \leq wt$$

PROOF :
Let y be the prefix of wt which is in $P \setminus PA$ and of maximal length with respect to this condition. Since $1 \in P \setminus PA$, then y exists. We must have $w < y \leq wt$ otherwise we would have $w \in P$. □

Proposition 4.1 *Let P be a finite subset of A^*. P is a solution for $S = 1 + w + w^2$ if and only if $P(w) \subseteq P$ and the following conditions hold*

1) for any $z \in A^+$, if $z \in P \cap Pw$ and $z \notin Pw^2$ then $(PA, z) = 1$ and a unique $i \in \{1, 2\}$ exists such that $(PAw^i, z) = 1$.
Furthermore, for any $z \in P \setminus PA$, $z \neq 1$, there exist $p \in P$, $a \in A$ with $pa \notin P$ and such that

2) $z = paw$ or $z = paw^2$

3) $paP(w) \cap P = \emptyset$*, and if* $paw \notin P$ *then* $paP(ww) \cap P = \emptyset$
4) $paw^3 \notin P$
5) if $z = paw^2$ *then* $paw^2 P(w) \subseteq P$
6) if $z = paw$ *then we have*
  (a) $z \in Pw^2 \Rightarrow z \in PAw^2$
  (b) if w *is a primitive word or* $paw^2 \in P \Rightarrow pawP(w) \subseteq P$
  (c) if w *is a non primitive word and* $paw^2 \notin P \Rightarrow$
    $pawP(w) \subseteq P$ *or* h *exists such that* $0 \leq h < k$,
    $paw\{p_0, \ldots, p_h\} \subseteq P$ *and* $paw^2\{p_{h+1}, \ldots, p_k\} \subseteq P$

Sketch of the proof:
Suppose that P verifies the conditions in Proposition 4.1. In order to show that P is a solution for any $y \in A^+$, we distinguish three cases which depend on the multiplicity of $y \in PS$:

$$(PS, y) = 1; \quad (PS, y) = 2; \quad (PS, y) = 3$$

and we verify that, in any case, the conditions $1) - 6)$ allow us to conclude $(PS, y) \leq (PAS, y)$. Conversely, suppose that P is a solution for $S = 1 + w + w^2$. Then, condition 2) is evident. Furthermore, let us prove that for any $z \in P \setminus PA$, $z \neq 1$, there exist $p \in P$, $a \in A$ with $pa \notin P$ and such that $paP(w) \cap P = \emptyset$ (which is the real crux of the matter). Indeed, using condition 2), there exist $p \in P$, $a \in A$ with $pa \notin P$ and such that $z = paw$ or $z = paw^2$. By contradiction, let h, $0 < h \leq k$ be such that $pap_h \in P$, and let $|p|$ be minimal with respect to this condition. Using Lemma 4.1, there exist t, $0 < t \leq h$, such that $pap_t \in P \setminus PA$. As P is a solution for S, we obtain

$$\begin{aligned} 1 + (Pw + Pw^2, pap_t) &= (P + Pw + Pw^2, pap_t) \leq \\ &\leq (PA, pap_t) + (PAw + PAw^2, pap_t) = \\ &= (PAw + PAw^2, pap_t) \end{aligned}$$

So $p' \in P, a' \in A, s' \in \{w, w^2\}$ exist such that

$$p'a' \notin P, \qquad pap_t = p'a's'$$

According to this relation we have two cases, but we observe that $s' = w$ implies $|ap_t| < |w|$, otherwise $p = p'a' \in P$. So $p_q \in P(w)$ exists such that $p = p'a'p_q \in P$, with $p'a' \in PA \setminus P$ and $|p'| < |p|$, a contradiction with the definition of p.
Thus $s' = w^2$, i.e., $pap_t = p'a'w^2 \in P \setminus PA$. Let us consider two situations:

$$\begin{array}{ll} (a) & p'a'w \in P \\ (b) & p'a'w \notin P \end{array}$$

Case (a): as $p'a'w^2 \in P \setminus PA$ and $p'a'w \in P$, we get

$$\begin{aligned} 2 &= (P + Pw, p'a'w^2) \leq (PS, p'a'w^2) \leq (PAS, p'a'w^2) = \\ &= (PA + PAw + PAw^2, p'a'w^2) = \\ &= 0 + (PAw, p'a'w^2) + 1 \Rightarrow (PAw, p'a'w^2) = 1 \Rightarrow p'a'p_k \in P \end{aligned}$$

a contradiction, being $|p'| < |p|$.
Case (b): obviously, $p_t \neq p_k$ (otherwise $p = p'a'w \in P$) and so $|ap_t| < |w|$. Then $p = p'a'wp_g$, $0 < g \leq k$. Thus, being $p'a'w \notin P$, once again using Lemma 4.1, p_l exists, with $0 < l \leq g \leq k$, such that $p'a'wp_l \in P \setminus PA$. Consequently

$$\begin{aligned} 1 &\leq (P + Pw + Pw^2, p'a'wp_l) \leq \\ &\leq (PA + PAw + PAw^2, p'a'wp_l) = (PAw + PAw^2, p'a'wp_l) \end{aligned}$$

Furthermore, $(PAw, p'a'wp_l) = 0$, otherwise $p'a'p_{l-1} \in P$ and $p'a' \notin P$, $|p'| < |p|$. Finally we get $(PAw^2, p'a'wp_l) = 1$, $(Pw^2, p'a'wp_l) = 0$. So $p'' \in P$, $a'' \in A$ exist such that $p''a'' \notin P$ and $p'a'wp_l = p''a''w^2$, i.e., $p'a'p_l = p''a''w$. Since $|a'p_l| < |w|$ (otherwise $p''a'' \in P$) we have $p' = p''a''p_r \in P$, with $p_r \in P(w)$. As $|p'| < |p|$, this is in contradiction with the definition of $|p|$.

The above-proved result allows us to obtain conditions 1) and $4) - 6)$. As an example, let us prove 1). Let $z \in A^+$ be such that $(P, z) = 1 = (Pw, z)$ and $(Pw^2, z) = 0$. We observe that if $(PAw + PAw^2, z) = 0$, then P cannot be a solution for S. If $(PAw + PAw^2, z) = 2$, since $z = p_1 w \in P$ with $p_1 \in P$, then we have $p_1 \in PA \cap PAw$. Then $p' \in P$, $a' \in A$ exist such that $p_1 = p'a'w$, $p'a' \notin P$ (otherwise $z \in Pw^2$). So, using the result proved above and applied to $p'a'$, we get $p'a'P(w) \cap P = \emptyset$, i.e., $p'a'w = p_1 \notin PA$, a contradiction. Thus, $(PAw + PAw^2, z) = 1$ and condition 1) holds. □

In the next proposition we will examine the second case in which $S = 1 + w_1 + w_2$, with $\{w_1, w_2\}$ a prefix code. We will report here the characterization of the solutions of $S = 1 + X$, with $X = \{w_1, \ldots, w_k\}$ being a prefix code and, for better readability, we will set $P_i = P(w_i)$. This characterization is a slight modification of a result given in [6] and does not directly give an algorithm for constructing the solutions P for S. On the contrary, the above-mentioned algorithm is suggested by the proof of Proposition 4.2 and it is an extension of the analogous algorithm for $S = 1 + w$. In particular, if $w_i \notin A^* w_j A^*$, for $i, j \in \{1, \ldots, k\}$, we can see that every solution P for S can be obtained from $P_1' + \ldots + P_k'$, with P_i' being a solution for $S = 1 + w_i$. Precisely, P is a solution if and only if P verifies Proposition 2.3 and for any $z \in P \setminus PA$, $z \neq 1$, $h \in \{1, \ldots, k\}$ exists such that $z = paw_h$, $paP_h \cap P = \emptyset$ and $zP_i \subseteq P$, $zw_i \notin P$, for every $i \in \{1, \ldots, k\}$. If $w_2 \in A^* w_1 A^*$, every solution P for $S = 1 + w_1 + w_2$

can be obtained either as above or as above but substituting some words in zP_2 with some other words.

Proposition 4.2 *P is a solution for $S = 1 + w_1 + \ldots + w_k$, with $\{w_1, \ldots, w_k\}$ a prefix code if and only if $1 = w_0 \in P$ and for all $z \in PS$:*

1. *$\nexists j \in \{0, \ldots, k\}$ such that $z \in Pw_j \setminus PAw_j$*

   or

   $\exists!\ j \in \{0, \ldots, k\}$ such that $z \in Pw_j \setminus PAw_j$

2. *$\forall j \in \{0, \ldots, k\}$ such that $z \in Pw_j \setminus PAw_j$: $\exists p \in P$, $a \in A$, $h \in \{0, \ldots, k\}$ such that $z = paw_h$, $pa \notin P$*

Sketch of the proof:

If P verifies conditions 1) and 2), then P is a solution for S, since these two conditions ensure that the multiplicity of z in PS is less than or equal to the multiplicity of z in PAS. Conversely, if P is a solution for S, then 2) holds with no hypothesis on S, whereas 1) can be obtained since $paP_h \cap P = \emptyset$, for any $z \in PAw_h \setminus Pw_h$ (this last condition can be proved by taking a counterexample z of minimal length and $p_n \in P(w_h)$ of minimal length such that $pap_n \in P \setminus PA$). $\square$

References

1. M. Anselmo, C. De Felice, A. Restivo, On some factorization problems, *Bull. Belg. Math. Soc.* **4** (1997) 25–43.
2. M. Anselmo, A. Restivo, On languages factorizing the free monoid, *IJAC* **4** (1996) 413–427.
3. J. Berstel, D. Perrin, *Theory of Codes*, Academic Press, New York (1985).
4. J. Berstel, C. Reutenauer, *Rational series and their languages*, EATCS Monogr. Theoret. Comput. Sci. **12**, Springer Verlag (1988).
5. V. Bruyère and M. Latteux, Variable-Length Maximal Codes, *Lecture Notes in Comput. Sci.* **1099** (1996) 24–47.
6. C. De Felice, *Construction et complétion de codes finis*, Thèse de 3ème cycle, Université de Paris VII (1984).
7. C. De Felice, Construction of a family of finite maximal codes, *Theoret. Comput. Sci.* **63** (1989) 157–184.
8. C. De Felice, A partial result about the factorization conjecture for finite variable-length codes, *Discrete Math.* 122 (1993) 137–152.
9. C. De Felice, On a property of the factorizing codes, *IJAC* (1998) to appear, 26 pages.

10. C. De Felice, C. Reutenauer, Solution partielle de la conjecture de factorisation des codes, C.R. Acad. Sc. Paris **302** (1986) 169–170.
11. C. De Felice, R. Zizza, Factorizing Codes and Krasner Factorizations, *mauscript* (1998) 31 pages.
12. M. Krasner, B. Ranulac, Sur une propriété des polynômes de la division du cercle, *C. R. Acad. Sc. Paris* **240** (1937) 397–399.
13. N. H. Lam, On codes having no finite completion, Inf. Théor. et App./ Theor. Inf. and App. **29** (1995) 145–155.
14. M. Lothaire, *Combinatorics on Words*, Encyclopedia of Math. and its Appl., Addison Wesley Publishing Company (1983).
15. 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.
16. D. Perrin, M. P. Schützenberger, A conjecture on sets of differences of integer pairs, *J. Comb. Theory B* **30** (1981) 91–93.
17. A. Restivo, On codes having no finite completions, *Discrete Math.* **17** (1977) 309–316.
18. C. Reutenauer, Non commutative factorization of variable-length codes, *J. Pure and Applied Algebra* **36** (1985) 167–186.
19. A. Salomaa, M. Soittola, *Automata-Theoretic Aspects of Formal Power Series*, Springer, Berlin Heidelberg New York (1978).
20. M. P. Schützenberger, Une théorie algébrique du codage, *Séminaire Dubreil-Pisot* 1955-56, exposé no 15 (1955) 24 pages.
21. L. Zhang, C. K. Gu, Two classes of factorizing codes - (p,p)-codes and $(4,4)$-codes, in *"Words, Languages and Combinatorics II"* (M. Ito, H. Jürgensen, eds.), World Scientific (1994) 477–483.