RAIRO-Inf. Theor. Appl. **39** (2005) 161-174
DOI: 10.1051/ita:2005010

SOME DECOMPOSITIONS OF BERNOULLI SETS AND CODES [*]

Aldo de Luca[1]

Abstract. A decomposition of a set X of words over a d-letter alphabet $A = \{a_1, \ldots, a_d\}$ is any sequence $X_1, \ldots, X_d, Y_1, \ldots, Y_d$ of subsets of A^* such that the sets X_i, $i = 1, \ldots, d$, are pairwise disjoint, their union is X, and for all i, $1 \leq i \leq d$, $X_i \sim a_i Y_i$, where $\sim$ denotes the commutative equivalence relation. We introduce some suitable decompositions that we call good, admissible, and normal. A normal decomposition is admissible and an admissible decomposition is good. We prove that a set is commutatively prefix if and only if it has a normal decomposition. In particular, we consider decompositions of Bernoulli sets and codes. We prove that there exist Bernoulli sets which have no good decomposition. Moreover, we show that the classical conjecture of commutative equivalence of finite maximal codes to prefix ones is equivalent to the statement that any finite and maximal code has an admissible decomposition.

Mathematics Subject Classification. 94A45.

Introduction

Prefix sets are sets X of words over a given finite alphabet such that no word of X is a proper initial part (or prefix) of another word of X. As is well known prefix sets play an essential role in Information Theory and in Computer Science. This is mainly due to the existence of a one-to-one correspondence between prefix sets and the sets of leaves of rooted trees. A prefix set is maximal if the corresponding

Keywords and phrases. Bernoulli sets, codes, decompositions, commutative equivalence.

[*] *The work for this paper has been supported by the Italian Ministry of Education under Project COFIN '01 "Linguaggi Formali e Automi: metodi, modelli e applicazioni".*

[1] Dipartimento di Matematica e Applicazioni dell'Università di Napoli "Federico II",
via Cintia, Complesso Universitario di Monte S. Angelo, 80126 Napoli, Italy and Istituto di Cibernetica "E. R. Caianiello" del CNR, 80078 Pozzuoli, Italy;
aldo.deluca@dma.unina.it

tree is complete, *i.e.*, the degree of all internal nodes of the tree is equal to the number of letters of the alphabet.

An important property satisfied by a finite maximal prefix set X is that for any positive Bernoulli distribution π on the alphabet A one has $\pi(X) = 1$. Any set X over A satisfying the preceding relation for all positive Bernoulli distributions is called a Bernoulli set (*cf.* [4]). For instance, any finite maximal code is a Bernoulli set.

In general a Bernoulli set X is not commutatively equivalent to a prefix set, *i.e.*, one cannot produce by permuting the letters in the words of X, a prefix set Y having the same cardinality of X. Some characterizations of sets which are commutatively equivalent to prefix sets are in [1,2,4]. We recall that the statement that a finite maximal code is commutatively equivalent to a prefix set is a classical still open conjecture of theory of codes formulated by Schützenberger at the end of '50s [6, 7].

In this paper we start by the fact that if X is a maximal prefix code, then for any letter $a \in A$, the set $Y_a = a^{-1}X$ is a maximal prefix set. Thus any maximal prefix code can be analyzed in terms of 'simpler' maximal prefix sets. Let us observe that if X is finite, then any set Y_a, $a \in A$, is a Bernoulli set. It is likely that a finite set is commutatively prefix only if it preserves in a suitable way this kind of property. For this reason we consider in Section 3 some decompositions of sets of words. More precisely, a decomposition of a set X of words over a d-letter alphabet $A = \{a_1, \ldots, a_d\}$ is any sequence $X_1, \ldots, X_d, Y_1, \ldots, Y_d$ of subsets of A^* such that the sets X_i, $i = 1, \ldots, d$, are pairwise disjoint, their union is X, and for all i, $1 \leq i \leq d$, X_i is commutatively equivalent to $a_i Y_i$.

A decomposition is called good if for all positive Bernoulli distributions π over A one has $\pi(Y_i) \leq 1$, $1 \leq i \leq d$. Good decompositions are considered in Section 4. In particular, it is shown that the property of being a Bernoulli set is preserved under good decompositions. Moreover, it is shown the existence of non-trivial Bernoulli sets which do not admit good decompositions; this implies that they are not commutatively prefix. However, there exist Bernoulli sets which are not commutatively prefix and have a good decomposition.

In Section 5 we consider admissible and normal decompositions. A decomposition is admissible when for all i, $1 \leq i \leq d$, the sets Y_i are codes or $Y_i = \{\epsilon\}$. A decomposition is normal if for all i, $1 \leq i \leq d$, the sets Y_i are prefix. A normal decomposition is admissible and an admissible decomposition is good. We prove that a set is commutatively prefix if and only if it has a normal decomposition. Finally, in Section 6 we show that the conjecture of Schützenberger is equivalent to the statement that any finite and maximal code has an admissible decomposition.

1. Preliminaries

Let A be a finite nonempty set, or *alphabet*, of cardinality $d > 0$ and A^* the free monoid generated by A. The elements of A are usually called *letters* and those of A^* *words*. The identity element of A^* is called *empty word* and denoted by ϵ. We set $A^+ = A^* \setminus \{\epsilon\}$.

A word $w \in A^+$ can be written uniquely as a sequence of letters as $w = w_1 w_2 \cdots w_n$, with $w_i \in A$, $1 \leq i \leq n$, $n > 0$. The integer n is called the *length* of w and denoted $|w|$. The length of ϵ is 0. For any $w \in A^*$ and $a \in A$, $|w|_a$ denotes the number of occurrences of the letter a in w.

Let $w \in A^*$. The word $u \in A^*$ is a *prefix* of w if there exists a word λ such that $w = u\lambda$. A prefix u of w is called *proper* if $u \neq w$. In the following a subset X of A^* will be simply called *set over* (the alphabet) A.

A *Bernoulli distribution* π over the alphabet A (*cf.*[1]) is any map

$$\pi : A \to \mathbb{R}_+,$$

where $\mathbb{R}_+$ is the set of non-negative real numbers, such that

$$\sum_{a \in A} \pi(a) = 1.$$

A Bernoulli distribution is *positive* if for all $a \in A$ one has $\pi(a) > 0$. We denote by $PBD(A)$, or simply PBD, the set of all positive Bernoulli distributions on A. A particular positive Bernoulli distribution is the *uniform* distribution π_u defined for any $a \in A$ as $\pi_u(a) = d^{-1}$.

If π is a Bernoulli distribution over A, then one can extend π to a morphism of A^* in the multiplicative monoid $\mathbb{R}_+$. Hence, $\pi(\epsilon) = 1$ and for all $u, v \in A^*$ one has:

$$\pi(uv) = \pi(u)\pi(v).$$

One can extend also π to sets X over A by setting: $\pi(\emptyset) = 0$ and for $X \neq \emptyset$

$$\pi(X) = \sum_{w \in X} \pi(w).$$

Let us observe that for some sets X the value $\pi(X)$ may be infinite. A set X over A is called *Bernoulli set* if for all $\pi \in PBD$:

$$\pi(X) = 1. \tag{1}$$

If X is a set over A, we denote by $\underline{\underline{X}}$ the *characteristic series* of X in commutative variables. If X is a finite set, the series $\underline{\underline{X}}$ becomes a polynomial. For instance, in the case of the sets $X = \{aa, ab, ba, bb\}$, $\{\epsilon\}$, and $\emptyset$, one has $\underline{\underline{X}} = a^2 + 2ab + b^2$, $\underline{\underline{\{\epsilon\}}} = 1$, and $\underline{\underline{\emptyset}} = 0$.

The following characterization of finite Bernoulli sets holds (*cf.*[1, 4]):

Proposition 1.1. *A finite set X over the alphabet A is a Bernoulli set if and only if $\underline{\underline{A}} - 1$ divides the polynomial $\underline{\underline{X}} - 1$, i.e.,*

$$\underline{\underline{X}} - 1 = P(\underline{\underline{A}} - 1), \tag{2}$$

where P is a polynomial of $\mathbb{Z}[A]$.

We recall the following property which is satisfied by finite Bernoulli sets [4]:

Proposition 1.2. *Let X be a finite Bernoulli set. For any $a \in A$ there exists a unique non-negative integer p such that $a^p \in X$.*

We consider in A^* the relation of *commutative equivalence* $\sim$ defined as follows: two words $u, v \in A^*$ are commutatively equivalent, and we write $u \sim v$, if

$$|u|_a = |v|_a, \quad \text{for any} \quad a \in A.$$

Two sets X and Y over A are commutatively equivalent, and we write $X \sim Y$, if there exists a bijection $\delta : X \to Y$, called *commutation map*, such that for any $x \in X$ one has $x \sim \delta(x)$. In terms of the commutative characteristic series one has that $X \sim Y$ if and only if

$$\underline{\underline{X}} = \underline{\underline{Y}}.$$

If X is a finite set over A, we denote by $\|X\|$ the quantity $\|X\| = \sum_{x \in X} |x|$. We call $\|X\|$ the *size* of X.

Let X be a set over A. For any letter a, we denote by $a^{-1}X$ the set

$$a^{-1}X = \{w \in A^* \mid aw \in X\}.$$

2. Codes and prefix sets

A set X over a d-letter alphabet $A = \{a_1, \ldots, a_d\}$ is called *code* over A if X is the base of a free submonoid of A^*, *i.e.*, any nonempty word of X^* can be uniquely factorized in terms of the elements of X. We note that according to the definition the empty subset of A^* is a code.

As is well known [1] if X is a code over A, then for any Bernoulli distribution π over A one has $\pi(X) \leq 1$ (generalized Kraft-McMillan inequality). In the case of the uniform distribution one has the classic Kraft-McMillan inequality

$$\sum_{x \in X} d^{-|x|} \leq 1.$$

A code X over A is *maximal* if it is not properly included in a larger code over the same alphabet A. The following proposition holds (*cf.* [1]):

Proposition 2.1. *Let X be a finite code. X is maximal if and only if it is a Bernoulli set.*

A set X over A is commutatively equivalent to a code if there exists a code Y such that $X \sim Y$. We observe (*cf.*[4]) that there exist finite Bernoulli sets which are not commutatively equivalent to any code.

A set X over A is called *prefix* if

$$X \cap XA^+ = \emptyset,$$

i.e., no word of X is a proper prefix of another word. For instance, the sets $X_1 = \emptyset$, $X_2 = \{\epsilon\}$, and $X_3 = \{a, ba, bb\}$ are prefix sets. We observe that all prefix sets $X \subseteq A^+$ are codes.

A prefix set is *maximal* if it is not properly included in a larger prefix set over the same alphabet. This is trivially equivalent to say (*cf.*[1]) that for any $w \in A^*$

$$wA^* \cap XA^* \neq \emptyset. \tag{3}$$

We remark that the set $\{\epsilon\}$ is a maximal prefix set. As is well known [1] a finite prefix code is maximal (as a prefix) if and only if it is maximal as code.

Let X be a subset of A^+. Trivially X can be uniquely written as

$$X = \bigcup_{i=1}^{d} a_i Y_i, \tag{4}$$

where the union is disjoint and $Y_i = a_i^{-1}X$, $1 \leq i \leq d$. The (maximal) prefix property is preserved passing from the set X to the sets Y_i, $1 \leq i \leq d$, and *vice versa*, as shown by the following lemma whose simple proof we report for the sake of completeness.

Lemma 2.2. *Let X be a set over the alphabet A. The set X is prefix if and only if the sets $Y_i = a_i^{-1}X$, $1 \leq i \leq d$, are prefix. Moreover, X is a maximal prefix code if and only if the sets Y_i, $1 \leq i \leq d$, are maximal prefix sets.*

Proof. If $X = \emptyset$ or $X = \{\epsilon\}$ the result is trivial. Thus we suppose that X is a non-empty prefix code. In such a case for any letter a, the set $Y = a^{-1}X$ is a prefix set. Indeed, if $Y = \emptyset$ or $Y = \{\epsilon\}$ we are done. Therefore, we suppose that $\emptyset \neq Y \subseteq A^+$. Let $y_1, y_2 \in Y$ and $\xi \in A^*$ be such that $y_1 = y_2\xi$, so that $ay_1 = ay_2\xi$. As $ay_1, ay_2 \in X$ it follows that $\xi = \epsilon$ and $y_1 = y_2$. Conversely, suppose that the sets Y_i, $1 \leq i \leq d$, are prefix sets and suppose that $x_1 = x_2\xi$ with $x_1, x_2 \in X$ and $\xi \in A^*$. This implies that there exists a letter a such that $x_1 = ay_1, x_2 = ay_2$ with $y_1, y_2 \in Y = a^{-1}X$. Therefore, $y_1 = y_2\xi$. This implies $\xi = \epsilon$. Hence, $y_1 = y_2$ and $x_1 = x_2$.

Let $a \in A$, $w \in A^*$, and $Y = a^{-1}X$. If X is a maximal prefix code, then from equation (3) one has $awA^* \cap XA^* \neq \emptyset$, so that $wA^* \cap YA^* \neq \emptyset$. Hence, Y is maximal prefix set. Conversely, let w be any word of A^+ and write $w = au$ with $a \in A$ and $u \in A^*$. Since by hypothesis the set $Y = a^{-1}X$ is a maximal prefix set by equation (3) one has $uA^* \cap YA^* \neq \emptyset$ and $auA^* \cap aYA^* \neq \emptyset$. Hence, for any $w \in A^*$ one has $wA^* \cap XA^* \neq \emptyset$. Since $X \neq \{\epsilon\}$ the result follows. □

Example 2.3. Let $X = \{a^2, aba, ab^2, ba, b^2\}$. The set X is a maximal prefix code. In this case $a^{-1}X = \{a, ba, b^2\}$ and $b^{-1}X = \{a, b\}$ are also maximal prefix codes. Let $X = \{a^2, aba, ab^2, b\}$. One has that X is a maximal prefix code and $a^{-1}X = \{a, ba, b^2\}$ and $b^{-1}X = \{\epsilon\}$ are maximal prefix sets.

Let us observe that the code property is not, in general, preserved passing from the set X to the sets $Y_i = a_i^{-1}X$, $1 \leq i \leq d$, or *vice versa*. For instance, in the case of the code $X = \{a, ab, bb\}$ one has that $a^{-1}X = \{\epsilon, b\}$ which is not a code. Conversely, the set $X = \{aa, ba, bb, aab, abb\}$ is not a code whereas $a^{-1}X = \{a, ab, bb\}$ and $b^{-1}X = \{a, b\}$ are codes.

A set X is *commutatively prefix* if there exists a prefix set Y such that $X \sim Y$. The following holds (*cf.* [1]):

Proposition 2.4. *A set X over the alphabet A is commutatively prefix if and only if the series*

$$\frac{\underline{\underline{X}} - 1}{\underline{\underline{A}} - 1}$$

has non-negative coefficients.

Let X be a finite Bernoulli set and set, by Proposition 1.1, $\underline{\underline{X}} - 1 = P(\underline{\underline{A}} - 1)$. From Proposition 2.4 one has that X is commutatively prefix if and only if the polynomial P has non-negative coefficients (in such a case we write $P \geq 0$).

In [4] some examples of finite Bernoulli sets which are not commutatively prefix are given. For instance, the set $X = \{a^4, ba^2, a^2b, aba, ba, b^3a, b^2\}$ is a Bernoulli set which is not commutatively prefix. In fact, in this case one has $\underline{\underline{X}} - 1 = P(a+b-1)$ with

$$P = ab^2 + ab - a^2b + b + a^3 + a^2 + a + 1.$$

3. Decompositions

Let X be a set over a d-letter alphabet $A = \{a_1, \ldots, a_d\}$. A *decomposition* of X is any sequence $X_1, \ldots, X_d, Y_1, \ldots, Y_d$ of $2d$ subsets of A^* such that

$$X = \bigcup_{i=1,\ldots,d} X_i, \quad X_i \cap X_j = \emptyset, \quad \text{for } i \neq j,$$

and for all i, $1 \leq i \leq d$, one has

$$X_i \sim a_i Y_i.$$

In the following we shall denote a decomposition of X by $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$.

It is clear that if $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$ is a decomposition of X, then also $(X_1, \ldots, X_d; Z_1, \ldots, Z_d)$ with $Z_i \sim Y_i$, $1 \le i \le d$, is a decomposition of X.

With each set $X \subseteq A^+$ one can associate the *natural decomposition*, already introduced by equation (4),

$$(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$$

where for any $a_i \in A$, $1 \le i \le d$, $X_i = X \cap a_i A^*$ and $Y_i = a_i^{-1} X$.

Example 3.1. Let X be the set over $A = \{a, b\}$

$$X = \{a^3, ab, ba, b^2a, a^3b\}.$$

A first decomposition of X is $(X_1, X_2; Y_1, Y_2)$ where $X_1 = \{a^3, ab\}$, $X_2 = \{ba, b^2a, a^3b\}$, and $Y_1 = \{a^2, b\}$, $Y_2 = \{a, ba, a^3\}$. A second decomposition is $(X_1', X_2'; Y_1', Y_2')$ where $X_1' = \{a^3, ab, a^3b\}$, $X_2' = \{b^2a, ba\}$, and $Y_1' = \{a^2, b, aba\}$, $Y_2' = \{a, ba\}$.

Lemma 3.2. *If the set X over A has a decomposition $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$, then*

$$\underline{\underline{X}} = \sum_{i=1}^{d} a_i \; \underline{\underline{Y_i}},$$

and for any positive Bernoulli distribution π one has

$$\pi(X) = \sum_{i=1}^{d} \pi(a_i)\pi(Y_i).$$

Proof. We can write

$$X = \bigcup_{i=1}^{d} X_i.$$

Since the union is disjoint, one has

$$\underline{\underline{X}} = \sum_{i=1}^{d} \underline{\underline{X_i}}$$

and

$$\pi(X) = \sum_{i=1}^{d} \pi(X_i).$$

Since for any i, $1 \le i \le d$, one has

$$\underline{\underline{X_i}} = a_i \underline{\underline{Y_i}} \quad \text{and} \quad \pi(X_i) = \pi(a_i)\pi(Y_i),$$

the result follows. □

Proposition 3.3. *Let X be a set over A having the decomposition $(X_1, \dots, X_d; Y_1, \dots, Y_d)$. If Z is a set commutatively equivalent to X and $\delta : Z \to X$ is the commutation map, then Z admits a decomposition $(Z_1, \dots, Z_d; Y_1, \dots, Y_d)$ where $Z_i = \delta^{-1}(X_i)$ for all $1 \le i \le d$.*

Proof. If $\delta : Z \to X$ is the commutation map, one has $Z = \delta^{-1}(X)$. For any $1 \le i \le d$ let us set $Z_i = \delta^{-1}(X_i)$. One has

$$Z = \bigcup_{i=1}^{d} Z_i$$

where the sets Z_i, $1 \le i \le d$, are pairwise disjoint. Moreover, $Z_i \sim X_i \sim a_i Y_i$ $(i = 1, \dots, d)$, so that $(Z_1, \dots, Z_d; Y_1, \dots, Y_d)$ is a decomposition of Z. □

4. Good decompositions

Let X be a set over the alphabet A. A decomposition $(X_1, \dots, X_d; Y_1, \dots, Y_d)$ is *good* if for all $\pi \in PBD$ one has

$$\pi(Y_i) \le 1, \quad 1 \le i \le d.$$

Let us observe that the natural decomposition of a prefix code X over A is a good decomposition of X. Indeed, by Lemma 2.2 the sets $Y_i = a_i^{-1}X$, $1 \le i \le d$, are prefix sets so that for any $\pi \in PBD$, one has $\pi(Y_i) \le 1$.

Example 4.1. The first decomposition of X in Example 3.1 is not a good decomposition. Indeed, let $\pi(a) = p$ and $\pi(b) = 1 - p$ with $0 < p < 1$. One has $\pi(Y_2) = p^3 + 2p - p^2$ and this quantity is greater than 1 for p near to 1. The second decomposition is good. Indeed, one has $\pi(Y_1') = 1 + 2p^2 - p - p^3 = 1 - p(p-1)^2$ and $\pi(Y_2') = 2p - p^2 = 1 - (p-1)^2$. These quantities are always < 1 for all $0 < p < 1$, so that this decomposition is good.

Lemma 4.2. *Let X be a set over the alphabet A having a good decomposition $(X_1, \dots, X_d; Y_1, \dots, Y_d)$. The set X is a Bernoulli set if and only if for all i, $1 \le i \le d$, Y_i is a Bernoulli set.*

Proof. By Lemma 3.2 for any $\pi \in PBD$ one has:

$$\pi(X) = \sum_{i=1}^{d} \pi(a_i)\pi(Y_i). \tag{5}$$

Since X is a Bernoulli set one has $\pi(X) = 1$. Moreover, the decomposition is good so that for all i, $1 \le i \le d$, $\pi(Y_i) \le 1$. From equation (5) it follows that for all i, $1 \le i \le d$, $\pi(Y_i) = 1$, *i.e.*, Y_i is a Bernoulli set. Conversely, if one supposes that for all i, $1 \le i \le d$, Y_i is a Bernoulli set, then from equation (5) for any $\pi \in PBD$ one has $\pi(X) = \pi(A) = 1$, *i.e.*, X is a Bernoulli set. □

Proposition 4.3. *Let X be a finite Bernoulli set over the alphabet A having a good decomposition $(X_1, \dots, X_d; Y_1, \dots, Y_d)$. One has for all i, $1 \leq i \leq d$,*

$$\underline{\underline{Y_i}} - 1 = P_i(\underline{\underline{A}} - 1), \;\; P_i \in \mathbb{Z}[A]$$

and

$$\underline{\underline{X}} - 1 = P(\underline{\underline{A}} - 1),$$

with

$$P = 1 + \sum_{i=1}^{d} a_i P_i.$$

Proof. Since X has the decomposition $(X_1, \dots, X_d; Y_1, \dots, Y_d)$, by Lemma 3.2 one has

$$\underline{\underline{X}} = \sum_{i=1}^{d} a_i \; \underline{\underline{Y_i}}. \tag{6}$$

Moreover, as the decomposition is good and X is a Bernoulli set, by Lemma 4.2 for any i, $1 \leq i \leq d$, Y_i is a Bernoulli set. By Proposition 1.1 one can write

$$\underline{\underline{Y_i}} = 1 + P_i(\underline{\underline{A}} - 1) \;\text{ with }\; P_i \in \mathbb{Z}[A], \;\; 1 \leq i \leq d.$$

By replacing this expression for $\underline{\underline{Y_i}}$, $1 \leq i \leq d$, in equation (6) one obtains

$$\underline{\underline{X}} - 1 = (1 + \sum_{i=1}^{d} a_i P_i)(\underline{\underline{A}} - 1),$$

which proves our assertion. □

Corollary 4.4. *There exist finite Bernoulli sets $\neq \{\epsilon\}$ on the alphabet $A = \{a, b\}$ which have no good decomposition.*

Proof. Suppose by contradiction that any finite Bernoulli set $\neq \{\epsilon\}$ over $\{a, b\}$ has a good decomposition. This would imply that any finite Bernoulli set over $\{a, b\}$ is commutatively prefix. In fact, let X be a Bernoulli set over $\{a, b\}$ of minimal size which is not commutatively prefix, so $X \neq \{\epsilon\}$. By Proposition 4.3 we can write, setting $a_1 = a$ and $a_2 = b$,

$$\underline{\underline{X}} = a_1 \underline{\underline{Y_1}} + a_2 \underline{\underline{Y_2}} \; .$$

Moreover,

$$\underline{\underline{Y_i}} - 1 = P_i(a_1 + a_2 - 1), \;\; i = 1, 2,$$

and

$$P = 1 + a_1 P_1 + a_2 P_2 \; .$$

Since Y_i, $i = 1, 2$, are Bernoulli sets of size less than the size $\|X\|$ one has that they are commutatively prefix. By Proposition 2.4 one has $P_1, P_2 \geq 0$. This implies that $P \geq 0$, so that by Proposition 2.4 one has that X is commutatively prefix which is a contradiction. Since there exist finite Bernoulli sets $\neq \{\epsilon\}$ over $\{a, b\}$ which are not commutatively prefix [4], the result follows. □

Example 4.5. Consider the Bernoulli set $X = \{a^4, ba^2, a^2b, aba, ba, b^3a, b^2\}$. We prove that X has no good decomposition. Indeed, suppose that X has a good decomposition $(X_1, X_2; Y_1, Y_2)$, where $X_1 \sim aY_1$ and $X_2 \sim bY_2$. By definition of decomposition one has that $a^4 \in X_1$ and $b^2 \in X_2$. If $ba \in X_2$, then $a, b \in Y_2$. By Lemma 4.2, Y_2 is a Bernoulli set so that it cannot contain other words. This implies that

$$Y_1 = \{a^3, ba, ab, b^3\},$$

which is not a Bernoulli set. This is a contradiction so that $ba \in X_1$ and $b \in Y_1$. If $b^3a \in X_1$, then $b^3 \in Y_1$ which is a contradiction since Y_1 is a Bernoulli set so that it cannot contain two different powers of the same letter (*cf.* Prop. 1.2). Hence, $b^3a \in X_2$ and Y_2 has to contain one word in the commutation class of b^2a. The words of X which remain are ba^2, a^2b, aba. Since the set Y_2 has to contain a power of a, one, and only one, of the preceding three words has to belong to X_2. This implies that $a^2 \in Y_2$. Therefore,

$$Y_2 \sim \{b, b^2a, a^2\}$$

which is not a Bernoulli set. Thus we have a contradiction and the assertion is proved.

Proposition 4.6. *Any set $X \subseteq A^+$ which is commutatively prefix has a good decomposition.*

Proof. By hypothesis there exists a prefix code Z such that $X \sim Z$. Now any prefix code has a decomposition $(Z_1, \ldots, Z_d; Y_1, \ldots, Y_d)$, where for all i, $1 \leq i \leq d$, the sets Y_i are prefix sets. By Proposition 3.3, the set X has a decomposition $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$. Since for all i, $1 \leq i \leq d$, and all $\pi \in PBD$, $\pi(Y_i) \leq 1$, the result follows. □

From the preceding proposition one has in particular that any Bernoulli set $\neq \{\epsilon\}$ which is commutatively prefix has a good decomposition. However, the converse is not generally true as shown by the following:

Example 4.7. Let X be the set $X = \{a^4, ba^2, a^2b, aba, ba, b^3a, b^2\}$ and consider the Bernoulli set $Y = aX \cup \{b\}$. Since X is a Bernoulli set, one has, trivially, that Y is a Bernoulli set having the good decomposition $(aX, \{b\}; X, \{\epsilon\})$. However, Y is not commutatively prefix. In fact, $\underline{\underline{Y}} - 1 = P(a + b - 1)$ with

$$P = a^2b^2 + ab + a^2b - a^3b + a^4 + a^3 + a^2 + a + 1.$$

5. Admissible and normal decompositions

Let X be a set over A. A decomposition $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$ of X is called *admissible* if for any i, $1 \leq i \leq d$, the set Y_i is a code or $Y_i = \{\epsilon\}$.

Let us observe that the natural decomposition of a prefix code is trivially an admissible decomposition.

Lemma 5.1. *An admissible decomposition of a set is a good decomposition.*

Proof. Since for any i, $1 \leq i \leq d$, the set Y_i is a code or $Y_i = \{\epsilon\}$ by the generalized Kraft-McMillan inequality and the fact that $\pi(\epsilon) = 1$, one has for any $\pi \in PBD$, $\pi(Y_i) \leq 1$, $i = 1, \ldots, d$, that proves the assertion. □

The converse of the preceding lemma is not generally true. For instance, the second decomposition $(X_1', X_2'; Y_1', Y_2')$ of the set X of Example 3.1 is admissible since Y_1' and Y_2' are codes. However, if we replace Y_1' with the set $Y_1'' \sim Y_1'$ given by $Y_1'' = \{a^2, b, ba^2\}$ one has that $(X_1', X_2'; Y_1'', Y_2')$ is a good decomposition of the set X which is not admissible.

Proposition 5.2. *Let X be a finite set over A having an admissible decomposition $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$. The set X is a Bernoulli set if and only if for all i, $1 \leq i \leq d$, the sets $Y_i \neq \{\epsilon\}$ are maximal codes.*

Proof. By Lemma 5.1 an admissible decomposition is a good decomposition. Hence, if X is a Bernoulli set, by Lemma 4.2 for all i, $1 \leq i \leq d$, the sets Y_i are Bernoulli sets. From Proposition 2.1, one derives that for all i, $1 \leq i \leq d$, if $Y_i \neq \{\epsilon\}$, then it is a maximal code. Conversely, if for all i, $1 \leq i \leq d$, $Y_i = \{\epsilon\}$ or Y_i is a maximal code, then from Proposition 2.1 for all $\pi \in PBD$ one has $\pi(Y_i) = 1$, which implies $\pi(X) = 1$, *i.e.*, X is a Bernoulli set. □

Remark 5.3. We observe that the necessary part of the preceding proposition holds true even if the set X is not finite. Indeed, if for all i, $1 \leq i \leq d$, the sets Y_i are Bernoulli sets, then for any $\pi \in PBD$ one has $\pi(Y_i) = 1$. In view of the generalized Kraft-McMillan inequality, this trivially implies that the sets $Y_i \neq \{\epsilon\}$ are maximal codes.

An admissible decomposition $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$ of a set X over the alphabet A is called *normal* if for all i, $1 \leq i \leq d$, the sets Y_i are prefix sets. Obviously, the natural decomposition of a prefix code is a normal decomposition.

Proposition 5.4. *A set $X \subseteq A^+$ has a normal decomposition if and only if it is commutatively prefix.*

Proof. Let X be a set over A having a normal decomposition

$$(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$$

where for all i, $1 \leq i \leq d$, the sets Y_i are prefix. One can write by Lemma 3.2

$$\underline{\underline{X}} = \sum_{i=1}^{d} a_i \, \underline{\underline{Y_i}}.$$

Since for all i, $1 \leq i \leq d$, the sets $a_i Y_i$ are prefix, the set

$$Z = \bigcup_{i=1}^{d} a_i Y_i$$

is a prefix code such that $\underline{\underline{Z}} = \underline{\underline{X}}$. This implies that X is commutatively prefix.

Conversely, suppose that $X \sim Z$ where Z is a prefix code and denote by δ the commutation map $\delta : X \to Z$. The prefix code Z admits the decomposition $(Z_1, \ldots, Z_d; Y_1, \ldots, Y_d)$ where for any i, $1 \leq i \leq d$, $Z_i = Z \cap a_i A^* = a_i Y_i$. By Proposition 3.3, X has the decomposition $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$ with $X_i = \delta^{-1}(Z_i)$ $(i = 1, \ldots, d)$. Since the sets Y_i $(i = 1, \ldots, d)$ are prefix sets, this decomposition is normal which proves the assertion. □

6. The Schützenberger conjecture

P. Shor has shown [9] the existence of a finite code S in a two-letter alphabet $\{a, b\}$ which is not commutatively prefix. The code S is the subset of a^*ba^* formed by the following 16 words:

$$\begin{array}{llll}
b & a^3b & a^8b & a^{11}b \\
ba & a^3ba^2 & a^8ba^2 & a^{11}ba \\
ba^7 & a^3ba^4 & a^8ba^4 & a^{11}ba^2 \\
ba^{13} & a^3ba^6 & a^8ba^6 & \\
ba^{14} & & &
\end{array}$$

Shor's result gave a negative answer to a conjecture (triangle conjecture) formulated by Perrin and Schützenberger in [6, 7] (see also [3, 5, 8]). Since S is not commutatively prefix, by Proposition 5.4 one has that Shor's code has no normal decomposition.

Proposition 6.1. *The Shor code admits a good decomposition.*

Proof. We consider the decomposition $(S_1, S_2; Y_1, Y_2)$ of S where $S_2 = \{b\}$ and $S_1 = S \setminus \{b\}$. Since $S_1 \sim aY_1$ and $S_2 \sim bY_2$ one has that $Y_2 = \{\epsilon\}$ and Y_1 is any set commutatively equivalent to the set Z given by the following 15 words

$$\begin{array}{llll}
b & a^2b & a^7b & a^{10}b \\
ba^6 & a^2ba^2 & a^7ba^2 & a^{10}ba \\
ba^{12} & a^2ba^4 & a^7ba^4 & a^{10}ba^2 \\
ba^{13} & a^2ba^6 & a^7ba^6 &
\end{array}$$

Let us now show that the preceding decomposition is good. Indeed, for any $\pi \in PBD$, setting $\pi(a) = p$, one has:

$$\pi(S) = \pi(S_1) + \pi(S_2) = \pi(S_1) + 1 - p = p\pi(Z) + 1 - p.$$

Since S is a code $\pi(S) \leq 1$, so that $\pi(Z) \leq 1$ which proves that the decomposition is good. □

Let us observe that any good decomposition $(S_1', S_2'; Y_1', Y_2')$ of S has to be such that $S_2' = S_2 = \{b\}$ and $Y_2' = Y_2 = \{\epsilon\}$. Indeed, if $Y_2 \subset Y_2'$, then $\pi(Y_2') > 1$ which is a contradiction. This implies $S_1' = S_1$ and $Y_1' \sim Y_1 \sim Z$.

Proposition 6.2. *There exists a code in a two letter alphabet which has no admissible decomposition.*

Proof. Let $T \subseteq a^*ba^*$ be a code of minimal size such that T is not commutatively prefix. Such a code exists since Shor's code $S \subseteq a^*ba^*$ is not commutatively prefix. Suppose that T has an admissible decomposition $(T_1, T_2; Y_1, Y_2)$. One would have

$$T_1 \sim aY_1 \quad \text{and} \quad T_2 \sim bY_2 \ ,$$

where Y_1 and Y_2 are codes or $Y_1 = \{\epsilon\}$ or $Y_2 = \{\epsilon\}$. At least one of the two sets Y_1 and Y_2 is $\neq \{\epsilon\}$, otherwise $T = \{a, b\}$. Let us suppose that $Y_1 \neq \{\epsilon\}$. Since $Y_1 \subseteq a^*ba^*$ and $\|Y_1\| < \|T\|$, one has that Y_1 is commutatively prefix. If $Y_2 \neq \{\epsilon\}$, then $Y_2 \subseteq a^*$, so that $Y_2 = \{a^i\}$, with $i > 0$ since Y_2 is a code. Thus, Y_2 is a prefix code. Hence, T has a normal decomposition that implies by Proposition 5.4 that T is commutatively prefix which is a contradiction. □

It would be interesting to prove that Shor's code is the code of minimal size included in a^*ba^* which is not commutatively prefix. Shor's code is not maximal and it is unknown if it is included in a finite maximal code on the alphabet $\{a, b\}$. However, Shor's code, as well as any finite code in a two-letter alphabet, is always included in a finite Bernoulli set [2].

It is still open the following problem which was formulated as a conjecture by Schützenberger at the end of '50s [1,6]:

Conjecture 6.3. Any finite and maximal code is commutatively prefix.

The following proposition relates Schützenberger's conjecture with decompositions of finite maximal codes:

Proposition 6.4. *The following statements are equivalent:*

1. *The Schützenberger conjecture is true.*
2. *Any finite and maximal code has a normal decomposition.*
3. *Any finite and maximal code has an admissible decomposition.*

Proof. 1.⇒ 2. By Proposition 5.4.

2.⇒ 3. Trivial, since any normal decomposition is an admissible decomposition.

3.⇒ 1. Suppose that the conjecture of Schützenberger is false. There would exist a finite maximal code on a suitable d-letter alphabet A which is not commutatively prefix. Let X be a finite and maximal code over A of minimal size which is not commutatively prefix. By Proposition 2.1, X is a Bernoulli set. Moreover, by hypothesis X has an admissible decomposition $(X_1, \ldots, X_d; Y_1, \ldots, Y_d)$ where for all i, $1 \leq i \leq d$, the sets $Y_i \neq \{\epsilon\}$ are maximal codes by Proposition 5.2. Since for all i, $1 \leq i \leq d$, $\|Y_i\| < \|X\|$ these sets are commutatively prefix. Therefore, the admissible decomposition of X is normal. By Proposition 5.4, X is commutatively prefix which is a contradiction. □

References

[1] J. Berstel and D. Perrin, *Theory of Codes.* Academic Press, New York (1985).

[2] A. Carpi and A. de Luca, Completions in measure of languages and related combinatorial problems. *Theor. Comput. Sci.* To appear.

[3] C. de Felice, On the triangle conjecture. *Inform. Process. Lett.* **14** (1982) 197–200.

[4] A. de Luca, Some combinatorial results on Bernoulli sets and codes. *Theor. Comput. Sci.* **273** (2002) 143–165.

[5] G. Hansel, Baïonnettes et cardinaux. *Discrete Math.* **39** (1982) 331–335.

[6] D. Perrin and M.P. Schützenberger, Un problème élémentaire de la théorie de l'Information, in *Theorie de l'Information*, Colloq. Internat. du CNRS No. 276, Cachan (1977) 249–260.

[7] D. Perrin and M.P. Schützenberger, A conjecture on sets of differences on integer pairs. *J. Combin. Theory Ser. B* **30** (1981) 91–93.

[8] J.E. Pin and I. Simon, A note on the triangle conjecture. *J. Comb. Theory Ser. A* **32** (1982) 106–109.

[9] P. Shor, A counterexample to the triangle conjecture. *J. Comb. Theory Ser. A* **38** (1985) 110–112.

To access this journal online:
www.edpsciences.org