RAIRO-Theor. Inf. Appl. 50 (2016) 81–91
DOI: 10.1051/ita/2016008

Available online at:
www.rairo-ita.org

CROSS-BIFIX-FREE SETS GENERATION *VIA* MOTZKIN PATHS *

Elena Barcucci[1], Stefano Bilotta[1], Elisa Pergola[1], Renzo Pinzani[1] and Jonathan Succi[1]

Abstract. Cross-bifix-free sets are sets of words such that no proper prefix of any word is a proper suffix of any other word. In this paper, we introduce a general constructive method for the sets of cross-bifix-free q-ary words of fixed length. It enables us to determine a cross-bifix-free words subset which has the property to be non-expandable.

Mathematics Subject Classification. 68R05, 68P30.

1. Introduction

A *cross-bifix-free set* of words (also known as *non-overlapping code*) is a set where, given any two elements of the set, possibly the same, any prefix of the first one is not a suffix of the second one and *vice versa* (from now on, by abuse of language, we will use the term prefix and suffix instead of proper prefix and proper suffix, respectively). Cross-bifix-free sets are involved in the study of frame synchronization which is an essential requirement in digital communication systems to establish and maintain a connection between a transmitter and a receiver.

Analytical approaches to the synchronization acquisition process and methods for the construction of sequences with the best aperiodic autocorrelation properties [6, 13, 16, 20] have been the subject of numerous analyses in the digital transmission.

The historical engineering approach started with the introduction of bifix, a name proposed by Massey as acknowledged in [17]. It denotes a factor that is both a prefix and suffix of a longer observed sequence.

In [13] the notion of a *distributed sequence* is introduced, where the synchronization word is not a contiguous sequence of symbols but is instead interleaved into the data stream. In [4] it is shown that the distributed sequence entails a simultaneous search for a set of synchronization words. Each word in the set of sequences is required to be bifix-free, moreover no prefix of any length of any word in the set is a suffix of any other word in the set. This property of the set of synchronization words was termed as *cross-bifix-free*.

Keywords and phrases. Codes, Motzkin paths.

* *This work has been partially supported by the PRIN project "Automi e linguaggi formali: aspetti matematici ed applicativi" and GNCS project "Strutture discrete con vincoli".*

[1] Dipartimento di Matematica e Informatica "U.Dini", Università degli Studi di Firenze, Viale G.B. Morgagni 65, 50134 Firenze, Italy. elena.barcucci@unifi.it; stefano.bilotta@unifi.it; elisa.pergola@unifi.it; renzo.pinzani@unifi.it; jons10@hotmail.it

The problem of determining such sets is also related to several other scientific applications, for instance in pattern matching [12], automata theory [7] and pattern avoidance theory [8].

Several methods for constructing cross-bifix-free sets have been recently proposed as in [2,9,11]. In particular, once the cardinality q of the alphabet and the length n of the words are fixed, a matter is the construction of a cross-bifix-free set with the cardinality as large as possible. An interesting method has been proposed in [2] (see also [3]) for words on a binary alphabet. This specific construction reveals interesting connections to the Fibonacci sequence of numbers. In a recent paper [11] the authors revisit the construction in [2] and generalize it obtaining cross-bifix-free sets having greater cardinality over an alphabet of any size q. They also show that their cross-bifix-free sets have a cardinality close to the maximum possible. To our knowledge this is the best result in the literature about the greatest size of cross-bifix-free sets. See also [10] for the "optimal cardinality whenever n divides q".

For the sake of completeness we note that an intermediate step between the original method [2] and its generalization [11] has been proposed in [9] and it is constituted by a different construction of binary cross-bifix-free sets based on lattice paths which allows to obtain greater values of cardinality if compared to the ones in [2].

In this study, we revisit the construction in [9]. We give a new construction of cross-bifix-free sets that generalizes the construction in [9] to q-ary alphabets, for each $q > 2$, by means of some particular lattice paths in the discrete plane called k-colored Motzkin paths [5]. This approach enables us to obtain cross-bifix-free sets having greater cardinality than the ones presented in [11], for the initial values of n. This new result extends the theory of cross-bifix-free sets and it could be used to improve some technical applications.

This paper is organized as follows. In Section 2 we give some preliminaries and describe the adopted notation. In Section 3 we present a new construction of cross-bifix-free sets in the q-ary alphabet and in Section 4 we analyze the sizes of the sets of our construction in comparison to the ones in the literature.

2. Basic definitions and notations

Let $\mathbb{Z}_q = \{0, 1, \ldots, q-1\}$ be an alphabet of q elements. A (finite) sequence of elements in $\mathbb{Z}_q$ is called (finite) *word*. The set of all words over $\mathbb{Z}_q$ having length n is denoted by $\mathbb{Z}_q^n$. A consecutive sequence of m element $a \in \mathbb{Z}_q$ is denoted by the short form a^m. Let $w \in \mathbb{Z}_q^n$, then $|w|_a$ denotes the number of occurrences of a in w, being $a \in \mathbb{Z}_q$, and $|w| = n$. Let $w = uzv$ then u is called a *prefix* of w and v is called a *suffix* of w. A *bifix* of w is a factor of w that is both its prefix and suffix. We recall that, for any word w we only consider prefixes and suffixes that are proper, that is, which have length strictly less than the length of w.

A word $w \in \mathbb{Z}_q^n$ is said to be *bifix-free* or *unbordered* [18] if and only if no prefix of w is also a suffix of w. Therefore, w is bifix-free if and only if $w = uzu$ implies that u is empty word. Obviously, a necessary condition for w to be bifix-free is that the first and the last letters of w must be different.

Example 2.1. In $\mathbb{Z}_2 = \{0, 1\}$, the word 111010100 of length $n = 9$ is bifix-free, while the word 101001010 contains two bifixes, 10 and 1010.

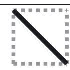
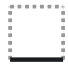
Let $\mathrm{BF}_q(n)$ denote the set of all bifix-free words of length n over an alphabet of fixed size q (for more details about this topic see [18]).

Given $q > 1$ and $n > 1$, two distinct words $w, w' \in \mathrm{BF}_q(n)$ are said to be *cross-bifix-free* [4] if and only if no strict prefix of w is also a suffix of w' and *vice versa*.

Example 2.2. The binary words 111010100 and 110101010 in $\mathrm{BF}_2(9)$ are cross-bifix-free, while the binary words 111001100 and 110011010 in $\mathrm{BF}_2(9)$ have the cross-bifix 1100.

A subset of $\mathrm{BF}_q(n)$ is said to be a *cross-bifix-free set* if and only if for each w, w', with $w \neq w'$, in this set, w and w' are cross-bifix-free. This set is said to be *non-expandable* on $\mathrm{BF}_q(n)$ if and only if the set obtained by adding any other word in $\mathrm{BF}_q(n)$ is not a cross-bifix-free set. The set having maximal cardinality is called a *maximal cross-bifix-free set* (*optimal non-overlapping code*) on $\mathrm{BF}_q(n)$.

TABLE 1. Equivalence between symbols and steps for $\mathbb{Z}_3 = \{0, 1, 2\}$.

Symbol	Step	Color	Representation
0	*fall*	–	
1	*rise*	–	
2	*level*	Black	

Let $C(n, q)$ denote the cardinality of the maximal cross-bifix-free set of length n over an alphabet of size q. In [14], it is proven that

$$C(n, q) \leq \frac{1}{n}\left(\frac{n-1}{n}\right)^{n-1} q^n. \tag{2.1}$$

In a recent paper [11] the authors provide a general construction of cross-bifix-free sets over a q-ary alphabet. Below, we recall such generation for the family of cross-bifix-free sets in $\mathbb{Z}_q^n$.

For any $2 \leq k \leq n-2$, the cross-bifix-free set $\mathcal{S}_{k,q}(n)$ in [11] is the set of all words $s = s_1 s_2 \dots s_n$ in $\mathbb{Z}_q^n$ that satisfy the following two properties:

1) $s_1 = \cdots = s_k = 0$, $s_{k+1} \neq 0$ and $s_n \neq 0$;
2) the factor $s_{k+2} \dots s_{n-1}$ does not contain k consecutive 0's.

Let

$$F_{k,q}(n) = \begin{cases} q^n & \text{if } 0 \leq n < k, \\ (q-1)\sum_{l=1}^{k} F_{k,q}(n-l) & \text{if } n \geq k, \end{cases} \tag{2.2}$$

be the sequence enumerating the words in $\mathbb{Z}_q^n$ avoiding k consecutive zero's [15]. Then, from the above definition of $\mathcal{S}_{k,q}(n)$, we have

$$|\mathcal{S}_{k,q}(n)| = (q-1)^2 F_{k,q}(n-k-2). \tag{2.3}$$

For any fixed n and q, the largest size of $|\mathcal{S}_{k,q}(n)|$ is denoted by $S(n, q)$ and it is given by the following expression as in [11]

$$S(n, q) = \max_{k=2,\dots,n-2} |\mathcal{S}_{k,q}(n)|. \tag{2.4}$$

This result allows to obtain non-expandable cross-bifix-free sets in the q-ary alphabet having cardinality close to the maximum.

In the present paper we introduce an alternative constructive method for the generation of cross-bifix-free set in $\mathbb{Z}_q$. Our approach is based on the study of lattice paths in the discrete plane and it moves from the construction in [9].

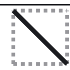
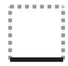
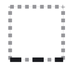
Each word $w \in \mathbb{Z}_q^n$ can be represented as a lattice path of $\mathbb{N}^2$ running from $(0, 0)$ to (n, h), with $-n \leq h \leq n$, having the following properties:

– the element 0 corresponds to a *fall step* running from (x, y) to $(x+1, y-1)$;
– the element 1 corresponds to a *rise step* running from (x, y) to $(x+1, y+1)$;
– the elements $2, \dots, q-1$ correspond respectively to a *colored level step* running from (x, y) to $(x+1, y)$ and it is labeled by one of the $q-2$ fixed colors.

For example, Tables 1 and 2 show an equivalence between elements and steps of lattice paths in the alphabets $\mathbb{Z}_3$ and $\mathbb{Z}_4$, respectively.

From now on, we will refer interchangeably to words or their graphical representations on the discrete plane, that are paths. The definition of bifix-free and cross-bifix-free can be easily extended to paths.

TABLE 2. Equivalence between symbols and steps for $\mathbb{Z}_4 = \{0, 1, 2, 3\}$.

Symbol	Step	Color	Representation
0	*fall*	–	
1	*rise*	–	
2	*level*	Black	
3	*level*	Red	

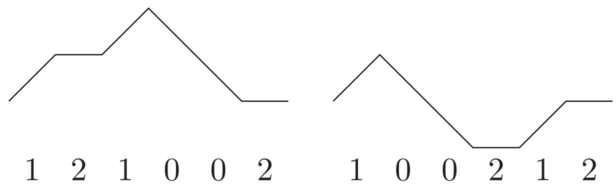


FIGURE 1. Words 121002, 100212 and the equivalent paths. The first one is a Motzkin word.

A *k-colored Motzkin path* of length n is a lattice path of $\mathbb{N}^2$ running from $(0, 0)$ to $(n, 0)$ that never goes below the x-axis and whose admitted steps are rise steps, fall steps and k-colored level steps (for more details about this topic see [5]).

For example, the left side of Figure 1 shows a Motzkin path in $\mathbb{Z}_3$ having length 6, while the path in its right side is not a Motzkin path since it crosses the x-axis.

We denote by $\mathcal{M}_k(n)$ the set of all k-colored Motzkin paths of length n, and let $M_k(n)$ be the size of $\mathcal{M}_k(n)$.

The following proposition can be easily generalized from the recurrence of the Motzkin numbers in [1] (case $k = 1$).

Proposition 2.3. *For any $n \geq 0$ and $k \geq 1$, $M_k(n)$ is given by the following expression*

$$M_k(n+1) = kM_k(n) + \sum_{i=0}^{n-1} M_k(i)M_k(n-1-i) \tag{2.5}$$

with $M_k(0) = 1$ and $M_k(1) = k$.

In [19], a generating function for $M_k(n)$ is derived as:

$$M_k(x) = \sum_{n\geq 0} M_n(k)x^n = \frac{1 - kx - \sqrt{(1-kx)^2 - 4x^2}}{2x^2}, \tag{2.6}$$

and the following formula, which is related to the well-known Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n}$ with $n \geq 0$, is also presented

$$M_k(n) = \sum_{r=0}^{[\frac{n}{2}]} \binom{n}{2r} C_r k^{n-2r}. \tag{2.7}$$

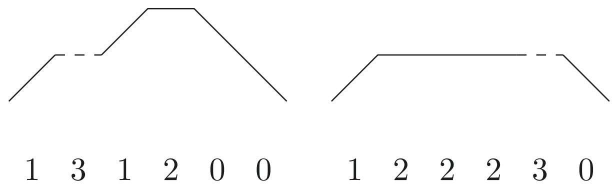


FIGURE 2. An example of elevated 2-colored Motzkin words.

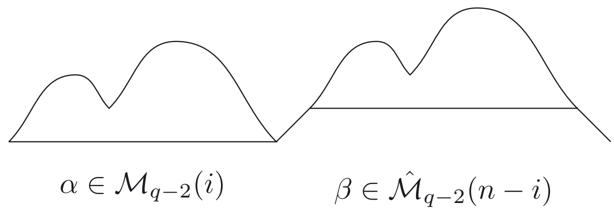


FIGURE 3. Graphical representation of the set $\mathcal{A}_q(n)$, $n \geq 3$.

A word $w \in \mathbb{Z}_q^n$ is called $(q-2)$-*colored Motzkin word* if the equivalent lattice path is a $(q-2)$-colored Motzkin path.

For our purposes, it is useful to denote by $\hat{\mathcal{M}}_{q-2}(n)$ the set of all *elevated* $(q-2)$-*colored Motzkin words* of length n, defined as:

$$\hat{\mathcal{M}}_{q-2}(n) = \{1\alpha 0 : \alpha \in \mathcal{M}_{q-2}(n-2)\}.$$

For example, in Figure 2 two words in $\hat{\mathcal{M}}_2(6)$ are depicted.

In the next section of the present paper we are interested in determining one among all the possible non-expandable cross-bifix-free sets of words of fixed length $n > 1$ on $\mathbb{Z}_q^n$ by means of $(q-2)$-colored Motzkin words. We denote this set by $\mathrm{CBFS}_q(n)$.

3. On the non-expandability of $\mathrm{CBFS}_q(n)$

In this section we define the set $\mathrm{CBFS}_q(n)$, with $q \geq 3$ and $n \geq 3$, which is formed by the union of three disjoint sets: a set of $(q-2)$-colored Motzkin paths of length n denoted by $\mathcal{A}_q(n)$, a set of paths, denoted by $\mathcal{B}_q(n)$, formed by a rise step followed by a $(q-2)$-colored Motzkin path of length $n-1$, and a set of paths, denoted by $\mathcal{C}_q(n)$, formed by a $(q-2)$-colored Motzkin path of length $n-1$ followed by a fall step.

Let

$$\mathcal{A}_q(n) = \bigcup_{0 \leq i \leq \lfloor \frac{n}{2} \rfloor} \left\{\alpha\beta : \alpha \in \mathcal{M}_{q-2}(i),\, \beta \in \hat{\mathcal{M}}_{q-2}(n-i)\right\} \setminus \left\{\alpha\beta : \alpha, \beta \in \hat{\mathcal{M}}_{q-2}\left(\frac{n}{2}\right)\right\}$$

be the set of words composed by a $(q-2)$-colored Motzkin word α of length i, and a elevated $(q-2)$-colored Motzkin word β of length $n-i$ (see Fig. 3). If n is even, we need to remove the words composed by two elevated subwords of the same length. On the other side, if n is odd, we assume the set $\left\{\alpha\beta : \alpha, \beta \in \hat{\mathcal{M}}_{q-2}\left(\frac{n}{2}\right)\right\}$ empty, since it does not exists any path of non-integer length.

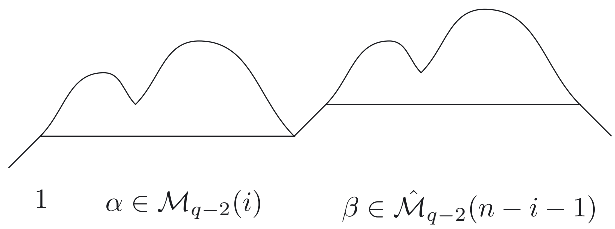


FIGURE 4. Graphical representation of the set $\mathcal{B}_q(n)$, $n \geq 3$.

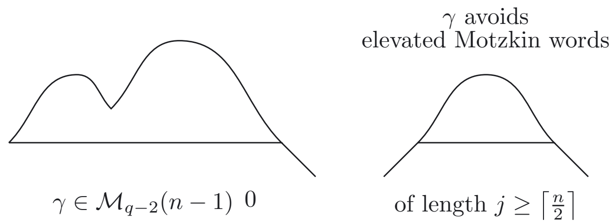


FIGURE 5. Graphical representation of the set $\mathcal{C}_q(n)$, $n \geq 3$.

Then, the enumeration of the set $\mathcal{A}_q(n)$ is given by the following expression

$$|\mathcal{A}_q(n)| = \left(\sum_{i=0}^{\lfloor n/2 \rfloor} M_{q-2}(i) M_{q-2}(n-i-2) \right) - \left(M_{q-2}\left(\frac{n}{2}-2\right)\right)^2. \tag{3.1}$$

Let

$$\mathcal{B}_q(n) = \bigcup_{0 \leq i \leq \lfloor \frac{n}{2} \rfloor - 1} \left\{ 1\alpha\beta : \alpha \in \mathcal{M}_{q-2}(i),\ \beta \in \hat{\mathcal{M}}_{q-2}(n-i-1) \right\}$$

be the set of words composed by a rise step, a $(q-2)$-colored Motzkin word α of length i, and a elevated $(q-2)$-colored Motzkin word β of length $n-i-1$ (see Fig. 4).

Then, the enumeration of the set $\mathcal{B}_q(n)$ is given by the following expression

$$|\mathcal{B}_q(n)| = \sum_{i=0}^{\lfloor n/2 \rfloor - 1} M_{q-2}(i) M_{q-2}(n-i-3). \tag{3.2}$$

Let

$$\mathcal{C}_q(n) = \left\{ \gamma 0 : \gamma \in \mathcal{M}_{q-2}(n-1),\ \gamma \neq u\beta v,\ \beta \in \hat{\mathcal{M}}_{q-2}(j),\ j \geq \left\lceil \frac{n}{2} \right\rceil \right\}$$

be the set of words composed by a $(q-2)$-colored Motzkin word γ of length $n-1$ that avoids elevated $(q-2)$-colored Motzkin words of length j, and a fall step (see Fig. 5).

Then, the enumeration of the set $\mathcal{C}_q(n)$ is given by the following expression

$$|\mathcal{C}_q(n)| = M_{q-2}(n-1) - \sum_{k=\lceil n/2 \rceil}^{n-1} \sum_{i=0}^{n-1-k} M_{q-2}(i) M_{q-2}(k-2) M_{q-2}(n-1-i-k). \tag{3.3}$$

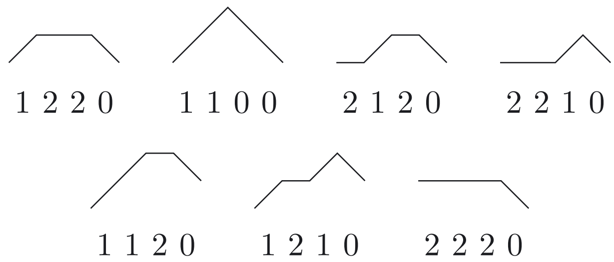


FIGURE 6. Graphical representation of the set $\mathrm{CBFS}_3(4)$.

Note that, in order to obtain the size $|\mathcal{C}_q(n)|$ we need to subtract from all words γ of length $n-1$ those containing a elevated Motzkin subword β of length greater than or equal to $\lceil n/2 \rceil$, and γ can contain one of those subwords at most. Then, for $k = \lceil n/2 \rceil, \ldots, n-1$ we need to remove the words $u\beta v$, with $u \in \mathcal{M}_{q-2}(i)$, $\beta \in \hat{\mathcal{M}}_{q-2}(k)$, $v \in \mathcal{M}_{q-2}(n-1-i-k)$ and $0 \le i \le n-1-k$.

At this point, we define the set $\mathrm{CBFS}_q(n)$ as follows

$$\mathrm{CBFS}_q(n) = \mathcal{A}_q(n) \cup \mathcal{B}_q(n) \cup \mathcal{C}_q(n)$$

that is the union of the above described sets. For instance, in Figure 6 the set $\mathrm{CBFS}_3(4)$ is depicted, where $\mathcal{A}_3(4) = \{1220, 1100, 2120, 2210\}$, $\mathcal{B}_3(4) = \{1120, 1210\}$ and $\mathcal{C}_3(4) = \{2220\}$.

Proposition 3.1. *The set* $\mathrm{CBFS}_q(n)$ *is a cross-bifix-free set on* $\mathrm{BF}_q(n)$*, for any* $q \ge 3$ *and* $n \ge 3$.

Proof. Let $w, w' \in \mathrm{CBFS}_q(n)$. Let u be a prefix of w, and v be a suffix of w' such that $|u| = |v|$. We need to check that in each case the prefix u does not match with the suffix v.

(1) Let $w \in \mathcal{A}_q(n)$ and $w' \in \mathcal{A}_q(n) \cup \mathcal{B}_q(n)$.
For each prefix u of w we have $|u|_0 \le |u|_1$ and if $|u| > \lfloor \frac{n}{2} \rfloor$, then $|u|_0 < |u|_1$. For each suffix v of w' we have $|v|_0 \ge |v|_1$ and if $|v| < \lfloor \frac{n+1}{2} \rfloor$, then $|v|_0 > |v|_1$.
Let $|u| = |v| = \ell$, if either $\ell < \lfloor \frac{n+1}{2} \rfloor$ or $\ell > \lfloor \frac{n}{2} \rfloor$, then u does not match with v. So we have to check the case $\lfloor \frac{n+1}{2} \rfloor \le \ell \le \lfloor \frac{n}{2} \rfloor$.
If n is odd, there does not exist an integer ℓ satisfying $\lfloor \frac{n+1}{2} \rfloor \le \ell \le \lfloor \frac{n}{2} \rfloor$, otherwise if n is even, the case $\lfloor \frac{n+1}{2} \rfloor \le \ell \le \lfloor \frac{n}{2} \rfloor$ is verified only for $\ell = \frac{n}{2}$. Therefore let n be even and $\ell = \frac{n}{2}$. In this case $|u|_0 \le |u|_1$ and $|v|_0 \ge |v|_1$. At this point u can match with v only if $|v|_0 = |v|_1$, and this can happen only if v is a elevated Motzkin word. Suppose now that $u = v$, so u should be a elevated Motzkin word too, and they have both length $\frac{n}{2}$. In this case, w should be a word composed of two elevated Motzkin subwords of the same length, but such a word does not exist in $\mathrm{CBFS}_q(n)$ since the set $\left\{\alpha\beta : \alpha, \beta \in \hat{\mathcal{M}}_{q-2}\left(\frac{n}{2}\right)\right\}$ is not included in it, thus u does not match with v.

(2) Let $w \in \mathcal{B}_q(n)$ and $w' \in \mathcal{A}_q(n) \cup \mathcal{B}_q(n)$.
For each prefix u of w we have $|u|_0 < |u|_1$, and for each suffix v of w' we have $|v|_0 \ge |v|_1$, thus u does not match with v.

(3) Let $w \in \mathcal{C}_q(n)$ and $w' \in \mathcal{A}_q(n) \cup \mathcal{B}_q(n)$.
For each prefix u of w we have $|u|_0 \le |u|_1$. For each suffix v of w' we have $|v|_0 \ge |v|_1$ and if $|v| < \lfloor \frac{n+1}{2} \rfloor$, then $|v|_0 > |v|_1$.
Let $|u| = |v| = \ell$. If $\ell < \lfloor \frac{n+1}{2} \rfloor$, then u does not match with v. So we have to check the case $\ell \ge \lfloor \frac{n+1}{2} \rfloor$. In this case v contains a elevated Motzkin subword of length $\lfloor \frac{n+1}{2} \rfloor = \lceil \frac{n}{2} \rceil$ at least, and u does not match with v, since u avoids such subwords.

(4) Let $w \in \mathrm{CBFS}_q(n)$ and $w' \in \mathcal{C}_q(n)$.

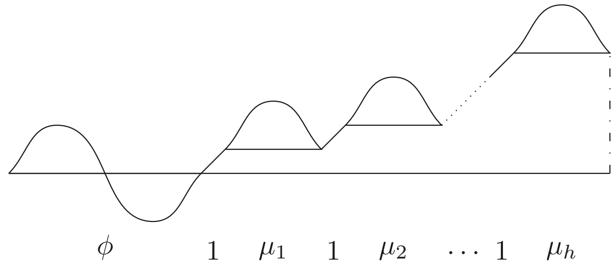


FIGURE 7. Graphical representation of w, in the case $h > 0$.

For each prefix u of w we have $|u|_0 \leq |u|_1$, and for each suffix v of w' we have $|v|_0 > |v|_1$, thus u cannot match with v.

We proved that $\mathrm{CBFS}_q(n)$ is a cross-bifix-free set on $\mathrm{BF}_q(n)$, for any $q \geq 3$ and $n \geq 3$. □

Proposition 3.2. *The set* $\mathrm{CBFS}_q(n)$ *is a non-expandable cross-bifix-free set on* $\mathrm{BF}_q(n)$*, for any* $q \geq 3$ *and* $n \geq 3$.

Proof. Let $w \in \mathrm{BF}_q(n) \setminus \mathrm{CBFS}_q(n)$ and $W = \mathrm{CBFS}_q(n) \cup \{w\}$. If w begins with 0 then W is not cross-bifix-free since any word in $\mathrm{CBFS}_q(n)$ ends with 0. If w ends with 1 then W is not cross-bifix-free since any word in $\mathcal{A}_q(n)$ begins with 1. If w ends with a letter $k \neq 0, 1$ then W is not cross-bifix-free since the suffix k of w matches, for instance, with the prefix k of the word $k^{n-1}0 \in \mathcal{C}_q(n)$. Consequently we have to consider w as a word beginning with a non-zero letter and ending with 0.

Let $h = |w|_1 - |w|_0$ be the ordinate of the last point of the path corresponding to w. We now need to distinguish three different cases: $h > 0$, $h < 0$ and $h = 0$.

If $h > 0$, w can be written as (see Fig. 7)

$$w = \phi \; 1 \; \mu_1 \; 1 \; \mu_2 \; \ldots \; 1 \; \mu_h,$$

where ϕ is a word satisfying $|\phi|_1 = |\phi|_0$ and not beginning with 0, and $\mu_1, \ldots, \mu_h$ are $(q-2)$-colored Motzkin words with μ_h non-empty as w ends with 0.

In this case, if $|\mu_h| = \ell \leq n-2$, considering for instance the word $u = 1\mu_h 2^{n-\ell-2}0 \in \mathcal{A}_q(n)$ we can clearly see that $1\mu_h$ is a cross-bifix between w and u, and then W is not cross-bifix-free. On the other hand, if $|\mu_h| = n-1$, then necessarily $h = 1$ and $w = 1\mu_1$. So, w can be written as $w = 1\alpha\beta$, where $\alpha \in \mathcal{M}_{q-2}(i)$, $\beta \in \hat{\mathcal{M}}_{q-2}(n-i-1)$ with $i > \lfloor \frac{n}{2} \rfloor$ (otherwise $w \in \mathcal{B}_q(n)$). In this case, for instance, the word $\beta 12^{i-1}0 \in \mathcal{A}_q(n)$ has a cross-bifix with w, thus W is not a cross-bifix-free-set.

If $h < 0$, w can be written as (see Fig. 8)

$$w = \mu_{-h} \; 0 \; \ldots \; \mu_2 \; 0 \; \mu_1 \; 0 \; \phi$$

where ϕ is a word satisfying $|\phi|_1 = |\phi|_0$ and ending with 0, and $\mu_1, \ldots, \mu_{-h}$ are $(q-2)$-colored Motzkin words with μ_{-h} non-empty as w begins with a non-zero letter.

In this case, if $|\mu_{-h}| = \ell \leq n-2$, considering for instance the word $u = 12^{n-\ell-2}\mu_{-h}0 \in \mathcal{A}_q(n)$ we can clearly see that $\mu_{-h}0$ is a cross-bifix between w and u, and then W is not cross-bifix-free. On the other hand, if $|\mu_{-h}| = n-1$, then necessarily $h = -1$ and $w = \mu_1 0$. So, w can be written as $w = \alpha\beta\delta 0$, where $\beta \in \hat{\mathcal{M}}_{q-2}(j)$ with $j \geq \lceil \frac{n}{2} \rceil$ (otherwise $w \in \mathcal{C}_q(n)$), and α, δ any two $(q-2)$-colored Motzkin words of the appropriate length. In this case, for instance, the word $2^{n-j-|\alpha|}\alpha\beta \in \mathcal{A}_q(n)$ has a cross-bifix with w, thus W is not a cross-bifix-free-set.

Finally, if $h = 0$, the path associated to w can either remain above x-axis or fall below it.

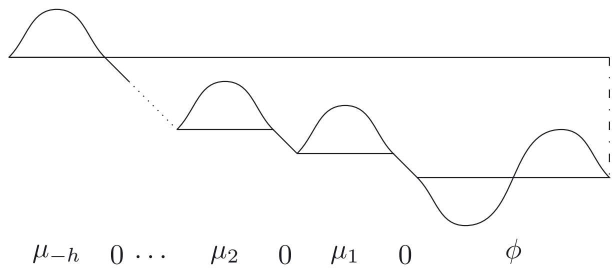


FIGURE 8. Graphical representation of w, in the case $h < 0$.

In the first case let i, with $\lfloor \frac{n}{2} \rfloor \leq i < n$, be the last x-coordinate of the path intercepting the x-axis. Notice that i can not be less than $\lfloor \frac{n}{2} \rfloor$, otherwise $w \in \mathcal{A}_q(n)$. We can write $w = \alpha\beta$, where α is a non-empty word in $\mathcal{M}_{q-2}(i)$ and $\beta \in \hat{\mathcal{M}}_{q-2}(n-i)$. We now need to take into consideration two different cases: $i = \lfloor \frac{n}{2} \rfloor$ and $i > \lfloor \frac{n}{2} \rfloor$. If $i = \lfloor \frac{n}{2} \rfloor$ then $\alpha \in \hat{\mathcal{M}}_{q-2}(\frac{n}{2})$, otherwise $w \in \mathcal{A}_q(n)$, so for instance, the word $2^{n/2}\alpha \in \mathcal{A}_q(n)$ has a cross-bifix with w. If $i > \lfloor \frac{n}{2} \rfloor$ then, for instance, the word $\beta 2^{i-1} 0 \in \mathcal{C}_q(n)$ has a cross-bifix with w, so that W is not a cross-bifix-free-set.

In the other case the path associated to w crosses the x-axis. Let i, with $0 < i < n$, be the first x-coordinate of the path crossing x-axis. We can write $w = \alpha 0 \phi$, where α is a non-empty word in $\mathcal{M}_{q-2}(i)$. In this case, for instance, the word $12^{n-i-2}\alpha 0 \in \mathcal{A}_q(n)$ has a cross-bifix with w, then W is not a cross-bifix-free-set.

We proved that $\mathrm{CBFS}_q(n)$ is a non-expandable cross-bifix-free set on $\mathrm{BF}_q(n)$, for any $q \geq 3$ and $n \geq 3$. □

4. Sizes of cross-bifix-free sets for small lengths

In this section we present some results concerning the size of $\mathrm{CBFS}_q(n)$ compared to the ones in [11].

For fixed n and q, we recall that the size of q-ary cross-bifix-free sets given in [11] is obtained by

$$S(n,q) = \max\{(q-1)^2 F_{k,q}(n-k-2) : 2 \leq k \leq n-2\} \tag{4.1}$$

which is proved to be nearly optimal.

In Table 3 the values of $|\mathrm{CBFS}_q(n)|$ and $S(n,q)$ for $3 \leq q \leq 6$ and $n \leq 16$ are shown. For the initial values of n, we can observe that the sizes obtained by our construction are greater than the size $S(n,q)$. In particular, the number of the initial values of n, for which $|\mathrm{CBFS}_q(n)|$ is greater than $S(n,q)$, grows together with q and this trend can be easily verified by experimental results.

In order to improve the values of the size $S(n,q)$ for the initial size of n, we can consider the following expression

$$S^*(n,q) = \max\{(q-1)^2 F_{k,q}(n-k-2) : 1 \leq k \leq n-2\}, \tag{4.2}$$

where k can assume also the value 1. When $k = 1$, in the case of small n and large q, we obtain cross-bifix-free sets having cardinality greater than the one proposed in [11]. A similar argument is also discussed in [10], where a construction giving the maximal cardinality $C(n,q) = \frac{1}{n}(\frac{n-1}{n})^{n-1}q^n$ is presented when n divides q. Such a particular case requires that the size q of the alphabet must be greater than the length n of the words, whereas $S^*(n,q)$ gives an exact cardinality for all possible values of q and n, with $n, q > 2$.

Table 4 shows the values of $|\mathrm{CBFS}_q(n)|$, $S^*(n,q)$ and $C(n,q)$ for $3 \leq q \leq 6$ and $n \leq 16$. Also in this situation, we can observe that the sizes obtained by our construction are greater than the size $S^*(n,q)$ in a range of values of n. In particular, the range of values of n, for which $|\mathrm{CBFS}_q(n)|$ is greater than $S^*(n,q)$, grows together with q and this trend can be easily verified by experimental results.

Table 3. Comparing the values from [11] with $\mathrm{CBFS}_q(n)$, for $3 \le q \le 6$.

n	$\lvert\mathrm{CBFS}_3(n)\rvert$	$S(n,3)$	$\lvert\mathrm{CBFS}_4(n)\rvert$	$S(n,4)$
3	4	4	9	9
4	**7**	4	**25**	9
5	**16**	12	**72**	36
6	**36**	32	**223**	135
7	87	88	**712**	513
8	210	240	**2 334**	1 944
9	535	656	**7 868**	7 371
10	1 350	1 792	26 731	27 945
11	3 545	4 896	93 175	105 948
12	9 205	13 376	324 520	401 679
13	24 698	36 544	1 157 031	1 522 881
14	65 467	99 840	4 104 449	5 773 680
15	178 375	272 768	14 874 100	21 889 683
16	480 197	745 216	53 514 974	82 990 089

n	$\lvert\mathrm{CBFS}_5(n)\rvert$	$S(n,5)$	$\lvert\mathrm{CBFS}_6(n)\rvert$	$S(n,6)$
3	16	16	25	25
4	**61**	16	**121**	25
5	**224**	80	**550**	150
6	**900**	384	**2 739**	875
7	**3 595**	1 856	**13 260**	5 125
8	**15 014**	8 960	**67 740**	30 000
9	**63 135**	43 264	**342 676**	175 625
10	**271 136**	208 896	**1 787 415**	1 028 125
11	**1 178 677**	1 008 640	**9 324 647**	6 018 750
12	**5 167 953**	4 870 144	**49 456 240**	35 234 375
13	22 986 100	23 515 136	**263 776 127**	206 265 625
14	102 403 229	113 541 120	**1 417 981 855**	1 207 500 000
15	463 098 075	548 225 024	**7 688 015 908**	7 068 828 125
16	2 089 302 415	2 647 064 576	**41 785 951 916**	41 381 640 625

Table 4. Comparing the values from $S^*(n,q)$ and $C(n,q)$ with $\mathrm{CBFS}_q(n)$, for $3 \le q \le 6$.

n	$\lvert\mathrm{CBFS}_3(n)\rvert$	$S^*(n,3)$	$C(n,3)$	$\lvert\mathrm{CBFS}_4(n)\rvert$	$S^*(n,4)$	$C(n,4)$
3	4	4	4	9	9	
4	7	8		25	27	27
5	16	16		72	81	
6	**36**	32		223	243	
7	87	88		712	729	
8	210	240		**2 334**	2 187	
9	535	656		**7 868**	7 371	
10	1 350	1 792		26 731	27 945	
11	3 545	4 896		93 175	105 948	
12	9 205	13 376		324 520	401 679	
13	24 698	36 544		1 157 031	1 522 881	
14	65 467	99 840		4 104 449	5 773 680	
15	178 375	272 768		14 874 100	21 889 683	
16	480 197	745 216		53 514 974	82 990 089	

n	$\lvert\mathrm{CBFS}_5(n)\rvert$	$S^*(n,5)$	$C(n,5)$	$\lvert\mathrm{CBFS}_6(n)\rvert$	$S^*(n,6)$	$C(n,6)$
3	16	16		25	25	32
4	61	64		121	125	
5	224	256	256	550	625	
6	900	1 024		2 739	3 125	3125
7	3 595	4 096		13 260	15 625	
8	15 014	16 384		67 740	78 125	
9	63 135	65 536		342 676	390 625	
10	**271 136**	262 144		1 787 415	1 953 125	
11	**1 178 677**	1 048 576		9 324 647	9 765 625	
12	**5 167 953**	4 870 144		**49 456 240**	48 828 125	
13	22 986 100	23 515 136		**263 776 127**	244 140 625	
14	102 403 229	113 541 120		**1 417 981 855**	1 220 703 125	
15	463 098 075	548 225 024		**7 688 015 908**	7 068 828 125	
16	2 089 302 415	2 647 064 576		**41 785 951 916**	41 381 640 625	

5. Conclusions and further developments

In this paper, we introduce a general constructive method for cross-bifix-free sets in the q-ary alphabet based upon the study of lattice paths on the discrete plane. This approach enables us to obtain the cross-bifix-free set $\mathrm{CBFS}_q(n)$ having greater cardinality than the ones proposed in [11], for the initial values of n.

Moreover, we prove that $\mathrm{CBFS}_q(n)$ is a non-expandable cross-bifix-free set on $\mathrm{BF}_q(n)$, *i.e.* $\mathrm{CBFS}_q(n) \cup \{w\}$ is not a cross-bifix-free set on $\mathrm{BF}_q(n)$, for any $w \in \mathrm{BF}_q(n) \backslash \mathrm{CBFS}_q(n)$.

The non-expandable property is obviously a necessary condition to obtain a maximal cross-bifix-free set on $\mathrm{BF}_q(n)$, anyway the problem of determine maximal cross-bifix-free sets is still open and no general solution has been found yet.

References

[1] M. Aigner, Motzkin numbers. *Eur. J. Combin.* **19** (1998) 663–675.
[2] D. Bajic, On construction of cross-bifix-free kernel sets, in *Proc. of Conference on 2nd MCM COST 2100* (2007) TD(07)237.
[3] D. Bajic and T. Loncar-Turukalo, A simple suboptimal construction of cross-bifix-free codes. *Cryptogr. Commun.* **6** (2014) 27–37.
[4] D. Bajic and J. Stojanovic, Distributed sequences and search process, in *Proc. of IEEE International Conference on Communications ICC 2004* (2004) 514–518.
[5] E. Barcucci, A. Del Lungo, E. Pergola and R. Pinzani, A construction for enumerating k-coloured Motzkin paths. *Lect. Notes Comput. Sci.* **959** (1995) 254–263.
[6] R.H. Barker, Group Synchronizing of Binary Digital Systems, Communication theory, London, Butterworth (1953) 273–287.
[7] J. Berstel, D. Perrin and C. Reutenauer, Codes and Automata. *Encycl. Math. Appl.* Cambridge University Press (2009).
[8] S. Bilotta, E. Grazzini, E. Pergola and R. Pinzani, Avoiding cross-bifix-free binary words. *Acta Inform.* **50** (2013) 157–173.
[9] S. Bilotta, E. Pergola and R. Pinzani, A new approach to cross-bifix-free sets. *IEEE Trans. Inform. Theory* **58** (2012) 4058–4063.
[10] S. Blackburn. Non-overlapping codes. *IEEE Trans. Inform. Theory* **61** (2015) 4890–4894.
[11] Y.M. Chee, H.M. Kiah, P. Purkayastha and C. Wang, Cross-bifix-free codes within a constant factor of optimality. *IEEE Trans. Inform. Theory* **59** (2013) 4668–4674.
[12] M. Crochemore, C. Hancart and T. Lecroq, Algorithms on Strings. Cambridge University Press (2007).
[13] A.J. De Lind Van Wijngaarden and T.J. Willink, Frame Synchronization Using Distributed Sequences. *IEEE Trans. Commun.* **48** (2000) 2127–2138.
[14] V.I. Levenshtein, Maximun number of words in codes without overlaps. *Probl. Inf. Transm.* **6** (1970) 355–357.
[15] C. Levesque, On mth order linear recurrences. *Fibonacci Quart.* **23** (1985) 290–293.
[16] J. Lindner, Binary sequences up to length 40 with best possible autocorrelation function. *Electron. Lett.* **11** (1975) 507.
[17] P.T. Nielsen, On the expected duration of a search for a fixed pattern in random data. *IEEE Trans. Inform. Theory* **29** (1973) 702–704.
[18] P.T. Nielsen. A note on bifix-free sequences. *IEEE Trans. Inform. Theory* **29** (1973) 704–706.
[19] A. Sapounakis and P. Tsikouras, On k-colored Motzkin words. *J. Integer Seq.* **7** (2004) 04.2.5.
[20] R.A. Scholtz, Frame synchronization techniques. *IEEE Trans. Commun.* **20** (1980) 1204–1213.

Communicated by D. Jamet.
Received March 24, 2016. Accepted March 24, 2016.