RAIRO-Theor. Inf. Appl. 46 (2012) 131–145
DOI: 10.1051/ita/2011116

Available online at:
www.rairo-ita.org

ON THE PRODUCT OF BALANCED SEQUENCES

Antonio Restivo[1] and Giovanna Rosone[1]

Abstract. The product $w = u \otimes v$ of two sequences u and v is a naturally defined sequence on the alphabet of pairs of symbols. Here, we study when the product w of two balanced sequences u, v is balanced too. In the case u and v are binary sequences, we prove, as a main result, that, if such a product w is balanced and $deg(w) = 4$, then w is an ultimately periodic sequence of a very special form. The case of arbitrary alphabets is approached in the last section. The partial results obtained and the problems proposed show the interest of the notion of product in the study of balanced sequences.

Mathematics Subject Classification. 68R15.

Introduction

The *balance* property is a fine tool in Combinatorics on Words and is used as a characteristic property of the well-known Sturmian sequences. Sturmian sequences were introduced in the forties by Morse and Hedlund (*cf.* [8]). They are defined as the infinite sequences having exactly $n+1$ distinct factors of length n. Their numerous properties have lead to a great development of many fields of research. Such a versatility explains also the existence of many equivalent definitions (see Chap. 2 of [7]). Furthermore, balanced sequences on more than two letters alphabets appear in the statement of the famous Fraenkel conjecture (*cf.* [5]) with links to Beatty sequences and to number theory (*cf.* [2,4,9]). Remark that, on the contrary to the binary case, the structure of balanced sequences on arbitrary alphabets is to a large extent unknown. Vuillon in [11] provides a thorough survey on the topic. The reader can also look at the references therein.

Keywords and phrases. Infinite sequences, Sturmian words, balance, product.

[1] University of Palermo, Dipartimento di Matematica e Informatica, Via Archirafi 34, 90123 Palermo, Italy. {restivo, giovanna}@math.unipa.it

Article published by EDP Sciences

© EDP Sciences 2011

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

Apart from their theoretical interest, balanced sequences have several applications. For instance, Altman *et al.* in [1] show the link between the notion of balanced sequences and the optimal scheduling in networks.

The notion of *product* of two sequences has been introduced by Salimov in [10], where the author studies the class of uniformly recurrent sequences such that the product of any of its members and each uniformly recurrent sequence is uniformly recurrent too.

In this paper, we study when the product $w = u \otimes v$ of two balanced sequences u and v is balanced too. In Section 2 we give a complete answer in the case of binary alphabets. In particular, we show that, if $deg(u) = deg(v) = 2$ and the product w is balanced, with $deg(w) = 4$, then w is an ultimately periodic sequence of a very special form. The case of larger alphabets is considered in the last section: it gives rise to new interesting problems, which suggest a new approach to investigate the structure of balanced sequences on more than two symbols.

1. Basic definitions

Let $\Sigma = \{a_1, a_2, \ldots, a_k\}$ be a finite *alphabet*. A *finite* word over Σ is a finite sequence of letters from Σ. The *empty word* ε is the empty sequence. Under the operation of concatenation, the set Σ^* of all finite words over Σ is a free monoid with identity element ε and set of generators Σ. A (right) *infinite* word w is a sequence indexed by $\mathbb{N}^+$ with values in Σ, *i.e.*, $w = b_1 b_2 b_3 \ldots$ with each $b_i \in \Sigma$. The set of all infinite words over Σ is denoted by Σ^ω.

An *ultimately periodic* infinite word can be written as $uv^\omega = uvvv \cdots$, for some $u, v \in \Sigma^*$, $v \neq \varepsilon$. If $u = \varepsilon$, then such a word is *periodic*. An infinite word that is not ultimately periodic is said to be *aperiodic*. For all notions and results not explicitly reported here we refer to [3,7]

Given a finite word $w = b_1 b_2 \cdots b_n$ on the alphabet Σ, with each $b_i \in \Sigma$, the length of w, denoted by $|w|$, is equal to n and we denote by $|w|_a$ the number of occurrences of a letter a in w. By convention, the *empty* word ε is the unique word of length 0. A finite word z is called a *factor* of a finite or infinite word w on Σ if $w = uzv$ for some word u, v. The set $F(w)$ is the set of factors of w. If w is infinite, we denote by $Ult(w)$ the set of all letters occurring infinitely often in w and by $deg(w)$ the *degree* of w (*i.e.*, the cardinality of the alphabet of all letters occurring infinitely often in the word itself), hence $deg(w) = |Ult(w)|$. In what follows, we suppose that w is an infinite sequence where all letters occur infinitely often in w.

A finite or infinite word is *balanced* if, for any two of its factors u, v with $|u| = |v|$, we have $||u|_a - |v|_a| \leq 1$ for any letter a, *i.e.* the number of a's in each of u and v differs by at most 1. A periodic infinite sequence v^ω is balanced if the finite word v is *circularly balanced*, that is if all its conjugates are balanced.

Remark 1.1. For a two-letter alphabet, being balanced is equivalent to being balanced with respect to one letter.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

Remark 1.2. If w is a balanced sequence over the alphabet Σ and $a \in \Sigma$, then there exists a non-negative integer k such that the number of letters between two consecutive a's is either k or $k+1$.

It is interesting to remark that, in the binary case, infinite balanced words are closely related to Sturmian sequences. Sturmian words can be defined in several different but equivalent ways (*cf.* [7]; Chap. 2). Some definitions are "combinatorial" and others of "geometrical" nature. With regard to the first type of definition a Sturmian word is a binary infinite word which is not ultimately periodic, of minimal complexity. From the "geometrical" point of view, the Sturmian words code discrete lines. In particular, a Sturmian word can be defined by considering the intersections with a squared-lattice of a semiline having a slope which is an irrational number. A vertical intersection is denoted by the letter 0, a horizontal intersection by 1 and the intersection with a corner by 01 or 10 (*cf.* [7]). If the semiline starts from the origin the corresponding Sturmian word is called *characteristic*. Characteristic Sturmian words can be constructed by a family of finite words called *standard words*, in the sense that every characteristic word is the limit of a sequence of standard words (*cf.* [7]). Standard words are defined as follows. Let $d_1, d_2, \ldots, d_n, \ldots$ be a sequence of natural numbers, with $d_i \geq 0$ and $d_i > 0$ for $i = 2, \ldots, n, \ldots$. Consider the following sequence $\{s_n\}_{n\geq 0}$ of words over the binary alphabet $\{0,1\}$: $s_0 = 1$, $s_1 = 0$ and $s_{n+1} = s_n^{d_n} s_{n-1}$ for $n \geq 1$. Each finite word s_n in the sequence is called a *standard* word. The sequence $\{s_n\}_{n\geq 0}$ converges to a limit s that is a *characteristic Sturmian* word. Moreover any characteristic Sturmian word is obtained in this way. The sequence $\{s_n\}_{n\geq 0}$ is called the approximating sequence of s and $(d_1, d_2, \ldots, d_n, \ldots)$ is the *directive sequence* of s. The best known example of such a word is the *Fibonacci sequence*, which is obtained by the directive sequence $1, 1, 1, 1, \ldots$.

2. Product of two binary sequences

Let us define a *product* of two infinite sequences (or words of the same length) $u = u_0 u_1 \cdots$ on Σ and $v = v_0 v_1 \cdots$ on Δ as the word

$$w = u \otimes v =< u_0, v_0 >< u_1, v_1 > \cdots$$

on the alphabet $\Lambda = \Sigma \times \Delta$.

In this section, we assume that $\Sigma = \Delta = \{0,1\}$. We study the product between two binary balanced sequences. An interesting question is: when a product of two balanced binary sequences is balanced too?

Notation 2.1. *We set* $\Sigma \times \Sigma = \Lambda$ *and code* $\langle 0,0\rangle = a$, $\langle 0,1\rangle = b$, $\langle 1,0\rangle = c$ *and* $\langle 1,1\rangle = d$. *If we denote by* f *the first component* x *and by* l *the last component* y *of* $\langle x, y\rangle$, *where* $x, y \in \Sigma$, *given a letter* $z \in \Lambda$, *we define the projections* $f(z)$ *and* $l(z)$ *as follows:*

- $f(a) = 0$ *and* $f(b) = 0$
- $f(c) = 1$ *and* $f(d) = 1$

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

- $l(a) = 0$ *and* $l(c) = 0$
- $l(b) = 1$ *and* $l(d) = 1$

Example 2.2. We consider the Fibonacci sequence f=0100101001001010$\cdots$ and the sturmian sequence $s = 00100010010001001001\cdots$ obtained by the directive sequence $2, 1, 1, 1, 1, 1, \ldots$.

$f:$	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0	$\cdots$
$s:$	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	$\cdots$
$w:$	a	c	b	a	c	a	d	a	a	d	a	a	c	b	c	a	a	d	a	c	b	$\cdots$

$w = f \otimes s = \langle 0,0\rangle\langle 1,0\rangle\langle 0,1\rangle\langle 0,0\rangle\cdots = acba\cdots$ is not a balanced sequence, because w has factors $u = aa$ and $v = cb$, for which $||u|_a - |v|_a| = 2$.

In the proof of the next result, we suppose that the letter 0 is the most frequent letter both in u and in v. Otherwise it suffices swapping the occurrences of the letter 0 with the occurrences of letter 1 either in u, or in v, or in both, and such a swapping corresponds to rename the letters in w. For instance, swapping the letter 0 with the letter 1 in u, and not in v, corresponds to swap a with c and b with d in w.

The main result of the paper is the following theorem.

Theorem 2.3. *Let u, v be two binary balanced sequences. If $w = u \otimes v$ is balanced and $deg(w) = 4$ then w is (ultimately) periodic and is a suffix of one of the following sequences:*

(i) $(adacb)^t(adabc)^\omega$
(ii) $(adabc)^t(adacb)^\omega$
(iii) $(adabacb)^t(adabcab)^\omega$
(iv) $(adabcab)^t(adabacb)^\omega$

where $t \in \mathbb{N}$.

The proof is obtained as a consequence of several lemmas.

In what follows, we suppose that $w = u \otimes v$, $deg(w) = 4$, the letter 0 is the most frequent letter both in u and v. As u is a binary balanced sequence, from Remark 1.2, it follows that two consecutive 1's in u are separated by h or $h + 1$ 0's. As v is a binary balanced sequence, from Remark 1.2, it follows that two consecutive 1's in v are separated by k or $k + 1$ 0's. Now, we suppose that $h \geq k$ and we prove that when $k \geq 2$, if w is balanced and $deg(w) = 4$ then it cannot be obtained as the product of two binary balanced sequences and hence $k = 1$.

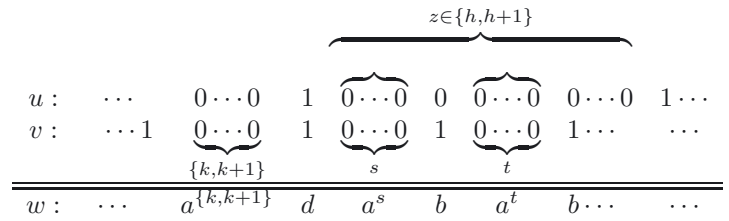
Hence, we suppose that $k \geq 2$ and distinguish the case where $h > k$ and the case where $h = k$. First, we suppose that $h > k$. Since $deg(w) = 4$, there exist infinitely many positions i, where begins the factor 10^s10^t1, with $s, t \in \{k, k + 1\}$ in v, and $\langle u_i, v_i\rangle = \langle 1, 1\rangle = d$. Let z be the number of 0's between the occurrence of 1 at position i in u and the next occurrence of 1 in u. So we suppose that $z = s + 1 + r \in \{h, h + 1\}$ for some $r \geq 0$. Clearly, since $h \geq k + 1$, it follows that if $s = k$ then $h \geq s + 1$, instead if $s = k + 1$ then $h \geq s$.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

Lemma 2.4. *If $k \geq 2$ and $z > s+1+t$ then w is not balanced.*

Proof. We have to prove that if between two consecutive 1's in u there are more than two 1's in v (equivalently, between two consecutive 1's in v there are not 1's in u or $z \geq s+t+2$), then $w = u \otimes v$ is not balanced.

We have the situation depicted below.

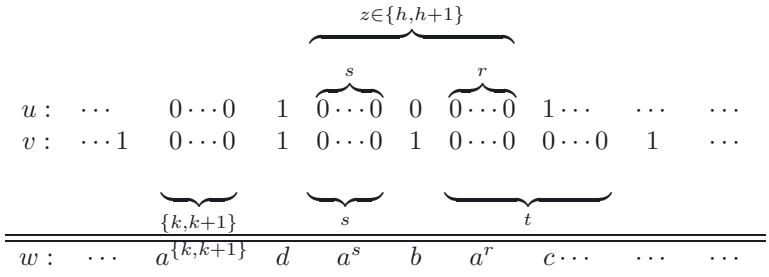


We consider the distinct cases:

(1) if $s = k$ and $t = k$ then it suffices to consider the factors ada^k and ba^kb to show that w is not balanced on b;
(2) if $s = k$ and $t = k+1$ then it suffices to consider the factors a^2da^k and $ba^{k+1}b$ to show that w is not balanced on b;
(3) if $s = k+1$ and $t = k$ then it suffices to consider the factors da^{k+1} and ba^kb to show that w is not balanced on b;
(4) if $s = k+1$ and $t = k+1$ then it suffices to consider the factors ada^{k+1} and $ba^{k+1}b$ to show that w is not balanced on b. □

Lemma 2.5. *If $k \geq 2$ and $z < s+1+t$ then w is not balanced.*

Proof. Recall that $z = s+1+r \in \{h, h+1\}$ for some $r \geq 0$. As $z < s+1+t$, then $r < t$ and we consider two cases: either $r \leq s-2$ or $s-2 < r < t$. If $r \leq s-2$ then we have the situation depicted below.



We consider the factor ba^rc of length $r+2$. As $s \geq r+2$, it follows that there exists the factor a^{r+2}, so we obtain that w is not balanced.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

If $s-1 \leq r < t$ then we have the situation depicted below.

$$\begin{array}{lccccccccccc}
u: & \cdots & & 0\cdots0 & 1 & \overbrace{0\cdots0}^{} & 0 & \overbrace{0\cdots0}^{r} & 1 & \overbrace{0\cdots0}^{} & 0 & \cdots \\
v: & \cdots 1 & & \underbrace{0\cdots0}_{\{k,k+1\}} & 1 & \underbrace{0\cdots0}_{s} & 1 & \underbrace{0\cdots0}_{r} & 0 & \underbrace{0\cdots0}_{m} & 1 & \cdots \\
\hline
w: & \cdots & & a^{\{k,k+1\}} & d & a^s & b & a^r & c & a^m & b & \cdots
\end{array}$$

where $t = r+1+m$ with $m \geq 0$. We consider the factor ba^rca^mb of length $t+2$. As there exists the factor $a^{(t+2)-s-1}da^s$ of length $t+2$, where $(t+2)-s-1 \leq k$ we obtain that w is not balanced. □

Remark 2.6. We observe that if $z < s+1+t$ and $z > s+1+t$ then w is not balanced, so we have $z = s+1+t$ and this contradicts the fact that $deg(w) = 4$.

Now we suppose that $h = k$ and we prove the following lemma.

Lemma 2.7. *If $k \geq 2$ and $h = k$ and u, v are not equal then w is not balanced.*

Proof. Since all letters occur infinitely often in w, there exist infinitely many positions i where $\langle u_i, v_i\rangle = \langle 1,1\rangle = d$. We have to prove that if, from the position i, begins the factor 10^s1 in v, with $s \in \{k, k+1\}$, and, the factor 10^m1 in u, where $m = s \pm 1 \in \{k, k+1\}$, then w is not balanced.

If $s = k$ and $m = k+1$, we have the situation depicted below.

$$\begin{array}{lccccccccc}
u: & \cdots & & 0\cdots0 & 1 & \overbrace{0\cdots0}^{k} & 0 & 1 & \cdots \\
v: & \cdots 1 & & \underbrace{0\cdots0}_{\{k,k+1\}} & 1 & \underbrace{0\cdots0}_{k} & 1 & 0 & \cdots \\
\hline
w: & \cdots & & a^{\{k,k+1\}} & d & a^k & b & c & \cdots
\end{array}$$

If $s = k+1$ and $m = k$, we have the situation depicted below.

$$\begin{array}{lccccccccc}
u: & \cdots & & 0\cdots0 & 1 & \overbrace{0\cdots0}^{k} & 1 & 0 & \cdots \\
v: & \cdots 1 & & \underbrace{0\cdots0}_{\{k,k+1\}} & 1 & \underbrace{0\cdots0}_{k} & 0 & 1 & \cdots \\
\hline
w: & \cdots & & a^{\{k,k+1\}} & d & a^k & c & b & \cdots
\end{array}$$

In these cases, as $k \geq 2$ we have that w is not balanced: there exist the factors bc or cb and aa. □

In order to prove the theorem we further need the following key lemma. We notice that some of the techniques exploited in its proof are also used in [1].

Lemma 2.8. *Let $w \in \{a,b,c,d\}^\omega$ be a balanced sequence, with $deg(w) = 4$, satisfying the following two conditions:*

- $\{aa, aca, bd, db, cd, dc, dd\} \cap F(w) = \emptyset$
- *$aba \in F(w)$ if and only if $\{cad, dac\} \cap F(w) = \emptyset$.*

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

Then w is (ultimately) periodic and is a suffix of one of the following sequences:

(i) $(adacb)^t(adabc)^\omega$
(ii) $(adabc)^t(adacb)^\omega$
(iii) $(adabacb)^t(adabcab)^\omega$
(iv) $(adabcab)^t(adabacb)^\omega$

where $t \in \mathbb{N}$.

Proof. Since $deg(w) = 4$, all letters occur infinitely often in w, so it exist infinitely many positions i such that the ith letter of w is d. By hypothesis, $\{bd, db, cd, dc, dd\} \cap F(w) = \emptyset$, the letter d must be preceded and followed by a in w. Hence $ada \in F(w)$.

In the sequel of the proof, we suppose that the infinite sequence w starts with the factor ada and we derive that w is one of the sequences (i), (ii), (iii) and (iv) in the statement of the lemma. If this is not the case, the same arguments lead to the conclusion that w is a suffix of one of these sequences.

We note that, from Remark 1.2, the following facts hold:
Fact 1. Any factor of length 3 contains either 1 or 2 occurrences of the letter a.
Fact 2. Any factor of length 3 contains at most an occurrence of the letter b or of the letter c.

Since, by hypothesis, $aa \notin F(w)$, then the factor ada in w can be followed either by the letter c (*Case 1*) or by the letter d (*Case 2*) or by the letter b (*Case 3*) in w.

(*Case 1*).
First, we suppose that we have the factor $adac$. As by hypothesis $\{cd, aca\} \cap F(w) = \emptyset$ and, from Remark 1.2, $cc \notin F(w)$, the factor $adac$ is followed by b. So we have the factor $adacb$. By *Fact 1*, this factor must be followed by the letter a. So we have the factor $adacba$, which, from Remark 1.2, is not followed by the letters a and b. So, if the factor $adacba$ is followed by the letter c, we have the factor $adacbac$. We observe that $(adacbac)^\omega$ is balanced, but $aca \notin F(w)$, hence we have to extend the factor, but it is easy to verify that this is not possible because $\{cb, cd, cc\} \cap F(w) = \emptyset$. Hence the factor $adacba$ must be followed by the letter d and then by letter a. If the factor $adacbada$ is followed by the letter c, we can obtain $(adacb)^\omega$, which is balanced: thesis (ii) with $t = 0$. If the factor $(adacb)^t ada$, with $t > 0$, is followed by the letter b, then the factor $(adacb)^t adab$ must be followed by the letter c (because $\{bd, bb\} \cap F(w) = \emptyset$ and, as $dac \in F(w)$, it follows that $aba \notin F(w)$). So, we have $(adacb)^t adabc$. Repeating the similar arguments, we have that this factor is followed by the factor ada. So, it is easy to verify that we can only obtain the balanced sequence $(adacb)^t(adabc)^\omega$: thesis (i).

(*Case 2*).
Now, we suppose that we have the factor $adad$. As, by hypothesis, $\{dd, dc, db\} \cap F(w) = \emptyset$, then we have the factor $adada$. So, we can obtain the factor $(ad)^s a$ for some positive integer s. Since, $(ad)^s a$ is a factor of w and w is a sequence on a four-letters alphabet, then the factor $(ad)^s a$ must be followed either by letter b or

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

by letter c. In both cases, w is not a balanced sequence, indeed, in the first case, we obtain a factor that contains dad and bca and, in the second case, we obtain a factor that contains dad and cba.

(*Case 3*).
Now, we suppose that we have the factor $adab$. This factor can be followed either by the letter c or by the letter a.

In the first case, we have the factor $adabc$ in w. By *Fact 1*, the factor $adabc$ is followed by the letter a. So, we have the factor $adabca$.

If the factor $adabca$ is followed by the letter d, repeating similar arguments, we can obtain either the balanced sequence $(adabc)^\omega$ or the balanced sequence $(adabc)^t(adacb)^\omega$, with $t \in \mathbb{N}$: thesis (i) with $t = 0$ and thesis (ii).

Now, we suppose that the factor $adabca \in F(w)$ is followed by the letter b. So, we have the factor $adabcab \in F(w)$, which is followed by the letter a, indeed, from Remark 1.2, as $adab \in F(w)$, it follows that $cabc \notin F(w)$, hence $adabcab$ cannot be followed by the letter c. By *Fact 2*, it cannot be followed by the letter b and, by hypothesis, it cannot be followed by the letter d. So we have the factor $adabcaba \in F(w)$.

Now, we see that the factor $adabcaba$ is followed by the letter d, so, repeating the similar arguments, we can obtain either $(adabcab)^\omega$ or $(adabcab)^t(adabacb)^\omega$, with $t \in \mathbb{N}$.

We prove that $adabcaba$ can only be followed by letter d, indeed $\{bab, aa\} \cap F(w) = \emptyset$ and, now, we prove that $adabcaba$ is not followed by the letter c. By contradiction, we suppose that $adabcaba$ is followed by the letter c. By hypothesis $adabcabac$ cannot be followed by the letters a, c and d. So, we have the factor $adabcabacb$. By Fact 1, the factor $adabcabacb$ is followed by the letter a. So, we have the factor $adabcabacba$. From Remark 1.2, as $adab \in F(w)$, it follows that $cbac \notin F(w)$, hence $adabcabacba$ cannot be followed by the letter c. By hypothesis, it cannot be followed by the letters a and b. So, we have the factor $adabcabacbad$. By hypothesis $\{dc, db, dd\} \cap F(w) = \emptyset$, so the factor $adabcabacbad$ is followed by the letter a. So, we have the factor $adabcabacbada$. We observe that the sequence $(adabcabacbada)^\omega$ is not balanced, so we have to extend it, but it is not possible, because $adabcabacbadab$ is not balanced (there are the factors $cabac$ and $badab$, for which $|cabac|_c - |badab|_c = 2$), $adabcabacbadac$ is not balanced (there are the factors $bcab$ and $adac$, for which $|bcab|_b - |adac|_b = 2$), $adabcabacbadad$ is not balanced (there are the factors bca and dad, for which $|dad| - |bca| = 2$) and $aa \notin F(w)$.

Hence, the factor $adabcaba$ is not followed by the letters a, b and c. So the factor $adabcaba$ is followed by the letter d. The factor $adabcabad$ is followed by the letter a. So, we have the factor $adabcabada$.

From Remark 1.2, as the factor $bcab \in F(w)$, in any factor of length 4 in w there must be at least one occurrence of the letter b. Hence, the factor $adabcabada$ is followed by the letter b. So, we have the factor $adabcabadab$.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

Since $\{bb, bd\} \cap F(w) = \emptyset$, the factor $adabcabadab$ is followed either by the letter c or by the letter a.

If the factor $adabcabadab$ is followed by the letter c, then we can obtain $(adabcab)^\omega$. We remark that this sequence is balanced: thesis (iii) with $t = 0$.

If $adabcabadab$ is followed by the letter a, we have the factor $adabcabadaba$ in w. From Remark 1.2, as $abcab \in F(adabcabadaba)$, the factor $adabcabadaba$ is not followed by the letter d. Moreover, it is not followed by the letter a (because $aa \notin F(w)$) and by the letter b ($ada \in F(w)$). So $adabcabadab$ is followed by the letter c and we have the factor $adabcabadabac$.

Now, as $\{aca, cd, cc\} \cap F(w) = \emptyset$, we have $adabcabadabacb$. As $\{bb, bd, cbd\} \cap F(w) = \emptyset$, this factor is followed by a, so we have $adabcabadabacba$. So, we have again $adabcabadabacbadab$.

As, there is the factor $abadaba$ in w, in any factor of length 7 in w, there cannot be two occurrences of letter c. So, $adabcabadabacbadab$ is not followed by the letter c. Moreover, $\{bd, bb\} \cap F(w) = \emptyset$, so $adabcabadabacbadab$ is followed by the letter a and we have $adabcabadabacbadaba$.

As $\{aa, bab\} \cap F(w) = \emptyset$ and, as the factor $bacba \in F(w)$, it follows that $dabad \notin F(w)$. So, we have $adabcabadabacbadabac$.

The factor $adabcabadabacbadabac$ is followed by the letter b. So, we have $adabcabadabacbadabacb$. We observe that the word $(adabcab)(adabacb)(adabacb)$ is balanced.

Hence we can obtain $(adabcab)^t(adabacb)^\omega$, with $t \in \mathbb{N}$, which is balanced: thesis (iv).

We have supposed that we have the factor $adab$ followed by the letter c, but it is possible that the factor $adab$ is followed by the letter a. So, we have the factor $adaba$. By hypothesis, if $aba \in F(w)$ then $\{cad, dac\} \cap F(w) = \emptyset$. So, repeating the similar arguments, we can obtain either $(adabacb)^\omega$ or $(adabacb)^t(adabcab)^\omega$, with $t \in \mathbb{N}$: thesis (iv) with $t = 0$ and thesis (iii). □

Now, we are ready to prove our main result.

Proof of Theorem 2.3. Assume that $k \geq 2$, then by using Lemmas 2.4, 2.5, Remark 2.6 and Lemma 2.7, we obtain that w is not balanced. So $k = 1$.

Now, we have to prove that when $k = 1$ then w is (ultimately) periodic and can assume one of the forms in the statements.

In order to prove it, we distinguish the cases where aa is a factor of w or not. First we suppose that $aa \in F(w)$.

As $k = 1$, from Remark 1.2, it means that two consecutive 1's in v are separated by 1 or 2 letters 0. So, we cannot have the factor 000 in v, hence we cannot have the block $010 \otimes 000 = aca$ and hence $aca \notin F(w)$. Now, we prove, by contradiction, that if $aa \in F(w)$ then $w = u \otimes v$ is not balanced. Suppose that w is balanced. As w is balanced on a four-letter alphabet, then there exists a position i such that the ith letter of w is c. As the factor $aa \in F(w)$, we observe that, from Remark 1.2, any factor of length 2, in w, contains at least an occurrence of letter a. So the letter c must be preceded and followed by a in w, but $aca \notin F(w)$, a contradiction.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

Now we suppose that $k = 1$ and the factor $aa \notin F(w)$. In order to obtain the thesis of theorem, we make the following observations which allow us to use Lemma 2.8.

As $k = 1$, from Remark 1.2, it means that two consecutive 1's in v are separated by 1 or 2 letters 0. So, we cannot have

- the factor 000 in v, hence we cannot have the block $010 \otimes 000 = aca$.
- the factor 11 in v, hence we cannot have the block $01 \otimes 11 = bd$ and the block $10 \otimes 11 = db$.

As $h \geq k$, it means that two consecutive 1's in u are separated by at least a letter 0. So, we cannot have

- the factor 11 in u, hence we cannot have the block $11 \otimes 10 = dc$ and the block $11 \otimes 01 = cd$.

So we cannot have the block $11 \otimes 11 = dd$. Hence $\{aca, bd, db, cd, dc, dd\} \cap F(w) = \emptyset$.

We note that, when $k = 1$, if the factor $aba \in F(w)$, then the factor $000 \in F(u)$, hence $h \geq 3$ and, from Remark 1.2, $101 \notin F(u)$. So we cannot have the blocks $101 \otimes 001 = cad$ and $101 \otimes 100 = dac$, and *vice versa*.

By these observations we obtain the conditions of Lemma 2.8:

- $\{aca, bd, db, cd, dc, dd\} \cap F(w) = \emptyset$
- $aba \in F(w)$ if and only if $\{cad, dac\} \cap F(w) = \emptyset$,

and, so we obtain that w is (ultimately) periodic and is a suffix of one of the following sequences:

(i) $(adacb)^t(adabc)^\omega$
(ii) $(adabc)^t(adacb)^\omega$
(iii) $(adabacb)^t(adabcab)^\omega$
(iv) $(adabcab)^t(adabacb)^\omega$

Therefore, we obtain the thesis. □

We have proved that the only balanced sequences on a four-letters alphabet, that can be obtained as a product of two balanced binary sequences, are (ultimately) periodic and they can assume one of the forms in Theorem 2.3. Next theorem shows that, on the contrary, any balanced sequence on a three-letters alphabet can be obtained as a product of two binary balanced sequences.

Theorem 2.9. *Any balanced sequence w on three letters can be obtained as the product of two binary balanced sequences u and v.*

Proof. Let $\{a, b, d\}$ be the alphabet of w.

From Notation 2.1, we know that $f(a) = f(b) = 0$ and $f(d) = 1$, so u can be obtained by w replacing the letters a, b with the letter 0 and the letter d with the letter 1.

As w is balanced on all letters, then the letter 1 is balanced in u. Since u is binary, from Remark 1.1, it follows that the letter 0 is balanced in u, hence u is balanced.

Moreover, we know that $l(a) = 0$ and $l(b) = l(d) = 1$, so v can be obtained by w replacing the letter a with the letter 0 and the letters b and d with the letter 1.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

As w is balanced on all letters, then the letter 0 is balanced in v. Since v is binary, from Remark 1.1, it follows that the letter 1 is balanced in v, hence v is balanced.

If $Ult(w)$ is different then the proof is similar. □

The proof of the following theorem gives a method to construct u by v to obtain the balanced sequence w where $w = u \otimes v$ and $deg(w) = 3$.

We first recall that an infinite periodic word w^ω has constant gaps if the number of letters between two occurrences of successive letter a_i of w^ω is constant for each i. For example $(abac)^\omega$ is a constant gap word and $(aba)^\omega$ is not constant gaps (*cf.* [11]). It is well known that the unique constant gap sequence G on two letters $\{a, b\}$ is $(ab)^\omega$ up to letter permutation. The following proposition has been proved by Hubert [6].

Proposition 2.10. *Let u be a balanced sequence on the alphabet $\{0, 1\}$. Construct a new sequence w by replacing in u, the subsequence of 0 by a constant gap sequence G on an alphabet Γ, and the subsequence of 1 by a constant gap sequence H on a disjoint alphabet Δ. Then w is balanced on the alphabet $\Gamma \cup \Delta$.*

Theorem 2.11. *For any binary balanced sequence v, one can construct a binary balanced sequence u such that $w = u \otimes v$ is balanced and $deg(w) = 3$.*

Proof. From Proposition 2.10 and Notation 2.1, the balanced sequence w can be obtained by v replacing the occurrences of the letter 1 with a constant gap sequence of letters b and d and the occurrences of the letter 0 with a. We remark that the unique constant gap sequence of two letters (b, d) is either $(bd)^\omega$ or $(db)^\omega$.

Now we construct u from v. We know that $f(a) = f(b) = 0$ and $f(d) = 1$ (see Notation 2.1), so u can be obtained by w replacing the occurrences of the letters a, b with the letter 0 and the occurrences of the letter d with the letter 1. As w is balanced on all letters, then the letter 1 in u is balanced. Since u is binary, from Remark 1.1, it follows that the letter 0 is balanced in u, hence u is balanced.

Clearly, this sequence w is not the unique balanced sequence that we can obtain by v, because we can use other constant gap sequences. □

3. Extensions and further work

In this section we consider the product of two balanced sequences on arbitrary alphabets. We give some preliminary results and we introduce some new problems. However, in such a general case, we are far to obtain a complete classification.

Let u, v be two balanced sequences, with $deg(u) = h$ and $deg(v) = k$, and suppose that their product $w = u \otimes v$ is balanced. From the definition

$$\max\{h, k\} \leq deg(w) \leq hk.$$

It would be interesting to consider the following problems:

(1) Does there exist, for each r, such that $\max\{h, k\} \leq r \leq hk$, at least one balanced sequence w having $deg(w) = r$?

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

(2) For each r, such that $\max\{h,k\} \leq r \leq hk$, characterize the class of the balanced sequences of degree r.

In the previous section we have answered to the problems (1) and (2) in the case $h = k = 2$. In particular we have shown that, for each r such that $2 \leq r \leq 4$, there exists a balanced sequence of degree r. Moreover, Theorem 2.3 characterizes the sequences of degree 4, Theorem 2.9 characterizes the sequences of degree 3, whereas the sequences of degree 2 are trivially obtained as product of two identical sequences.

The next step is to give an answer in the case $h = 2$ and $k = 3$. The following example shows that, if $deg(u) = 2$ and $deg(3) = 3$, then we can obtain at least one balanced sequence of degree r, for each r such that $3 \leq r \leq 6$.

Example 3.1. Let $\Sigma = \{0,1\}$ and $\Delta = \{0,1,2\}$. For simplicity of notation, we set $\Delta \times \Sigma = \Lambda$ and code $\langle 0,0\rangle = a$, $\langle 0,1\rangle = b$, $\langle 1,2\rangle = c$, $\langle 0,2\rangle = d$, $\langle 1,1\rangle = e$ and $\langle 1,0\rangle = f$, so $\Lambda = \{a,b,c,d,e,f\}$. A balanced sequence w, where $deg(w) = 3$, can be obtained in the following way:

$$\begin{array}{ll} u: & (0\ \ 0\ \ 1)^\omega \\ v: & (0\ \ 1\ \ 2)^\omega \\ \hline w: & (a\ \ b\ \ c)^\omega \end{array}$$

A balanced sequence w, where $deg(w) = 4$, can be obtained in the following way:

$$\begin{array}{ll} u: & (0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1)^\omega \\ v: & (0\ \ 1\ \ 2\ \ 0\ \ 1\ \ 2)^\omega \\ \hline w: & (a\ \ b\ \ d\ \ a\ \ b\ \ c)^\omega \end{array}$$

A balanced sequence w, where $deg(w) = 5$, can be obtained in the following way:

$$\begin{array}{ll} u: & (0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 1)^\omega \\ v: & (0\ \ 1\ \ 2\ \ 1\ \ 0\ \ 1\ \ 2)^\omega \\ \hline w: & (a\ \ b\ \ d\ \ e\ \ a\ \ b\ \ c)^\omega \end{array}$$

A balanced sequence w, where $deg(w) = 6$, is obtained in the following way:

$$\begin{array}{ll} u: & (0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 1)^\omega \\ v: & (0\ \ 1\ \ 2\ \ 0\ \ 1\ \ 2)^\omega \\ \hline w: & (a\ \ e\ \ d\ \ f\ \ b\ \ c)^\omega \end{array}$$

Previous examples give an answer to the problem (1) in the case $h = 2$ and $k = 3$. However the problem (2) remains open. In general, problems (1) and (2) remain open for arbitrary values of h and k.

A special aspect of problem (1) is to determine the minimum or maximum degree of the product. We here remark that if $mcd(h,k) = 1$, then the maximum degree of the product $w = u \otimes v$ is hk. Indeed if $\Sigma = \{0,1,\ldots,h-1\}$,

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

$\Delta = \{0, 1, \ldots, k-1\}$, $u = (01 \cdots (h-1))^\omega$ and $v = (01 \cdots (k-1))^\omega$, the product $u \otimes v$ is a balanced sequence of degree hk.

The product of two sequences provides a useful method to obtain balanced sequences on a given alphabet, from balanced sequences on smaller alphabets. We show that such a method generalizes an analogous technique introduced by Hubert. Indeed, by using the result of Hubert (*cf.* Prop. 2.10), one can obtain a balanced sequence w from a balanced sequence u, where $|Ult(u)| < |Ult(w)|$. The main idea is to periodically replace the occurrences of a letter $a \in Ult(u)$ by a constant gap sequence on an alphabet Δ, where $Ult(w) = Ult(u) \cup \Delta$. Hubert states that this replacement can be repeated for each letter $a \in Ult(u)$, but the condition in order to have a balanced sequence on each letter of v is that the constant gap sequences are on disjoint alphabets. The next proposition states that when a balanced sequence w is obtained by using Hubert's method then w can be also obtained as product of two balanced sequences.

Proposition 3.2. *Let u be a balanced sequence and let $a \in Ult(u)$. Let w be the balanced sequence obtained by replacing the occurrences of a in u by a constant gap sequence y such that $Ult(y) \cap Ult(u) = \emptyset$. Then there exists a balanced sequence v on $|Ult(y)| + 1$ letters such that $w = u \otimes v$. Moreover, $|Ult(w)| = |Ult(y)| + |Ult(u)| - 1$.*

Proof. We construct the balanced sequence v by replacing in u the occurrences of a letter $a \in Ult(u)$ by a constant gap sequence y such that $Ult(y) \cap Ult(u) = \emptyset$ and by replacing each letter $b \in Ult(u)$, $b \neq a$, with a new letter $x \notin Ult(y) \cup Ult(u)$. So $Ult(v) = Ult(y) \cup \{x\}$. From Proposition 2.10, the sequence v is balanced on the letters belonging to $Ult(y)$, so we have to prove that v is balanced on x. In order to prove it, we consider the binary sequence z obtained by replacing the occurrences of the letter a of u by the letter 0 and by replacing the occurrences of other letters belonging to $Ult(u)$ by the letter 1. From Remark 1.2, the sequence z is balanced. The occurrences of the letter 1 in z correspond to the occurrences of the letter x in v. Since z is balanced, it follows that v is balanced on c. □

Example 3.3. We suppose that $u = (fbfefbf)^\omega$, $y = (adac)^\omega$ and w is obtained by replacing the occurrences of the letter f in u by the constant gap y, so $w = (abdeabc)^\omega$. Now, we can obtain v by replacing the occurrences of the letter f of u by the constant gap sequence y and by replacing the occurrences of the letters b and e by letter x. So, we obtain $v = (axdxaxc)^\omega$. Now, if we code $\langle f, a\rangle = a$, $\langle b, x\rangle = b$, $\langle f, c\rangle = c$, $\langle f, d\rangle = d$ and $\langle e, x\rangle = e$, we have that:

$u:$	$(f$	b	f	e	f	b	$f)^\omega$
$v:$	$(a$	x	d	x	a	x	$c)^\omega$
$w:$	$(a$	b	d	e	a	b	$c)^\omega$

where u, v and w are (periodic) balanced sequences. We observe that the sequence v is balanced, indeed the sequence $z = (0101010)^\omega$, obtained by replacing the letters b and e in u by 1, is balanced. Remark that the sequence w is the same as the one obtained in Example 3.1. However the two products are different.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

This shows that a balanced sequence can be obtained in different ways as product of balanced sequences. Moreover, the different pairs of balanced sequences are over alphabets of different size.

Proposition 3.2 shows that Hubert's method can be expressed in terms of product of two sequences. However the reverse is not true. In Example 3.3 we translate the Hubert's method in terms of product. On the contrary, the product in Example 3.1 cannot be expressed in terms of Hubert's method. However in Example 3.1 and in Example 3.3 we obtain the same sequence $(abdeabc)^\omega$. So one can say that this sequence can be obtain both by a product and by Hubert's method. The following example shows that there are sequences that can be obtained as a product and not by the Hubert's method.

Example 3.4. The Fraenkel sequence w on 3 letters can be obtained as product of two balanced sequences u and v.

$$\begin{array}{lccccccc} u: & (0 & 1 & 0 & 0 & 0 & 1 & 0)^\omega \\ v: & (0 & 0 & 0 & 1 & 0 & 0 & 0)^\omega \\ \hline w: & (a & b & a & c & a & b & a)^\omega \end{array}$$

We observe that the sequence w cannot be obtained by Hubert's method.

Let us now consider the special problem whether any balanced sequence can be obtained as the product of two balanced sequences over smaller alphabets.

Theorem 2.9 shows that any balanced sequence on a three-letter alphabet can be obtained as the product of two binary balanced sequences. On the contrary, we cannot obtain all balanced sequences on four letters as the product of two balanced sequences on smaller alphabets. We call *indecomposable* these sequences. For instance, one can verify that the sequences $(abdacbadbacbadbc)^\omega$, $(abcababcabd)^\omega$, $(abacaabacad)^\omega$ and $(acabaabadab)^\omega$ are indecomposable. We observe that the last three sequences are quoted by Altman et al. in [1] as primitive sequences, because they cannot be obtained by using Hubert's method. We leave open the problem to characterize the indecomposable sequences.

A natural related problem is to classify a balanced sequence w according to the degrees of the balanced sequences u and v such that $w = u \otimes v$. We say that w is of the type (h, k) if $deg(u) = h$ and $deg(v) = k$. The indecomposable balanced sequences of degree r are of type $(r, 1)$.

For instance, if we consider the balanced sequences on four-letter alphabets, we have that the sequences of type $(2, 2)$ are given in Theorem 2.3, and the sequences of type $(4, 1)$ are the indecomposable ones. An example of sequence of type $(2, 3)$ is given in Example 3.1. The following example provides a sequence of type $(3, 3)$.

Example 3.5. We suppose that $\Sigma = \Delta = \{0, 1, 2\}$ and $\Lambda = \{a, b, c, d\}$, and we code $\langle 0, 0\rangle = a$, $\langle 1, 1\rangle = b$, $\langle 2, 0\rangle = c$ and $\langle 1, 2\rangle = d$. A balanced sequence w of

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press

type $(3,3)$, where $deg(w)=4$, can be obtained in the following way:

$$\begin{array}{lllll} u: & (0 & 1 & 2 & 1)^\omega \\ v: & (0 & 1 & 0 & 2)^\omega \\ \hline\hline w: & (a & b & c & d)^\omega \end{array}$$

w is a (periodic) balanced sequence of the type $(3,3)$ with $deg(w)=4$.

Acknowledgements. The authors wish to thank anonymous referees for their helpful suggestions that have improved the presentation of the paper.

References

[1] E. Altman, B. Gaujal and A. Hordijk, Balanced sequences and optimal routing. *J. ACM* **47** (2000) 752–775.
[2] N. Chekhova, P. Hubert and A. Messaoudi, Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci. *J. Théor. Nombres Bordeaux* **13** (2001) 371–394.
[3] C. Choffrut and J. Karhumaki, Combinatorics of words, in G. Rozenberg and A. Salomaa eds., *Handbook of Formal Language Theory* **1**. Springer-Verlag, Berlin (1997).
[4] S. Ferenczi and C. Mauduit, Transcendence of numbers with a low complexity expansion. *J. Number Theory* **67** (1997) 146–161.
[5] A.S. Fraenkel, Complementing and exactly covering sequences. *J. Combin. Theory Ser. A* **14** (1973) 8–20.
[6] P. Hubert, Suites équilibrées (french). *Theor. Comput. Sci.* **242** (2000) 91–108.
[7] M. Lothaire, *Algebraic Combinatorics on Words*. Cambridge University Press (2002).
[8] M. Morse and G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.* **62** (1940) 1–42.
[9] R.N. Risley and L.Q. Zamboni, A generalization of sturmian sequences: combinatorial structure and transcendence. *Acta Arith.* **95** (2000) 167–184.
[10] P.V. Salimov, On uniform recurrence of a direct product. *Discrete Math. Theoret. Comput. Sci.* **12** (2010) 1–8.
[11] L. Vuillon, Balanced words. *Bull. Belg. Math. Soc.* **10** (2003) 787–805.

Communicated by G. Richomme.
Received November 2, 2010. Accepted July 11, 2011.

https://doi.org/10.1051/ita/2011116 Published online by Cambridge University Press