Journal of Automata, Languages and Combinatorics **u** (v) w, x–y

GENERATING THE PSEUDO-POWERS OF A WORD

Lila Kari[0]
Department of Computer Science, The University of Western Ontario
London, Ontario, N6A 5B7 Canada
e-mail: `lila@csd.uwo.ca`

and

Manasi Kulkarni
Department of Computer Science, The University of Western Ontario
London, Ontario, N6A 5B7 Canada
e-mail: `mkulkar3@uwo.ca`

ABSTRACT

The notions of power of word, periodicity and primitivity are intrinsically connected to the operation of catenation, that dynamically generates word repetitions. When considering generalizations of the power of a word, other operations will be the ones that dynamically generate such pseudo-repetitions. In this paper we define and investigate the operation of θ-catenation that gives rise to the notions of θ-power (pseudo-power) and θ-periodicity (pseudo-periodicity). We namely investigate the properties of θ-catenation, its connection to the previously defined notion of θ-primitive word, briefly explore closure properties of language families under θ-catenation and language operations involving this operation, and propose Abelian catenation as the operation that generates Abelian powers of words.

Keywords: pseudo-power, θ-power, pseudo-primitive, pseudo-periodic, weakly periodic

1. Introduction

Periodicity and primitivity of words are fundamental properties in combinatorics on words and formal language theory. Their wide-ranging applications include pattern-matching algorithms (see e.g. [3], and [4]) and data-compression algorithms (see, e.g., [27]). Sometimes motivated by their applications, these classical notions have been modified or generalized in various ways. A representative example is the "weak periodicity" of [5] whereby a word is called *weakly periodic* if it consists of repetitions of words with the same Parikh vector. This type of period was also called *Abelian period* in [2]. Carpi and de Luca extended the notion of periodic words to that of periodic-like words, according to the extendability of factors of a word [1]. Czeizler,

[0]Corresponding author. This research was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to L.K.

Kari, and Seki have proposed and investigated the notion of *pseudo-primitivity* (and pseudo-periodicity) of words in [6, 20], motivated by the properties of information encoded as DNA strands. In addition,

Indeed, one of the particularities of information encoded as DNA strands is that a word u over the DNA alphabet $\{A, C, G, T\}$ contains basically the same information as its Watson-Crick complement, denoted here by $\theta(u)$. This led to natural as well as theoretically interesting extensions of various notions in combinatorics on words and formal language theory such as pseudo-palindrome [7], pseudo-commutativity [18], as well as hairpin-free and bond-free languages (e.g., [17, 19, 25, 13, 16]). In this context, Watson-Crick complementarity has been modeled mathematically by an antimorphic involution θ over an alphabet Σ, i.e., a function that is an antimorphism, $\theta(uv) = \theta(v)\theta(u)$, $\forall u, v \in \Sigma^*$, and an involution, $\theta(\theta(x)) = x$, $\forall x \in \Sigma^*$. In [6], a word w is called *θ-primitive*, or pseudo-primitive, if we cannot find any word u that is strictly shorter than w such that w can be written as repetitions of u and $\theta(u)$. A word w is called a θ-power or pseudo-power if $w \in \{u, \theta(u)\}^+$ for some $u \in \Sigma^+$, and is called θ-periodic or pseudo-periodic if it can be written as two or more repetitions of a word u and its image under θ.

The static notions of the power of word, period of a word, and primitive word are intrinsically connected to the operation of catenation, that dynamically generates word repetitions. In the case of generalizations of the notion of power of a word (primitive word), other operations will be the ones that dynamically produce such generalized powers, [26, 21, 10, 14, 22, 9].

In this paper we define and investigate the operation of θ-catenation that gives rise to the notion of θ-power (pseudo-power) and θ-periodicity (pseudo-periodicity). We namely investigate the properties of θ-catenation (Section 3), its connection to the previously defined notion of θ-primitive word (Section 4), briefly explore closure properties of language families under θ-catenation and language operations involving this operation (Section 5), and conclude by proposing Abelian catenation as the operation that generates Abelian powers of words (Section 6).

2. Basic definitions and notations

An alphabet Σ is a finite non-empty set of symbols. Σ^* denotes the set of all words over Σ, including the empty word λ. Σ^+ is the set of all non-empty words over Σ. The length of a word $u \in \Sigma^*$ (i.e. number of symbols in the word) is denoted by $|u|$. A word $u \in L$ is said to be length-minimal if for all $w \in L$, $|w| \geq |u|$. $|u|_a$ denotes the number of occurrences of a letter a in u. The complement of a language $L \subseteq \Sigma^*$ is $L^c = \Sigma^* \backslash L$.

An *involution* is a function $\theta : \Sigma^* \to \Sigma^*$ with the property that θ^2 is identity. θ is called a *morphism* if for all words $u, v \in \Sigma^*$ we have that $\theta(uv) = \theta(u)\theta(v)$, and an *antimorphism* if $\theta(uv) = \theta(v)\theta(u)$.

A word is called *primitive* if it cannot be expressed as a power of another word. Similarly, [6], a word is called as *θ-primitive* if it cannot be expressed as a non-trivial θ-power of another word. A *θ-power* of u is a word of the form $u_1 u_2 \cdots u_n$ for some

$n \geq 1$, where $u_1 = u$ and for any $2 \leq i \leq n$, u_i is either u or $\theta(u)$. Also, *θ-primitive root* of w denoted by $\rho_\theta(w)$ is the shortest word t such that w is a θ-power of t.

The *left quotient* of a word u by a word v is defined by

$$v^{-1}u = w \text{ iff } u = vw,$$

and the *right quotient* of u by v,

$$uv^{-1} = w \text{ iff } u = wv.$$

A language $L \subseteq \Sigma^+$ is said to be a prefix code if $L \cap L\Sigma^+ = \emptyset$. For all other concepts related to formal language theory and combinatorics on words, the reader is referred to [11] and [23].

A *binary word operation with right identity*, [12, 26], (shortly *bw-operation)* is defined as a mapping $\circ : \Sigma^* \times \Sigma^* \longrightarrow 2^{\Sigma^*}$ with $u \circ \lambda = \{u\}$. Furthermore, $L_1 \circ L_2 = \bigcup_{u \in L_1, v \in L_2}(u \circ v)$ and $L_1 \circ \emptyset = \emptyset \circ L_2 = \emptyset$ for any two languages L_1 and L_2. The *iterated bw-operation* $\circ^i$ for $i \geq 1$ and languages L_1 and L_2 is defined as $L_1 \circ^0 L_2 = L_1$ and $L_1 \circ^i L_2 = (L_1 \circ^{i-1} L_2) \circ L_2$. The *$i$-th $\circ$-power* of a non-empty language L is defined as $L^{\circ(0)} = \{\lambda\}$, and $L^{\circ(i)} = L \circ^{i-1} L$ for $i \geq 1$. If $\circ$ is the operation of catenation, then $L^0 = \{\lambda\}$, $L^1 = L$ and $L^n = L^{n-1}L$, corresponding to the usual notions of power of a language.

A non-empty word w is called *$\circ$-primitive* if $w \in u^{\circ(i)}$ for some word $u \in \Sigma^+$ and $i \geq 1$ yields $i = 1$ and $w = u$.

The *$+$-closure* of a non-empty language L with respect to a bw-operation $\circ$, denoted by $L^{\circ(+)}$, is defined as $L^{\circ(+)} = \cup_{k \geq 1} L^{\circ(k)}$. A language L is *$\circ$-closed* if $u, v \in L$ imply $u \circ v \subseteq L$. A bw-operation is called *plus-closed* if for any non-empty language L, $L^{\circ(+)}$ is $\circ$-closed.

Given a non-empty language L, a word u is a *right $\circ$-residual* of L if $w \circ u \subseteq L$ for all $w \in L$, i.e., $L \circ u \subseteq L$. Let $\rho_\circ(L)$ denote the set of all right o-residuals of L, i.e., $\rho_\circ(L) = \{u \in \Sigma^* | \forall w \in L, (w \circ u) \subseteq L\}$. Note that $\rho_\circ(\emptyset) = \emptyset$ and $\lambda \in \rho_\circ(L)$ for any non-empty language L.

The *$\circ$-left-quotient*, denoted by $\triangleleft_\circ$, is defined as

$$L_1 \triangleleft_\circ L_2 = \{w \in \Sigma^* | (L_2 \circ w) \cap L_1 \neq \emptyset\}.$$

3. θ-catenation

We introduce a new bw-operation (binary word operation with right identity) called *θ-catenation* which generates pseudo-powers, that is, θ-powers where θ is a morphic or antimorphic involution. In this section we will give a formal definition of *θ-catenation* and discuss some of its properties. Note that, unless otherwise specified, θ is any morphic or antimorphic involution.

Definition 1 Given a morphic or antimorphic involution θ on Σ^* and any two words $u, v \in \Sigma^*$, we define the binary operation *θ-catenation* as

$$u \odot v = \{uv, u\theta(v)\}.$$

For example, consider the DNA alphabet $\Sigma = \{A, G, C, T\}$ and its associated antimorphic involution defined by $\theta(A) = T, \theta(T) = A, \theta(C) = G$ and $\theta(G) = C$. If $u = ATC$ and $v = GCTA$ then

$$u \odot v = \{ATCGCTA, ATCTAGC\}.$$

The operation of θ-*catenation* can be generalized to languages in the usual way.

Note that for any (anti)morphic involution θ, the operation of θ-*catenation* has a right identity since $u \odot \lambda = \{u\}$ for all $u \in \Sigma^*$.

A bw-operation $\circ$ is called *length-increasing* if for any $u, v \in \Sigma^+$ and $w \in u \circ v$, $|w| > \max\{|u|, |v|\}$. The operation of θ-*catenation* is length-increasing since, if $w \in u \odot v = \{uv, u\theta(v)\}$ then $|w| = |u| + |v| >\max\{|u|, |v|\}$.

A bw-operation $\circ$ is called *propagating* if for any $u, v \in \Sigma^*$, $a \in \Sigma$ and $w \in u \circ v$, $|w|_a = |u|_a + |v|_a$. The operation of θ-*catenation* is clearly not propagating. However, a similar property does hold. We will namely call a bw-operation $\circ$ θ-*propagating* if for any $u, v \in \Sigma^*$, $a \in \Sigma$ and $w \in u \circ v$, $|w|_{a,\theta(a)} = |u|_{a,\theta(a)} + |v|_{a,\theta(a)}$. (The mapping which counts number of a's and $\theta(a)$'s together is the *characteristic function on the alphabet* Σ defined in [6].)

Proposition 1 For a given (anti)morphic involution θ of Σ^*, the operation of θ-*catenation* is θ-propagating.

Proof. Let $u, v \in \Sigma^*$ and let $w \in u \odot v = \{uv, u\theta(v)\}$. If $w = uv$ then the required equality clearly holds.

If $w = u\theta(v)$, we have

$$\begin{aligned} |w|_{a,\theta(a)} &= |u|_{a,\theta(a)} + |\theta(v)|_{a,\theta(a)} \\ &= |u|_{a,\theta(a)} + (|\theta(v)|_a + |\theta(v)|_{\theta(a)}) \\ &= |u|_{a,\theta(a)} + (|v|_{\theta(a)} + |v|_a) \\ &= |u|_{a,\theta(a)} + |v|_{a,\theta(a)}. \end{aligned}$$

□

A bw-operation $\circ$ satisfies the *left-identity* condition if $\lambda \circ L = L$ for any language $L \subseteq \Sigma^*$. Note that, in general, the operation of θ-*catenation* does not satisfy the left-identity condition. However, there exists languages of Σ^* which satisfy this condition, such as the language of θ-palindromes $P_\theta = \{u \in \Sigma^* | u = \theta(u)\}$ for which $\lambda \odot P_\theta = P_\theta$.

A bw-operation $\circ$ is called *left-inclusive* if for any three words $u, v, w \in \Sigma^*$ we have

$$(u \circ v) \circ w \supseteq u \circ (v \circ w)$$

and is called *right-inclusive* if

$$(u \circ v) \circ w \subseteq u \circ (v \circ w).$$

If θ is a morphic involution then the θ-*catenation* is trivially associative. However, if θ is an antimorphic involution then θ-catenation is not associative in general, and not even right- or left-inclusive . The following proposition provides necessary and sufficient conditions for associativity to hold in the antimorphic case. To prove Proposition (2), we will make use of the following Lemmas from [24].

Lemma 1 *Let $u, v \in \Sigma^+$. Then $uv = vu$ implies that u and v are powers of a common word.*

Lemma 2 *If $u^m = v^n$ and $m, n \geq 1$, then u and v are powers of a common word.*

Proposition 2 Let $\odot$ denote the operation of θ-*catenation* associated with an antimorphic involution θ of Σ^*. Given words $u, v, w \in \Sigma^*$ we have $(u \odot v) \odot w = u \odot (v \odot w)$ if and only if v and w are powers of the same θ-palindromic word.

Proof. For the direct implication, let us assume that $(u \odot v) \odot w = u \odot (v \odot w)$, i.e., $\{uvw, u\theta(v)w, uv\theta(w), u\theta(v)\theta(w)\} = \{uvw, uv\theta(w), u\theta(w)\theta(v), uw\theta(v)\}$, i.e. $\{u\theta(v)w, u\theta(v)\theta(w)\} = \{u\theta(w)\theta(v), uw\theta(v)\}$.

Case 1 : $u\theta(v)\theta(w) = u\theta(w)\theta(v)$ and $u\theta(v)w = uw\theta(v)$ implies $\theta(wv) = \theta(vw)$ and $\theta(v)w = w\theta(v)$ which further implies $wv = vw$ and $\theta(v)w = w\theta(v)$, respectively. So, according to Lemma (1), v and w are powers of a common word, as well as w and $\theta(v)$ are powers of a common word. This means, v, w and $\theta(v)$ are all powers of a common word, say p. So, we have $v = p^i$, $w = p^j$ and $\theta(v) = p^k$ for some $i, j, k \geq 1$. It implies, $\theta(v) = \theta(p)^i = p^k$, which further implies $i = k$ and $p = \theta(p)$. Hence v and w are powers of the same θ-palindromic word p.

Case 2 : $u\theta(v)w = u\theta(w)\theta(v)$ and $u\theta(v)\theta(w) = uw\theta(v)$ implies

$$\theta(v)w = \theta(w)\theta(v) \tag{1}$$

and

$$\theta(v)\theta(w) = w\theta(v). \tag{2}$$

Let us catenate $\theta(v)$ to the right of Equation (2). It will give, $\theta(v)\theta(w)\theta(v) = w\theta(v)\theta(v)$, which in turn along with Equation (1) implies

$$\theta(v)\theta(v)w = w\theta(v)\theta(v). \tag{3}$$

According to Lemma (1) w and $(\theta(v))^2$ are powers of a common word, say p. So, we will get $w = p^i$ and $(\theta(v))^2 = p^j$ for some $i, j \geq 1$. Now, according to Lemma (2) $\theta(v)$ and p are powers of a common word, say q. So, we get

$$p = q^l, \theta(v) = q^m \text{ and } w = q^n \text{ for } l, m, n \geq 1. \tag{4}$$

Substituting Equation (4) in the Equation (1) we get

$$q^m q^n = \theta(q^n) q^m \tag{5}$$

which implies that $q = \theta(q)$, i.e. q is a θ-palindromic word and v and w are powers of q.

Conversely, suppose v and w are powers of the same θ-palindromic word, say p. This implies, $v = p^i, w = p^j$ for $i, j \geq 1$ and $p = \theta(p)$, which further implies

$$\theta(v) = (\theta(p))^i = p^i \text{ and } \theta(w) = p^j. \tag{6}$$

Now, we know that, $(u \odot v) \odot w = \{uvw, u\theta(v)w, uv\theta(w), u\theta(v)\theta(w)\}$ and $u \odot (v \odot w) = \{uvw, uv\theta(w), u\theta(w)\theta(v), uw\theta(v)\}$. If we compare these two expressions, we are left to show that $\{u\theta(v)w, u\theta(v)\theta(w)\} = \{u\theta(w)\theta(v), uw\theta(v)\}$, which is clear from Equation (6).

□

In the previous section, we have seen the definition of i-th $\circ$-power of a non-empty language L. The following Lemma and its Corollary clarify this definition in the case of any bw-operation.

Lemma 3 *Given a bw-operation $\circ$, we have*

$$L^{\circ(0)} = \{\lambda\},$$

$$L^{\circ(1)} = L,$$

$$L^{\circ(n)} = L^{\circ(n-1)} \circ L, \ \forall n \geq 2.$$

Proof. Fistly, $L^{\circ(0)} = \{\lambda\}$ by definition. Secondly, $L^{\circ(1)} = L \circ^0 L = L$. Thirdly, for $n \geq 2$ we have $L^{\circ(n)} = L \circ^{n-1} L = (L \circ^{n-2} L) \circ L = L^{\circ(n-1)} \circ L$.

□

Corollary 4 *Given a bw-operation $\circ$, we have*

$$u^{\circ(0)} = \lambda,$$

$$u^{\circ(1)} = u,$$

$$u^{\circ(n)} = u^{\circ(n-1)} \circ u, \ \forall n \geq 2.$$

The following lemma characterizes the form of the words in $L^{\odot(n)}$ when the operation that is applied iteratively is the *θ-catenation*.

Lemma 5 *If $\odot$ denotes the operation of θ-catenation associated to a morphic or antimorphic involution θ of Σ^* then for $n \geq 1$,*

$$L^{\odot(n)} = \{uv_1v_2\cdots v_{n-1} | u \in L, v_i \in L \cup \theta(L), 0 \leq i \leq n-1\}.$$

In particular, when $n = 1$ we have $L^{\odot(1)} = L$.

Proof. We will prove this by induction on n.

For $n = 1$, $L^{\odot(1)} = L \odot^0 L = L$.

For $n = 2$, $L^{\odot(2)} = LL \cup L\theta(L) = \{uv | u \in L, v \in L \cup \theta(L)\}$.

Assume that the result is true for an arbitrary $k \geq 2$, i.e.,

$$L^{\odot(k)} = \{uv_1v_2\cdots v_{k-1} | u \in L, v_i \in L \cup \theta(L), 1 \leq i \leq k-1\}.$$

For $k+1 \geq 2$ the last equation of Lemma (3) holds and, together with the induction hypothesis we have $L^{\odot(k+1)} = L^{\odot(k)} \odot L = \{uv_1v_2\cdots v_{k-1} | u \in L, v_i \in L \cup \theta(L), 1 \leq i \leq k-1\} \odot L = \{uv_1v_2\cdots v_k | u \in L, v_i \in L \cup \theta(L), 1 \leq i \leq k\}$.

□

The following Corollary demonstrates that, in the same way the operation of catenation dynamically generates regular powers of words, the operation of θ-*catenation* is the one that generates the θ-powers of a word.

Corollary 6 *If $\odot$ denotes the operation of θ-catenation associated to a morphic or antimorphic involution θ of Σ^*, then every word $w \in u^{\odot(n)}$, $n \geq 1$, is of the form*

$$w = uv_1v_2\cdots v_{n-1}$$

where $v_i \in \{u, \theta(u)\}$ for $0 \leq i \leq n-1$. In particular, for $n = 1$ we have $w = u$.

The following Proposition relates the number of occurrences of a letter a and $\theta(a)$ in a word to the number of occurences of a and $\theta(a)$ of its $\circ$-power.

Proposition 3 If $\circ$ is θ-*propagating* bw-operation, then for any $w \in u^{\circ(n)}$, $|w|_{a,\theta(a)} = n \cdot |u|_{a,\theta(a)}$, for $n \geq 1$.

Lemma 7 *If $\circ$ is an associative bw-operation and $L \subseteq \Sigma^*$, $L \neq \emptyset$, we have*

$$L^{\circ(m)} \circ L^{\circ(n)} = L^{\circ(m+n)} \text{ for } m, n \geq 1.$$

Proof.

$$\begin{aligned}
L^{\circ(m+n)} &= L^{\circ(m+(n-1))} \circ L \\
&= (L^{\circ(m+(n-2))} \circ L) \circ L \\
&= L^{\circ(m+(n-2))} \circ (L \circ L) \\
&= L^{\circ(m+(n-2))} \circ L^{\circ(2)} \\
&= L^{\circ(m+(n-3))} \circ L^{\circ(3)} = ... \\
&= L^{\circ(m)} \circ L^{\circ(n)}.
\end{aligned}$$

□

Lemma (7) does not hold in general for operations that are not associative. However, in the case of θ-*catenation*, when θ is an antimorphic involution, one of the inclusions in Lemma (7) holds, even though θ-*catenation* is not right- or left-inclusive. As a consequence, as seen in Corollary (9), θ-*catenation* is plus-closed.

Lemma 8 *If $\odot$ is the operation of θ-catenation associated with any morphic or antimorphic involution θ of Σ^* and $L \subseteq \Sigma^*$ is a nonempty language, then*

$$L^{\odot(m)} \odot L^{\odot(n)} \subseteq L^{\odot(m+n)},\ \forall m, n \geq 1.$$

Proof. If θ is a morphic involution then the operation of θ-*catenation* is associative and the inclusion holds by Lemma (7).

If θ is an antimorphic involution then, by Lemma (5), for every $n \geq 1$ we have

$$L^{\odot(n)} = \{uv_1v_2\cdots v_{n-1} | u \in L, v_i \in L \cup \theta(L), 0 \leq i \leq n-1\}.$$

Let $x \in L^{\odot(m)}$ and $y \in L^{\odot(n)}$ for some $m, n \geq 1$. Then by Corollary (5) $x = uv_1v_2\cdots v_{m-1}$ and $y = u'v'_1v'_2\cdots v'_{n-1}$ for some $u, u' \in L$, $v_i, v'_i \in L \cup \theta(L)$, $0 \leq i \leq m-1$ and $0 \leq j \leq n-1$. By the definition of *θ-catenation*,

$$x \odot y = \{uv_1v_2\cdots v_{m-1}u'v'_1v'_2\cdots v'_{n-1}, uv_1v_2\cdots v_{m-1}\theta(v'_{n-1})\ldots\theta(u')\},$$

which is a word in $L^{\odot(m+n)}$.

□

Corollary 9 *The operation of θ-catenation is plus-closed for morphic as well as antimorphic involutions θ.*

A non-empty language $L \subseteq \Sigma^*$ is called $\circ$-*free* if $(L^{\circ(+)} \circ L) \cap L = \emptyset$. In the case of *$\theta$-catenation*, for example, if $L \subseteq \Sigma^*$ and $R = \{uv_1v_2...v_k | u \in L, v_i \in L \cup \theta(L), k \geq 1, 1 \leq i \leq k\}$ then, if $L \cap R = \emptyset$, L is $\odot$-free. The following lemma provides more examples of $\odot$-free languages.

Proposition 4 Given a morphic or antimorhic involution θ over Σ, and the operation $\odot$ (*θ-catenation*), any prefix code is $\odot$-free.

Proof. Let $L \subseteq \Sigma^*$ be a prefix code, and assume that L is not $\odot$-free. Then there exist $w \in L$, $u \in L^{\odot(+)}$ and $v \in L$ such that $w \in u \odot v = \{uv, u\theta(v)\}$. By the definition of *$\theta$-catenation* and Lemma (5), w is of the form $\alpha\beta_1\beta_2\ldots\beta_{n-1}v$ or $\alpha\beta_1\beta_2\ldots\beta_{n-1}\theta(v)$, where $\alpha \in L$ and $\beta_i \in L \cup \theta(L)$, $1 \leq i \leq n-1$, $n \geq 2$. This is a contradiction to the fact that L is a prefix code.

□

The converse of the previous Proposition does not hold, as shown by the following example.

Example Let $\Sigma = \{A, G, C, T\}$, $\theta(A) = T, \theta(G) = C$, $L = \{AG, TT, AGCA\}$. The language L is $\odot$-free, but not a prefix code.

Another way of obtaining $\odot$-free languages is given by means of the left θ-quotient. The *left θ-quotient* of two languages $L_1, L_2 \subseteq \Sigma^*$ is defined as

$$L_1 \lhd_{\odot} L_2 = \{w \in \Sigma^* | \ (L_2 \odot w) \cap L_1 \neq \emptyset\}.$$

Lemma 10 *If θ is a morphic involution then the left θ-quotient is given by*

$$u \lhd_{\odot} v = \{v^{-1}u, \theta(v)^{-1}\theta(u)\}$$

and if θ is an antimorphic involution then the left θ-quotient is given by

$$u \lhd_{\odot} v = \{v^{-1}u, \theta(u)\theta(v)^{-1}\}.$$

Proof. Let θ be a morphic involution and let $w \in (u \lhd_{\odot} v)$. This implies $(v \odot w) \cap \{u\} \neq \emptyset$, that is $u \in \{vw, v\theta(w)\}$, which further implies $w \in \{v^{-1}u, \theta(v)^{-1}\theta(u)\}$.

Let θ be an antimorphic involution and let $w \in (u \lhd_{\odot} v)$. This implies $(v \odot w) \cap \{u\} \neq \emptyset$, that is $u \in \{vw, v\theta(w)\}$, which further implies $w \in \{v^{-1}u, \theta(u)\theta(v)^{-1}\}$.

□

Lemma 11 *Let θ be a morphic or antimorphic involution over Σ and let L be a language in Σ^*. If L closed under left θ-quotient then L is not $\odot$-free.*

Proof. $\lhd_{\odot}(L, L) = \{w \in \Sigma^* | (L \odot w) \cap L \neq \emptyset\}$. As L is $\lhd_{\odot}$-closed, $\lhd_{\odot}(L, L) \subseteq L$, which implies that $(L \odot L) \cap L \neq \emptyset$ which, since $L \subseteq L^{\odot(+)}$, further implies that L is not $\odot$-free.

□

4. θ-Primitive Words

In this section we show that if the operation under consideration is θ-catenation, denoted by $\odot$, then the $\odot$- primitive words coincide with the θ-primitive words defined in section (2). We study some properties of such θ-primitive words. Recall the following result from [12].

Proposition 5 [12] Let $\circ$ be plus-closed and length-increasing. Then for every word $w \in \Sigma^+$ there exists a $\circ$-primitive word u and an integer $n \geq 1$ such that $w \in u^{\circ(n)}$.

The following results (Proposition 6, Lemma 13, and Proposition 7) are similar to analogous results in [26], involving propagating bw-operations.

Proposition 6 Let $\circ$ be plus-closed and θ-propagating. Then for every word $w \in \Sigma^+$ there exists a $\circ$-primitive word u and a unique integer $n \geq 1$ such that $w \in u^{\circ(n)}$.

Proof. Every θ-propagating bw-operation is length-increasing. Now, by Proposition (5), for every word $w \in \Sigma^+$ there exists a $\circ$-primitive word u and an integer $n \geq 1$ such that $w \in u^{\circ(n)}$. Consider $a \in \Sigma$ such that $|u|_{a,\theta(a)} \neq 0$. Since $\circ$ is θ-propagating, for any $w_1 \in u^{\circ(m)}$ with $m \neq n$, by Proposition (3), we get $|w_1|_{a,\theta(a)} = m|u|_{a,\theta(a)} \neq n|u|_{a,\theta(a)} = |w|_{a,\theta(a)}$. Thus $w \notin u^{\circ(m)}$ for any $m \neq n$. Hence n is such an unique integer.

□

A $\circ$-primitive word $u \in \Sigma^+$ such that $w \in u^{\circ(n)}$ for some $n \geq 1$, is called a $\circ$-root of w. In general, a word may not have a unique $\circ$-root. However, if $\circ$ is the operation of θ-catenation, then every word $w \in \Sigma^+$ has an unique $\odot$-root, also called θ-root, denoted by $\rho_\theta(w)$. The uniqueness of the θ-root of a word was demonstrated by the following theorem (corollary of Theorems 13 and 14 from [6]).

Theorem 12 *If θ is a morphic or antimorphic involution on Σ^* then for any word $w \in \Sigma^+$ there exists a unique θ-primitive word $t \in \Sigma^+$ such that $w \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(w) = t$.*

Lemma 13 *Let Σ be an alphabet with $|\Sigma| \geq 2$ and $\circ$ be plus-closed and θ-propagating bw-operation. If a word $w \in \Sigma^+$ is not $\circ$-primitive, then for any $a \neq b$, $a, b \in \Sigma$ we have that $|w|_{a,\theta(a)}$ and $|w|_{b,\theta(b)}$ have a common factor $n > 1$.*

Proof. If w is not $\circ$-primitive, then according to Proposition (5), $w \in u^{\circ(n)}$ for some $\circ$-primitive word $u \in \Sigma^+$ and $n > 1$. Since $\circ$ is θ-propagating and Proposition (3) holds, $|w|_{a,\theta(a)} = n \cdot |u|_{a,\theta(a)}$ for all $a \in \Sigma$. Similarly, $|w|_{b,\theta(b)} = n \cdot |u|_{b,\theta(b)}$. Hence, for any $a, b \in \Sigma$, we have that $|w|_{a,\theta(a)}$ and $|w|_{b,\theta(b)}$ have the common factor $n > 1$.

□

Proposition 7 Let Σ be an alphabet with $|\Sigma| \geq 3$ and $\circ$ be plus-closed and θ-propagating bw-operation. If $w \in \Sigma^+$, $a \in \Sigma$, $w \notin \{a, \theta(a)\}^+$, then there is an integer $m \geq 1$ such that all the words $v_1 \in (w \circ w^{m-1}a)$, $v_2 \in (aw^{m-1} \circ w)$, $v_3 = w^m a$ and $v_4 = aw^m$ are $\circ$-primitive.

Proof. For $w \in \Sigma^+$, let $m = \prod_{b \in \Sigma, |w|_{b,\theta(b)} \neq 0} |w|_{b,\theta(b)}$. For any $a \in \Sigma$, suppose $w \notin \{a, \theta(a)\}^+$. Such a word exists since $|\Sigma| \geq 3$. Let $v_1 \in (w \circ w^{m-1}a)$, $v_2 \in (aw^{m-1} \circ w)$, $v_3 = w^m a$ and $v_4 = aw^m$. If $b \notin \{a, \theta(a)\}$ is a letter occurring in w, $|v_1|_{a,\theta(a)} = |v_2|_{a,\theta(a)} = |v_3|_{a,\theta(a)} = |v_4|_{a,\theta(a)} = m \cdot |w|_{a,\theta(a)} + 1$ whereas $|v_1|_{b,\theta(b)} = |v_2|_{b,\theta(b)} = |v_3|_{b,\theta(b)} = |v_4|_{b,\theta(b)} = m \cdot |w|_{b,\theta(b)}$. As the number of occurrences of a together with $\theta(a)$ respectively the number of occurrences of b together with $\theta(b)$ in each v_i, $i = 1, 2, 3, 4$, are relatively prime, by Lemma (13), v_1, v_2, v_3 and v_4 are $\circ$-primitive words.

□

In the remainder of the section we will investigate some properties of θ-primitive words.

Definition 2 [12] A language $L \subseteq \Sigma^*$ is called right-$\circ$-dense (resp. left-$\circ$-dense) if for each $w \in \Sigma^+$, there exists $u \in \Sigma^*$ such that $(w \circ u) \cap L \neq \emptyset$ (resp. $(u \circ w) \cap L \neq \emptyset$).

If $\circ$ is the catenation of words, then the right and left $\circ$-dense languages are called right and left dense languages, respectively. Let $Q_\circ(\Sigma)$ denote the set of all $\circ$-primitive words over Σ.

Proposition 8 If Σ is an alphabet with $|\Sigma| \geq 3$ and $\circ$ is plus-closed and θ-propagating bw-operation, then $Q_\circ(\Sigma)$ is right and left $\circ$-dense.

Proof. For each $w \in \Sigma^+$, since $|\Sigma| \geq 3$, there exists $a \in \Sigma$ such that $w \notin \{a, \theta(a)\}^+$. As $\circ$ is plus-closed and θ-propagating, by Proposition (7), there exists $m \geq 1$, such that $(w \circ w^{m-1}a) \in Q_\circ(\Sigma)$ and $(aw^{m-1} \circ w) \in Q_\circ(\Sigma)$. This proves that $Q_\circ(\Sigma)$ is right and left $\circ$-dense.

□

Next, we show that the set of θ-primitive words $Q_\odot(\Sigma)$ is right and left dense.

Proposition 9 Let the operation of θ-catenation $\odot$ associated to morphic or anti-morphic involution θ be plus-closed θ-propagating and let $|\Sigma| \geq 3$. Then $Q_{\odot}(\Sigma)$ is right and left dense.

Proof. Let $w \in \Sigma^{+}$. If $w \in \{a, \theta(a)\}^{+}$ and $b \in \Sigma$ such that $b \notin \{a, \theta(a)\}$, then, $|wb|_{a,\theta(a)} = |bw|_{a,\theta(a)} = m \geq 1$. Also, $|wb|_{b,\theta(b)} = |bw|_{b,\theta(b)} = 1$, hence by Lemma (13) $wb \in Q_{\odot}(\Sigma)$ and $bw \in Q_{\odot}(\Sigma)$. If $w \notin \{a, \theta(a)\}^{+}$, then by Proposition (7), $w^{m}a \in Q_{\odot}(\Sigma)$ and $aw^{m} \in Q_{\odot}(\Sigma)$ for some $m \geq 1$. This proves that $Q_{\odot}(\Sigma)$ is right and left dense.

□

Proposition 10 Let $\circ$ be a plus-closed and θ-propagating bw-operation and $L \subseteq \Sigma^{+}$ a non-empty $\circ$-closed language such that L^{c} is also $\circ$-closed. Let $F(L)$ be the set of length-minimal words of L and $P_{\circ}(L) = L \cap Q_{\circ}(\Sigma)$. Then

1. If $w \in L$ and if u is a $\circ$-root of w, then $u \in L$.
2. If $L^{'}$ is a $\circ$-closed language containing $P_{\circ}(L)$, then $L \subseteq L^{'}$.
3. Every word $w \in F(L)$ is $\circ$-primitive.

Proof.
1. Since u is a $\circ$-root of w, $w \in u^{\circ(n)}$, for some $n \geq 1$. If $u \in L^{c}$, then, since L^{c} is $\circ$-closed, $u^{\circ(n)} = (u \circ^{n-1} u) \subseteq L^{c}$ and therefore, $w \in L^{c}$, which is a contradiction. Hence $u \in L$.
2. Let $w \in L$, then there are two possibilities, either $w \in P_{\circ}(L)$ or $w \notin P_{\circ}(L)$. If $w \in P_{\circ}(L)$, then $w \in L^{'}$ as $P_{\circ}(L) \subseteq L^{'}$. If $w \notin P_{\circ}(L)$ then w is not $\circ$-primitive. That means there exists a $\circ$-primitive word u and $n \in \mathbb{N}$ such that $w \in u^{\circ(n)}$. But as u is $\circ$-primitive, $u \in P_{\circ}(L) \subseteq L^{'}$, so $w \in L^{'}$. So, we have showed that in both cases $L \subseteq L^{'}$.
3. Assume that $w \in F(L)$ is not $\circ$-primitive. Then by Proposition (5), $w \in u^{\circ(n)}$, for some $\circ$-primitive word u and $n > 1$. By (1), $u \in L$.
   Case 1: There is no $a \in \Sigma$ such that $\theta(a) = a$. Then, as Proposition (3) holds true,
   $$|w| = \frac{1}{2} \sum_{a \in \Sigma, a \neq \theta(a)} |w|_{a,\theta(a)} > \frac{1}{2} \sum_{a \in \Sigma, a \neq \theta(a)} |u|_{a,\theta(a)} = |u|$$
   which contradicts the fact that $w \in F(L)$.
   Case 2: There exists $a \in \Sigma$ such that $\theta(a) = a$. Then as Proposition (3) holds true,
   $$\begin{aligned} |w| &= \sum_{a \in \Sigma, a = \theta(a)} |w|_{a,\theta(a)} + \frac{1}{2} \sum_{a \in \Sigma, a \neq \theta(a)} |w|_{a,\theta(a)} \\ &> \sum_{a \in \Sigma, a = \theta(a)} |u|_{a,\theta(a)} + \frac{1}{2} \sum_{a \in \Sigma, a \neq \theta(a)} |u|_{a,\theta(a)} = |u| \end{aligned}$$
   which contradicts the fact that $w \in F(L)$.

□

5. Closure Properties and Language Equations

In this section we will briefly discuss the closure properties of families of languages under θ-*catenation* and explore language equations involving this operation.

Proposition 11 The families of regular, context-free and context-sensitive languages are closed under the operation of θ-*catenation*.

Binary word operations can be extended naturally to binary language operations by defining,

$$L_1 \diamond L_2 = \bigcup_{u \in L_1, v \in L_2} (u \diamond v).$$

Language equations of type $L \diamond Y = R$ and $X \diamond L = R$, where $\diamond$ is an invertible binary word operation and L and R are two given languages have been extensively studied, e.g., in [15]. Finding the solutions to such equations involves the concept of "right inverse" and "left inverse" of an operation.

Definition 3 [15] Let $\circ$ and $\diamond$ be two binary word operations. The operation $\diamond$ is said to be the right-inverse of the operation $\circ$ if for all words u, v, w over the alphabet Σ the following relation holds:

$$w \in (u \circ v) \text{ iff } v \in (u \diamond w).$$

Definition 4 [15] Let $\circ$ and $\bullet$ be two binary word operations. The operation $\bullet$ is said to be the left-inverse of the operation $\circ$ if for all words u, v, w over the alphabet Σ, the following relation holds:

$$w \in (u \circ v) \text{ iff } u \in (w \bullet v).$$

Proposition (12) and (13) find the right and left inverses of θ-*catenation* for θ morphic as well as antimorphic. Given a bw-operation $\circ$, the reverse of this operation, denoted by $\circ'$, is defined as

$$u \circ' v = v \circ u.$$

Proposition 12 If θ is a morphic or antimorphic involution then the right-inverse of the operation of θ-catenation $\odot$ is the reverse left θ-quotient.

Proof. Let θ be a morphic involution, and let $w \in u \odot v$. Then either $w = uv$ or $w = u\theta(v)$. By the definition of left quotient, $w = uv$ implies that $v = u^{-1}w$. Also, $w = u\theta(v)$ which implies that $\theta(w) = \theta(u)v$ and thus that $v = \theta(u)^{-1}\theta(w)$. This shows that $v \in \{u^{-1}w, \theta(u)^{-1}\theta(w)\} = u \lhd'_{\odot} w$. The converse is similar.

Let θ be an antimorphic involution and let $w \in u \odot v$. Then either $w = uv$ or $w = u\theta(v)$. By the definition of left quotient, $w = uv$ implies that $v = u^{-1}w$. Also, $w = u\theta(v)$ implies that $\theta(w) = v\theta(u)$. Then, by the definition of right quotient, $\theta(w) = v\theta(u)$ which implies that $v = \theta(w)\theta(u)^{-1}$. This shows that $v \in \{u^{-1}w, \theta(w)\theta(u)^{-1}\} = u \lhd'_{\odot} w$. The converse is similar.

□

Proposition 13 Let θ be a morphic or antimorphic involution, and let the binary word operation $\bullet$ be defined as $w \bullet v = \{wv^{-1}, w\theta(v)^{-1}\}$. Then θ-catenation and $\bullet$ are left inverses of each other.

Proof. Let $w \in u \odot v$. Then either $w = uv$ or $w = u\theta(v)$. By definition of right quotient, $w = uv$ implies $u = wv^{-1}$. Also, $w = u\theta(v)$ implies $u = w\theta(v)^{-1}$. This shows that $u \in \{wv^{-1}, w\theta(v)^{-1}\} = w \bullet v$. The converse is similar.

□

The preceding results provide tools to solve language equations involving the operation of θ-catenation. The following two propositions are consequences of more general results from [15].

Proposition 14 Let L, R be languages over an alphabet Σ. If the equation $L \odot Y = R$ has a solution Y, then the language $R^{'} = (L \lhd^{\prime}_{\odot} R^c)^c$ is also a solution of the equation. Moreover, $R^{'}$ includes all the other solutions of the equation (set inclusion).

Corollary 14 *Let L be a language in Σ^*. If the equation $L \odot Y = L$ has a solution, then $\rho_{\odot}(L)$, the set of all right $\odot$-residuals of L is a solution, which moreover includes all the other solutions to the equation.*

Proof. By the previous proposition, if a solution to the equation $L \odot Y = L$ exists, then also $R^{'} = (L \lhd^{\prime}_{\odot} L^c)^c = (L^c \lhd_{\odot} L)^c$ is a solution. By a result in [12], for any language $L \subseteq \Sigma^*$ and bw-operation $\circ$, the set of all right $\circ$ residuals of L, denoted by $\rho_{\circ}(L)$, equals $(\lhd_{\circ}(L^c, L))^c$, which proves the statement of the corollary.

□

Proposition 15 Let L,R be languages over an alphabet Σ. If the equation $X \odot L = R$ has a solution $X \subseteq \Sigma^*$, then also the language $R^{'} = (R^c \lhd^{\prime}_{\odot} L)^c$ is a solution of the equation. Moreover, $R^{'}$ includes all the other solutions of the equation (set inclusion).

6. Conclusions and future work

This paper proposes and investigates the operation of θ-catenation, that generates the pseudo-powers (θ-powers) of a word. An avenue of further research is to determine and investigate operations that generate other types of generalized powers. One such type is the Abelian power, [8] defined as follows.

A word w is a *k-th Abelian power* if $w = u_1 u_2 \cdots u_k$ for some $u_1, u_2, \cdots u_k$, $u_i \in \Sigma^+$, $1 \leq i \leq k$, such that for all $1 \leq i, j \leq k$, $\pi(u_i) = \pi(u_j)$, where $\pi(u)$ denotes the set of all words obtained by permuting the letters of u. A word w is *Abelian primitive* if w fails to be a k-th Abelian power for every $k \geq 2$. A word u is an *Abelian root* of w if $w = uu_1u_2 \cdots u_{k-1}$ for some $u_1 \cdots u_{k-1} \in \Sigma^+$ with $\pi(u) = \pi(u_i)$ for all $1 \leq i \leq k-1$. Unlike words that are not primitive or not θ-primitive, a word that is not Abelian primitive may have several Abelian roots.

We can now define a bw-operation $\boxdot$, called *Abelian-catenation*, as $u \boxdot v = u\pi(v)$. For example, if we consider the alphabet $\Sigma = \{a, b, c\}$ and the words $u = acba$ and $v = bcc$, then

$$u \boxdot v = \{acbabcc, acbacbc, acbaccb\}.$$

The operation of *Abelian-catenation* is length-increasing as well as propagating, but its neither left-inclusive nor right-inclusive and therefore is not plus-closed.

Note that the operation of Abelian-catenation generates *Abelian-powers*. Indeed, if $w \in u^{\boxdot(k)}$, for $k \geq 1$, then $w = uv_1v_2 \cdots v_{k-1}$, where $v_i \in \{\pi(u)\}$, for $1 \leq i \leq k-1$.

References

[1] A. Carpi, A. de Luca, Periodic-like words, periodicity, and boxes. *Acta Informatica* **37** (2001) 8, 597–618.

[2] S. Constantinescu, L. Ilie, Fine and Wilf's theorem for Abelian periods. *Bulletin of the EATCS* **89** (2006), 167–170.

[3] M. Crochemore, C. Hancart, T. Lecroq, *Algorithms on Strings*. Cambridge University Press, 2007.

[4] M. Crochemore, W. Rytter, *Jewels of Stringology*. World Scientific, 2002.

[5] L. J. Cummings, W. F. Smyth, Weak repetitions in strings. *J. Combinatorial Mathematics and Combinatorial Computing* **24** (1997), 33–48.

[6] E. Czeizler, L. Kari, S. Seki, On a special class of primitive words. *Theoretical Computer Science* **411** (2010), 617 – 630.

[7] A. de Luca, A. de Luca, Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science* **362** (2006) 13, 282 – 300.

[8] M. Domaratzki, N. Rampersad, Abelian primitive words. In: G. Mauri, A. Leporati (eds.), *Developments in Language Theory*. Lecture Notes in Computer Science 6795, Springer Berlin Heidelberg, 2011, 204–215.

[9] P. Dömösi, G. Horváth, M. Ito, K. Shikishima-Tsuji, Some periodicity of words and Marcus contextual grammars. *Vietnam Journal of Mathematics* **34** (2006), 381–387.

[10] P. Gawrychowski, F. Manea, R. Mercaş, D. Nowotka, C. Tiseanu, Finding pseudo-repetitions. *Leibniz International Proceedings in Informatics* **20** (2013), 257–268.

[11] J. E. Hopcroft, J. D. Ullman, *Formal Languages and their Relation to Automata*. Addison-Wesley Longman Inc., 1969.

[12] H. K. Hsiao, C. C. Huang, S. S. Yu, Word operation closure and primitivity of languages. *J.UCS* **8** (2002) 2, 243–256.

[13] S. Hussini, L. Kari, S. Konstantinidis, Coding properties of DNA languages. In: N. Jonoska, N. Seeman (eds.), *Proc. of DNA7*. Lecture Notes in Computer Science 2340, Springer, 2002, 57–69.

[14] M. Ito, G. Lischke, Generalized periodicity and primitivity for words. *Mathematical Logic Quarterly* **53** (2007) 1, 91–106.

[15] L. Kari, On language equations with invertible operations. *Theoretical Computer Science* **132** (1994), 129–150.

[16] L. Kari, S. Konstantinidis, P. Sosík, Bond-free languages: Formalizations, maximality and construction methods. *International Journal of Foundations of Computer Science* **16** (2005), 1039–1070.

[17] L. Kari, E. Losseva, S. Konstantinidis, P. Sosík, G. Thierrin, A formal language analysis of DNA hairpin structures. *Fundamenta Informaticae* **71** (2006), 453–475.

[18] L. Kari, K. Mahalingam, Watson-Crick conjugate and commutative words. In: M. H. Garzon, H. Yan (eds.), *Proc. of DNA13*. Lecture Notes in Computer Science 4848, Springer-Verlag, 2008, 273–283.

[19] L. Kari, S. Seki, On pseudoknot-bordered words and their properties. *Journal of Computer and System Sciences* **75** (2009), 113 − 121.

[20] L. Kari, S. Seki, An improved bound for an extension of Fine and Wilf's theorem and its optimality. *Fundamenta Informaticae* **101** (2010), 215–236.

[21] L. Kari, G. Thierrin, Word insertions and primitivity. *Utilitas Mathematica* **53** (1998), 49–61.

[22] G. Lischke, Primitive words and roots of words. *Acta Universitatis Sapientiae* **3** (2011), 5–34.

[23] M. Lothaire, *Combinatorics on Words*. Cambridge University Press, 1997.

[24] R. C. Lyndon, M. P. Schutzenberger, The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.* **9** (1962), 289–298.

[25] G. Paun, G. Rozenberg, T. Yokomori, Hairpin languages. *Int. J. Found. Comput. Sci.* **12** (2001), 837–847.

[26] S.-S. Yu, *Languages and Codes*. Tsang Hai Book Publishing Co., 2005.

[27] J. Ziv, A. Lempel, A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory* **23** (1977) 3, 337–343.