International Journal of Foundations of Computer Science

Multipseudoperiodic words

Sarah Desmeules, Alexandre Blondin Massé, Sylvain Hallé and Sébastien Gaboury

Laboratoire d'informatique formelle
,Département d'informatique et mathématique,
Université du Québec à Chicoutimi, 555, boul. de l'Université,
Chicoutimi, Québec G7H 2B1, Canada
zeldasolutions@hotmail.com; ablondin@uqac.ca; shalle@uqac.ca; sgabour1@uqac.ca
http://thales.math.uqam.ca/ blondin

Received (Day Month Year)
Accepted (Day Month Year)
Communicated by (xxxxxxxxxx)

We consider words over an arbitrary alphabet admitting multiple pseudoperiods according to permutations. We describe the conditions under which such a word exists. Moreover, a natural generalization of Fine and Wilf's Theorem is proved. Finally, we introduce and describe a new family of words sharing properties with the so-called *central* words. In particular, under some simple conditions, we prove that these words are pseudopalindromes, a result consistent with the fact that central words are palindromes.

Keywords: Pseuperiods; Fine and Wilf's Theorem, Permutations.

1. Introduction

The study of *periodic functions*, in particular trigonometric functions, goes back to medieval times. It is widely known that the understanding of periodic functions is fundamental in many areas of physics that range from signal processing to economics and mechanics. Particular cases of periodic functions are their discrete counterpart, the so-called *periodic sequences* or *periodic words*. They are of great interest in bioinformatics since repetitions in DNA sequences reveal critical structural information [11].

In 1965, Fine and Wilf published their article "Uniqueness theorems for periodic functions" [10], which answered the following question: What is the minimum length a finite sequence admitting two periods p and q must have so that it also admits $\gcd(p, q)$ as a period? They proved that $p + q - \gcd(p, q)$ is sufficient and that this bound is tight. In the next five decades, Fine and Wilf's result has been extensively studied and its applications are numerous. For instance, it turns out that the worst cases of the well-known Knuth-Morris-Pratt string search algorithm are exactly the maximal counter-examples (also called *central words*) of Fine and Wilf's Theorem [15].

An exhaustive survey of the consequences and applications of Fine and Wilf's Theorem may be found in [20]. It has been generalized to more than two periods [3, 14, 12, 4], to multi-dimensional words [21, 19] and to pseudoperiods [24, 5, 2]. Words of maximum length admitting periods p and q but not $\gcd(p, q)$ have been studied as well. In a sequence of two articles [24, 23], Tijdeman and Zamboni provide an algorithm to generate and prove that these words are palindromes. An alternate proof of this latter fact can be found in [12] as well.

In 1994, De Luca and Mignosi considered the palindromes occurring in standard Sturmian words and introduced the so-called family of *central words*. They proved that those words are exactly the maximum ones admitting two periods p and q, with $\gcd(p, q) = 1$, but not with period 1 [8]. In [6], De Luca studied words generated by *iterated palindromic closure* — an operator that allows one to generate infinite words having infinitely many palindromic prefixes — and proved that central words are obtained by iterated palindromic closure. More recently, de Luca and De Luca defined a family of words called *generalized pseudostandard words* that are generated by *iterated pseudopalindromic closure*. In [2], the authors have shown that generalized pseudostandard words present *pseudoperiodic* properties. The main purpose of this article is to further study this concept of *pseudoperiod*. It shall also be noted that this is an extended journal version of a conference paper presented to Development in Language Theory held at Taipei, Taiwan in August, 2012 [18].

2. Definitions and notation

In this section, we introduce the definitions and notation used in the following sections.

2.1. *Words*

All the basic terminology about words is taken from M. Lothaire [16, 17]. In the following, Σ is a finite *alphabet* whose elements are called *letters*. By *word* we mean a finite sequence of letters $w : [0..n-1] \to \Sigma$, where $n \in \mathbb{N}$. The length of w is $|w| = n$ and $w[i]$ denotes its i-th letter. The set of words of length n over Σ is denoted Σ^n. The *empty* word is denoted by ε and its length is 0.

The free monoid generated by Σ is defined by $\Sigma^* = \bigcup_{n \geq 0} \Sigma^n$. The k-th power of w is defined recursively by $w^0 = \varepsilon$ and $w^k = w^{k-1}w$. Given a word $w \in \Sigma^*$, a *factor* u of w is a word $u \in \Sigma^*$ such that $w = xuy$, with $x \in \Sigma^*$ and $y \in \Sigma^*$. If $x = \varepsilon$ (resp. $y = \varepsilon$) then u is called a *prefix* (resp. *suffix*). The set of all factors of w is denoted by $\mathrm{Fact}(w)$, and $\mathrm{Pref}(w)$ is the set of all its prefixes. An *antimorphism* is a map $\varphi : \Sigma^* \to \Sigma^*$ such that $\varphi(uv) = \varphi(v)\varphi(u)$ for any word $u, v \in \Sigma^*$. A useful example is the *reversal* of $w \in \Sigma^n$ defined by $\widetilde{w} = w_{n-1}w_{n-2}\cdots w_0$. It is also convenient to denote the reversal operator by R. A *palindrome* is a word that reads the same forward and backward, i.e. a word w such that $w = \widetilde{w}$.

2.2. *Permutations*

A function $\sigma : \Sigma \to \Sigma$ is called a *permutation* if it is a bijection. We recall that permutations can be decomposed as a product of cycles. As usual, we shall use the *cycle notation*. For instance, $\sigma = (021)(3)$ means that $0 \mapsto 2$, $2 \mapsto 1$, $1 \mapsto 0$ and $3 \mapsto 3$. The *identity* permutation is denoted by I. The inverse of a permutation σ is denoted by σ^{-1} while σ^n stands for $\sigma \cdot \sigma \cdot \ldots \cdot \sigma$ (n times), where the product corresponds to the composition of functions. The *order* of a permutation σ, denoted by $\mathrm{ord}(\sigma)$ is the least integer n such that $\sigma^n = I$. When the order is 2, the permutation is called an *involution*.

2.3. *Pseudopalindromes*

Recently, several works have been devoted to the study of pseudopalindromes [7], a natural generalization of palindromes. Given an involutory antimorphism ϑ, the word w is called *ϑ-palindrome* if $\vartheta(w) = w$. When $\vartheta = R$, the definition coincides with that of usual palindromes. It is easy to see that $\vartheta = R \circ \sigma$ for some involutory permutation σ.

In [7], the authors introduce an operator called *pseudopalindromic closure*, which generalize the so-called *palindromic closure* [7]. Let ϑ be an involutory antimorphism, w be a word, with $w = up$ where p is the longest ϑ-palindromic suffix of w (it exists since ε is an ϑ-palindrome). Then the *ϑ-palindromic closure of w* is defined by $w^{\oplus_\vartheta} = up\vartheta(u)$. In other words, $w^{\oplus_\vartheta}$ is the shortest ϑ-palindrome having w as a prefix.

2.4. *Generalized pseudostandard words*

Also in [7], the authors introduce a family of words called *generalized pseudostandard words* generated by pseudopalindromic closure. More precisely, let $\Theta = \vartheta_1\vartheta_2\vartheta_3 \cdots \vartheta_n$ be a finite sequence of involutory antimorphisms and w be a word of length n. Let $^{\oplus_i}$ be the ϑ_i-palindromic closure operator, for $1 \le i \le n$. The operator ψ_Θ is defined as follows:

$$\psi_{\vartheta_1\vartheta_2\cdots\vartheta_n}(w) = \begin{cases} \varepsilon & \text{if } n = 0; \\ \left(\psi_{\vartheta_1\vartheta_2\cdots\vartheta_{n-1}}(w_1w_2\cdots w_{n-1})w_n\right)^{\oplus_n} & \text{otherwise.} \end{cases}$$

This definition extends naturally to infinite words.

Example 1. [7] *Let $\Sigma = \{0, 1\}$, R be the reversal operator and E be the antimorphism that swaps the letters 0 and 1. Then the Thue-Morse word is exactly*

$\psi_{(ER)^\omega}(01^\omega)$.

$$\begin{aligned}\psi_E(0) &= 01\\ \psi_{ER}(01) &= 0110\\ \psi_{ERE}(011) &= 01101001\\ \psi_{ERER}(0111) &= 0110100110010110\\ &\vdots\end{aligned}$$

2.5. *Pseudoperiodicity*

Not much is known about generalized pseudostandard words. In [2], the authors provide an efficient and nontrivial algorithm to generate all such words in the binary case. The key idea of their approach is to use the pseudoperiods induced by the overlapping of the successive pseudopalindromic prefixes corresponding to the iterated pseudopalindromic closure. For example, the fourth Thue-Morse prefix in the previous example is pseudoperiodic since 01101001 is followed by 10010110, which is the same word with the 0's and 1's swapped.

A *period* of a word w is an integer k such that $w[i] = w[i+k]$, for all $i < |w| - k$. In particular, every $k \geq |w|$ is a period of w. An important result about periods is due to Fine and Wilf.

Theorem 2 (Fine and Wilf [10]) *Let w be a word having p and q for periods. If $|w| \geq p + q - \gcd(p,q)$, then $\gcd(p,q)$ is also a period of w.*

A natural generalization of period is the following:

Definition 3. [2] *Let Σ be an alphabet, w some word over Σ, σ some involutory permutation over Σ and p some positive integer. Then p is called a σ-period of w if $w[i+p] = \sigma(w[i])$. for $i = 1, 2, \ldots, n-p$.*

It is well-known that overlapping palindromes yield periodicity [6, 1]. One can ask what is the result of overlapping pseudopalindromes. This question is answered in [2] as follows:

Proposition 4. [2] *Let u be a finite word, v be a ϑ_1-palindrome and w be a ϑ_2-palindrome for some involutory antimorphisms ϑ_1 and ϑ_2 such that $uv = w$. Then q has the $(\sigma_1 \circ \sigma_2)$-period $|u|$, where σ_1 and σ_2 are the permutations associated with ϑ_1 and ϑ_2.*

It shall be noted that an alternate, non equivalent definition of pseudoperiods have already been introduced in [5] as follows: Let w be some word over an alphabet Σ and ϑ be an involutory antimorphism. Then the positive integer p is called a ϑ-*period* of the word w if $w = u_1 u_2 \cdots u_n$, where $|u_i| = p$ and $u_i \in \{w, \vartheta(w)\}$ for $i = 1, 2, \ldots, n$.

Example 5. *Let $\Sigma = \{A, C, G, T\}$ and ϑ be the antimorphism such that $\vartheta(A) = T$ and $\vartheta(C) = G$. Then 4 is a ϑ-period of the word*

$$ACAGCTGTCTGTACAGACAG = u \cdot \vartheta(u) \cdot \vartheta(u) \cdot u \cdot u,$$

where $u = ACAG$.

In this paper, we only consider pseudoperiods as in Definition 3.

2.6. *Graphs*

A *graph* is a couple $G = (V, E)$, where $E \subseteq \mathcal{P}_2(V)$ is the set of unordered pairs of elements in V. The elements of the sets V and E are called respectively *vertices* and *edges*. We say that the edge $e \in E$ is *incident* to the vertex $v \in V$ if $e = \{u, v\}$ for some vertex u. The *degree* of a vertex v is its number of incident edges, i.e. $\deg(v) = \mathrm{Card}\{u \in V \mid \{u, v\} \in E\}$. A *path* of G is a sequence of vertices $(v_1, v_2, \ldots, v_k)$, where k is a nonnegative integer, and such that $\{v_i, v_{i+1}\} \in E$ for $i = 1, 2, \ldots, k-1$. Let $\sim$ be the relation on V defined by $u \sim v$ if and only if there exists a path between u and v. Clearly, $\sim$ is an equivalence relation. The equivalence classes of the relation $\sim$ are called *connected components* of G. Given a graph $G = (V, E)$ and $U \subseteq V$, the *subgraph induced by U* is the graph $G = (U, E \cap \mathcal{P}_2(U))$. The subgraphs induced by the connected components are often simply called connected components.

3. Fine and Wilf's Theorem

In [2], the authors state the following theorem:

Theorem 6. [2] *Let w be a finite word over a binary alphabet Σ. Let p be a σ_1-period of w and q be a σ_2-period of w, where $(\sigma_1, \sigma_2) \neq (I, I)$ is a pair of permutations of Σ. If $|w| \geq p + q$, then $\gcd(p, q)$ is a e-period of w, where e is the swap of letters of Σ.*

As a first step, we generalize Theorem 6 for arbitrary alphabet and arbitrary permutations:

Theorem 7. *Let p, q be two positive integers and σ_1, σ_2 two permutations such that σ_1 and σ_2^{-1} commute. Then any word w of length at least $p+q$ admitting p as a σ_1-period and q as a σ_2-period also admits $\gcd(p, q)$ as a σ-period, where $\sigma = \sigma_1^x \sigma_2^{-y}$ and x, y are any integers such that $\gcd(p, q) = xp - yq$.*

Proof. Some part of the proof is similar to the one found in [2]. However, since it generalizes it and for the sake of completeness, we include the whole proof below.

Let $g = \gcd(p, q)$. The solutions of the Diophantine equation

$$g = xp - yq \tag{1}$$

are well-known and the integers x and y are also called *Bezout coefficients*. It is easy to show that x and y have the same sign. Without loss of generality, assume

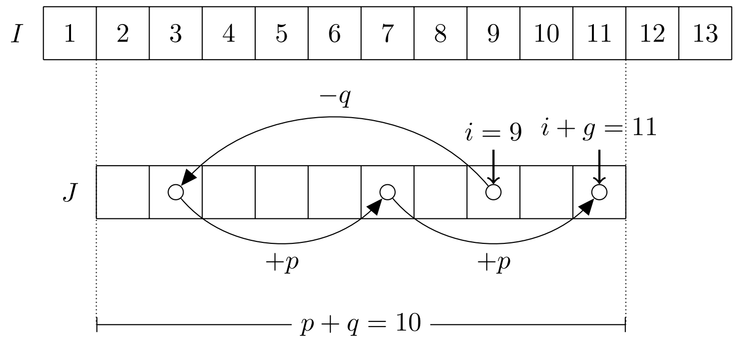


Fig. 1. Illustration of the proof of Theorem 7, when $|w| = 13$, $p = 4$, $q = 6$, $g = 2$ and $i = 9$. The sets I and J denote respectively the whole set of positions and an arbitrary window containing i and $i+g$. The traveled positions i_0, i_1, …, i_k are represented as well as the type of moves, which is coded by the sequence $d_1, d_2, \ldots, d_k$. This illustrates the fact that the position $i+g$ can always be reached from position i by steps of $+p$ and $-q$.

that $x, y > 0$. Let $I = \{1, 2, \ldots, |w|\}$ be the set of indices of the word w. Let i be an integer such that $1 \leq i \leq |w| - g$. We show that $w[i+g] = \sigma(w[i])$. For this purpose, let $J \subseteq I$ be some subset of consecutive integers containing both i and $i+g$ and satisfying $|J| = p + q$. Finally, let $k = x + y$. We define two finite sequences d_1, d_2, …, d_k and i_0, i_1, …, i_k as follows:

(i) $i_0 = i$;

(ii) $d_{j+1} = \begin{cases} p & \text{if } i_j + p \in J, \\ -q & \text{if } i_j - q \in J; \end{cases}$

(iii) $i_{j+1} = i_j + d_{j+1}$.

Figure 1 illustrates this idea on a word w of length 13, with $p = 4$, $q = 6$ and $i = 9$.

We prove the following four claims:

(1) The sequence d_1, d_2, …, d_k is well-defined;
(2) For $j = 0, 1, \ldots, k$, we have $i_j \in J$;
(3) The sequence d_1, d_2, …, d_k contains exactly x occurrences of p and y occurrences of q;
(4) g is a σ-period of w;

(1) To prove that the sequence d_1, d_2, …, d_k is well-defined, it suffices to prove that exactly one of the conditions $i_j + p \in J$ and $i_j - q \in J$ is verified. First, assume that both conditions hold, i.e. $i_j + p \in J$ and $i_j - q \in J$. Since J contains consecutive integers, then $|J| \geq (i_j + p) - (i_j - q) + 1 = p + q + 1$, which is impossible. Next, assume that none of the conditions hold, i.e. $i_j + p \notin J$ and $i_j - q \notin J$. Then $|J| \leq (i_j + p - 1) - (i_j - q + 1) + 1 = p + q - 1$, since J contains consecutive integers, another contradiction.

(2) This follows directly from the fact that the sequence d_1, d_2, ..., d_k is well-defined.

(3) To prove that $i_k = i + g$, we need to prove that the word $d = d_1 d_2 \cdots d_k$ has exactly x occurrences of p and y occurrences of $-q$. We argue once again by contradiction. Without loss of generality, assume that d contains more than x occurrences of p and let $d_1 d_2 \cdots d_j d_{j+1}$ be the shortest prefix of d containing exactly $x+1$ occurrences of p. In particular, $d_{j+1} = p$, so that $i_{j+1} = i_j + d_{j+1} = i_j + p$. Moreover, since $d_1 d_2 \cdots d_j$ contains exactly x occurrences of p, we have $i_j = i + xp - y'q$ for some integer $y' < y$. On the other hand, $i + g = i + xp - yq$. But $i_j - q = i + xp - (y'+1)q$ and, since $y' < y' + 1 \leq y$, we conclude that $i + g \leq i_j - q \leq i_j$. Consequently, $i_j - q \in J$, and we already know that $i_{j+1} = i_j + p \in J$, contradicting the fact that $d_1 d_2 \cdots d_k$ is uniquely determined.

(4) Let $s = s_1 s_2 \cdots s_k$ be the sequence of permutations defined by $s_j = \sigma_1$ if $d_j = p$ and $s_j = \sigma_2^{-1}$ if $d_j = (-q)$. Since the sequence i_0, i_1, ..., i_k is contained in I, starts with i and ends with $i + g$, it follows that

$$w[i+g] = (s_k \circ s_{k-1} \circ \ldots \circ s_2 \circ s_1)(w[i]).$$

But σ_1 and σ_2^{-1} commute and $|s|_{\sigma_1} = x$ while $|s|_{\sigma_2^{-1}} = y$, which implies $w[i+g] = (\sigma_1^x \circ \sigma_2^{-y})(w[i]) = \sigma(w[i])$. □

Example 8. *Let* $w = 012012012012$, $\sigma_1 = (021)$, *and* $\sigma_2 = (012)$. *One can see that* $p = 5$ *is a* σ_1*-period of* w *and* $q = 7$ *is a* σ_2*-period of* w. *We can observe that* σ_1 *and* σ_2^{-1} *commute, and that* $g = \gcd(5,7) = 1$ *is a* σ*-period of* w, *for* $\sigma = (012)$.

One notices that the bound $|w| \geq p + q$ is tight, as illustrated by the following example:

Example 9. *Let* $\Sigma = \{0,1\}$, $e = (01)$ *and* $w = 000111$. *Then* 3 *and* 4 *are both* e*-periods of* w, *but* $1 = \gcd(3,4)$ *is not an* e*-period of* w, *although* $|w| \geq 6 = 3 + 4 - \gcd(3,4)$.

On the other hand, in Theorem 7, it is assumed that the word w admits two σ-periods, but there is no guarantee that such a word exists:

Example 10. *Let* $\Sigma = \{0,1,2\}$, $\sigma_1 = (012)$ *and* $\sigma_2 = (01)(2)$. *Let* $p = 4$ *and* $q = 3$. *We prove that there does not exist any word of length* 7 *admitting* p *as a* σ_1*-period and* q *as a* σ_2*-period. Indeed, assume that such a word* w *exists. First, we suppose that* $w[1] = 0$. *Then using the two pseudoperiods, one gets* $w[4] = 1$, $w[7] = 0$, $w[3] = 2$, $w[6] = 2$, $w[2] = 1$, $w[5] = 0$, *so that* $w = 0121020$, *which is absurd, since* $w[1] = 0$ *and* $w[5] = 0$ *contradicts the fact that* $p = 4$ *is a* σ_1*-period. One obtains similar contradictions by assuming* $w[1] = 1$ *and* $w[1] = 2$. *Hence, there is no word* w *of length* 7 *such that* 4 *is a* σ_1*-period of* w *and* 3 *is a* σ_2*-period of* w.

However, under some conditions, we are guaranteed that two σ-periods coexist in words of any length:

Proposition 11. *Let p, q be to positive integers and σ_1, σ_2 two permutations such that $\sigma_2 = \sigma_1^n$ and $\mathrm{ord}(\sigma_1)$ divides $q - pn$. Then for any positive integer m, there exists a word w of length m admitting p as a σ_1-period and q as a σ_2-period.*

Proof. It suffices to show that the permutation σ of Theorem 7 is independent of the Bezout coefficients x and y. Let x', y' be two integers such that $\gcd(p, q) = xp - yq = x'p - y'q$, i.e. the couple (x', y') is also a solution of Equation (1). It is known that (x', y') can be expressed with respect to the particular solution (x, y). More precisely, there exists some integer t such that $x' = x + qt$ and $y' = y + pt$. Since σ_1 and σ_2^{-1} commute, it follows from the proof of Theorem 7 that

$$\begin{aligned}
\sigma_1^{x'}\sigma_2^{-y'} &= \sigma_1^{x+qt}\sigma_2^{-y-pt} \\
&= (\sigma_1^{x}\sigma_2^{-y})\sigma_1^{qt}\sigma_2^{-pt} \\
&= (\sigma_1^{x}\sigma_2^{-y})\sigma_1^{qt}\sigma_1^{-npt} \\
&= (\sigma_1^{x}\sigma_2^{-y})\sigma_1^{(q-pn)t} \\
&= (\sigma_1^{x}\sigma_2^{-y}),
\end{aligned}$$

since $\mathrm{ord}(\sigma_1)$ divides $q - pn$. □

The proof of Proposition 11 reveals that the condition $\sigma_1^{qt}\sigma_2^{-pt} = I$ is enough for constructing a word of length m having p as a σ_1-period and q as a σ_2-period. Clearly, $\sigma_2 = \sigma_1^n$ and $\mathrm{ord}(\sigma_1) \mid q - pn$ implies $\sigma_1^{qt}\sigma_2^{pt}$, but it is not simple to verify if the converse is true, since solving equations on permutations is a hard problem. Empirical results for alphabets of size 4 and 5 suggest that these two conditions are necessary.

4. Pseudocentral words

In this last section, we consider words admitting multiple pseudoperiods. Of particular interest are nontrivial words having multiple periods. It is well-known that the length of such words can be computed by generalizing the Euclidean division.

Theorem 12 (Holub, [13]) *Let P be a set of positive integers pairwise coprime and $m = \min(P) > 1$. Let*

$$Q = \{q - m \mid q \in P, q \neq m\} \cup \{m\}.$$

Then the maximal length $n(P)$ of a nontrivial word with periods in P is given by the following recursive formula

$$n(P) = m + \max\{n(Q), m - 1\}, \tag{2}$$

where $n(Q)$ is the maximum length of a nontrivial word with periods Q and is defined as 0 if $1 \in Q$.

The main objective of this section is to generalize Theorem 12 to pseudoperiods. For $i = 1, 2, \ldots, k$, let σ_i be a permutation and p_i be a positive integer. First, it is worth mentioning that the permutations needs not be involutory, as shown by the next example:

Example 13. *Let* $w = 01213102131012$, $\sigma_1 = (032)(1)$ *and* $\sigma_2 = (023)(1)$. *Then* $p_1 = 9$ *is a* σ_1*-period of* w *and* $p_2 = 7$ *is a* σ_2*-period of* w. *Moreover,* w *is a* σ*-palindrome, where* $\sigma = (02)(1)(3)$.

On the other hand, in some cases, for fixed values of p_1, p_2, σ_1 and σ_2, one has two different words, one of which is a pseudopalindrome while the other is not:

Example 14. *Let* $\sigma_1 = (023)(1)$, $\sigma_2 = (03)(1)(2)$, $p_1 = 4$ *and* $p_2 = 7$. *Moreover, let* $w = 001222133$ *and* $w' = 321303102$. *Then* p_1 *is a* σ_1*-period of both* w *and* w' *and* p_2 *is a* σ_2*-period of both* w *and* w'. *But* w *is a* ϑ*-palindrome, for* $\vartheta = R \circ (03)(1)(2)$ *and* w' *is not a pseudopalindrome.*

Under some conditions, however, we are guaranteed that the resulting word is indeed a pseudopalindrome.

Theorem 15. *Let* p_1, p_2, ..., p_k *be* k *positive integers such that* $gcd(p_i, p_j) = 1$ *for* $1 \leq i, j \leq k$, $i \neq j$. *Let* w *be a word of such that* p_1 *is a* σ_1*-period,* p_2 *is a* σ_2*-period,* $\ldots$, p_n *is a* σ_n*-period of* w, *with* $|w| = n(P)$, *where* σ_1, σ_2, ..., σ_n *are involutions.*

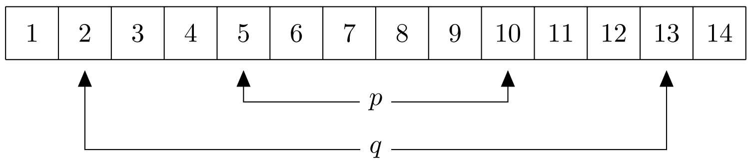
(i) If p_1, p_2, ..., p_n *are odd,* Lp *is even and* $\sigma = \sigma_1 = \sigma_2 = ... = \sigma_n$, *then* w *is a* σ*-palindrome;*
(ii) If one of the periods is even and its σ *is the identity, the other periods are odd and their* σ *are the same and if* Lp *is even, then* w *is a* σ*-palindrome;*
(iii) If one of the periods is even and its σ *is the identity, the other periods are odd and their* σ *are the same and if* Lp *is odd, then* w *is a palindrome.*

In order to prove Theorem 15, we need to introduce a convenient representation. For the remainder of this section, let $P = \{p_1, p_2, \ldots, p_k\}$, where p_i is a positive integer for $i = 1, 2, \ldots, k$. Moreover, let $S = \{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ be a set of permutations for $i = 1, 2, \ldots, k$. We assume that the numbers p_i are enumerated in increasing order and that the permutation σ_i is associated with number p_i, for $i = 1, 2, \ldots, k$. Finally, let n be a positive integer.

Definition 16. *The* period graph *of order* n *induced by* P *is the graph* $G = (V, E)$, *where* $V = \{1, 2, \ldots, n\}$ *and* $\{u, v\} \in E$ *if and only if* $|u - v| \in P$.

Roughly speaking, V is the set of positions in some word w of length n and there is an edge between two vertices if one of the pseudoperiods propagates between two positions. Figure 2 illustrates this concept for $p_1 = 5$ and $p_2 = 11$.

It should be noted that this construction has already been suggested in [12], where the author worked with equivalence classes that are exactly the connected



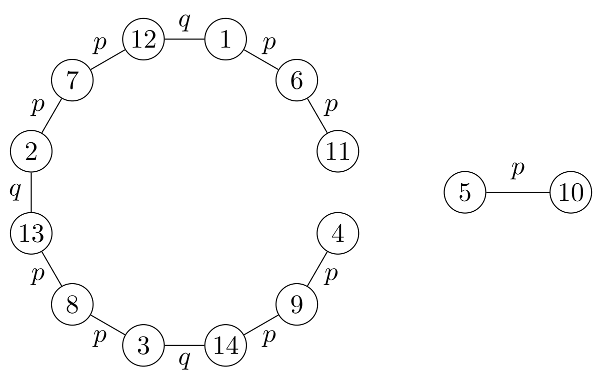


Fig. 2. The graph obtained for $n = 14$, $k = 2$, $p_1 = 5$ and $p_2 = 11$. Each connected component present palindromic properties. For instance, $11 + 4 = 6 + 9 = \ldots = 2 + 13 = 5 + 10$, since $\bar{\cdot}$ is an automorphism (Theorem 19). Also, since the numbers p_1 and p_2 are both odd and n is even, each connected component has a central edge, namely $(2, 13)$ and $(5, 10)$, which induces the palindromicity on the vertices as well as on the edge labels.

components of the graph G. The graph G has a very specific structure that we describe in the following lemmas.

Definition 17. *Let u be a vertex of G, $n = |V|$ and $p \in P$. We define $\overline{u} = n+1-u$.*

The next fact is easily observed.

Lemma 18. *Let $u \in V$. If $u + p \in V$, then $\overline{u} - p \in V$.*

Proof. Follows from Definition 17 and the definition of G. □

Also, the graph G presents symmetric properties.

Theorem 19. *The map $\bar{\cdot}$ is a graph automorphism.*

Proof. Clearly, $\bar{\cdot}$ is a bijection. Let $\{u, v\} \in E$. Then $|u - v| \in P$. But $\overline{u} - \overline{v} = (n + 1 - u) - (n + 1 - v) = v - u \in P$, so that $\{\overline{u}, \overline{v}\} \in E$ as well. □

When the order n of G is sufficiently large, it is clear that G is connected. Therefore, the number $n(P)$ given by Equation (2) might be interpreted as the maximum value such that G is not connected.

As observed in [12], words of length $n(P)$ having periods in P are palindromes. In term of graph theory, it translates into showing that complementary positions u and $\overline{u}$ belong to the same connected component.

Lemma 20. *Assume that the order of G is $n(P)$ and let $u \in V$. Then both u and $\overline{u}$ belong to the same connected component of G.*

Proof. This proof is adaptated from [12], but is formulated according to graphs instead of equivalence classes.

First, we prove that for any connected component C, we have $\min(C) = \overline{\max(C)}$. Without loss of generality, we may suppose that $\min(C) > \overline{\max(C)}$ since the assumption $\min(C) < \overline{\max(C)}$ would lead to a similar reasoning. Let G' be the graph of length $n(P)+1$. By definition of $n(P)$, we know that G' is connected. Therefore, there exists a path in G' starting from $\min(C)$ and ending with $n(P)+1$:

$$\min(C) = u_0 \to u_1 \to \ldots \to u_k = n(P) + 1$$

such that $\min(C) < u_i < n(P) + 1$ for $i = 1, 2, \ldots, k-1$. This implies that there exists a path

$$\min(C) = u_0 \to u_1 \to \ldots \to u_{k-1}$$

in G, where $u_i > \min(C)$ for $i = 1, 2, \ldots, k$. Since we have assumed that $\min(C) > \overline{\max(C)}$, then $u_i > \overline{\max(C)}$ for $i = 0, 1, \ldots, k$. Consider now the path

$$x_0 \to x_1 \to \ldots \to x_k$$

of G' defined by $x_i = u_i - \overline{\max(C)}$ for $i = 0, 1, \ldots, k$. Notice that $x_i > 0$. Since $u_i < n(P) + 1$, then $x_i \leq n(P) - \overline{\max(C)}$. Therefore,

$$x_i \leq n(P) - [n(P) + 1 - \max(C)] = \max(C) - 1 < \max(C)$$

for $i = 0, 1, \ldots, k$. Since $0 < x_i < \max(C)$, this means that the path $x_0 \to x_1 \to \ldots \to x_k$ is in G as well. But $\max(C) < n(P) + 1$, so that $\overline{\max(C)} > 0$. Therefore $x_i < u_i$ for $i = 0, 1, \ldots, k$. In particular, $x_0 < \min(C)$, but x_0 and $\min(C)$ are in the same connected component C, which is impossible.

To conclude, let C be any connected component of G and $u \in C$. Then there exists a path from u to $\min(C)$ of the form

$$u = y_0 \to y_1 \to \ldots \to y_\ell = \min(C).$$

Since $\bar{\cdot}$ is an automorphism, there also exists a path

$$\overline{u} = \overline{y_0} \to \overline{y_1} \to \ldots \to \overline{y_\ell} = \max(C).$$

But both $\min(C)$ and $\max(C)$ belong to C, which implies that both u and $\overline{u}$ are in C as well. □

When handling multiple periods, there exists a nontrivial word for any values of P. On the other hand, as it has been observed before, pseudoperiods may not be compatible. The next lemma provides a criterion for deciding if such a word exists.

Lemma 21. *Let G be a graph of length $n(P)$ having at least one cycle. If σ is the same involution for each odd periods and if σ_k is the identity for p_k the even period of the graph, then the result of all the permutations on the cycle is the identity.*

Proof. Let u be a vertex of the cycle. Starting from u and following the cycle, the path returns to u. The parity of u must be the same at the beginning and at the end. Knowing that, the cycle must have an even number of odd edges. The other edges will be even. Since σ is an involution, by repeating it an even number of times, the result will be the identity. The even edges do not matter since their permutation is the identity. Thus, the result is the identity. □

Proof of Theorem 15. Let u be a vertex of G. By lemma 18 and lemma 21, we know that u and $\overline{u}$ are in the same connected component. There is at least one path between u and $\overline{u}$. In the case where there are more than one path, G possesses cycles. As explained by lemma 21, cycles do not create contradictions with other paths. Thus, we can use only one path to prove the theorem.

(i) If $n(P)$ is even, $u + \overline{u}$ is odd. The parity of u is not the same as the parity of $\overline{u}$. Having only odd periods and starting from u, we need an odd number of edges to go to $\overline{u}$. Consequently, there is a central edge in each connected component. Since symmetric vertices are the result of the same composition of permutations, they all verify $w[i] = \sigma(w[n(P) + 1 - i])$, where σ is the permutation on the central edge.

(ii) Like the first case (i), if $n(P)$ is even, $u + \overline{u}$ is odd. The parity of u is not the same as the parity of $\overline{u}$. Moreover, since $n(P)$ is even and because the central vertex of a connected component is always $(n(P) + 1)/2$, all the connected components have a central edge. The central edge connects two vertices u_k and $\overline{u_k}$. Their parity is different so the central edge is necessarily odd. Knowing that and since symmetric vertices are the result of the same composition of permutation, they all verify $w[i] = \sigma(w[n + 1 - i])$, where is the permutation on the central edge.

(iii) If $n(P)$ is odd, $u + \overline{u}$ is even. The parity of u is the same as the parity of $\overline{u}$. There will be a central vertex in C, one of the connected components. Also, there is only one connected component with a central vertex because the only possible central vertex $u = \overline{u}$ is $(n(P) + 1)/2$. By induction, every pair of vertex at the same distance from the central vertex in C are the same letters since the same permutations are applied. All the symmetric vertices of C are palindromic. The other connected component will have an even number of vertices. There is a central edge connecting two vertices of same parity. The central edge must be even. Like the case (i) and (ii), symmetric vertices verify $w[i] = \sigma(w[n + 1 - i])$ where σ is the identity. Thus, $w[i] = w[n + 1 - i]$ and symmetric vertices are palindromic. Also, since there is only one even period, the graph will have two connected components. □

5. Properties of period graphs

There are some other interesting characteristics of the graph having length $n(P)$.

Lemma 22. *Assume that $gcd(p, q) = 1$ for any $p, q \in P$ and that n is the maximum length for a word w to have all pseudoperiods in P such that 1 is not a pseudoperiod. Then, $n \leq p_1 + p_2 - 2$.*

Proof. (i) In [12], the author proposes an algorithm to find the longest word having periods p_1, p_2, ..., p_n but not period $gcd(p_1, p_2, \ldots, p_n)$. Let p_1 and p_2 be the smallest periods of $p_1, p_2, ...p_n$, $p_1 < p_2$. In the case where $p_i > p_1 + p_2$ for $3 \leq i \leq n$, we use only p_1 and p_2 in Holubs algorithm which can be replaced by the Euclidean division. We obtain:

$$\begin{aligned} p_2 &= x_0 p_1 + r_1 \\ p_1 &= x_1 r_1 + r_2 \\ &\ldots \\ r_{j-2} &= x_{j-1} r_{j-1} + 1 \\ r_{j-1} &= x_{j-1} \end{aligned}$$

$$p_1 + p_2 + r_1 + \ + r_{j-1} = x_0 p_1 + r_1 + x_1 r_1 + r_2 + \ + x_{j-1} r_{j-1} + 1 + x_{j-1}$$

$$p_1 + p_2 = x_0 p_1 + x_1 r_1 + \ + x_{j-1} r_{j-1} + 1 + x_{j-1}$$

$$p_1 + p_2 = m_{k-1} + 1 + \sum_{i=1}^{k-1} m_i$$

$$p_1 + p_2 2 = n(P)$$

(ii) In the case where p_3, p_4 ... p_n are used in Holubs algorithm, we use them when they are less than p_1 or p_2. So the result of the algorithm will be less or equal to the first case (i). □

Lemma 23. *Let p_1, p_2, ..., p_n be n arbitrary periods such that $gcd(p_i, p_j) = 1$ for $1 \leq i, j \leq n$, $i \neq j$. Let p_1 and p_2 be the smallest periods, $p_1 < p_2$.*

(i) If $n(P) \neq 2p_1 - 1$, then the number of vertices having degree one is $2(p_1 + p_2 - n(P))$. The vertices are $\overline{p_2}, \overline{p_2} + 1, , p_1$ and $\overline{p_1}, \overline{p_1} + 1, , p_2$.

(ii) If $n(P) = 2p_1 - 1$, then the number of vertices having degree one is $2(p_1 - 1 + p_2 - n(P))$. The vertices are $\overline{p_2}, \overline{p_2} + 1, , p_1 - 1$ and $\overline{p_1 - 1}, \overline{p_1 - 1} + 1, , p_2$. Also, the vertex p_1 is having degree 0.

Proof. First of all, the case $n(P) < 2p_1 - 1$ is impossible because $n(P) \leq p_1 + p_2 2$, $p_1 < p_2$.

(i) By Lemma 22, $n(P) \leq p_1 + p_2 2$, where $n(P)$ is the maximal length of a word which has periods p_1, p_2, ..., p_n and not period $gcd(p_i, p_j)$ for $1 \leq i, j \leq n$, $i \neq j$. Therefore, $\overline{p_2}$ has only one edge which has label p_1. The following vertices have one edge too, until we reach p_1. A similar reasoning can be done with $\overline{p_1}$, which has p_2 as only labeled edge. Moreover, since $n(P) \neq 2p_1 - 1$, the vertex p_1 has one edge. The number of vertices having at most degree one is even and it is given by

$$p_1 + 1 - (n(P)p_2 + 1) + p_2 + 1 - (n(P)p_1 + 1)$$
$$2(p_1 + p_2 - n(P))$$

(ii) When $n(P) = 2p_1 - 1$, then the vertex p_1 does not have any edge. Thus, p_1 is a connected component of G with only one vertex. Like the case (i), all the other vertices $\overline{p_2}, \overline{p_2} + 1, , p_1 - 1$ and $\overline{p_1 - 1}, \overline{p_1 - 1} + 1, , p_2$ still have degree one. The number of vertices having degree one is given by

$$p_1 - (n(P)p_2 + 1) + p_2 + 1 - (n(P)p_1 + 1)$$
$$2(p_1 - 1 + p_2 - n(P))$$

□

Lemma 24. *Let p_1, p_2 ... p_n, n arbitrary periods such that $gcd(p_i, p_j) = 1$ for $1 \leq i, j \leq n$, $i \neq j$. The number of vertices having degree k for $k > 2$ is given by $2(p_{k+1} - p_k)$.*

Proof. Let u_k be a vertex of G such that $u_k = n(P) - p_k$, $u_k > 0$.We know that u_k has k labeled edges which are $p_1, p_2, ..., p_k$. Then all the vertices $u_i < u_k$ have at least k edges too. The vertex p_{k+1} and all vertices $> p_{k+1}$ also haver at least k edges. The number of vertices having degree k or more is $2(n(P) - p_k)$. By doing the same reasoning, the number of vertices having degree $k+1$ or more is $2(n(P) - p_{k+1})$. The number of vertices having only degree k is

$$2(n(P) - p_k) - 2(n(P) - p_{k+1})$$
$$2(p_{k+1} - p_k)$$

□

Theorem 25. *Let G be a graph with only 2 connected components and let p, q, r be three arbitrary periods of G such that $gcd(p, q) = gcd(q, r) = gcd(p, r) = 1$, $p < q < r$. If the difference between the number of vertices having degree 1 and the number of vertices having degree 3 is less than 3, then G has at least one cycle. If not, then G does not have any cycle.*

Proof. We can use the lemma 23 and the lemma 24 to find the number of vertices having degree 1 and 3. The lemma 22 proved that $n(P) \leq p + q - 2$. Therefore, there is always at least 4 vertices having degree one. Knowing that, it is impossible to have cycles without vertices of degree 3. By theorem 19, the number of vertices having degree 3 is always even.

(i) In the case where $p \neq 2n(P) - 1$, if G does not have any cycle or vertices having degree 3, there is 4 vertices having degree one. By adding 2 vertices having degree 3 without creating any cycle, 2 new vertices of degree one are added. For each vertex having degree 3 added, a new vertex having degree one is added too. The difference between the number of vertices having degree one and those having degree 3 is always 4.

(ii) In the case where $p = 2n(P) - 1$ and G has no cycle, then by following the same reasoning, the difference between the number of vertices having degree one and those having degree 3 is 3.

(iii) When G has cycles, at least 2 vertices having degree 3 connect together. Because of that, two vertices of degree one disappear. Thus, the difference between the number of vertices having degree 1 and the number of vertices having degree 3 is less than 3. □

6. Concluding remarks and future work

The enumeration of all solutions for alphabet of size up to 5 was obtained by computer exploration using the open-source software MiniSat [9]. The exploration of the graphs were driven using the open-source software Sage [22].

Fine and Wilf's Theorem is known for having many applications and generalizations. The results in this paper are unique in the sense that they also deal with permutations. Moreover, we show that many results hold even if the permutations are not involutory.

The next step would be to consider the multipseudoperiodicity property to describe generalized pseudostandard words as sketched in Subsection 2.4. Another interesting area of study would also consist in generalizing the theorem 15 presented here for different permutations. For that purpose, the graph representation seems a convenient starting point.

References

[1] A. Blondin Massé, S. Brlek, A. Garon, and S. Labbé. Palindromes and local periodicity. In *Words 2009, 7th Int. Conf. on Words*, 2009. electronic proceedings.

[2] A. Blondin Massé, G. Paquin, and L. Vuillon. A Fine and Wilf's theorem for pseudoperiods and Justin's formula for generalized pseudostandard words. In *8e Journes montoises d'informatique thorique*, septembre 2010.

[3] M.Gabriella Castelli, Filippo Mignosi, and Antonio Restivo. Fine and Wilf's theorem for three periods and a generalization of Sturmian words. *Theoretical Computer Science*, 218(1):83–94, 1999.

[4] Sorin Constantinescu and Lucian Ilie. Generalised Fine and Wilf's theorem for arbitrary number of periods. *Theoretical Computer Science*, 339(1):49–60, 2005.

[5] E. Czeizler, E. Czeizler, L. Kari, and S. Seki. An extension of the Lyndon Schützenberger result to pseudoperiodic words. In *DLT '09: Proceedings of the 13th International Conference on Developments in Language Theory*, pages 183–194, Berlin, Heidelberg, 2009. Springer-Verlag.

[6] A. de Luca. Sturmian words: structure, combinatorics, and their arithmetics. *Theoretical Computer Science*, 183:45–82, 1997.

[7] A. de Luca and A. De Luca. Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science*, 362(1-3):282–300, 2006.

[8] A. de Luca and F. Mignosi. Some combinatorial properties of sturmian words. *Theoretical Computer Science*, 136(2):361–385, 1994.

[9] Niklas Eén and Niklas Sörensson. An extensible SAT-solver. In Enrico Giunchiglia and Armando Tacchella, editors, *SAT*, volume 2919 of *Lecture Notes in Computer Science*, pages 502–518. Springer, 2003.

[10] N.J. Fine and H.S. Wilf. Uniqueness theorems for periodic functions. *Proceeding of American Mathematical Society*, 16:109–114, 1965.

[11] Dan Gusfield. *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge University Press, New York, NY, USA, 1997.

[12] S. Holub. On multiperiodic words. *RAIRO - Theoretical Informatics and Applications*, 40:583–591, 2006.

[13] S. Holub. Corrigendum: On multiperiodic words. *RAIRO - Theoretical Informatics and Applications*, 45:467–469, 2011.

[14] Jacques Justin. On a paper by Castelli, Mignosi, Restivo. *RAIRO - Theoretical Informatics and Applications*, 34:373–377, 2000.

[15] Donald E. Knuth, J.H. Morris, and Vaughan R. Pratt. Fast pattern matching in strings. *SIAM Journal of Computing*, 6(2):323–350, 1977.

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

[17] M. Lothaire. *Applied Combinatorics on Words*. Cambridge University Press, Cambridge, 2005.

[18] Alexandre Blondin Massé, Sébastien Gaboury, and Sylvain Hallé. Pseudoperiodic words. In Hsu-Chun Yen and Oscar H. Ibarra, editors, *Developments in Language Theory*, volume 7410 of *Lecture Notes in Computer Science*, pages 308–319. Springer, 2012.

[19] Filippo Mignosi, Antonio Restivo, and Pedro V. Silva. On Fine and Wilf's theorem for bidimensional words. *Theor. Comput. Sci.*, 292:245–262, January 2003.

[20] J. Shallit. 50 years of Fine and Wilf, 2011. Waterloo Workshop in Computer Algebra, Canada, `http://www.cs.uwaterloo.ca/~shallit/Talks/wilf3.pdf`.

[21] R. J. Simpson and R. Tijdeman. Multi-dimensional versions of a theorem of Fine and Wilf and a formula of Sylvester. *Proceedings of the American Mathematical Society*, 131(6):1661–1671, 2003.

[22] W. A. Stein et al. *Sage Mathematics Software (Version 4.3.3)*. The Sage Development Team, 2010. `http://www.sagemath.org`.

[23] R. Tijdeman and L. Q. Zamboni. Fine and Wilf words for any periods. *Indag. Math*, 14:200–3, 2003.

[24] Robert Tijdeman and Luca Q. Zamboni. Fine and Wilf words for any periods II. *Theoretical Computer Science*, 410(30-32):3027–3034, 2009.