

# Computing Weak Periods of Partial Words

## EXTENDED ABSTRACT\*

F. Blanchet-Sadri<sup>1</sup>

Taktin Oey<sup>2</sup>

Timothy D. Rankin<sup>3</sup>

### Abstract

Fine and Wilf's well-known theorem states that any word having periods  $p, q$  and length at least  $p + q - \gcd(p, q)$  also has  $\gcd(p, q)$ , the greatest common divisor of  $p$  and  $q$ , as a period. Moreover, the length  $p + q - \gcd(p, q)$  is critical since counterexamples can be provided for shorter words. This result has since been extended to partial words, or finite sequences that may contain a number of "do not know" symbols or "holes." More precisely, any partial word  $u$  with  $H$  holes having *weak* periods  $p, q$  and length at least the so-denoted  $l_H(p, q)$  also has *strong* period  $\gcd(p, q)$  provided  $u$  is not  $(H, (p, q))$ -*special*. This extension was done for one hole by Berstel and Boasson (where the class of  $(1, (p, q))$ -special partial words is empty), for two or three holes by Blanchet-Sadri and Hegstrom, and for an arbitrary number of holes by Blanchet-Sadri. In this paper, we further extend these results, allowing an arbitrary number of weak periods. In addition to speciality, the concepts of *intractable* period sets and *interference* between periods play a role.

## 1 Introduction

Notions and techniques related to periodic structures in words find applications in virtually every area of theoretical and applied computer science, notably in text processing [11, 12], data compression [28, 30], coding [2], computational biology [15], string searching and pattern matching algorithms [11, 27]. Repeated patterns and related phenomena in words have played over the years a central role in the development of combinatorics on words, and have been highly valuable tools for the design and analysis of algorithms [9, 21].

A well-known theorem of Fine and Wilf [14] states that any word having periods  $p$  and  $q$  and length at least  $p + q - \gcd(p, q)$  also has  $\gcd(p, q)$ , the greatest common divisor of  $p, q$ , as a period. Moreover, the length  $p + q - \gcd(p, q)$  is critical since counterexamples can be provided for shorter lengths. This periodicity result on words has been generalized in many ways. For instance, extension to more than two periods are given in [8, 10, 17, 29]. In particular, Constantinescu and Ilie [10] give an extension of Fine and Wilf's theorem for an arbitrary number of periods and prove that their bounds are optimal.

In many practical applications, such as DNA sequence analysis, repetitions admit a certain variation between copies of the repeated pattern because of errors due to mutation, experiments, etc. Approximate repeated patterns, or repetitions where errors are allowed, are playing a central role in different variants of string searching and pattern matching problems. *Partial words*, or sequences of symbols from a finite alphabet that may have some "do not know" symbols also called "holes," have acquired great importance in this context [19, 20, 25]. Partial words are useful in a new generation of

---

\*This material is based upon work supported by the National Science Foundation under Grant No. DMS-0452020. We thank the referees of a preliminary version of this paper for their very valuable comments and suggestions.

<sup>1</sup>Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402-6170, USA, [blanchet@uncg.edu](mailto:blanchet@uncg.edu)

<sup>2</sup>Department of Mathematics, Harvard University, One Oxford Street, Cambridge, MA 02138, USA

<sup>3</sup>Department of Mathematics, Duke University, Box 90320, Durham, NC 27708-0320, USA

pattern matching algorithms that search for local similarities between sequences. In this area, they are called “spaced seeds,” the best current tool for similarity search between biological sequences, and a lot of work has been dedicated to their influence on the algorithms’ performance [7, 13, 16, 18, 22, 23, 24]. There is a very strong connection that has been established between combinatorics on words and algorithms for computing spaced seeds which enabled computation of the latter several orders of magnitude faster than previous algorithms [16].

Fine and Wilf’s result has also been generalized to partial words in which case there are two notions of periodicity: one is that of *strong period* and the other is that of *weak period*. First, any partial word  $u$  with  $H$  holes and having strong periods  $p, q$  and length at least the so-denoted  $L_H(p, q)$  has also strong period  $\gcd(p, q)$ . The study of the optimal bounds  $L_H(p, q)$  was initiated in [26] and pursued in [4, 6]. Second, any partial word  $u$  with  $H$  holes and having weak periods  $p, q$  and length at least the so-denoted  $l_H(p, q)$  has also strong period  $\gcd(p, q)$  provided  $u$  is not  $(H, (p, q))$ -special. This extension was done for one hole by Berstel and Boasson in their seminal paper [1] where the class of  $(1, (p, q))$ -special partial words is empty; for two or three holes by Blanchet-Sadri and Hegstrom [5]; and for an arbitrary number of holes by Blanchet-Sadri [3]. In this paper, we further extend these results, allowing an arbitrary number of holes and weak periods. Along with speciality, the concepts of intractable period sets and interference between periods play a role. In addition, we have established a World Wide Web server interface at

[www.uncg.edu/mat/research/finewilf2](http://www.uncg.edu/mat/research/finewilf2)

for automated use of a program which, given a tuple of weak periods  $\mathbf{p} = (p_1, \dots, p_m)$  and a number of holes  $H$ , outputs the critical length  $l_H(\mathbf{p})$  as well as a counterexample word for one less than the critical length if applicable.

The contents of our paper are as follows: In Section 2, we review some preliminaries. In Section 3, we discuss the case of partial words with two weak periods, generalize the formula of Berstel and Boasson to partial words with one hole and arbitrarily many weak periods, and give an argument for the case of three weak periods and two holes. In Section 4, we discuss counterexample forms and interference. In Section 5, we provide formulas for the case of partial words with  $m$  weak periods and an odd number of holes  $H$  satisfying  $H < m - 1$ ,  $m - 1 \leq H < 2m - 3$ , and  $H = 2m - 3$ . In Section 6, we discuss intractable weak period sets. In Section 7, we provide some refinements. There, we show that our formulas are attained exactly in all but an asymptotically diminishing fraction of exception cases for arbitrarily many weak periods and odd numbers of holes. Finally, we conclude with some remarks.

## 2 Basic concepts

Let  $A$  be an *alphabet*, or a finite nonempty set whose elements are referred to as *letters*. A *string* or *word*  $u$  over  $A$  is a finite sequence or concatenation of letters from  $A$ . The number of letters in  $u$ , or *length* of  $u$ , is denoted by  $|u|$ . The unique word of length 0, denoted by  $\varepsilon$ , is called the *empty word*. A word of length  $n$  over  $A$  can be defined by a total function  $u : \{0, \dots, n - 1\} \rightarrow A$  and is usually represented as  $u = a_0a_1 \dots a_{n-1}$  with each  $a_i \in A$ . For any word  $u$ ,  $u[i..j)$  is the *factor* of  $u$  that starts at position  $i$  and ends at position  $j - 1$ . In particular,  $u[0..j)$  is the *prefix* of  $u$  of length  $j$ , and  $u[|u| - i..|u|)$  is the *suffix* of  $u$  of length  $i$ . For a word  $u$ , the powers of  $u$  are defined inductively by  $u^0 = \varepsilon$  and, for any  $n \geq 1$ ,  $u^n = uu^{n-1}$ . The set of all words over  $A$  of finite length (greater than or equal to zero) is denoted by  $A^*$ . It is a monoid under the associative operation of concatenation or product of words ( $\varepsilon$  serves as the identity) and is referred to as the *free monoid* generated by  $A$ .

A *partial word*  $u$  of length  $n$  over  $A$  is a partial function  $u : \{0, \dots, n - 1\} \rightarrow A$ . The *domain*  $D(u)$  of  $u$  is defined to be the set of positions  $i$  such that  $u(i)$  is defined (or  $u(i) \in A$ ), and the *set of holes*  $H(u)$  of  $u$  the set of positions  $i$  such that  $u(i)$  is undefined. A (*full*) word over  $A$  is a

partial word over  $A$  with an empty set of holes. For convenience, we will refer to a partial word over  $A$  as a word over the enlarged alphabet  $A_\diamond = A \cup \{\diamond\}$ , where  $\diamond \notin A$  represents a “do not know” symbol or “hole.” So a partial word  $u$  of length  $n$  (or  $|u|$ ) over  $A$  can be viewed as a total function  $u : \{0, \dots, n-1\} \rightarrow A_\diamond$  where  $u(i) = \diamond$  whenever  $i \in H(u)$ . For example,  $u = a\diamond bbc\diamond cb$  is a partial word of length 8 where  $D(u) = \{0, 2, 3, 4, 6, 7\}$  and  $H(u) = \{1, 5\}$ . We can thus define for partial words concepts such as concatenation, powers, etc. in a trivial way.

A (*strong*) *period* of a partial word  $u$  over  $A$  is a positive integer  $p$  such that  $u(i) = u(j)$  whenever  $i, j \in D(u)$  and  $i \equiv j \pmod{p}$ . In this case  $u$  is called *p-periodic*. A *weak period* of  $u$  is a positive integer  $p$  such that  $u(i) = u(i+p)$  whenever  $i, i+p \in D(u)$ . In this case  $u$  is called *weakly p-periodic*. For instance, the partial word  $abb\diamond bbcb$  is weakly 3-periodic but is not 3-periodic. Throughout this paper, when we say that a partial word  $u$  has weak period set  $\mathbf{p} = (p_1, \dots, p_m)$ , we mean that  $u$  has weak periods  $p_1, \dots, p_m$  satisfying  $p_1 < \dots < p_m$ . The problem studied in this paper involves finding the minimum (critical) length a partial word with weak periods  $p_1, \dots, p_m$  and  $H$  holes must have in order to guarantee that it has  $\gcd(p_1, \dots, p_m)$  as a strong period, assuming the word is not  $(H, (p_1, \dots, p_m))$ -special (speciality refers to words where some letters are artificially cut off from the rest of the word by holes). In other words, any partial word with weak periods  $p_1, \dots, p_m$ ,  $H$  holes, and length  $\geq l_H(p_1, \dots, p_m)$  will be  $\gcd(p_1, \dots, p_m)$ -periodic if it is not  $(H, (p_1, \dots, p_m))$ -special, but if  $l_H(p_1, \dots, p_m) > 0$  there exists a partial word of length  $l_H(p_1, \dots, p_m) - 1$  with weak periods  $p_1, \dots, p_m$  and  $H$  holes that is neither  $\gcd(p_1, \dots, p_m)$ -periodic nor  $(H, (p_1, \dots, p_m))$ -special as a counterexample. If no counterexamples exist, meaning every partial word with weak periods  $p_1, \dots, p_m$  and  $H$  holes is either  $\gcd(p_1, \dots, p_m)$ -periodic or  $(H, (p_1, \dots, p_m))$ -special regardless of length, the critical length is zero.

The motivation for the exclusion of  $(H, \mathbf{p})$ -special partial words is as follows. There are some placements of holes for which a partial word of unbounded length can be constructed that is not  $\gcd(p_1, \dots, p_m)$ -periodic. For example, let  $m = 3$  and  $\mathbf{p} = (2, 5, 6)$ , and consider the sequence  $(u_i)_{i>0}$  where  $u_i = aba\diamond aa\diamond a^i$ . It is possible to increase the length of  $u$  without limit by increasing  $i$  indefinitely because the  $b$  is effectively isolated away from the rest of the word by the holes 2, 5, and 6 positions to its right. In order for the notion of a critical length to apply, it is necessary to introduce the notion of a word being  $(H, (p_1 \dots p_m))$ -special.

The following definitions, the purposes of which will become clear in later use, generalize the notions of isolation and speciality for two weak periods introduced in [3].

**Definition 2.1.** *Let  $u$  be a partial word over an alphabet  $A$ ,  $S$  be a nonempty proper subset of  $D(u)$ , and  $p_k$  be a weak period of  $u$ .*

1. *The set  $S$  is 1-isolated with respect to  $p_k$  if the following hold:*

- (a) **Left** *If  $i \in S$  and  $i \geq p_k$ , then  $i - p_k \in S$  or  $i - p_k \in H(u)$ .*
- (b) **Right** *If  $i \in S$ , then  $i + p_k \in S$  or  $i + p_k \in H(u)$ .*

2. *The set  $S$  is 2-isolated with respect to  $p_k$  if the following hold:*

- (a) **Left** *If  $i \in S$ , then  $i - p_k \in S$  or  $i - p_k \in H(u)$ .*
- (b) **Right** *If  $i \in S$ , then  $i + p_k \in S$  or  $i + p_k \in H(u)$ .*

3. *The set  $S$  is 3-isolated with respect to  $p_k$  if the following hold:*

- (a) **Left** *If  $i \in S$ , then  $i - p_k \in S$  or  $i - p_k \in H(u)$ .*
- (b) **Right** *If  $i \in S$  and  $i < |u| - p_k$ , then  $i + p_k \in S$  or  $i + p_k \in H(u)$ .*

**Definition 2.2.** *Let  $u$  be a partial word over an alphabet  $A$ ,  $S$  be a nonempty proper subset of  $D(u)$ , and  $\mathbf{p} = (p_1, \dots, p_m)$  be a weak period set of  $u$ . The set  $S$  is isolated if it is 1-, 2-, or 3-isolated with respect to every  $p_k$ . For  $1 \leq i \leq 3$ , the set  $S$  is  $i$ -isolated if it is  $i$ -isolated with respect to the largest weak period,  $p_m$ .*

**Definition 2.3.** Let  $u$  be a partial word over an alphabet  $A$  with  $H$  holes and with weak period set  $\mathbf{p} = (p_1, \dots, p_m)$ . The word  $u$  is called  $(H, \mathbf{p})$ -special if there exist  $0 \leq j < \gcd(\mathbf{p})$  and a nonempty isolated proper subset of  $D(u) \cap N_j$ , where  $N_j = \{j + k \gcd(\mathbf{p}) \mid k \in \mathbb{Z}\}$ .

**Lemma 2.4.** If a partial word with weak period set  $\mathbf{p} = (p_1, \dots, p_m)$  is  $(H, \mathbf{p})$ -special, then  $H \geq m$ .

We now define the concept of critical length.

**Definition 2.5.** Let  $\mathbf{p} = (p_1, \dots, p_m)$  be a tuple of positive integers and  $H$  be a nonnegative integer. The **critical length**  $l_H(\mathbf{p})$  is the smallest nonnegative integer  $n$  such that any non- $(H, \mathbf{p})$ -special partial word  $u$  with weak periods  $p_1, \dots, p_m$ ,  $|H(u)| = H$  and length  $\geq n$  is  $\gcd(p_1, \dots, p_m)$ -periodic.

We take from here on the alphabet  $A$  to be arbitrarily large, so that “there exists a word  $u$ ” is taken to refer to such a word existing over some finite alphabet  $A$  of unspecified size. With the exception of a few unusual cases (see Section 6), the problem can be reduced to the cases where  $\gcd(p_1, \dots, p_m) = 1$ , as evinced by the following result.

**Lemma 2.6.** For any positive integer  $d$ , if  $l_H(p_1, \dots, p_m) > 0$ , then

$$l_H(dp_1, \dots, dp_m) = dl_H(p_1, \dots, p_m)$$

We now associate a partial word  $u$  having weak period set  $\mathbf{p} = (p_1, \dots, p_m)$  with a graph. As in [5] we will freely use the essential model of  $u$  as a graph,  $G_{\mathbf{p}}(u) = (V, E)$ , where  $V$  is the set  $D(u)$  comprising the defined positions of  $u$ , and  $E = \bigcup_{1 \leq k \leq m} (E_{1,k} \cup E_{2,k})$  where  $E_{1,k} = \{\{i, i - p_k\} \mid i, i - p_k \in V\}$  and  $E_{2,k} = \{\{i, i + p_k\} \mid i, i + p_k \in V\}$ . For  $0 \leq j < \gcd(\mathbf{p})$ , the subgraph of  $G_{\mathbf{p}}(u)$  corresponding to  $D(u) \cap N_j$  will be denoted by  $G_{\mathbf{p}}^j(u)$ . Whenever  $\gcd(\mathbf{p}) = 1$ ,  $G_{\mathbf{p}}^0(u)$  is just  $G_{\mathbf{p}}(u)$ . A *component* of  $u$  refers to a component of its graph. Note that  $u$  is  $g$ -periodic if and only if it has  $g$  components, corresponding to the  $g$  cosets  $j + lg \mid l \in \mathbb{Z}$  for  $0 \leq j < g$ , so that we may call  $u$  *connected* if it is 1-periodic. Finally, all the subsets  $S$  described above in the characterization of different types of isolation are clearly components or unions of components of  $u$ . Placing a hole in a partial word corresponds to removing a vertex and all its incident edges from its associated graph, which may increase the number of components. Unless stated otherwise, the graph associated with a partial word refers to the graph once all the vertices corresponding to holes in the word are removed (along with their incident edges). We will refer to the *leftmost* element of a component  $C$  as the smallest index of any position in  $u$  whose corresponding vertex in the graph of  $u$  is contained in  $C$ . Likewise, the *rightmost* element of  $C$  is the largest index of any position in  $u$  whose corresponding vertex in the graph of  $u$  is contained in  $C$ .

The following reduction makes it possible to determine whether a potential counterexample word with weak periods  $p_1, \dots, p_m$  is  $(H, (p_1, \dots, p_m))$ -special simply by looking at the word.

**Proposition 2.7.** Let  $u$  be a partial word of length  $n$  with weak period set  $\mathbf{p} = (p_1, \dots, p_m)$ . Then  $u$  is  $(H, \mathbf{p})$ -special if and only if a component  $C$  of  $u$  that is a nonempty proper subset of  $D(u) \cap N_j$  for some  $0 \leq j < \gcd(\mathbf{p})$ , where  $N_j = \{j + k \gcd(\mathbf{p}) \mid k \in \mathbb{Z}\}$  for nonnegative integers  $k$ , exists such that either the leftmost element of  $C$  is at least  $p_m$  or the rightmost element of  $C$  is at most  $n - 1 - p_m$ .

Generally, in potential counterexample words, two positions will contain the same letter if and only if they are in the same component. This proposition makes it easy to tell if such a word is  $(H, \mathbf{p})$ -special at a glance. For instance, if  $\mathbf{p} = (2, 3, 5)$ , the word  $aa\circ a\circ\circ\circ bab$  is  $(4, \mathbf{p})$ -special since the component of  $b$ 's is 3-isolated (the leftmost  $b$  is in a position greater than or equal to 5), but the word  $\circ\circ aba\circ\circ aba$  is not  $(4, \mathbf{p})$ -special. The following corollary will also be useful.

**Corollary 2.8.** If a partial word  $u$  with weak periods  $p_1, \dots, p_m$  has a subword  $w$  of length  $\geq p_m$  that is  $\gcd(p_1, \dots, p_m)$ -periodic, then either  $u$  is  $\gcd(p_1, \dots, p_m)$ -periodic or  $u$  is  $(H, (p_1, \dots, p_m))$ -special.

It should be noted that if  $\gcd(p_1, \dots, p_m) = 1$ , the corollary states that if  $u$  has a unary (1-periodic) subword of length  $\geq p_m$ , either  $u$  is unary or  $u$  is  $(H, (p_1, \dots, p_m))$ -special.

When the types of isolation described above are excluded, we have that the critical length always exists. As the word length increases without bound, eventually there must be at least  $\max(l_0, p_m)$  filled positions between a pair of consecutive holes (or between a hole and the edge of the word), causing the whole word to be either  $\gcd(\mathbf{p})$ -periodic or special by Corollary 2.8.

### 3 Cases of $H$ and $m$

In this section, we discuss the case of  $m = 2$  and arbitrary  $H$ , then the case of  $H = 1$  and arbitrary  $m$ , and finally the case of  $H = 2$  and  $m = 3$ . We first recall the case of one hole.

**Theorem 3.1** ([1]). *Let  $p$  and  $q$  satisfy  $p < q$ , and let  $u$  be a partial word with one hole. If  $u$  is weakly  $p$ -periodic and weakly  $q$ -periodic and  $|u| \geq p + q$ , then  $u$  is  $\gcd(p, q)$ -periodic.*

The case of partial words with an arbitrary number of holes generalizes the preceding result since the class of  $(1, (p, q))$ -special partial words is empty.

**Theorem 3.2** ([3]). *Let  $p$  and  $q$  satisfy  $p < q$ , and let  $u$  be a non  $(H, (p, q))$ -special partial word with  $H$  holes. If  $u$  is weakly  $p$ -periodic and weakly  $q$ -periodic, then the following hold:*

- *If  $H$  is even and  $|u| \geq \frac{H+2}{2}(p+q) - \gcd(p, q)$ , then  $u$  is  $\gcd(p, q)$ -periodic.*
- *If  $H$  is odd and  $|u| \geq \frac{H+1}{2}(p+q)$ , then  $u$  is  $\gcd(p, q)$ -periodic.*

It has been shown that the bounds in Theorem 3.2 are in general critical.

**Proposition 3.3** ([3]). *Let  $p$  and  $q$  satisfy  $q = np + r$  with  $n > 0$  and  $0 < r < p$ , and let  $H > 0$ . Except for a few cases,*

$$l_H(p, q) = \begin{cases} \frac{H+2}{2}(p+q) - \gcd(p, q) & \text{if } H \text{ is even} \\ \frac{H+1}{2}(p+q) & \text{if } H \text{ is odd} \end{cases}$$

The few cases referred to in the preceding proposition are: the case when  $q - p = \gcd(p, q)$ , the critical length being  $\gcd(p, q)$  less than the stated one, and the case where  $2q = 3p$  which is *intractable* with “skeleton” word  $\diamond aba \diamond$  for  $p = 2$  and  $q = 3$  (see Proposition 3.5, where the “skeleton” word is  $a \diamond b$ , for what is meant).

The case where  $q = np$  with  $n > 1$  is not trivial. While any word with strong period  $p$  is certainly periodic for any multiple of  $p$ , the same is not true for weak periods. This case can, however, be reduced to the case where  $\gcd(p, q) = 1$ , or equivalently,  $p = 1$ . Any word with a weak period of 1 has the property that any subword with no holes is necessarily unary, so the holes effectively divide the word into sections each being some power of one letter. Obviously the case where  $H = 0$  is trivial since then weak periodicity implies strong periodicity.

**Proposition 3.4.** *Let  $p$  and  $q$  satisfy  $q = np$  for some integer  $n > 4$ , and let  $H > 0$ . Then*

$$l_H(p, q) = \begin{cases} \frac{H+1}{2}(p+q) & \text{if } H \text{ is odd} \\ \frac{H+2}{2}(p+q) - 3p & \text{if } H \text{ is even and } H \neq 2 \\ 2q & \text{if } H = 2 \end{cases}$$

While a general formula for the critical length for the case where  $p$  divides  $q$  has just been found, it fails in some small cases (hence the assumption in the theorem that  $n > 4$ ). For  $n = 2$ , it is not difficult to see that any partial word with weak periods 1 and 2 and  $H$  holes will be either 1-periodic or  $(H, (1, 2))$ -special since a hole and any letter adjacent to it form a unary subword of length  $q = 2$ . Thus,  $l_H(1, 2) = l_H(p, 2p) = 0$ . The case where  $n = 3$  is somewhat strange.

**Proposition 3.5.** *Let  $p$  and  $q$  satisfy  $q = 3p$ , and let  $H > 0$ . Then*

$$l_H(p, q) = \begin{cases} p + q & \text{if } H \leq 4p - 3 \\ 0 & \text{otherwise} \end{cases}$$

Finally, if  $n = 4$ , we have the following.

**Proposition 3.6.** *Let  $p$  and  $q$  satisfy  $q = 4p$ , and let  $H > 0$ . Then*

$$l_H(p, q) = \begin{cases} (2H + 4)p & \text{if } H > 1 \\ p + q & \text{if } H = 1 \end{cases}$$

As the case of two weak periods has been discussed, we assume henceforth that  $|\mathbf{p}| \geq 3$ . The case  $H = 0$  being covered by [10], we also restrict our attention to  $H > 0$ . The next result applies to partial words with one hole and in fact constitutes a generalization of Theorem 3.1.

**Theorem 3.7.** *Let  $\mathbf{p} = (p_1, \dots, p_m)$  ( $m > 2$ ), and let  $d = \gcd(p_1, \dots, p_{m-1})$ . Then*

$$l_1(p_1, \dots, p_m) = \begin{cases} \max(p_m + d, l_1(p_1, \dots, p_{m-1})) & \text{if } d \nmid p_m \\ l_1(p_1, \dots, p_{m-1}) & \text{if } d \mid p_m \end{cases}$$

*This can also be written as*

$$l_1(p_1, \dots, p_m) = \max_{\substack{2 < k \leq m \\ \gcd(p_1, \dots, p_{k-1}) \mid p_k}} (l_1(p_1, p_2), p_k + \gcd(p_1, \dots, p_{k-1}))$$

*for  $p_2 \neq 2p_1$ .*

For the next logical step, partial words with two holes, we consider the least case that has not yet been done, that is, the case with three weak periods  $p$ ,  $q$  and  $r$  where  $p < q < r$  (when drawing graphs,  $p$  will be represented horizontally and  $q$  vertically). As in our further analyses there is one main dichotomy: between the weak period sets  $(p, q, r)$  for which  $\gcd(p, q) = 1$  and for which  $\gcd(p, q) > 1$ . This becomes apparent mostly in the proof of the following proposition.

**Proposition 3.8.** *The critical length for weak period set  $(p, q, r)$  is*

$$l_2(p, q, r) = \begin{cases} 0 & \text{if } q = 2p \text{ and } r = 3p \\ q + r - \gcd(p, q, r) & \text{if } r = 2p \\ \max(r + p, 2(p + q) - \gcd(p, q, r)) & \text{if } p \mid r - q \\ \max(2q, r + 2p - \gcd(p, q, r)) & \text{if } p \mid q, p \nmid r - q, r \neq 2p \\ \max(2q, r + p) & \text{if } p \nmid q, p \nmid r - q, r \neq 2p \\ & \text{or } p \mid q, p \mid r \end{cases}$$

Note that the other cases occur relatively infrequently compared to the fifth. A few optimal words are worthy of interest. Those of length  $2q - 1$  and  $r + p - 1$ , for instance, are given by the so-called ‘‘pseudo-special forms’’

$$a^{q-p-1} \diamond a^{p-1} b a^{p-1} \diamond a^{q-p-1}$$

$$a^{p-1} b a^{p-1} \diamond a^{q-p-1} \diamond a^{r-q-1}$$

where there is a single, felicitously placed  $b$  and holes marking either  $p$  to the left and right of it or  $p$  and  $q$  to one side. In either case the next smallest periods, i.e. in the first case  $q$  left and right,

in the second  $p$  left and  $r$  right, fall just outside the word bounds, so that the word indeed achieves its maximal length with the given number of holes.

Moreover, analogously to  $H = 1$  above, the case when  $d = \gcd(p, q) > 1$  involves forming cycles between the  $d$  cosets, connecting via the period of  $r$ . The graph will be connected as long as  $r$  may be added to the lowest position of each coset (likewise the greatest position of each coset) and the second-lowest position of all but one coset (similarly for the second-greatest positions). This is because in such a situation there will be only a single link between the cosets  $\langle d \rangle - 1$  and  $\langle d \rangle + r - 1$ , but a double link between all other cosets  $\langle d \rangle + j$  and  $\langle d \rangle + j + r$ , making the graph connected as long as each coset is itself connected: unless both holes are in the same coset, still ensuring all of them connect cyclically, neither hole will serve a connection between cosets except if between the  $\langle d \rangle - 1$  or  $\langle d \rangle + r - 1$  cosets, the loss of one connection being insufficient to disconnect the entire graph. Thus we have the bound of  $r + 2d - 1$ , exceeded in turn by the bound of  $r + p$  unless  $d = p$  for obvious reasons.

It is important to note that the optimal word corresponds to there being another single link between  $\langle d \rangle - 2$  and  $\langle d \rangle + r - 2$ , with the two components being  $\langle d \rangle + r - 1, \langle d \rangle + 2r - 1, \dots, \langle d \rangle - 2$  and  $\langle d \rangle + r - 2, \langle d \rangle + 2r - 2, \dots, \langle d \rangle - 1$  with holes at  $d - 2$  and  $d - 1$ . This stands in contrast to the situations with 1 or 3 holes, wherein the alternate component of the optimal word consists of the single coset  $\langle d \rangle - 1$  that also contains the word's holes. The importance of having an alternate component reside in a single coset is evident when we try to impose greater periods: the absence of links to that coset signifies that a difference of the largest period  $r$  to every (non-hole) position in the coset goes outside the word. This property allows this word to remain a counterexample even as larger periods are added, as opposed to the 2-hole situation in which there may be  $r$ -links between cosets in either component. Due to this and other complications, we will restrict our attention in the next two sections to odd numbers of holes.

A summary of the generalizations of the above cases are then as follows:

| Case                                  | Bound                                   | Section |
|---------------------------------------|---|---------|
| $q = 2p$ and $r = 3p$                 | 0                                       | 6       |
| $r = 2p$                              | $q + r - \gcd(p, q, r)$                 | 5       |
| $p \mid r - q$                        | $\max(r + p, 2(p + q) - \gcd(p, q, r))$ | 5       |
| $p \mid q, p \nmid r - q, r \neq 2p$  | $\max(2q, r + 2p - \gcd(p, q, r))$      | 4, 5    |
| $p \nmid q, p \nmid r - q, r \neq 2p$ | $\max(2q, r + p)$                       | 4, 5    |
| $p \mid q, p \mid r$                  | $\max(2q, r + p)$                       | 5       |

## 4 Counterexample forms and interference

For most weak period sets  $\mathbf{p} = (p_1, \dots, p_m)$  with a reasonably small number of holes, the critical length and the form of the maximal counterexample are predictable.

“Pseudo-special” form: one  $b$ . This is the most common maximal counterexample form for  $H \leq 2m - 2$ . It contains one position with a  $b$  which is “isolated” from the rest of the word by holes. The critical lengths represented by the pseudo-special forms are  $p_k + p_{H+2-k}$ . The length of the word itself is one less. For example, if  $\mathbf{p} = (2, 6, 9, 14)$  and  $H = 3$ , then  $aba \diamond aaaa \diamond aa \diamond aaaa$  has length  $p_1 + p_4 - 1 = 15$  while  $aaa \diamond aba \diamond aaaa \diamond aa$  has length  $p_2 + p_3 - 1 = 14$ . Every word of pseudo-special form has the property that the largest period cannot be added or subtracted from the position containing the  $b$  without falling outside of the word.

Except in cases such as  $\mathbf{p} = (1, 2, 3, \dots)$ , the pseudo-special word is always a valid counterexample if  $H < m$ . If  $m \leq H \leq 2m - 2$ , the pseudo-special word is *almost* always valid. Consider the period set  $\mathbf{p} = (3, 4, 10, 11)$  with  $H = 4$  holes. In theory, the maximal pseudo-special word should be  $aaaaa \diamond aabaa \diamond aaaaa$ , which has length  $p_3 + p_3 - 1 = 10 + 10 - 1 = 19$ , but it fails. Why?

“Separated coset” form:  $b$ 's in a modular equivalence class. The general counterexample form for  $H = 1$  has an equivalence class of  $b$ 's and holes (see Theorem 3.7). It is actually a specific case of

the separated coset form. A counterexample word in the separated coset form (with an odd number of holes) has length  $p_k + \frac{H+1}{2} \gcd(p_1, \dots, p_{k-1}) - 1$ , where  $k > 2$  and  $\gcd(p_1, \dots, p_{k-1}) \nmid p_k$ . For example, take  $\mathbf{p} = (3, 6, 9, 15, 22, 23)$  with  $H = 5$ . The separated coset form also has the property that the largest period cannot be added or subtracted from any position containing a  $b$  without falling outside of the word.

The separated coset form fails when

$$H > \lfloor \frac{p_k}{\gcd(p_1, \dots, p_{k-1})} \rfloor + \frac{H+1}{2}$$

The example  $\mathbf{p} = (4, 8, 9, 10)$  with  $H = 5$  illustrates this phenomenon of *annihilation*.

The following results hold.

**Proposition 4.1.** *Let  $H < m$ . For an  $1 - O(N^{1-m})$  fraction of weak period sets  $\mathbf{p} = (p_1, \dots, p_m)$  with  $p_m \leq N$ , in particular when*

$$\mathbf{p} \neq (d, 2d, \dots, \frac{H+3}{2}d, \dots)$$

*for some  $d$ , a word can be constructed of length  $p_k + p_{H+2-k} - 1$  for each  $1 \leq k \leq \frac{H}{2} + 1$  which has weak periods  $p_1, \dots, p_m$  and is not  $\gcd(\mathbf{p})$ -periodic.*

**Lemma 4.2.** *Consider the set  $P_N$  of all possible weak period sets  $\mathbf{p} = (p_1, \dots, p_m)$  such that  $1 \leq p_1 < \dots < p_m \leq N$ , and let  $S_{1,N} \subset P_N$  such that  $\frac{|S_{1,N}|}{|P_N|} = O(N^{-s})$ , respectively  $\Omega(N^{-s})$  for some  $s < m - 1$  as  $N \rightarrow \infty$  and  $\gcd(p_1, \dots, p_m) = 1$  for all  $(p_1, \dots, p_m) \in S_{1,N}$ . The subsets  $P_{j,N} = \{\mathbf{p} \mid j \text{ divides } \gcd(\mathbf{p})\}$  and  $S_{j,N} = \{\mathbf{p} \mid (\frac{p_1}{j}, \dots, \frac{p_m}{j}) \in S_{1,N}\}$  of  $P_N$  then satisfy  $\frac{|S_N|}{|P_N|} = O(N^{-s})$ , respectively  $\Omega(N^{-s})$  for  $S_N := \bigcup_j S_{j,N}$ .*

**Proposition 4.3.** *Let  $m \leq H \leq 2m - 3$ . For an  $1 - O(N^{-\lfloor \frac{m}{2} \rfloor})$  fraction of weak period sets  $\mathbf{p} = (p_1, \dots, p_m)$  with  $p_m \leq N$  a word can be constructed of length  $p_k + p_{H+2-k} - 1$  for each  $H + 2 - m \leq k \leq \frac{H+1}{2}$  which has weak periods  $p_1, \dots, p_m$ , is not  $(H, \mathbf{p})$ -special, and is not  $\gcd(\mathbf{p})$ -periodic.*

Referring to Proposition 4.3, the reader is invited to check that  $\mathbf{p} = (2, 3, 7, 8)$  is an example when the pseudo-special word actually fails.

**Proposition 4.4.** *Let  $H < 2m - 3$  be odd. For each weak period set  $\mathbf{p} = (p_1, \dots, p_m)$ , a word can be constructed of length  $p_k + \frac{H+1}{2} \gcd(p_1, \dots, p_{k-1})$  that has weak periods  $p_1, \dots, p_m$ , is not  $(H, \mathbf{p})$ -special, and is not  $\gcd(\mathbf{p})$ -periodic, for any  $\frac{H+3}{2} < k \leq m$  such that  $\gcd(p_1, \dots, p_{k-1})$  does not divide  $p_k$ .*

That these counterexample words are indeed maximal for their weak period sets, at least most of the time, is the content of our main theorem.

**Theorem 4.5.** *The critical length for an  $1 - O(N^{-1})$  fraction of weak period sets  $(p_1, \dots, p_m)$  with  $p_m \leq N$  and  $H \leq 2m - 3$  odd is  $\max(S, T)$  for*

$$S = \max_{\max(1, H+2-m) \leq k \leq \frac{H}{2} + 1} (p_k + p_{H+2-k})$$

$$T = \max_{\substack{H+1 < k < m \\ \gcd(p_1, \dots, p_{k-1}) \nmid p_k}} (p_k + \frac{H+1}{2} \gcd(p_1, \dots, p_{k-1}))$$

We perform a series of reductions to the case when  $H = 2m - 3$ , then prove the bound of  $p_m + p_{m-1}$  for that case.

We have already seen some instances of *shortening interference* in the cases where the pseudo-special word fails or the separated coset is annihilated, where the critical length is smaller than expected (such as  $\mathbf{p} = (2, 3, 7, 8)$  as mentioned above). There can also be *lengthening interference* where periods combine in such a way to allow for unusual counterexample forms longer than the typical forms. In general, for lengthening interference to occur, one period (usually the largest) must be the sum of two smaller periods (such as  $\mathbf{p} = (5, 8, 11, 16)$  with  $H = 3$ ).

## 5 Case of arbitrary $m$ and odd $H$ with $H \leq 2m - 3$

In this section, we discuss partial words with  $m$  weak periods and an odd number of holes  $H$ .

If  $H < m - 1$ , then the critical length is computed as follows.

**Theorem 5.1.** *The critical length for every weak period set  $\mathbf{p} = (p_1, \dots, p_m)$  when  $H < m - 1$  is*

$$l_H(\mathbf{p}) = \begin{cases} \max(p_m + \frac{H+1}{2}d, l_H(p_1, \dots, p_{m-1})) & \text{if } d \nmid p_m \\ l_H(p_1, \dots, p_{m-1}) & \text{if } d \mid p_m \end{cases}$$

where  $d = \gcd(p_1, \dots, p_{m-1})$ .

This situation is analogous to that found for 1 hole, where there is a basic ‘‘pseudo-special bound’’ for low numbers of periods (predictably, 2 in this case) and after which new periods affect the bound only when they change the greatest common divisor of all previous periods.

If  $m - 1 \leq H < 2m - 3$ , then the critical length is as stated in the next theorem.

**Theorem 5.2.** *The critical length for an  $1 - O(N^{-1})$  fraction of weak period sets  $\mathbf{p} = (p_1, \dots, p_m)$  with  $p_m \leq N$  when  $m - 1 \leq H < 2m - 3$  is*

$$l_H(\mathbf{p}) = \begin{cases} \max(p_m + p_{H+2-m}, p_m + \frac{H+1}{2}d, l_H(p_1, \dots, p_{m-1})) & \text{if } d \nmid p_m \\ \max(p_m + p_{H+2-m}, l_H(p_1, \dots, p_{m-1})) & \text{if } d \mid p_m \end{cases}$$

where  $d = \gcd(p_1, \dots, p_{m-1})$ .

The following proposition shows that the first option above can in fact be removed without changing the order of magnitude of exception cases.

**Proposition 5.3.** *An  $1 - O(N^{m-H-2})$  fraction of weak period sets  $(p_1, \dots, p_m)$  with  $p_m \leq N$  and  $d = \gcd(p_1, \dots, p_{m-1}) \nmid p_m$  satisfy  $p_{H+2-m} \geq \frac{H+1}{2}d$ .*

If  $H = 2m - 3$ , then the following result holds.

**Theorem 5.4.** *The critical length for an  $1 - O(N^{-1})$  fraction of weak period sets  $(p_1, \dots, p_m)$  with  $p_m \leq N$  when  $H = 2m - 3$  is  $p_{m-1} + p_m$ .*

Experimental data suggests, however, that this theorem can be strengthened.

**Conjecture 5.5.** *The critical length for all weak period sets  $(p_1, \dots, p_m)$  not satisfying  $p_i + p_j = p_k$  for some  $1 \leq i, j, k \leq m$  when  $H = 2m - 3$  is  $p_{m-1} + p_m$ .*

It would then follow that the hypotheses of Theorem 4.5 can be extended thus we conjecture the following.

**Conjecture 5.6.** *The critical length for all weak period sets  $(p_1, \dots, p_m)$  with*

$$p_{\frac{H+3}{2}} > \frac{H+3}{2} \gcd(p_1, \dots, p_{\frac{H+3}{2}})$$

*not satisfying  $p_k + p_{k'} = p_l + p_{l'}$  for some  $0 \leq k, k', l, l' \leq m$  when  $H \leq 2m - 3$  is odd and  $p_0 = 0$  is given by  $\max(S, T')$ , where*

$$S = \max_{\max(1, H+2-m) \leq k \leq \frac{H}{2} + 1} (p_k + p_{H+2-k})$$

*as before and*

$$T' = \max_{\substack{\frac{H+3}{2} < k < m \\ \gcd(p_1, \dots, p_{k-1}) \mid p_k}} (p_k + \frac{H+1}{2} \gcd(p_1, \dots, p_{k-1}))$$

Note that all the  $O(N^{-s})$  exclusions except those in Theorem 5.4 above have fit this description. The first exception made in the previous conjecture is of a simple enough form as to generalize to arbitrarily many holes.

## 6 Intractable weak period sets

For most weak period sets, as the number of holes increases without bound, the critical length does so as well. However, for some dense weak period sets, the critical length has a finite maximum that can never be surpassed regardless of the number of holes. For these sets, which we will refer to as intractable sets, once the number of holes gets too large, the critical length becomes zero.

**Definition 6.1.** *A weak period set  $(p_1, \dots, p_m)$  is intractable if there exists an integer  $N$  such that for any number of holes  $H$ ,  $l_H(p_1, \dots, p_m) \leq N$ .*

A few intractable sets have already been discussed in this paper. For example, the weak period set  $(2, 3)$  is intractable, which can be seen from the definition by choosing  $N \geq 6$ . Another example is  $\mathbf{p} = (2, 4, 5)$ . We have

$$\begin{aligned} l_0(\mathbf{p}) &= 6 & ababa \\ l_1(\mathbf{p}) &= 7 & \diamond ababa \\ l_2(\mathbf{p}) &= 8 & \diamond\diamond ababa \\ l_3(\mathbf{p}) &= 9 & \diamond\diamond\diamond ababa \\ l_4(\mathbf{p}) &= 10 & \diamond\diamond\diamond ababa\diamond \\ l_5(\mathbf{p}) &= 12 & a\diamond ba\diamond ab\diamond a \\ l_6(\mathbf{p}) &= 13 & \diamond a\diamond ba\diamond ab\diamond a \\ l_7(\mathbf{p}) &= 14 & \diamond a\diamond ba\diamond ab\diamond a\diamond \end{aligned}$$

But for  $H \geq 8$ , the critical length is 0, that is, any word that is not 1-periodic is  $(H, \mathbf{p})$ -special.

The following proposition shows a property all intractable sets have in common.

**Proposition 6.2.** *If a weak period set  $\mathbf{p} = (p_1, \dots, p_m)$  is intractable, then for every weak period  $p_k$  with  $1 \leq k \leq m$  and every integer  $i$  such that  $1 \leq i < p_k$  and  $\gcd(\mathbf{p})$  divides  $i$ , there exists a weak period  $p_j \in \mathbf{p}$  that is congruent to either  $i$  or  $p_k - i$  modulo  $p_k$ .*

**Corollary 6.3.** *For any given  $m$ , there are only finitely many weak period sets  $\mathbf{p} = (p_1, \dots, p_m)$  with  $\gcd(\mathbf{p}) = 1$  that are intractable.*

The converse of Proposition 6.2 is not true. The weak period set  $\mathbf{p} = (2, 3, 6, 7, 8, 9, 12, 13)$  satisfies the condition that for every weak period  $p_k \in \mathbf{p}$  and every integer  $i$  such that  $1 \leq i < p_k$ , there exists a weak period  $p_j \in \mathbf{p}$  that is congruent to either  $i$  or  $p_k - i$  modulo  $p_k$ ; however, the word  $u = ab\diamond\diamond ca\diamond\diamond bc\diamond\diamond ab\diamond\diamond ca\diamond\diamond bc\diamond\diamond \dots$  has every weak period in  $\mathbf{p}$  and can be extended indefinitely as the number of holes increases, showing  $\mathbf{p}$  is not intractable.

## 7 Refinements and remarks

We first discuss order-analytic significance of known results. That each of the bounds in Theorem 4.5 is order-analytically significant, occurring for a non-diminishing proportion of weak period sets  $1 \leq p_1 < \dots < p_m \leq N$  as  $N \rightarrow \infty$ , is the content of the following observation.

**Corollary 7.1.** *Each of the formulae of Theorem 4.5 are attained in an  $\Omega(1)$  fraction of weak period sets  $(p_1, \dots, p_m)$ .*

We now discuss order-analytic significance of unknown results.

**Proposition 7.2.** *There is an  $\Omega(N^{-s})$  proportion of weak period set exceptions  $1 \leq p_1 < \dots < p_m \leq N$  to the lower bound in Theorem 4.5, where  $s = m - \frac{H+1}{2}$  and  $m \leq H + 1$ .*

**Corollary 7.3.** *There is an  $\Omega(N^{-s})$  proportion of weak period set exceptions  $1 \leq p_1 < \dots < p_m \leq N$  to the lower bound in Theorem 4.5, where  $s = \min(m - \frac{H+1}{2}, \frac{H+1}{2})$ .*

While experimental data seems to indicate that this bound may be improved, i.e. that the true value of  $s$  is greater, nothing in this regard has been understood. In particular, it is not clear why  $s$  would attain a non-integral value when  $H = 2m - 5$ , though this is the hypothesis best evinced by the evidence.

In general there is much that remains to be understood about these “interference cases,” and they seem at a glance to obey no set order. For example, while one might think that the bound for two weak periods and arbitrarily many holes given by Blanchet-Sadri [3] would be an upper bound for the length of a counterexample word with weak periods  $(p_1, \dots, p_m)$  when applied to the first two weak periods  $(p_1, p_2)$ , because the additional weak periods would only impose additional restrictions on its position labels, these weak periods in fact prevent an otherwise  $(p_1, p_2)$ -special word from being  $(p_1, \dots, p_m)$ -special, in the case of  $m = 6, H = 5$ , weak periods 3, 4, 8, 12, 13, 16, the old bound being  $3(3 + 4) = 21$  and a counterexample word of that length  $aabaa \diamond aaaa \diamond aaaa \diamond aabaa$  – this in fact taking the form found in the previous proposition.

In summary, we have studied the minimal length  $l_H(\mathbf{p})$  for any so-called non  $(H, \mathbf{p})$ -special partial word with  $H$  holes and weak period set  $\mathbf{p} = (p_1, \dots, p_m)$  to have also strong period  $\gcd(\mathbf{p})$ . The values  $(m, H)$  studied so far are: The value  $(2, 0)$  was studied by Fine and Wilf [14]. Then the value  $(2, 1)$  by Berstel and Boasson [1]; the values  $(2, 2)$  and  $(2, 3)$  by Blanchet-Sadri and Hegstrom [5], and the values  $(2, H)$  for  $H$  arbitrary by Blanchet-Sadri [3]; the value  $(3, 0)$  by Castelli, Mignosi and Restivo [8], and the values  $(m, 0)$  for  $m$  arbitrary by Justin [17], Tijdeman and Zamboni [29], and Constantinescu and Ilie [10]. Finally, values of the form  $(m, H)$  for  $m$  arbitrary and  $H$  arbitrary odd by this paper.

## References

- [1] Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. *Theoret. Comput. Sci.* **218** (1999) 135–141
- [2] Berstel, J., Perrin, D.: *Theory of Codes*. Academic Press, Orlando, FL (1985)
- [3] Blanchet-Sadri, F.: *Algorithmic Combinatorics on Partial Words*. Chapman & Hall/CRC Press, Boca Raton, FL (2007)
- [4] Blanchet-Sadri, F., Bal, D., Sisodia, G.: Graph connectivity, partial words, and a theorem of Fine and Wilf. *Inform. Comput.* **206** (2008) 676–693 ([www.uncg.edu/mat/research/finewilf3](http://www.uncg.edu/mat/research/finewilf3))
- [5] Blanchet-Sadri, F., Hegstrom, R.A.: Partial words and a theorem of Fine and Wilf revisited. *Theoret. Comput. Sci.* **270** (2002) 401–419

- [6] Blanchet-Sadri, F., Mandel, T., Sisodia, G.: Computing strong periods in partial words. Preprint ([www.uncg.edu/cmp/research/finewilf4](http://www.uncg.edu/cmp/research/finewilf4))
- [7] Buhler, J., Keich, U., Sun, Y.: Designing seeds for similarity search in genomic DNA. *J. Comput. System Sci.* **70** (2005) 342–363
- [8] Castelli, M.G., Mignosi, F., Restivo, A.: Fine and Wilf’s theorem for three periods and a generalization of Sturmian words. *Theoret. Comput. Sci.* **218** (1999) 83–94
- [9] Choffrut, C., Karhumäki, J.: Combinatorics of Words. In Rozenberg, G., Salomaa, A. (eds.): *Handbook of Formal Languages*. Vol. 1. Ch. 6. Springer-Verlag, Berlin (1997) 329–438
- [10] Constantinescu, S., Ilie, L.: Generalised Fine and Wilf’s theorem for arbitrary number of periods. *Theoret. Comput. Sci.* **339** (2005) 49–60
- [11] Crochemore, M., Hancart, C., Lecroq, T.: *Algorithms on Strings*. Cambridge University Press (2007)
- [12] Crochemore, M., Rytter, W.: *Jewels of Stringology*. World Scientific, NJ (2003)
- [13] Farach-Colton, M., Landau, G.M., Sahinalp, S.C., Tsur, D.: Optimal spaced seeds for approximate string matching. In Caires, L., Italiano, G.F., Monteiro, L., Palanidessi, C., Yung, M. (eds.): *ICALP 2005, Lecture Notes in Comput. Sci.* Vol. 3580. Springer-Verlag, Berlin (2005) 1251–1262
- [14] Fine, N.J., Wilf, H.S.: Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.* **16** (1965) 109–114
- [15] Gusfield, D.: *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press, Cambridge (1997)
- [16] Ilie, L.: Multiple spaced seeds for homology search. *Bioinformatics* **23** (2007) 2969–2977
- [17] Justin, J.: On a paper by Castelli, Mignosi, Restivo. *Theoret. Inform. Applic.* **34** (2000) 373–377
- [18] Keich, U., Li, M., Ma, B., Tromp, J.: On spaced seeds for similarity search. *Discrete Appl. Math.* **138** (2004) 253–263
- [19] Kolpakov, R., Kucherov, G.: Finding approximate repetitions under hamming distance. *Theoret. Comput. Sci.* **33** (2003) 135–156
- [20] Landau, G.M., Schmidt, J.P., Sokol, D.: An algorithm for approximate tandem repeats. *J. Comput. Biology* **8** (2001) 1–18
- [21] Lothaire, M.: *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge (2002)
- [22] Ma, B., Tromp, J., Li, M.: PatternHunter: faster and more sensitive homology search. *Bioinformatics* **18** (2002) 440–445
- [23] Nicolas, F., Rivals, E.: Hardness of optimal spaced seed design. *CPM 2005, 16th Annual Symposium on Combinatorial Pattern Matching* (2005) 144–155
- [24] Noé, L., Kucherov, G.: Improved hit criteria for DNA local alignment. *BMC Bioinformatics* **5** (2004)
- [25] Schmidt, J.P.: All highest scoring paths in weighted grid graphs and their application to finding all approximate repeats in strings. *SIAM J. Comput.* **27** (1998) 972–992
- [26] Shur, A.M., Gamzova, Y.V.: Partial words and the periods’ interaction property. *Izvestiya RAN* **68** (2004) 199–222
- [27] Smyth, W.F.: *Computing Patterns in Strings*. Pearson Addison-Wesley (2003)
- [28] Storer, J.A.: *Data Compression: Methods and Theory*. Computer Science Press, Rockville, MD (1988)
- [29] Tijdeman, R., Zamboni, L.: Fine and Wilf words for any periods. *Indag. Math.* **14** (2003) 135–147
- [30] Ziv, J., Lempel, A.: A universal algorithm for sequential data compression. *IEEE Trans. Inform. Theory* **23** (1977) 337–343