

## ABELIAN PERIODS, PARTIAL WORDS, AND AN EXTENSION OF A THEOREM OF FINE AND WILF \*

FRANCINE BLANCHET-SADRI<sup>1</sup>, SEAN SIMMONS<sup>2</sup>, AMELIA TEBBE<sup>3</sup>  
AND AMY VEPRASKAS<sup>4</sup>

**Abstract.** Recently, Constantinescu and Ilie proved a variant of the well-known periodicity theorem of Fine and Wilf in the case of two relatively prime abelian periods and conjectured a result for the case of two non-relatively prime abelian periods. In this paper, we answer some open problems they suggested. We show that their conjecture is false but we give bounds, that depend on the two abelian periods, such that the conjecture is true for all words having length at least those bounds and show that some of them are optimal. We also extend their study to the context of partial words, giving optimal lengths and describing an algorithm for constructing optimal words.

**Mathematics Subject Classification.** 68R15, 68Q25.

---

*Keywords and phrases.* Combinatorics on words, Fine and Wilf's theorem, partial words, abelian periods, periods, optimal lengths.

\* *This material is based upon work supported by the National Science Foundation under Grant Nos. DMS-0754154 and DMS-1060775. The Department of Defense is also gratefully acknowledged. Part of this paper was presented at JM 2010 [9].*

*This article was submitted within the framework of the "Journées Montoises 2010".*

<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, Massachusetts Institute of Technology, Building 2, Room 236, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA

<sup>3</sup> Department of Mathematics, University of Illinois, 1409 W. Green Street, Urbana, IL 61801, USA

<sup>4</sup> Department of Mathematics, The University of Arizona, 617 N. Santa Rita Ave., P.O. Box 210089 Tucson, AZ 85721-0089, USA

## 1. INTRODUCTION

Computing periods in words has important applications in data compression, string searching and pattern matching algorithms. The notion of *period* is central in combinatorics on words. Although there are many fundamental results on periods of words, the one of Fine and Wilf is perhaps the best known [18]. It states that any word having two periods  $p, q$  and length at least  $p+q-\gcd(p, q)$  also has the greatest common divisor of  $p$  and  $q$ ,  $\gcd(p, q)$ , as a period. The length  $p+q-\gcd(p, q)$  is *optimal* since there are examples of shorter words that have periods  $p$  and  $q$  but are not  $\gcd(p, q)$ -periodic [11]. Extensions of Fine and Wilf's result to more than two periods are given in [10, 12, 20, 26]. In particular, Constantinescu and Ilie [12] extend Fine and Wilf's result to words having an arbitrary number of periods and prove that their lengths are optimal. Fine and Wilf's periodicity theorem has been generalized to *partial words*, or finite sequences of symbols over a finite alphabet that may have some don't care symbols or holes [3, 4, 6, 7, 19, 23–25].

The notion of *abelian period*, a generalization of the one of period (see Def. 1.1), was recently introduced by Constantinescu and Ilie. Letting  $A = \{a_1, a_2, \dots, a_k\}$  be an alphabet, the number of occurrences of the letter  $a_i \in A$  in a word  $w$  over  $A$  is denoted by  $|w|_{a_i}$ . The length of  $w$  is  $|w| = \sum_{1 \leq i \leq k} |w|_{a_i}$  and the *Parikh vector* of  $w$  is  $\|w\| = (|w|_{a_1}, |w|_{a_2}, \dots, |w|_{a_k})$ . Note that, for two words  $u$  and  $v$ ,  $\|u\| = \|v\|$  means that  $u$  is a permutation of  $v$  and  $\|u\| \leq \|v\|$  means that  $u$  can be obtained from  $v$  by permuting and, possibly, deleting some of  $v$ 's letters.

**Definition 1.1** [13]. A word  $w$  over an alphabet  $A$  has *abelian period*  $p$  if  $w = u_0 u_1 u_2 \dots u_m u_{m+1}$ , where  $m \geq 1$ ,  $|u_1| = |u_2| = \dots = |u_m| = p$ ,  $|u_0| > 0$ , and  $\|u_0\| \leq \|u_1\| = \|u_2\| = \dots = \|u_m\| \geq \|u_{m+1}\|$ .

For example, the word *bbaaabaabaaba* has abelian period 4 since it can be factorized as *b.baaa.baaa.abaa.ba*. Here, we have used “.” to separate the factors of the word for showing it has abelian period 4 (in the paper, we also use “|” to separate the factors). In [17], Fici et al. show that a word of length  $n$  can have  $O(n^2)$  distinct abelian periods and present a number of algorithms for computing all the abelian periods of a given word. Abelian periods also appear in the literature under the names of *weak repetitions* or *abelian powers* when  $u_0$  and  $u_{m+1}$  are the empty word and  $m > 1$  [14]. Several recent works relate to these notions in both the context of ordinary words and the context of partial words (see, for example, [1, 2, 5, 8, 15, 16, 21, 22]).

Constantinescu and Ilie prove a variant of Fine and Wilf's theorem in the case of two relatively prime abelian periods, while they conjecture that any word having two non-relatively prime abelian periods  $p, q$  has at most cardinality  $\gcd(p, q)$  (or the word contains at most  $\gcd(p, q)$  distinct letters) [13]. More precisely, they prove that any word having two coprime abelian periods  $p, q$  and length at least  $2pq - 1$  has also  $\gcd(p, q) = 1$  as a period. Among a number of problems they suggest, we investigate the following: (1) Is the length  $2pq - 1$  optimal? (2) Is it true that from  $\gcd(p, q) = d$ ,  $d > 1$ , it follows that the word has at most cardinality  $d$ ?

In this paper, we answer Problem (1) affirmatively and Problem (2) negatively. However, we prove that it is true that from  $\gcd(p, q) = d, d > 1$ , it follows that the word has at most cardinality  $d$  if the word is “long enough”, and we give bounds, that depend on  $p$  and  $q$ , on the length. We also extend Constantinescu and Ilie’s result to the context of partial words, giving optimal lengths and describing an algorithm for constructing optimal partial words (a partial word  $w$  with  $h$  holes and having abelian periods  $p, q$  is optimal if the length of  $w$  is one less than the optimal length for the parameters  $h, p$  and  $q$ , and the cardinality of  $w$  is  $\gcd(p, q) + 1$ ). In addition, we have created a World Wide Web server interface which is located at [www.uncg.edu/cmp/research/finewilf6](http://www.uncg.edu/cmp/research/finewilf6) for automated use of a program which constructs an optimal partial word with abelian periods  $p, q$  and  $h$  holes. For  $p$  and  $q$  with  $\gcd(p, q) > 1$ , the program produces an optimal partial word for the case where the periods “match up”.

## 2. NOTATION AND TERMINOLOGY

In this section, we review basic definitions on partial words.

An *alphabet*  $A$  is a non-empty finite set of letters. A *partial word* over  $A$  is a finite sequence over the augmented alphabet  $A_\diamond = A \cup \{\diamond\}$ , where  $\diamond \notin A$  plays the role of a don’t care symbol or hole. More precisely, a partial word  $u$  of length  $n$  (or  $|u|$ ) over  $A$  is a function  $u : \{0, \dots, n - 1\} \rightarrow A_\diamond$ . For  $0 \leq i < n$ , if  $u(i) \in A$ , then  $i$  belongs to the *domain* of  $u$ , denoted by  $i \in D(u)$ , and if  $u(i) = \diamond$ , then  $i$  belongs to the *set of holes* of  $u$ , denoted by  $i \in H(u)$ . We refer to a partial word with an empty set of holes as a (*full*) *word*. The *empty* partial word is the sequence of length zero and is denoted by  $\varepsilon$ . The *cardinality* of a partial word  $u$  is the number of distinct letters in  $u$ . For example,  $ab\diamond bbab\diamond$  has cardinality two since it contains two distinct letters  $a$  and  $b$ . The set of all full (respectively, partial) words over  $A$  of finite length is denoted by  $A^*$  (respectively,  $A_\diamond^*$ ).

For any partial 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  $u[|u| - j..|u|]$  is the *suffix* of  $u$  of length  $j$ . A *period* of  $u$  is a positive integer  $p$  such that  $u(i) = u(j)$  whenever  $i, j \in D(u)$  and  $i \equiv j \pmod p$  (in such a case,  $u$  is *p-periodic*).

If  $u$  and  $v$  are two partial words of equal length, then  $u$  is *contained in*  $v$ , denoted by  $u \subset v$ , if  $u(i) = v(i)$  for all  $i \in D(u)$ . The partial words  $u$  and  $v$  are *compatible*, denoted by  $u \uparrow v$ , if there exists a partial word  $w$  such that  $u \subset w$  and  $v \subset w$ .

When  $w$  is a partial word over  $A = \{a_1, \dots, a_k\}$ , the number of occurrences of  $a_i$  in  $w$  is denoted by  $|w|_{a_i}$ , while the *Parikh vector* of  $w$  by  $\|w\| = (|w|_{a_1}, \dots, |w|_{a_k})$ .

**Definition 2.1.** A partial word  $w$  over an alphabet  $A$  has *abelian period*  $p$  if  $w = u_0u_1u_2 \dots u_mu_{m+1}$ , where  $m \geq 1, |u_1| = |u_2| = \dots = |u_m| = p, 0 < |u_0| \leq p, |u_{m+1}| \leq p$ , and there exists a full word  $v$  over  $A, |v| = p$ , such that for all  $0 \leq i \leq m + 1, \|u_i\| \leq \|v\|$ .

Let  $u_0u_1 \dots u_{m+1}$  and  $v_0v_1 \dots v_{n+1}$  be factorizations of a partial word  $w$  into abelian periods  $p$  and  $q$ , respectively. We say that the periods  $p$  and  $q$  *match up* if the equality  $u_0u_1 \dots u_i = v_0v_1 \dots v_j$  holds for some integers  $i \leq m, j \leq n$ . For example, the partial word  $a.b|a.ab.|ab.a|b.a\diamond.|ab.a|b$  has the abelian periods  $p = 2$  and  $q = 3$  that do match up. Here  $u_0 = a, u_1 = ba, u_2 = u_3 = u_4 = ab, u_5 = a\diamond, u_6 = u_7 = ab, v_0 = ab, v_1 = aab, v_2 = aba, v_3 = ba\diamond, v_4 = aba,$  and  $v_5 = b$  (there are actually two matching points: one after  $u_2$  and  $v_1$  and the other after  $u_5$  and  $v_3$ ). However, the word  $ab.aaa|b.aaab.a|aab.aba|a.aaab.b|aaa.ab$  has the abelian periods  $p = 4$  and  $q = 6$  that do not match up.

### 3. RELATIVELY PRIME ABELIAN PERIODS

Constantinescu and Ilie’s result is stated as follows.

**Theorem 3.1** [13]. *If a word  $w$  has abelian periods  $p$  and  $q$  which are relatively prime and  $|w| \geq 2pq - 1$ , then  $w$  has period  $\gcd(p, q) = 1$ .*

Constantinescu and Ilie proved that the length  $2pq - 1$  is an upper bound but they did not prove that it is optimal. But it is! Indeed, in Section 4 we give an algorithm for constructing non-unary words of length  $2pq - 2$  that have abelian periods  $p$  and  $q$  for any coprime positive integers  $p, q$ . For instance, on input  $p$  and  $q = p + 1$ , our algorithm outputs the optimal word

$$a^{p-1}.b|a^{p-1}.ab|a^{p-2}.a^2b| \dots |a.a^{p-1}b.|a^{p-1}b.a|a^{p-2}b.a^2| \dots |ab.a^{p-1}|b.a^{p-1}$$

of length  $2pq - 2$ .

Here, we repeat Constantinescu and Ilie’s proof from [13] since it contains the ideas that we use later for our own results. For convenience, we adopt their notation. To prove their theorem, they first calculate how many letters in a word  $w$  with abelian periods  $p$  and  $q$ , where  $p$  and  $q$  are relatively prime and  $p < q$ , are needed for the two periods to first match up. If  $u_0u_1 \dots u_{m+1}$  and  $v_0v_1 \dots v_{n+1}$  are factorizations of  $w$  into abelian periods  $p$  and  $q$ , respectively, they calculate how many letters are needed for the periods to match up or for the equality  $u_0u_1 \dots u_i = v_0v_1 \dots v_j$  to hold for some integers  $i, j \leq m$ . They conclude that the periods match up at or before  $pq - 1$  letters. After the first matching, all other matchings occur  $pq$  letters after the previous one. So, a word of length  $2pq - 1$  or greater has at least two matchings.

To calculate this they first write each  $v_i, 1 \leq i \leq n$ , in terms of  $u$ ’s. They set  $v_i = x_iu_{b_i+1}u_{b_i+2} \dots u_{b_i+1}y_i$ , where  $x_i$  is a suffix of  $u_{b_i}$ ,  $y_i$  is a prefix of  $u_{b_i+1}$ , and  $u_{b_i}$  is the first  $u$  such that  $|u_0u_1 \dots u_{b_i}| \geq |v_0v_1 \dots v_{i-1}|$ . This notation is made clearer by Figure 1. So, by definition,  $|x_i| < p$ . Both  $|x_i| + |y_i| \equiv q \pmod p$  and  $|x_{i+1}| + |y_i| = p$  hold. Subtracting the first from the second we get  $|x_{i+1}| \equiv |x_i| - q \pmod p$  and, by induction on  $r, r \geq 1$ , we obtain  $|x_{i+r-1}| \equiv |x_i| - (r-1)q \pmod p$ . In the case where  $i = 1$  we get  $|x_r| \equiv |x_1| - (r-1)q \pmod p$ . Letting  $r = ((|x_1| (q^{-1} \pmod p)) \pmod p) + 1$  we obtain  $|x_r| \equiv 0 \pmod p$ . So  $x_r = \varepsilon$  and  $r \leq p$ .

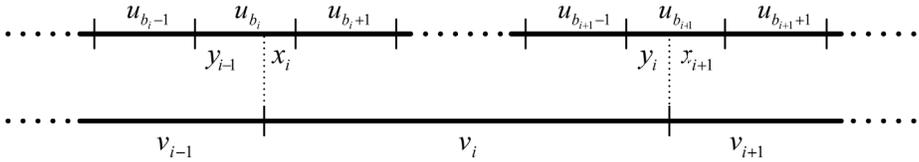


FIGURE 1. The abelian periods of  $w$ .

Hence  $v_0v_1 \dots v_{r-1} = u_0u_1 \dots u_{b_r}$  where  $|v_0v_1 \dots v_{r-1}| = |v_0| + (r - 1)q$  and  $1 \leq |v_0| \leq q$ . Since  $r \leq p$  we get  $|v_0| + (r - 1)q \leq pq$ . However, if  $|v_0| = q$  then  $|x_{r-1}| \equiv |x_0| - (r - 1)q \pmod p$  which implies  $|x_{r-1}| \equiv 0 \pmod p$  and so  $x_{r-1} = \varepsilon$ . This means that  $v_0v_1 \dots v_{r-2} = u_0u_1 \dots u_{b_r}$  and  $|v_0v_1 \dots v_{r-2}| = |v_0| + (r - 2)q \leq q(p - 1)$ . So, if  $|v_0| = q$  we obtain the equality  $q$  letters sooner and so the value is largest when  $|v_0| \neq q$ . This implies, however, that  $|v_0| + (r - 1)q \leq pq - 1$ . So the first matching occurs at or before  $pq - 1$  letters. Note that the first matching occurs at exactly  $pq - 1$  letters when  $|u_0| = p - 1$  and  $|v_0| = q - 1$ .

For any integers  $i$  and  $j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ,  $\alpha = \|v_i\|$  and  $\beta = \|u_j\|$  have the same non-zero components. Further, since there are  $q$  letters in  $v_i$ , the sum of the non-zero components in  $\alpha$  is equal to  $q$ . Denote  $\alpha_l$  (respectively,  $\beta_l$ ) to be the number of times the letter  $a_l$  occurs within one abelian  $q$ -period (respectively,  $p$ -period). Now the number of times  $a_l$  occurs in the subword  $v_rv_{r+1} \dots v_{r+p-1} = u_{b_r+1}u_{b_r+2} \dots u_{b_r+q}$ , which is the subword between the first two matchings of  $p$  and  $q$ , is  $\alpha_l p = \beta_l q$  times. Combining these facts, if  $\alpha$  has more than one non-zero component, then some component, say  $\alpha_i$ , is less than  $q$ . So  $\frac{p}{q} = \frac{\beta_i}{\alpha_i}$  which implies that  $\frac{p}{q}$  is reducible, a contradiction. Hence  $\alpha$  can only contain one non-zero component, so  $w$  has period 1.

Using similar logic, we now extend Theorem 3.1 to apply to partial words.

**Theorem 3.2.** *Let  $w$  be a partial word with an arbitrary number of holes  $h$ . If  $w$  has abelian periods  $p$  and  $q$  which are relatively prime and  $|w| \geq (h + 2)pq - 1$ , then  $w$  has period 1.*

*Proof.* As mentioned earlier, since  $\gcd(p, q) = 1$  the abelian periods  $p$  and  $q$  first match up at or before  $pq - 1$  letters and the subsequent matches occur every  $pq$  letters later. A partial word  $w$  with  $|w| \geq (h + 2)pq - 1$  contains at least  $h + 2$  matchings of  $p$  and  $q$ , and  $h + 1$  subwords between these matchings. Denoting as above the first matching by  $w_1 = v_0v_1 \dots v_{r-1} = u_0u_1 \dots u_{b_r}$ , for  $0 \leq i \leq h$  let

$$w_{i+2} = v_{r+ip}v_{r+ip+1} \dots v_{r+(i+1)p-1} = u_{b_r+iq+1}u_{b_r+iq+2} \dots u_{b_r+(i+1)q}$$

be the subword between the  $(i + 1)$ st and  $(i + 2)$ nd matching points of  $p$  and  $q$ . Since  $w$  has only  $h$  holes, one of the subwords  $w_2, w_3, \dots, w_{h+2}$  does not contain any hole. Examining this subword which is full, we get by the argument given in the proof of Theorem 3.1 that the Parikh vector of any  $u_l$  or  $v_l$  within this subword

cannot contain more than one non-zero component. So for any  $u_i$  and  $v_i$  in  $w$  we have  $\|u_i\| \leq \|u_l\|$  and  $\|v_i\| \leq \|v_l\|$ . Therefore all  $u_i$  and  $v_i$  in  $w$  contain at most one non-zero component, so  $w$  has period 1. □

Further, we claim that the length  $(h + 2)pq - 1$  is optimal for  $h$  holes as our algorithm in Section 4 constructs non-unary partial words with  $h$  holes of length  $(h + 2)pq - 2$  that have abelian periods  $p$  and  $q$  for any coprime positive integers  $p, q$ .

#### 4. CONSTRUCTING OPTIMAL PARTIAL WORDS

A partial word  $w$  with  $h$  holes and having abelian periods  $p$  and  $q$  is *optimal* if the length of  $w$  is one less than the optimal length for the parameters  $h, p$  and  $q$ , and the cardinality of  $w$  is  $\gcd(p, q) + 1$ .

We start our discussion with optimal full words. First, suppose that  $p < q$  and  $\gcd(p, q) = 1$ . We would like to construct a word  $w$  over the alphabet  $A = \{a, b\}$  such that  $w$  has abelian periods  $p$  and  $q$ , and  $w$  has length  $2pq - 2$ . Let  $\alpha_a$  and  $\alpha_b$  be the number of times the letters  $a$  and  $b$ , respectively, occur within one abelian  $q$ -period and let  $\beta_a$  and  $\beta_b$  be the number of times the letters  $a$  and  $b$ , respectively, occur within one abelian  $p$ -period.

For our word to have optimal length, the periods  $p$  and  $q$  must match up after exactly  $pq - 1$  letters. So  $|u_0| = |u_{m+1}| = p - 1$  and  $|v_0| = |v_{n+1}| = q - 1$ , where  $u_0u_1 \dots u_{m+1}$  and  $v_0v_1 \dots v_{n+1}$  are factorizations of  $w$  into abelian periods  $p$  and  $q$ , respectively. For simplicity we assume  $\beta_a \geq \beta_b$  and, whenever possible, we place  $a$ 's before  $b$ 's. For example, letting  $p = 3$  and  $q = 7$ , the word

$$w = aa.aab.b|aa.aab.ab|a.aab.aab.|aab.aab.a|ab.aab.aa|b.aab.aa$$

of length  $2pq - 2 = 40$  is optimal. We can write  $w = w_1w_2$ , where  $w_1 = w[0..pq - 1) = w_{1,9}w_{1,8}w_{1,7}w_{1,6}w_{1,5}w_{1,4}w_{1,3}w_{1,2}w_{1,1}$  and  $w_2 = w[pq - 1..2pq - 2) = w_{2,1}w_{2,2}w_{2,3}w_{2,4}w_{2,5}w_{2,6}w_{2,7}w_{2,8}w_{2,9}$ , where  $w_{1,1} = w_{2,1} = aab$ ,  $w_{1,2} = w_{2,2} = aab$ ,  $w_{1,3} = w_{2,3} = a$ ,  $w_{1,4} = w_{2,4} = ab$ , etc. More generally, subwords in  $w$  are created by both the  $p$  and  $q$ -periods as follows: if we look at the two subwords on each side of the first matching point, denote the first subword to the left of the matching  $w_{1,1}$  and the first subword to the right  $w_{2,1}$  and continue this labeling outward. We write  $w_1 = \text{rev}_{p,q}(w_2)$ . Note that in  $w_2$ , we have  $\beta_bq - \alpha_bp = \pm 1$ . So, in order to construct an optimal word the key is to determine for which values of  $\beta_b$  and  $\alpha_b$  we have  $\beta_bq - \alpha_bp = \pm 1$ .

Further, we can extend this idea to construct an optimal word  $w$  with abelian periods  $p = dp'$  and  $q = dq'$  such that  $\gcd(p, q) = d$  in the case where  $p$  and  $q$  have matching points. In this case, the key is to determine for which values of  $\alpha$  and  $\beta$  we have  $\beta q' - \alpha p' = \pm 1$ . Algorithm 1 gives a construction for optimal words when  $p$  and  $q$  have matching points.

---

**Algorithm 1** Constructing an optimal partial word for two abelian periods

---

Input: Non-negative integer  $h$  and positive integers  $p$  and  $q$ ,  $p < q$

Output: An optimal partial word  $w$  with  $h$  holes, abelian periods  $p$  and  $q$ , and length  $(h + 2) \text{lcm}(p, q) - 2$

1.  $d \leftarrow \text{gcd}(p, q)$ ,  $p' \leftarrow \frac{p}{d}$  and  $q' \leftarrow \frac{q}{d}$
  2. Find smallest positive integer  $\beta$  and corresponding positive integer  $\alpha$  such that  $\beta q' - \alpha p' = \pm 1$
  3. Define Parikh vectors for periods  $p$  and  $q$  with distinct letters  $a_1, \dots, a_{d+1}$   
 $U \leftarrow (p', p', \dots, p', p' - \beta, \beta)$  and  $V \leftarrow (q', q', \dots, q', q' - \alpha, \alpha)$
  4. Generate subword  $w_2$  of  $w$  from position  $\text{lcm}(p, q) - 1$  up to position  $2 \text{lcm}(p, q) - 3$ 
    - (a)  $U' \leftarrow U$  //  $U'$  represents the number of each letter left to be filled into the current  $p$ -period
    - (b)  $w_2 \leftarrow \varepsilon$  and  $L \leftarrow 0$
    - (c) while  $L < \text{lcm}(p, q) - q$ 
      - $V' \leftarrow V$  //  $V'$  represents the number of each letter left to be filled into the current  $q$ -period
      - $w_2 \leftarrow w_2 u'$  //  $u'$  is some word with  $\|u'\| = U'$
      - $l \leftarrow |u'|$  //  $l$  represents the number of letters added to the current  $q$ -period
      - $L \leftarrow L + l$  and  $V' \leftarrow V' - U'$  and  $U' \leftarrow U'$
      - while  $l + p < q$ 
        - $w_2 \leftarrow w_2 u$  //  $u$  is some word with  $\|u\| = U$
        - $l \leftarrow l + p$  and  $L \leftarrow L + p$  and  $V' \leftarrow V' - U'$
        - $w_2 \leftarrow w_2 v'$  and  $U' \leftarrow U' - V'$  //  $v'$  is some word with  $\|v'\| = V'$
        - $L \leftarrow L + |v'|$
    - (d)  $V' \leftarrow V$
    - (e)  $w_2 \leftarrow w_2 u'$  and  $l \leftarrow |u'|$  //  $u'$  is some word with  $\|u'\| = U'$
    - (f)  $V' \leftarrow V' - U'$  and  $U' \leftarrow U'$
    - (g) while  $l + p < q$ 
      - $w_2 \leftarrow w_2 u$  //  $u$  is some word with  $\|u\| = U$
      - $l \leftarrow l + p$  and  $V' \leftarrow V' - U'$
    - (h) for  $i = 1$  to  $d + 1$ 
      - for  $j = 1$  to  $\min\{\#_{a_i}(U'), \#_{a_i}(V')\}$ 
        - $w_2 \leftarrow w_2 a_i$
  5.  $w_1 \leftarrow \text{rev}_{p', q'}(w_2)$  and  $w \leftarrow w_1 w_2 (\diamond w_2)^h$
- 

**Theorem 4.1.** *Given as input a non-negative integer  $h$  and positive integers  $p$  and  $q$ ,  $p < q$ , Algorithm 1 outputs an optimal partial word  $w$  with  $h$  holes, abelian periods  $p$  and  $q$ , and length  $(h + 2) \text{lcm}(p, q) - 2$  in  $O((d + 1)|w|)$  time. Moreover, Algorithm 1's complexity is exponential in the input data, which has size  $\log(p) + \log(q) + \log(h)$ .*

*Proof.* Algorithm 1 outputs optimal partial words with  $h$  holes when  $p$  and  $q$  match up by constructing the subword  $w_2$  after the first matching and then concatenating  $w_1 = \text{rev}_{p', q'}(w_2)$  with  $w_2$ , and then with  $(\diamond w_2)^h$ . Algorithm 1 effectively constructs the word, so its complexity is exponential in the size of the input.  $\square$

Letting  $p = 2$  and  $q = 3$ , the binary word  $a.b|a.ab.|ab.a|b.a\circ.|ab.a|b.a$  with one hole of length  $3pq - 2 = 16$  with abelian periods  $p$  and  $q$  is first constructed. Algorithm 1 then starts with the above word as the base word and adds  $(\circ.|ab.a|b.a)^{h-1}$ . For  $h = 3$  we get

$$a.b|a.ab.|ab.a|b.a\circ.|ab.a|b.a\circ.|ab.a|b.a\circ.|ab.a|b.a$$

More generally, on input  $p, q = p + 1$  and  $h$ , our algorithm outputs the optimal word  $w_1w_2(\circ.|w_2)^h$  of length  $(h + 2) \text{lcm}(p, q) - 2$ , where

$$\begin{aligned} w_1 &= a^{p-1}.b|a^{p-1}.ab|a^{p-2}.a^2b|\dots|a.a^{p-1}b.| \\ w_2 &= a^{p-1}b.a|a^{p-2}b.a^2|\dots|ab.a^{p-1}|b.a^{p-1} \end{aligned}$$

**Remark 4.2.** The position in which a hole is placed within a subword contained between two matching points to construct an optimal partial word does not matter so long as the hole represents letter  $a$  in terms of  $p$  and letter  $b$  in terms of  $q$ , say, where  $a, b$  are distinct. To see this, if we add one more letter to an optimal full word, creating a second matching point between  $p$  and  $q$ , we have  $\beta_a q' - \alpha_a p' = \pm 1$  and  $\beta_b q' - \alpha_b p' = \mp 1$ . So, when we add a hole this way,  $\beta_a q' - \alpha_a p' + \beta_b q' - \alpha_b p' = 0$ , we can complete the first matching and continue the word from the new matching. Otherwise, we still cannot construct an optimal partial word longer than a full one. The same applies for each hole that we add.

The following illustrates all the possible positions of the hole:

$$\begin{array}{ll} a.b|a.ab.|ab.a|b.a\circ.|ab.a|b.a & a.b|a.ab.|ab.a|\circ.ab.|ab.a|b.a \\ a.b|a.ab.|ab.\circ|a.ab.|ab.a|b.a & a.b|a.ab.|a\circ.b|a.ab.|ab.a|b.a \end{array}$$

**Remark 4.3.** Modifying Algorithm 1 to output all the optimal partial words for two abelian periods first involves permuting the letters within each  $p$ - or  $q$ -period of any word it currently outputs. Our algorithm now gives precedence to the first letter available to complete each  $p$ - or  $q$ -period. Second, it involves putting holes in all possible ways according to Remark 4.2.

### 5. NON-RELATIVELY PRIME ABELIAN PERIODS

As observed in [13], Fine and Wilf’s theorem cannot in general be extended to non-relatively prime abelian periods. That is, if  $\text{gcd}(p, q) = d, d > 1$ , then the two abelian periods  $p$  and  $q$  cannot impose the abelian period  $d$  no matter how long the word is. For example, the infinite word  $(aabbcc.abc|abc.aabbcc)^\omega$  has abelian periods  $p = 6$  and  $q = 9$  but does not have abelian period  $\text{gcd}(p, q) = 3$  (note that if  $v$  is a non-empty finite word, then we denote by  $v^\omega$  the unique infinite word  $w$  such that  $w$  has period  $|v|$  and  $w(0)\dots w(|v| - 1) = v$ ).

**Conjecture 5.1** [13]. *If a word  $w$  has abelian periods  $p$  and  $q$  with  $\text{gcd}(p, q) = d, d > 1$ , then  $w$  has at most cardinality  $d$ .*

There are words for which this conjecture does not hold, hence it is false. For example, the word

$$a^{p'}b^{p'-1}.ac|a^{p'-1}b^{p'-1}.a^2bc|a^{p'-2}b^{p'-2}.a^3b^2c|a^{p'-3}b^{p'-3}.a^4b^3c|\dots|ab.a^{p'}b^{p'-1}c.|$$

has abelian periods  $p = 2p'$  and  $q = 2q' = 2(p' + 1) = p + 2$  and cardinality  $3 = \gcd(p, q) + 1$ .

In this section, we prove however that if a word  $w$  has abelian periods  $p$  and  $q$  with  $\gcd(p, q) = d, d > 1$ , and  $|w| \geq L$  for some  $L$  (that depends on  $p$  and  $q$ ), then  $w$  has at most cardinality  $d$  (see Thms. 5.5, 5.11 and 5.18). We say that the length (or the bound)  $L$  is *optimal* if there are examples of words with length  $L - 1$  that have abelian periods  $p$  and  $q$  with  $\gcd(p, q) = d, d > 1$ , and have cardinality at least  $d + 1$ .

For the remainder of this section, we assume that all words are full and that  $p, q$  are integers satisfying  $p < q, \gcd(p, q) = d, d > 1$ , and  $p = dp', q = dq'$  (we can assume that  $p' > 1$ ). Here  $u_0u_1 \dots u_{m+1}$  and  $v_0v_1 \dots v_{n+1}$  are factorizations of  $w$  into abelian periods  $p$  and  $q$ , respectively.

We will need a result from number theory, which we state now.

**Lemma 5.2.** *Let  $\alpha, \beta \in \mathbb{N}$  be two coprime integers such that  $1 < \alpha < \beta$ . Then for all  $0 \leq \mu < \beta$ , there exist  $s, t \in \mathbb{N}$  such that  $0 \leq s < \beta, 0 \leq t < \alpha$ , and  $s\alpha - t\beta = \mu$ .*

*Proof.* By the Euclidean Algorithm, there exist integers  $s_0$  and  $t_0$  such that  $s_0\alpha - t_0\beta = 1 = \gcd(\alpha, \beta)$  with  $|s_0| < \beta$  and  $|t_0| < \alpha$ . Either  $s_0, t_0 < 0$  or  $s_0, t_0 > 0$ . If the former, then let  $s = s_0 + \beta$  and  $t = t_0 + \alpha$  so that  $s, t > 0$ ; equality is preserved because  $s_0\alpha - t_0\beta = (s - \beta)\alpha - (t - \alpha)\beta = s\alpha - t\beta$ . And if the latter, then let  $s = s_0$  and  $t = t_0$ .

Simply by multiplying both sides of the equation  $s\alpha - t\beta = 1$  by  $\mu$ , where  $0 \leq \mu < \beta$ , we get  $(\mu s)\alpha - (\mu t)\beta = \mu$ . First, if  $\mu s < \beta$  and  $\mu t < \alpha$ , then we are done. Second, if  $\mu s \geq \beta$  and  $\mu t \geq \alpha$ , then let  $s^{(1)} = \mu s - \beta$  and  $t^{(1)} = \mu t - \alpha$  and notice

$$s^{(1)}\alpha - t^{(1)}\beta = (\mu s - \beta)\alpha - (\mu t - \alpha)\beta = (\mu s)\alpha - (\mu t)\beta$$

Again, if  $s^{(1)} < \beta$  and  $t^{(1)} < \alpha$ , we are done. If  $s^{(1)} \geq \beta$  and  $t^{(1)} \geq \alpha$ , let  $s^{(2)} = s^{(1)} - \beta$  and  $t^{(2)} = t^{(1)} - \alpha$ , and so on.

Assume at some point in this process that  $s^{(i)} < \beta$  but  $t^{(i)} \geq \alpha$ . Then since  $-t^{(i)}\beta \leq -\alpha\beta$ ,

$$\mu = s^{(i)}\alpha - t^{(i)}\beta \leq s^{(i)}\alpha - \alpha\beta < \beta\alpha - \alpha\beta = 0 \leq \mu$$

which is a contradiction. On the other hand, if we have  $s^{(i)} \geq \beta$  but  $t^{(i)} < \alpha$ , then  $s^{(i)}\alpha \geq \beta\alpha$ , and so

$$\mu = s^{(i)}\alpha - t^{(i)}\beta \geq \beta\alpha - t^{(i)}\beta \geq \beta\alpha - (\alpha - 1)\beta = \beta > \mu$$

Thus the case where only one of  $s$  and  $t$  is out of bounds is impossible. Because we may always reduce  $(\mu s)\alpha - (\mu t)\beta$ , the lemma follows. □

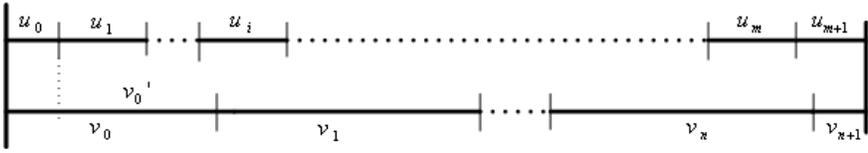


FIGURE 2. The subword  $u_1u_2 \dots u_mu_{m+1} = v'_0v_1 \dots v_nv_{n+1}$ .

Let  $s\alpha - t\beta = 1$  be as in the Proof of Lemma 5.2. Note that if  $s', \mu < \beta, t' < \alpha$ , and  $s'\alpha - t'\beta = \mu$ , then  $s' = \mu s \pmod{\beta}$  and  $t' = \mu t \pmod{\alpha}$ . For example, let  $\alpha = 3$  and  $\beta = 7$ . Then  $(-2)3 - (-1)7 = 1$ , and set  $s = -2 + 7 = 5$  and  $t = -1 + 3 = 2$ , so that  $(5)3 - (2)7 = 1$ . Let  $\mu = 5$ . Then

$$\begin{aligned} 5 &= (25)3 - (10)7 = (25 - 7)3 - (10 - 3)7 = (18 - 7)3 - (7 - 3)7 \\ &= (11 - 7)3 - (4 - 3)7 = (4)3 - (1)7 \end{aligned}$$

and  $s' = 4 = 25 \pmod{7}$  and  $t' = 1 = 10 \pmod{3}$  satisfy the conditions of Lemma 5.2.

**Lemma 5.3.** *For a word  $w$  with abelian periods  $p$  and  $q$  such that  $\gcd(p, q) = d$ ,  $d > 1$ , and  $|w| \geq \text{lcm}(p, q) - 1$ ,  $p$  and  $q$  match up if and only if  $\left| |u_0| - |v_0| \right| = \mu d$  for some integer  $\mu \geq 0$ .*

*Proof.* Let us suppose that  $\left| |u_0| - |v_0| \right| = \mu d$  for some integer  $\mu \geq 0$ . Since our argument does not depend on which length is greater, we may assume that  $|v_0| \geq |u_0|$ . Consider the subword  $w' = u_1 \dots u_{m+1} = u'_0u'_1 \dots u'_{m+1}$  where  $u'_0 = \varepsilon, u'_1 = u_1, \dots, u'_{m+1} = u_{m+1}$ , and  $w' = v'_0v'_1 \dots v'_{n+1}$  where  $v'_0 = v_0(|u_0|..|v_0|), v'_1 = v_1, \dots, v'_{n+1} = v_{n+1}$ . This is illustrated by Figure 2. Note that  $|v'_0| = \mu d$  and since  $|v'_0| < |v_0| \leq q = q'd$ , we get that  $0 \leq \mu < q'$ . Thus, periods  $p$  and  $q$  match up if there exist non-negative integers  $s$  and  $t, s \leq m$  and  $t \leq n$ , such that  $sp = \mu d + tq$ , where the  $s$   $p$ -periods of the matching end at length  $sp = d(sp')$  and the  $t$   $q$ -periods of the matching end at length  $\mu d + tq = d(\mu + tq')$ . The lengths  $sp$  and  $\mu d + tq$  are equal when  $sp'$  and  $\mu + tq'$  are equal, which is possible for any  $0 \leq \mu < q'$  by Lemma 5.2. In the other case where  $|v_0| \leq |u_0|$ , the problem boils down to equating  $\mu + sp'$  and  $tq'$ , which is similarly always possible.

The bound on  $|w|$  is found by maximizing  $|u_0|$  and  $|v_0|$ . From above, we see the length before the first matching is  $\mu d + sp$  or  $\mu d + tq$ , depending on whether  $|u_0|$  or  $|v_0|$  is larger. Because  $s \leq q' - 1$  and  $t \leq p' - 1$  by Lemma 5.2, if we choose  $\mu$  such that  $q' - p' = \mu$  then we can achieve the maximum for both  $|u_0|$  and  $|v_0|$ . Additionally, under this circumstance we have  $(q' - 1)p = q'p - p = p'q - q + \mu d = (p' - 1)q + \mu d$ , as required. Therefore the longest length before  $p$  and  $q$  match is

$$(q - 1) + (p' - 1)q = (p - 1) + (q' - 1)p = \text{lcm}(p, q) - 1$$

which proves the backward direction of the lemma.

For the forward direction, assume that  $p$  and  $q$  match up. This is equivalent to  $|u_0| + sp = |v_0| + tq$ , where we restrict  $s$  and  $t$  as in the premise, reformulated as

$sp - tq = |v_0| - |u_0|$ . Then since  $d$  divides both  $p$  and  $q$ , any linear combination of  $p$  and  $q$  will also be divisible by  $d$ . So  $|v_0| - |u_0| = sp - tq \equiv 0 \pmod d$ . Therefore the lemma holds in both directions.  $\square$

5.1. THE CASE WHERE  $\| |u_0| - |v_0| \|$  IS A MULTIPLE OF  $d$

We start with a word  $w$  having abelian periods  $p$  and  $q$ ,  $\gcd(p, q) = d > 1$ , with factorizations  $u_0 \dots u_{m+1}, v_0 \dots v_{n+1}$  of  $p$  and  $q$ , respectively. Under the conditions that  $\| |u_0| - |v_0| \| = \mu d$  for some integer  $\mu \geq 0$  and  $|w| \geq 2 \operatorname{lcm}(p, q) - 1$ , we will first show that  $w$  contains at least two matchings of  $p$  and  $q$ . Using them along with ideas of Section 3 related to number of occurrences of letters in  $p$ - and  $q$ -periods, we will then proceed by contradiction to prove that  $w$  has at most  $d$  distinct letters. In this case, the length  $2 \operatorname{lcm}(p, q) - 1$  turns out to be optimal.

**Lemma 5.4.** *If a word  $w$  has abelian periods  $p$  and  $q$  with  $\gcd(p, q) = d, d > 1, |w| \geq 2 \operatorname{lcm}(p, q) - 1$ , and  $\| |u_0| - |v_0| \| = \mu d$  for some integer  $\mu \geq 0$ , then the abelian periods  $p$  and  $q$  have at least two matchings.*

*Proof.* By Lemma 5.3,  $p$  and  $q$  have their first matching before length  $\operatorname{lcm}(p, q) - 1$ . Their next matching occurs  $\operatorname{lcm}(p, q)$  letters later, and the result follows.  $\square$

**Theorem 5.5.** *If a word  $w$  has abelian periods  $p$  and  $q$  with  $\gcd(p, q) = d, d > 1$ , and  $\| |u_0| - |v_0| \| = \mu d$  for some integer  $\mu \geq 0$ , then  $w$  has at most cardinality  $d$  for  $|w| \geq 2 \operatorname{lcm}(p, q) - 1$ .*

*Proof.* By Lemma 5.4,  $w$  contains at least two matchings of  $p$  and  $q$ . Now, we use these matchings to prove that the cardinality of  $w$  is at most  $d$ . Suppose  $w$  has cardinality  $d + 1$ . Let  $p = dp', q = dq'$ , for  $p', q'$  coprime. After  $p$  and  $q$  first match up our second matching occurs  $q'p = p'q = \operatorname{lcm}(p, q)$  letters later. As in Section 3, let  $v_0v_1 \dots v_{r-1} = u_0u_1 \dots u_{b_r}$  be the first matching, where  $r$  and  $b_r$  are positive integers. Then the next matching is  $v_0 \dots v_{r-1}v_rv_{r+1} \dots v_{r+p'-1} = u_0 \dots u_{b_r}u_{b_r+1}u_{b_r+2} \dots u_{b_r+q'}$ . Consider  $v_r \dots v_{r+p'-1} = u_{b_r+1} \dots u_{b_r+q'}$ . For a letter  $a_l$  in  $w$  let  $\alpha_l$  represent the number of times that letter occurs in one  $q$ -period and  $\beta_l$  represent the number of times that letter occurs in one  $p$ -period. Then we have  $\alpha_l p' = \beta_l q'$ . This implies that  $\frac{\alpha_l}{\beta_l} = \frac{q'}{p'}$ . But  $\gcd(p', q') = 1$  so we must have  $q' \mid \alpha_l$  and  $p' \mid \beta_l$ . Therefore, for  $\alpha_l \neq 0$  and  $\beta_l \neq 0$  we must have  $\alpha_l \geq q'$  and  $\beta_l \geq p'$ . Let our letters be indexed such that  $a_1, \dots, a_{d+1}$  are the letters with non-zero components. So we have  $q = \sum_{l=1}^{d+1} \alpha_l \geq (d+1)q'$  and  $p = \sum_{l=1}^{d+1} \beta_l \geq (d+1)p'$ . This gives  $p \geq (d+1)p'$  and  $q \geq (d+1)q'$ , a contradiction. Hence the cardinality of  $w$  is at most  $d$ .  $\square$

This bound is also optimal. For example, the word

$$a^{p'}b^{p'-1}.ac|a^{p'-1}b^{p'-1}.a^2bc|a^{p'-2}b^{p'-2}.a^3b^2c|a^{p'-3}b^{p'-3}.a^4b^3c|\dots|ab.a^{p'}b^{p'-1}c.|a^{p'}b^{p'-1}c.ab|\dots|a^4b^3c.a^{p'-3}b^{p'-3}|a^3b^2c.a^{p'-2}b^{p'-2}|a^2bc.a^{p'-1}b^{p'-1}|ac.a^{p'}b^{p'-1}$$

of length  $2\text{lcm}(p, q) - 2 = 2(q - 1 + (p' - 1)q)$ , which can be constructed by Algorithm 1, has abelian periods  $p = 2p'$  and  $q = 2q' = 2(p' + 1) = p + 2$  and cardinality  $3 = \text{gcd}(p, q) + 1$ .

Based on Algorithm 1, we give some closed forms. Letting  $p, q, p', q', d, \gamma$  be positive integers such that  $p = dp', q = dq' = \gamma p + d$ , and  $\text{gcd}(p, q) = d$ , define

$$\begin{aligned} v_0 &= a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} (.a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} .a_1 \dots a_{d-1} a_{d+1} | \\ v_i &= |a_1^{p'-i} \dots a_{d-i}^{p'-i} a_d^{p'-i} (.a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} . \\ &\quad a_1^{i+1} \dots a_{d-1}^{i+1} a_d^i a_{d+1} | \\ v_{p'} &= |(a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^\gamma .a_1 \dots a_{d-1} a_{d+1} | \\ v_{p'+j} &= |a_1^{p'-j} \dots a_{d-1}^{p'-j} a_d^{p'-1-j} a_{d+1} (.a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} . \\ &\quad a_1^{j+1} \dots a_{d-1}^{j+1} a_d^j | \\ v_{2p'-1} &= |a_1 \dots a_{d-1} a_{d+1} (.a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} .a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} \end{aligned}$$

for  $0 < i < p'$  and  $0 < j < p' - 1$ .

**Theorem 5.6.** *If  $\|u_0\| - \|v_0\| = \mu d$  for some integer  $\mu \geq 0$ , then  $w = u_0 u_1 \dots = v_0 v_1 \dots v_{2p'-1}$  is an optimal word of cardinality  $d + 1$ , length  $2\text{lcm}(p, q) - 2$ , having abelian periods  $p$  and  $q$ , that can be constructed by Algorithm 1.*

*Proof.* Since  $q' = \gamma p' + 1$ , we have from Algorithm 1 that  $\beta = 1$  and  $\alpha = \gamma$ , and therefore  $\|u_1\| = (p', p', \dots, p', p' - 1, 1)$  and  $\|v_1\| = (q', q', \dots, q', q' - \gamma, \gamma)$ . Also, the subword of length  $q$  after the matching, denoted  $v_z$ , contains  $\gamma$  full  $p$ -periods followed by one of each of the letters  $a_1, \dots, a_{d-1}, a_d$ . So,

$$v_z = (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^\gamma a_1 \dots a_{d-1} a_d.$$

Since  $q = \gamma p + d$  and from the proof of Theorem 5.5 we know that  $w$  contains only one matching of  $p$  and  $q$ ,  $w$  can contain only two full  $q$ -periods,  $v_{z-1}$  and  $v_z$ , containing  $\gamma$  full  $p$ -periods. Based on  $\|u_1\|$  and  $\|v_1\|$ , we find that

$$v_{z+1} = a_1^{p'-1} a_2^{p'-1} \dots a_{d-1}^{p'-1} a_d^{p'-2} a_{d+1} (a_1^{p'} a_2^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} a_1^2 \dots a_{d-1}^2 a_d^2.$$

By induction, we can show that for  $0 \leq i \leq p' - 2$ ,

$$v_{z+i} = a_1^{p'-i} \dots a_{d-1}^{p'-i} a_d^{p'-1-i} a_{d+1} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} a_1^{i+1} \dots a_{d-1}^{i+1} a_d^{i+1}.$$

Thus,

$$v_{z+p'-2} = a_1^2 \dots a_{d-1}^2 a_d a_{d+1} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} a_1^{p'-1} \dots a_{d-1}^{p'-1} a_d^{p'-1}.$$

So based on  $\|u_1\|$ , we know  $v_{z+p'-1}$  must begin with  $a_1 \dots a_{d-1} a_{d+1}$ . In order to complete the  $q$ -period, we must add  $q' - 1 = \gamma p'$  each of the letters  $a_1, \dots, a_{d-1}$ ,  $\gamma(p' - 1) + 1$   $a_d$ 's, and  $\gamma - 1$   $a_{d+1}$ 's. Since we can only add  $\gamma - 1$   $a_{d+1}$ 's, we can only add  $\gamma - 1$  full  $p$ -periods. Now, in order to complete our  $q$ -period  $v_{z+p'-1}$ , we still need  $\gamma p' - (\gamma - 1)p' = p'$  each of the letters  $a_1, \dots, a_{d-1}$ ,  $\gamma(p' - 1) + 1 - (\gamma - 1)(p' - 1) =$

$p'$   $a_d$ 's, and no  $a_{d+1}$ 's. So, we add all the letters from the next  $p$ -period, except  $a_{d+1}$ . We now have  $q'$  each of  $a_1, \dots, a_{d-1}, \gamma p' - \gamma a_d$ 's, and  $\gamma a_{d+1}$ 's. So from  $\|u_1\|$  we can only add an  $a_{d+1}$ , but from  $\|v_1\|$  we can only add an  $a_d$ . These conflict so our word must end here with

$$v_{z+p'-1} = a_1 \dots a_{d-1} a_{d+1} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1}$$

and  $|v_z \dots v_{z+p'-1}| = (p' - 1)q + q - 1$ . We can similarly argue that for  $0 \leq i \leq z - 1$ ,  $v_i = a_1^{p'-i} \dots a_{d-i}^{p'-i} a_d^{p'-i} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} a_1^{i+1} \dots a_{d-1}^{i+1} a_d^i a_{d+1}$ , and also that  $v_0 = a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} a_1 \dots a_{d-1} a_{d+1}$ , and  $|v_0 \dots v_{z-1}| = (p' - 1)q + q - 1$ . This gives us  $z = p'$  and  $|v_0 \dots v_{2p'-1}| = 2p'q - 2 = 2 \text{lcm}(p, q) - 2$ .  $\square$

The following corollary states the  $d = 1$  case of the previous theorem.

**Corollary 5.7.** *Let  $p, q, \gamma$  be positive integers such that  $q = \gamma p + 1$  and  $\text{gcd}(p, q) = 1$ . Then the word  $w = v_0 v_1 \dots v_{2p-1} =$*

$$\underbrace{a^{p-1} (.a^{p-1}b)^{\gamma-1} .b}_{v_0} | \dots | \underbrace{a^{p-i} (.a^{p-1}b)^{\gamma-1} .a^i b}_{v_i, 0 < i < p} | \dots | \underbrace{(.a^{p-1}b)^\gamma .a}_{v_p} | \dots | \underbrace{a^{p-1-i} b (.a^{p-1}b)^{\gamma-1} .a^{i+1}}_{v_{p+i}, 0 < i < p-1} | \dots | \underbrace{b (.a^{p-1}b)^{\gamma-1} .a^{p-1}}_{v_{2p-1}}$$

*is an optimal word of cardinality 2, length  $2pq - 2$ , having abelian periods  $p$  and  $q$ , that can be constructed by Algorithm 1.*

5.2. THE CASE WHERE  $\|u_0\| - |v_0\|$  IS NOT A MULTIPLE OF  $d$  ( $q = \gamma p + d$ )

We start with a word  $w$  having abelian periods  $p$  and  $q$ ,  $\text{gcd}(p, q) = d > 1$ , with factorizations  $u_0 \dots u_{m+1}, v_0 \dots v_{n+1}$  of  $p$  and  $q$ , respectively. Under the conditions that  $\|u_0\| - |v_0\| \neq \mu d$  for any integer  $\mu \geq 0$ ,  $|w| \geq 2 \text{lcm}(p, q) - 2$ , and  $q = \gamma p + d = \gamma p' d + d$  for some integer  $\gamma \geq 1$ , we will first show that  $w$  contains at least  $2p' - 1$  full  $q$ -periods and there must be one of them, say  $v_i$ , that contains  $\gamma$  full  $p$ -periods, i.e.,  $v_i = x_i u_{b_i+1} \dots u_{b_i+\gamma} y_i$  according to the notation of Section 3 (we assume that  $v_i$  is the first such  $q$ -period). We will then suppose, towards a contradiction, that  $w$  has  $d + 1$  distinct letters, and consequently at least one of these letters, say  $a$ , has a number of occurrences in a  $q$ -period which is exactly  $\gamma$  times its number of occurrences in a  $p$ -period. Arguing on the number of  $a$ 's in the  $x$  and  $y$  segments preceding and following  $x_i$  and  $y_i$ , we will be able to assume without loss of generality that  $v_i = v_{p'}$ . Finally, arguing on the number of occurrences of each of the  $d + 1$  letters in the  $x$  and  $y$  segments preceding and following  $x_{p'}$  and  $y_{p'}$ , a contradiction will be reached.

**Lemma 5.8.** *For a word  $w$  with abelian periods  $p$  and  $q$ , where  $\text{gcd}(p, q) = d > 1$  and  $\|u_0\| - |v_0\| \neq \mu d$  for any integer  $\mu \geq 0$ , if  $w$  contains at least  $p'$  full  $q$ -periods and  $q = \gamma p + d$  for some integer  $\gamma \geq 1$ , then there exists at least one  $q$ -period in  $w$  that contains  $\gamma$  full  $p$ -periods.*

*Proof.* We show that at least one of  $v_1, \dots, v_{p'}$  contains  $\gamma$  full  $p$ -periods. Suppose  $v_1$  contains only  $\gamma - 1$  full  $p$ -periods. Then, as in Section 3,  $v_1$  can be written as  $x_1 u_{b_1+1} \dots u_{b_1+\gamma-1} y_1$ , where  $x_1$  is a suffix of  $u_{b_1}$  and  $y_1$  is a prefix of  $u_{b_1+\gamma}$ . Since  $|x_1|, |y_1| < p$  and  $|x_1 u_{b_1+1} \dots u_{b_1+\gamma-1} y_1| = \gamma p + d$  we have that  $d < |x_1|, |y_1| < p$ . So, let  $|x_1| = d + \delta$  and  $|y_1| = p + d - |x_1| = p - \delta$  for some integer  $0 < \delta < p - d$ .

Now consider  $v_2 = x_2 \dots y_2$  factorized similarly, where  $|x_2| = p - |y_2| = \delta$ . If  $|x_2| \leq d$  then  $v_2$  contains  $\gamma$  full  $p$ -periods. Otherwise,  $|x_2| > d$  and suppose  $v_2$  contains only  $\gamma - 1$  full  $p$ -periods. In this case,  $|y_2| = \gamma p + d - (\gamma - 1)p - |x_2| = p + d - \delta > d$ .

By induction, if none of  $v_1, \dots, v_{p'-1}$  contains  $\gamma$  full  $p$ -periods, then  $|x_i| = \delta - (i - 2)d$  and  $|y_i| = p - \delta + (i - 1)d$  for  $1 \leq i \leq p' - 1$ . We get  $|x_{p'}| = p - |y_{p'-1}| = \delta - (p' - 2)d = \delta - p + 2d < p - d - p + 2d = d$  and  $v_{p'}$  contains  $\gamma$  full  $p$ -periods.  $\square$

**Theorem 5.9.** *For a word  $w$  with abelian periods  $p$  and  $q$ , where  $\gcd(p, q) = d > 1$  and  $\|u_0\| - |v_0| \neq \mu d$  for any integer  $\mu \geq 0$ , if  $|w| \geq 2 \operatorname{lcm}(p, q) - 2$  and  $q = \gamma p + d$  for some integer  $\gamma \geq 1$ , then  $w$  has cardinality at most  $d$ .*

*Proof.* Suppose  $w$  has cardinality  $d + 1$  and  $q = \gamma p + d = \gamma d p' + d$ . Since  $|w| \geq 2 \operatorname{lcm}(p, q) - 2$ ,  $w$  contains at least  $2p' - 1$  full  $q$ -periods, i.e.,  $n \geq 2p' - 1$ .

By Lemma 5.8, some  $q$ -period must contain  $\gamma$  full  $p$ -periods. Let  $v_i$  be this first  $q$ -period containing  $\gamma$  full  $p$ -periods. We can write  $v_i$  as  $x_i u_{b_i+1} \dots u_{b_i+\gamma} y_i$  and  $v_{i-1}$  as  $x_{i-1} u_{b_{i-1}+1} \dots u_{b_{i-1}+\gamma-1} y_{i-1}$ . Since we have  $\gamma$  full  $p$ -periods in  $v_i$ , there must exist a letter  $a$  such that  $\alpha_a = \gamma \beta_a$  (otherwise,  $\alpha_{a_l} > \gamma \beta_{a_l}$  for all  $l \in \{1, \dots, d + 1\}$  and  $q$ , which is the length of  $v_i$ , would be bigger than  $\gamma p + d$ ). We assume without loss of generality that  $a = a_1$ . Thus,  $|x_i|_a = |y_i|_a = 0$ . Further, since  $y_{i-1} x_i$  forms a full  $p$ -period,  $|y_{i-1}|_a = \beta_a$ , which implies  $|x_{i-1}|_a = 0$ . As we work backwards through  $w$ , this pattern continues with  $|y_{i-i'}|_a = \beta_a$  and  $|x_{i-i'}|_a = 0$ . Note that  $v_{i-p'}$ , if it were a full  $q$ -period, would be another  $q$ -period containing  $\gamma$  full  $p$ -periods (but then we would have that  $|y_{i-p'}|_a = \beta_a$  and  $\alpha_a = |v_{i-p'}|_a \geq (\gamma + 1)\beta_a > \beta_a$ , which would be a contradiction). We can similarly argue that our word must end with the  $v_{i+p'}$  subword. Since  $w$  contains  $2p' - 1$  full  $q$ -periods, either  $v_{p'-1}$  or  $v_{p'}$  must be a subword containing  $\gamma$  full  $p$ -periods. Without loss of generality, we let  $v_{p'}$  be this subword.

Let  $v_{p'} = x_{p'} u_{b_{p'}+1} \dots u_{b_{p'}+\gamma} y_{p'}$ , where  $x_{p'}$  is a suffix of  $u_{b_{p'}}$  and  $y_{p'}$  is a prefix of  $u_{b_{p'}+\gamma+1}$ . Let  $|x_{p'}|_{a_l} + |y_{p'}|_{a_l} = g_l$  and  $|x_{p'}|_{a_l} = h_l$ . So,  $|y_{p'}|_{a_l} = g_l - h_l$ ,  $|v_{p'}|_{a_l} = \gamma \beta_l + g_l$ , and  $\sum_{l=1}^{d+1} g_l = d$ . To see the latter, note that  $|v_{p'}| = q = \gamma p + d = \gamma p + |x_{p'}| + |y_{p'}|$  and so  $d = |x_{p'}| + |y_{p'}|$ .

Then  $|y_{p'-1}|_{a_l} = \beta_l - h_l$  and  $|x_{p'-1}|_{a_l} = g_l + h_l$ . By induction, we get  $|x_{p'-j}|_{a_l} = j g_l + h_l$ . We also have  $|x_{p'+1}|_{a_l} = \beta_l - g_l + h_l$  and  $|y_{p'+1}|_{a_l} = 2g_l - h_l$ . By induction, we get  $|y_{p'+j}|_{a_l} = (j + 1)g_l - h_l$ . So,  $\beta_l \geq |x_1|_{a_l} = (p' - 1)g_l + h_l = g_l p' - g_l + h_l$  and  $\beta_l \geq |y_{2p'-1}|_{a_l} = g_l p' - h_l$ . Since  $0 \leq h_l \leq g_l$ ,

$$\beta_l \geq \max \{g_l p' - g_l + h_l, g_l p' - h_l\} \geq g_l p' - \left\lfloor \frac{g_l}{2} \right\rfloor.$$

Note that each letter with  $\alpha_l = \gamma \beta_l$  must satisfy  $\beta_l \geq 1$ .

Suppose  $\alpha_l = \gamma\beta_l$  for  $l = 1, \dots, h$  (in which cases  $g_l = 0$ ) and  $\alpha_l > \gamma\beta_l$  for  $l = h + 1, \dots, d + 1$  (in which cases  $g_l > 0$ ). Note that  $d = \sum_{l=1}^{d+1} g_l = \sum_{l=h+1}^{d+1} g_l$ . Then

$$p = \sum_{l=1}^{d+1} \beta_l \geq h + \sum_{l=h+1}^{d+1} \beta_l \geq h + \sum_{l=h+1}^{d+1} (g_l p' - \lfloor \frac{g_l}{2} \rfloor) = h + dp' - \sum_{l=h+1}^{d+1} \lfloor \frac{g_l}{2} \rfloor$$

and since  $p = dp'$ , we get  $\sum_{l=h+1}^{d+1} \lfloor \frac{g_l}{2} \rfloor \geq h$ .

Suppose  $h = 1$ . Then we have  $d$  letters with  $\alpha_l > \gamma\beta_l$ . So, for  $h + 1 \leq l \leq d + 1$ ,  $g_l = 1$ , and  $\sum_{l=h+1}^{d+1} \lfloor \frac{g_l}{2} \rfloor = 0 < 1 = h$ , a contradiction. Thus  $h > 1$ , and a word with  $d + 1$  letters has at most  $2p' - 2$  full  $q$ -periods, *i.e.*,  $n \leq 2p' - 2$ , a contradiction. □

Based on Algorithm 1, we give some closed forms. Letting  $p, q, p', q', d, \gamma$  be positive integers such that  $p = dp', q = dq' = \gamma p + d$ , and  $\gcd(p, q) = d > 1$ , define

$$\begin{aligned} v_0 &= a_1^{p'} a_2^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} a_2 \dots a_{d-1} a_{d+1} | \\ v_i &= |a_1^{p'+1-i} a_2^{p'-i} \dots a_d^{p'-i} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} \\ &\quad a_1^i a_2^{i+1} \dots a_{d-1}^{i+1} a_d^i a_{d+1} | \\ v_{p'} &= |a_1 (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^\gamma a_2 a_3 \dots a_{d-1} a_{d+1} | \\ v_{p'+j} &= |a_1^{p'+1-j} a_2^{p'-j} \dots a_{d-1}^{p'-j} a_d^{p'-1-j} a_{d+1} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} \\ &\quad a_1^j a_2^{j+1} \dots a_d^{j+1} | \\ v_{2p'-1} &= |a_1^2 a_2 \dots a_{d-1} a_{d+1} (a_1^{p'} \dots a_{d-1}^{p'} a_d^{p'-1} a_{d+1})^{\gamma-1} \\ &\quad a_1^{p'-1} a_2^{p'} \dots a_{d-1}^{p'-1} a_d^{p'-1} \end{aligned}$$

for  $0 < i < p'$  and  $0 < j < p' - 1$ .

**Theorem 5.10.** *If  $||u_0| - |v_0|| \neq \mu d$  for any integer  $\mu \geq 0$ , then  $w = u_0 u_1 \dots = v_0 v_1 \dots v_{2p'-1}$  is an optimal word of cardinality  $d + 1$ , length  $2 \operatorname{lcm}(p, q) - 3$ , having abelian periods  $p$  and  $q$ , that can be constructed by Algorithm 1.*

*Proof.* The proof is similar to that of Theorem 5.6. □

### 5.3. THE CASE WHERE $||u_0| - |v_0||$ IS NOT A MULTIPLE OF $d$ ( $q = \gamma p + rd$ )

Here the situation becomes more complicated. Write  $q = \gamma p + rd$  for some  $\gamma > 0$  and some  $r$  where  $0 \leq rd < p$ . Assume that  $d \neq p$ , since the case  $d = p$  is trivial. Similarly, the case  $r = 0$  is trivial. This implies that  $\gcd(p', r) = 1$ , since  $d \gcd(p', r)$  divides both  $p$  and  $q$ .

We begin with an initial bound on the length, given in Theorem 5.11.

**Theorem 5.11.** *For a word  $w$  with abelian periods  $p$  and  $q$ , where  $\gcd(p, q) = d > 1$  and  $||u_0| - |v_0|| \neq \mu d$  for any integer  $\mu \geq 0$ , if  $|w| > 2(p' - 1) \operatorname{lcm}(p, q) - 2$ , then  $w$  has cardinality at most  $d$ .*

*Proof.* There exists an integer  $s$  such that  $0 < s < p'$  and  $rs = tp' + 1$  for some integer  $t$ . Then note that  $w$  has abelian period  $sq$ , where  $sq = (t + s\gamma)p + d$ . It then follows by Theorem 5.9 that since  $\gcd(sq, p) = d$  and since  $|w| > 2(p' - 1)\text{lcm}(p, q) - 2 \geq 2s\text{lcm}(p, q) - 2 = 2\text{lcm}(p, sq) - 2$  that  $w$  has cardinality at most  $d$ .  $\square$

Theorem 5.18 below provides an improved bound. Start with  $u_0u_1 \dots u_{m+1}$  and  $v_0v_1 \dots v_{n+1}$ , two factorizations of a word  $w$  into abelian periods  $p$  and  $q$ , respectively, where  $\gcd(p, q) = d > 1$  and  $\|u_0\| - \|v_0\| \neq \mu d$  for any integer  $\mu \geq 0$ . Recall that for each  $i$ ,  $1 \leq i \leq n$ , we can write  $v_i = x_iu_{b_i+1} \dots u_{b_{i+1}-1}y_i$ , where  $|x_i| < p$  and  $|y_i| < p$ . In case  $|v_0| = q$ , write  $v_0$  similarly.

- Our first step will be to show that for  $1 \leq i \leq n$ ,  $|x_i| \equiv |x_1| - (i - 1)(|x_1| + |y_1|) \pmod p$ . This will be implied by Corollary 5.14 below. Using  $|v_1| = \gamma p + rd$  and  $\gcd(r, p') = 1$ , we will show the existence of some  $i$ ,  $1 \leq i \leq p'$ , so that  $|x_i| \equiv |x_1| - (i - 1)rd \pmod p$  and  $|x_i| < d$ .
- Our second step will be to consider the word  $v = v_i \dots v_{n+1}$ , with  $|v_i| = q$ . Our bound  $|w| \geq \text{lcm}(p, q) + pq - 1$  will imply that  $n - i \geq p - 1$ , and consequently  $v$  will have factorizations into abelian periods  $p$  and  $q$  that satisfy some conditions. Corollary 5.17 below will then imply that  $v$  contains at most  $d$  distinct letters, and so will  $w$ .

We begin with the following lemma showing that the number of occurrences of letter  $a$  in  $x_i$ , where  $a$  occurs in  $w$ , can be expressed by some function  $f : \mathbb{Z}^+ \rightarrow \mathbb{N}$ .

**Lemma 5.12.** *Let  $w$  be a word with abelian periods  $p$  and  $q$ , where  $\gcd(p, q) = d > 1$  and  $\|u_0\| - \|v_0\| \neq \mu d$  for any integer  $\mu \geq 0$ . There exists a function  $f : \mathbb{Z}^+ \rightarrow \mathbb{N}$  so that, for  $i$ , where  $1 \leq i \leq n$ , and  $a \in A$  such that  $a$  occurs in  $w$ , the equality*

$$|x_i|_a = f(i)|u_1|_a - (i - 1)(|x_1|_a + |y_1|_a) + |x_1|_a \tag{1}$$

*holds. Furthermore, if  $i = n + 1$  and  $|y_{i-1}x_i| = p$  then equality (1) also holds.*

*Proof.* We proceed by induction on  $i$ . Equality (1) holds trivially when  $i = 1$  by letting  $f(i) = 0$ , so assume that it holds for  $1, 2, \dots, i - 1$ . Then note that  $|x_1|_a + |y_1|_a + (b_2 - b_1 - 1)|u_1|_a$

$$\begin{aligned} &= |x_1u_{b_1+1} \dots u_{b_2-1}y_1|_a \\ &= |x_{i-1}u_{b_{i-1}+1} \dots u_{b_i-1}y_{i-1}|_a \text{ since } |v_1|_a = |v_{i-1}|_a \\ &= |x_{i-1}|_a + |y_{i-1}|_a + (b_i - b_{i-1} - 1)|u_1|_a \\ &= c'|u_1|_a - (i - 2)(|x_1|_a + |y_1|_a) + |x_1|_a + |y_{i-1}|_a \text{ by ind. hyp.} \end{aligned}$$

for some integer  $c'$ . This implies that

$$c|u_1|_a + (i - 1)(|x_1|_a + |y_1|_a) - |x_1|_a = |y_{i-1}|_a$$

where  $c$  is some integer.

Since  $y_{i-1}x_i = u_{b_i}$ , we have

$$(1 - c)|u_1|_a - (i - 1)(|x_1|_a + |y_1|_a) + |x_1|_a = |u_1|_a - |y_{i-1}|_a = |x_i|_a.$$

Note that since  $|u_1|_a > 0$ ,  $(i - 1)(|x_1|_a + |y_1|_a) \geq 0$ , and  $|x_1|_a \geq 0$  we get that  $1 - c \geq 0$ . Thus set  $f(i) = 1 - c$ , and the claim follows.  $\square$

We get the following corollaries giving relationships on the length of  $x_i$ .

**Corollary 5.13.** *For  $1 \leq i \leq n$ , the equality*

$$|x_i| = f(i)|u_1| - (i - 1)(|x_1| + |y_1|) + |x_1|$$

holds, where  $f$  is defined as in Lemma 5.12.

*Proof.* This follows from Lemma 5.12 when we sum over all  $a \in A$ .  $\square$

**Corollary 5.14.** *For  $1 \leq i \leq n$ ,  $|x_i| \equiv |x_1| - (i - 1)(|x_1| + |y_1|) \pmod p$ . Furthermore, for  $1 \leq i + p' \leq n$ ,  $|x_i| = |x_{i+p'}|$ .*

*Proof.* Since by Corollary 5.13,  $|x_i| = f(i)|u_1| - (i - 1)(|x_1| + |y_1|) + |x_1| = f(i)p - (i - 1)(|x_1| + |y_1|) + |x_1|$ , the first claim follows easily. Then note that this implies  $|x_{i+p'}| \equiv |x_i| \pmod p$  (here we use the fact that  $|v_1| = q = \gamma p + rd$  implies  $|x_1| + |y_1| \equiv rd \pmod p$ ). Since  $0 \leq |x_l| < p$  for all  $l$ , we get  $|x_i| = |x_{i+p'}|$ .  $\square$

The next lemma gives some of the  $f$  values.

**Lemma 5.15.** *Assume that  $s \geq 0$  is an integer and that  $|x_1| + |y_1| = rd$ . If  $f$  is defined at  $sp' + 1$ , then  $f(sp' + 1) = sr$ .*

*Proof.* Let  $i = sp' + 1$ . Note that  $|x_i| = |x_{1+sp'}| = |x_{1+(s-1)p'}| = \dots = |x_{1+p'}| = |x_1|$  by Corollary 5.14. Thus we get by Corollary 5.13 that

$$|x_1| = |x_i| = f(i)p - (i - 1)rd + |x_1|$$

so that  $f(i)p = spr$ , and thus  $f(i) = sr$ .  $\square$

Using the previous lemma, we next prove, under some conditions on  $w$ 's factorization into abelian period  $q$ , some relationships on  $|x_0| + |y_0|$  and  $|x_0|_a + |y_0|_a$ , for all  $a \in A$ .

**Lemma 5.16.** *Assume that  $|v_0| = q$ ,  $|x_0| < d$ ,  $n \geq p - 1$ . Then  $|x_0| + |y_0| = rd$  and  $r|u_1|_a - p'(|x_0|_a + |y_0|_a) = 0$  for all  $a \in A$ .*

*Proof.* Without loss of generality we can assume that  $n = p - 1$  and  $v_{n+1} = \varepsilon$ . If  $A = \{a_1, \dots, a_k\}$ , then let  $c_i = |u_1|_{a_i} - |u_{m+1}|_{a_i}$ . Define

$$x_p = a_1^{c_1} \dots a_k^{c_k}.$$

Then  $wx_p$  has abelian periods  $p$  and  $q$ .

Note that for each  $i$ ,  $0 \leq i \leq n$ , either  $|x_i| + |y_i| = rd$  or  $|x_i| + |y_i| = p + rd$ . Since  $|y_0| < p$  by assumption and  $|x_0| < d$ , it follows that  $|x_0| + |y_0| < p + d \leq p + rd$ , so it must be that  $|x_0| + |y_0| = rd$ .

Assume that  $r|u_1|_a - p'(|x_0|_a + |y_0|_a) \neq 0$  for some  $a \in A$ . Assume  $r|u_1|_a - p'(|x_0|_a + |y_0|_a) < 0$ , the case where  $r|u_1|_a - p'(|x_0|_a + |y_0|_a) > 0$  being similar. Then if  $i = p$  we get that, since  $|x_0|_a < d$ ,

$$\begin{aligned} 0 &\leq |x_i|_a = f(i)|u_1|_a - (i - 1)(|x_0|_a + |y_0|_a) + |x_0|_a \\ &< f(i)|u_1|_a - (i - 1)(|x_0|_a + |y_0|_a) + d \\ &= d(r|u_1|_a - p'(|x_0|_a + |y_0|_a)) + d \text{ by Lemma 5.15} \\ &\leq -d + d = 0. \end{aligned}$$

This is a contradiction, so the claim follows □

We can deduce, as a corollary, that if  $w$ 's factorization into abelian period  $q$  satisfies the conditions of Lemma 5.16, then  $w$  contains at most  $d$  distinct letters.

**Corollary 5.17.** *If  $w$  is as in Lemma 5.16, then  $w$  has cardinality at most  $d$ .*

*Proof.* By Lemma 5.16, we get that  $r|u_1|_a - p'(|x_0|_a + |y_0|_a) = 0$  for all  $a \in A$ . Since  $\gcd(p', r) = 1$  this implies that  $r$  divides  $|x_0|_a + |y_0|_a$ , so that either  $|x_0|_a + |y_0|_a = 0$  or  $|x_0|_a + |y_0|_a \geq r$ . Note that if  $|x_0|_a + |y_0|_a = 0$  then it must be the case that  $|u_1|_a = 0$ , so that  $a$  does not occur in  $w$ . However, if  $B$  is the set of all letters that occur in  $w$  then

$$r|B| = \sum_{a \in B} r \leq \sum_{a \in B} |x_0|_a + |y_0|_a = |x_0| + |y_0| = rd.$$

This implies that  $|B| \leq d$ , which is what we wanted. □

We are now ready to prove our bound.

**Theorem 5.18.** *For a word  $w$  with abelian periods  $p$  and  $q$ , where  $\gcd(p, q) = d > 1$  and  $\|u_0\| - \|v_0\| \neq \mu d$  for any integer  $\mu \geq 0$ , if  $|w| \geq \text{lcm}(p, q) + pq - 1$ , then  $w$  has cardinality at most  $d$ .*

*Proof.* It is worth noting that  $n > p'$ . We know by Corollary 5.14 that for  $1 \leq i \leq n$ ,  $|x_i| \equiv |x_1| - (i - 1)(|x_1| + |y_1|) \equiv |x_1| - (i - 1)rd \pmod p$  (the latter equivalence can be deduced since  $|v_1| = q = \gamma p + rd$  and so  $|x_1| + |y_1| \equiv rd \pmod p$ ). The fact that  $\gcd(r, p') = 1$  implies that there is an  $i$ ,  $1 \leq i \leq p'$ , so that  $c \equiv |x_1| - (i - 1)rd \equiv |x_i| \pmod p$  for some  $c$ , where  $0 \leq c < d$ . Moreover, since  $0 \leq |x_i| < p$ , this implies that  $|x_i| = c$ . Therefore consider the word  $v = v_i \dots v_n v_{n+1}$ , where (since  $\frac{pq}{d} = \text{lcm}(p, q)$ )

$$\begin{aligned} |v| &= |w| - |v_0 \dots v_{i-1}| \geq |w| - |v_0| - (p' - 1)q \\ &\geq |w| - q + 1 - (p' - 1)q \geq \text{lcm}(p, q) + pq - 1 - q + 1 - \text{lcm}(p, q) + q = pq. \end{aligned}$$

Note that  $v$  is a word with abelian periods  $p$  and  $q$ . Indeed, we can write  $v = v'_0 v'_1 \dots v'_{n_1} v'_{n_1+1}$ , where  $v'_0 = v_i$ ,  $v'_1 = v_{i+1}$ , and so on. Then  $n_1 \geq p - 1$ . Similarly, we can write  $v = u'_0 u'_1 \dots u'_{n_2} u'_{n_2+1}$  where  $u'_0 = x_i$ ,  $u'_1 = u_{b_i+1}$ ,  $u'_2 = u_{b_i+2}$ , and so on. Then we can write each  $v'_i$  so that  $x'_i u'_{c_i+1} \dots u'_{c_{i+1}-1} y'_i$  as before. Moreover, note that  $x'_0 = x_i$ , so  $|x'_0| < d$ . Therefore by Corollary 5.17 we get that  $v$  contains at most  $d$  distinct letters, and thus  $w$  contains at most  $d$  distinct letters. □

## 6. CONCLUSION

We proved that for a full word  $w$  with abelian periods  $p$  and  $q$  such that  $\gcd(p, q) = d$ ,  $d > 1$ ,  $p$  and  $q$  match up if and only if  $\|u_0\| - \|v_0\| = \mu d$  for some integer  $\mu \geq 0$ . We then proved that if a word  $w$  has abelian periods  $p$  and  $q$  with  $\gcd(p, q) = d$ ,  $d > 1$ , and  $p$  and  $q$  match up, then  $w$  has at most cardinality  $d$  for  $|w| \geq 2\text{lcm}(p, q) - 1$  (see Thm. 5.5). We believe that the optimal length for words where the abelian periods do not match up is shorter than the one for when they do match up.

**Conjecture 6.1.** *If a word  $w$  has abelian periods  $p$  and  $q$  with  $\gcd(p, q) = d$ ,  $d > 1$ , and  $\|u_0\| - \|v_0\| \neq \mu d$  for any integer  $\mu \geq 0$ , then  $w$  has at most cardinality  $d$  for  $|w| \geq 2\text{lcm}(p, q) - 2$ .*

If Conjecture 6.1 is true, then a word  $w$  having abelian periods  $p$  and  $q$  with  $\gcd(p, q) = d$ ,  $d \geq 1$ , has at most cardinality  $d$  for  $|w| \geq 2\text{lcm}(p, q) - 1$  (see Thms. 3.1 and 5.5). We did prove Conjecture 6.1 true when  $q = \gamma p + d$  for some integer  $\gamma \geq 1$  (see Thm. 5.9).

Further, if Conjecture 6.1 is true, then Algorithm 1, when  $h = 0$ , gives a construction for all optimal words. Indeed, if Conjecture 6.1 is true, the Parikh vectors that create optimal length words for when the abelian periods do not match up are the same as the Parikh vectors for when they do match up. In other words, for any  $p$  and  $q$ , if we calculate their Parikh vectors based on Algorithm 1, we can construct an optimal word of length  $2\text{lcm}(p, q) - 3$  in which  $p$  and  $q$  do not match up. For instance, the words  $w$  of Theorem 5.10 have abelian periods  $p$  and  $q = \gamma p + d$  and length  $2\text{lcm}(p, q) - 3$ .

*Acknowledgements.* A research assignment from the University of North Carolina at Greensboro for F. Blanchet-Sadri is gratefully acknowledged. Some of this assignment was spent at the LIAFA: Laboratoire d'Informatique Algorithmique: Fondements et Applications of Université Paris Denis Diderot, Paris, France. We thank Professor Pál Dömösi for bringing Constantinescu and Ilie's paper [13] to our attention. We thank Ian Coley from Northwestern University for helping us correct an error in an earlier version of this paper. We also thank the referees of preliminary versions of this paper for their very valuable comments and suggestions.

## REFERENCES

- [1] S.V. Avgustinovich, A. Glen, B.V. Halldórsson and S. Kitaev, On shortest crucial words avoiding abelian powers. *Discrete Appl. Math.* **158** (2010) 605–607.
- [2] S.V. Avgustinovich, J. Karhumäki and S. Puzynina, On abelian versions of the critical factorization theorem. In *JM 2010, 13ièmes Journées Montoises d'Informatique Théorique*, Amiens, France (2010).
- [3] J. Berstel and L. Boasson, Partial words and a theorem of Fine and Wilf. *Theoret. Comput. Sci.* **218** (1999) 135–141.
- [4] F. Blanchet-Sadri, *Algorithmic Combinatorics on Partial Words*. Chapman & Hall/CRC Press, Boca Raton, FL (2008).

- [5] F. Blanchet-Sadri, J.I. Kim, R. Mercas, W. Severa, S. Simmons and D. Xu, Avoiding abelian squares in partial words. *J. Combin. Theory Ser. A* **119** (2012) 257–270.
- [6] F. Blanchet-Sadri, T. Mandel and G. Sisodia, Periods in partial words: An algorithm. *J. Discrete Algorithms* **16** (2012) 113–128.
- [7] F. Blanchet-Sadri, T. Oey and T. Rankin, Fine and Wilf’s theorem for partial words with arbitrarily many weak periods. *Internat. J. Foundations Comput. Sci.* **21** (2010) 705–722.
- [8] F. Blanchet-Sadri, S. Simmons and D. Xu, Abelian repetitions in partial words. *Adv. Appl. Math.* **48** (2012) 194–214.
- [9] F. Blanchet-Sadri, A. Tebbe and A. Veprauskas, Fine and Wilf’s theorem for abelian periods in partial words. In *JM 2010, 13ièmes Journées Montoises d’Informatique Théorique*, Amiens, France (2010).
- [10] M.G. Castelli, F. Mignosi and A. Restivo, Fine and Wilf’s theorem for three periods and a generalization of Sturmian words. *Theoret. Comput. Sci.* **218** (1999) 83–94.
- [11] C. Choffrut and J. Karhumäki, Combinatorics of Words. In *Handbook of Formal Languages*, edited by G. Rozenberg and A. Salomaa, Springer-Verlag, Berlin Vol. 1 (1997) 329–438.
- [12] S. Constantinescu and L. Ilie, Generalised Fine and Wilf’s theorem for arbitrary number of periods. *Theor. Comput. Sci.* **339** (2005) 49–60.
- [13] S. Constantinescu and L. Ilie, Fine and Wilf’s theorem for abelian periods. *Bull. Eur. Assoc. Theor. Comput. Sci.* **89** (2006) 167–170.
- [14] L.J. Cummings and W.F. Smyth, Weak repetitions in strings. *J. Combin. Math. Combin. Comput.* **24** (1997) 33–48.
- [15] J. Currie and A. Aberkane, A cyclic binary morphism avoiding abelian fourth powers. *Theoret. Comput. Sci.* **410** (2009) 44–52.
- [16] M. Domaratzki and N. Rampersad, Abelian primitive words. In *DLT 2011, 15th International Conference on Developments in Language Theory*, Milano, Italy, *Lect. Notes Comput. Sci.* Vol. 6795 edited by G. Mauri and A. Leporati. Springer-Verlag, Berlin, Heidelberg (2011) 204–215.
- [17] G. Fici, T. Lecroq, A. Lefebvre and E. Prieur-Gaston, Computing abelian periods in words. PSC 2011, Prague Stringology Conference, Prague, Czech Republic, (2011) 184–196.
- [18] N.J. Fine and H.S. Wilf, Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.* **16** (1965) 109–114.
- [19] V. Halava, T. Harju and T. Kärki, Interaction properties of relational periods. *Discrete Math. Theoret. Comput. Sci.* **10** (2008) 87–112.
- [20] J. Justin, On a paper by Castelli, Mignosi, Restivo. *Theoret. Inform. Appl.* **34** (2000) 373–377.
- [21] V. Keränen, Abelian squares are avoidable on 4 letters. In *ICALP 1992, 19th International Colloquium on Automata, Languages and Programming*, *Lect. Notes Comput. Sci.* Vol. 623 edited by W. Kuich. Springer-Verlag, Berlin (1992) 241–52.
- [22] A.V. Samsonov and A.M. Shur, On abelian repetition threshold. In *JM 2010, 13ièmes Journées Montoises d’Informatique Théorique*, Amiens, France (2010).
- [23] A.M. Shur and Y.V. Gamzova, Partial words and the interaction property of periods. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya* **68** (2004) 191–214.
- [24] A.M. Shur and Y.V. Konovalova, On the periods of partial words. In *MFCS 2001, 26th International Symposium on Mathematical Foundations of Computer Science*, *Lect. Notes Comput. Sci.* Vol. 2136 edited by J. Sgall, A. Pultr and P. Kolman. London, UK, Springer-Verlag. (2001) 657–665.
- [25] W. F. Smyth and S. Wang, A new approach to the periodicity lemma on strings with holes. *Theoret. Comput. Sci.* **410** (2009) 4295–4302.
- [26] R. Tijdeman and L. Zamboni, Fine and Wilf words for any periods. *Indagationes Math.* **14** (2003) 135–147.

Communicated by O. Serre.

Received November 2, 2010. Accepted January 21, 2013.