Fine and Wilf's Theorem for k-Abelian Periods ⋆

Juhani Karhumäki, Svetlana Puzynina, and Aleksi Saarela

Turku Centre for Computer Science TUCS and Department of Mathematics
University of Turku, FI-20014 Turku, Finland
{karhumak,svepuz,amsaar}@utu.fi

Abstract. Two words u and v are *k-abelian equivalent* if they contain the same number of occurrences of each factor of length k and, moreover, start and end with a same factor of length $k-1$, respectively. This leads to a hierarchy of equivalence relations on words which lie properly in between the equality and abelian equality.
The goal of this paper is to analyze Fine and Wilf's periodicity theorem with respect to these equivalence relations. A crucial question here is to ask how far two "periodic" processes must coincide in order to guarantee a common "period". Fine and Wilf's theorem characterizes this for words. Recently, the same was done for abelian words. We show here that for k-abelian periods the situation resembles that of abelian words: In general, there are no bounds, but the cases when such bounds exist can be characterized. Moreover, in the cases when such bounds exist we give nontrivial upper bounds for these, as well as lower bounds for some cases. Only in quite rare cases (in particular for $k = 2$) we can show that our upper and lower bounds match.

1 Introduction

In 1965, Fine and Wilf proved their famous periodicity theorem [1]. It tells exactly how long a word with two periods p and q can be without having the greatest common divisor of p and q as a period. Many variations of the theorem have been considered. For example, there are several articles on generalizations for more than two periods, see e.g. [2] and [3]. Periods of partial words were studied in e.g. [4], [5] and [6]. Periodicity with respect to an involution was considered in [7]. Particularly interesting in the context of this article is the variation related to abelian equivalence. This was first considered by Constantinescu and Ilie in 2006 [8]. They proved an upper bound in the case of relatively prime periods and stated that otherwise there are no upper bounds. Blanchet-Sadri, Tebbe and Veprauskas [9] gave an algorithm showing the optimality of the above bound, although they only proved the correctness of the algorithm in some cases.

In this paper the k-abelian versions of periodicity are studied. Two words are called k-abelian equivalent if they contain the same number of occurrences of each factor of length k, if their prefixes of length $k-1$ are the same and if

⋆ Supported by the Academy of Finland under grants 137991 (FiDiPro) and 251371 and by Russian Foundation of Basic Research (grants 10-01-00424, 12-01-00448)

their suffixes of length $k-1$ are the same. For k-abelian equivalence the problem is similar but more complicated than for abelian equivalence. Again, there does not always exist a bound: If $\gcd(p,q) > k$, then there are infinite words having k-abelian periods p and q but not $\gcd(p,q)$. In all other cases a finite upper bound for the length of such words is obtained. In the case $k = 2$ and in some other special cases an exact variant of Fine and Wilf's theorem can be given. In the general case, however, the problem seems to be rather intricate. Nontrivial upper bounds in the general case and lower bounds in some special cases are proved but many open questions about the behavior of the problem remain.

2 Preliminaries

We study words over a non-unary alphabet Σ. For a general reference on combinatorics on words, see e.g. [10].

The *length* of a word $w \in \Sigma^*$ is denoted by $|w|$ and the product of n copies of w by w^n. If $w = tuv$, then u is a *factor* of w. If $|t| = 0$, then u is a *prefix* of w, and if $|v| = 0$, then u is a *suffix* of w. The notation $u \leq w$ is used to mean that u is a prefix of w. The prefix and suffix of length $m \leq |w|$ are denoted by $\mathrm{pref}_m(w)$ and $\mathrm{suff}_m(w)$. The number of occurrences of a factor u in w is denoted by $|w|_u$ and the *reversal* of w by w^R. Occasionally we will also consider right-infinite words. Then $tu^\omega = tuuu\ldots$ means the word consisting of t followed by infinitely many copies of u.

Let $w = a_1 \ldots a_n$, where $a_1, \ldots, a_n \in \Sigma^*$. A positive integer p is a *period* of w if $a_{i+p} = a_i$ for every $i \in \{1, \ldots, n-p\}$. Equivalently, p is a period if there is a word u of length p, a prefix u' of u and a number m such that $w = u^m u'$.

Now we state Fine and Wilf's periodicity theorem, which was proved in [1].

Theorem 2.1 (Fine and Wilf). *Let $p, q > \gcd(p,q) = d$. Let w have periods p and q. If $|w| \geq p + q - d$, then w has period d. There are words of length $p + q - d - 1$ that have periods p and q but not period d.*

Two words u and v are *abelian equivalent* if $|u|_a = |v|_a$ for every letter a.

If there are abelian equivalent words $u_0, \ldots, u_{n+1}$ of length p and a non-negative integer $r \leq p-1$ such that

$$w = \mathrm{suff}_r(u_0)u_1 \ldots u_n \mathrm{pref}_{|w|-np-r}(u_{n+1}),$$

then w has *k-abelian period* p. If $r = 0$, then w has *initial* abelian period p.

In [8] it is proved that a word of length $2pq-1$ having relatively prime abelian periods p and q has also period 1. The authors also conjectured that this bound is optimal.

Theorem 2.2 (Constantinescu and Ilie). *Let $p, q > \gcd(p,q) = 1$. Let w have abelian periods p and q. If*

$$|w| \geq 2pq - 1,$$

then w is unary.

In [9] an algorithm constructing optimal words was described, and a proof of correctness was provided for some pairs (p, q).

Initial abelian periods were not considered in [8] and [9] but from the proofs it is quite easy to see that that the value $2pq - 1$ could be replaced with pq if the periods are assumed to be initial.

Let k be a positive integer. Two words u and v are *k-abelian equivalent* if the following conditions hold:

- $|u|_t = |v|_t$ for every word t of length k,
- $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$ and $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$ (or $u = v$, if $|u| < k-1$ or $|v| < k - 1$).

We can replace the conditions with a single one and get an equivalent definition:

- $|u|_t = |v|_t$ for every word t of length at most k.

It is easy to see that k-abelian equivalence implies k'-abelian equivalence for every $k' < k$. In particular, it implies abelian equivalence, which is the same as 1-abelian equivalence. For more on k-abelian equivalence, see [11] and [12].

We define k-abelian periodicity similarly to abelian periodicity: If there are k-abelian equivalent words $u_0, \dots, u_{n+1}$ of length p and a non-negative integer $r \leq p - 1$ such that

$$w = \mathrm{suff}_r(u_0)u_1 \dots u_n \mathrm{pref}_{|w|-np-r}(u_{n+1}),$$

then w has *k-abelian period* p with *offset* r. If $r = 0$, then w has *initial* k-abelian period p. Notice that if w has a k-abelian period p, then so has every factor of w and w^R. If w has initial k-abelian period p, then so has every prefix of w.

In this article we are mostly interested in initial k-abelian periods. Many of our results could be generalized for noninitial periods, but these generalizations are more complicated and the bounds are worse.

Example 2.3. The initial abelian periods of $w = babaaabaabb$ are $5, 7, 8, 9, 10, \dots$. In addition, w has abelian periods 3 and 6.

If k is large compared to p, then k-abelian period p is also an ordinary period.

Lemma 2.4. *If w has a k-abelian period $p \leq 2k - 1$, then it has period p.*

Proof. Words of length $\leq 2k - 1$ are k-abelian equivalent iff they are equal. $\square$

Let $k \geq 1$ and let $p, q \geq 2$ be such that neither of p and q divides the other. Let $d = \gcd(p, q)$. We define $L_k(p, q)$ to be the length of the longest word that has initial k-abelian periods p and q but does not have initial k-abelian period d. If there are arbitrarily long such words, then $L_k(p, q) = \infty$.

The following two questions can be asked:

- For which values of k, p and q is $L_k(p, q)$ finite?
- If $L_k(p, q)$ is finite, how large is it?

If w is a word of length pq/d that has initial k-abelian periods p and q but does not have initial k-abelian period d, then also the infinite word w^ω has initial k-abelian periods p and q but does not have initial k-abelian period d. So either $L_k(p,q) < pq/d$ or $L_k(p,q) = \infty$.

The first question is answered exactly in Sect. 3: $L_k(p,q)$ is finite if and only if $d \leq k$. The second question is answered exactly if $k = 2$. This is done in Sect. 4. In Sect. 5, nontrivial upper bounds are proved for $L_k(p,q)$ in the case $d \leq k$.

The same questions can be asked also for non-initial k-abelian periods. Again, infinite words exist if and only if $d > k$, but the proof is omitted here.

The following lemma, stated here without proof, shows that the size of the alphabet is not important in our considerations (if there are at least two letters).

Lemma 2.5. *If there is a word w that has k-abelian periods p and q but that does not have k-abelian period $d = \gcd(p,q)$, then there is a binary word of length $|w|$ that has k-abelian periods p and q but that does not have k-abelian period d.*

3 Existence of Bounds

In this section we characterize when $L_k(p,q)$ is finite: If $\gcd(p,q) > k$, then $L_k(p,q) = \infty$ by Theorem 3.1, otherwise $L_k(p,q) < pq/d$ by Theorem 3.4.

Theorem 3.1. *Let $p, q > \gcd(p,q) = d > k$. There is an infinite word that has initial k-abelian periods p and q but that does not have k-abelian period d.*

Proof. If $k = 1$, then $a^d bba^{d-2}(ba^{d-1})^\omega$ is such a word, and if $k > 1$, then $a^{2d-k-1}ba^{k-1}b(a^{d-1}b)^\omega$ is such a word. These words have initial k-abelian periods id for all $i > 1$ and hence initial k-abelian periods p and q. $\square$

Assume that a word has k-abelian periods p, q. If $p, q \leq 2k-1$, then they are ordinary periods, so Theorem 2.1 can be used. If $p \leq 2k-1$ but $q > 2k-1$, then we get the following result that is similar to Theorem 2.1 but slightly worse.

Theorem 3.2. *Let $p < 2k$ and $p \leq q$. Let w have k-abelian periods p and q. If*

$$|w| \geq 2p + 2q - 2k - 1 \qquad \text{and} \qquad |w| \geq 2q - 1,$$

then w has period $\gcd(p,q)$.

Proof. If $q \leq 2k-1$, then the claim follows from Lemma 2.4 and Theorem 2.1, so let $q \geq 2k$. By Lemma 2.4, p is a period. Let q be k-abelian period with offset r. Because $|w| \geq 2q - 1$, there is an integer j such that

$$0 \leq r + jq \leq \frac{|w|}{2} \leq r + (j+1)q \leq |w|.$$

Then there are words t, u, s such that $|t| = r + jq$, $|u| = q$ and $w = tus$. Because

$$|st| = |w| - q \geq 2p + q - 2k - 1 \geq 2p - 1,$$

one of t and s has length at least p. The other has length at least $\lceil |w|/2 \rceil - q \geq p-k$. It follows that w has a factor $v = t'us'$, where $|t's'| = p-1$, s' is a prefix of s and of $\mathrm{pref}_{k-1}(u)$ and t' is a suffix of t and of $\mathrm{suff}_{k-1}(u)$. Then v has periods p and q. By Theorem 2.1, v has period $\gcd(p, q)$. Because w has period p and its factor of length p has period $\gcd(p, q)$, w has period $\gcd(p, q)$. □

Lemma 3.3. *If w has a k-abelian period p and some factor of w of length $2p-1$ has at most k factors of length k, then w has a period $d \leq k$ that divides p.*

Proof. If $p \leq k$, then we can set $d = p$ by Lemma 2.4, so let $p > k$. There are k-abelian equivalent words $u_0, \dots, u_n$ of length p such that $w = tu_1 \dots u_{n-1}s$, where t is a suffix of u_0 and s is a prefix of u_n. Every factor v of w of length $2p-1$ has a factor of the form $v' = t'u_ms'$, where t' is a suffix of every u_i, s' is a prefix of every u_i and $|t's'| = k-1$. Every factor of w of length k is a factor of v'. Because v can be selected so that it has at most k factors of length k, it follows that also w has at most k factors of length k. Thus w has a period $d_1 \leq k$. By Theorem 3.2, w has period $\gcd(d_1, p)$. □

Theorem 3.4. *Let w have initial k-abelian periods p and q, $d = \gcd(p, q) < p, q$ and $d \leq k$. If*

$$|w| \geq \mathrm{lcm}(p, q),$$

then w has period d.

Proof. If $p \leq k$ or $q \leq k$, then the claim follows from Theorem 3.2, so let $p, q > k$. Let $p = dp'$ and $q = dq'$ and let w' be the prefix of w of length $pq/d = p'q'd$. There is a word u of length p and a word v of length q such that w' is k-abelian equivalent with $u^{q'}$ and $v^{p'}$. Let s be the common prefix of u and v of length $k-1$. If $x \in \Sigma^k$, then

$$|w's|_x = |u^{q'}s|_x = q'|us|_x \qquad \text{and} \qquad |w's|_x = |v^{p'}s|_x = p'|vs|_x.$$

Thus $|w's|_x$ is divisible by both p' and q', so it is divisible by $p'q'$. In particular, it is either 0 or at least $p'q'$. Because

$$\sum_{x \in \Sigma^k} |w's|_x = |w's| - k + 1 = |w'| = p'q'd,$$

there can be at most d factors $x \in \Sigma^k$ such that $|w's|_x \geq p'q'$. This means that $w's$ can have at most d different factors of length k. By Lemma 3.3, w has a period $d_1 \leq k$ that divides p. By Theorem 3.2, w has period $\gcd(d_1, q)$. This divides d, so w has period d. □

4 Initial 2-Abelian Periods

In this section the exact value of $L_2(p, q)$ is determined. We start with upper bounds and then give matching lower bounds. First we state the following lemma that is very useful also later in the general k-abelian case.

Lemma 4.1. *Let $p = dp'$, $q = dq'$, $\gcd(p', q') = 1$ and $p', q' \geq 2$. For every i satisfying $0 < |i| < \min\{p', q'\}$ there are numbers $m_i \in \{1, \ldots, q' - 1\}$ and $n_i \in \{1, \ldots, p' - 1\}$ such that*

$$n_i q - m_i p = id. \tag{1}$$

The notation of Lemma 4.1 is used in this section and in the later sections, that is, m_i and n_i are always numbers such that (1) holds. The equalities $n_1 q = m_1 p + d$ and $m_{-1} p = n_{-1} q + d$ are particularly important.

Upper Bounds. The following lemma gives an upper bound in the 2-abelian case.

Lemma 4.2. *Let $p, q \geq 2$ and $\gcd(p, q) = 1$. Let w have initial 2-abelian periods p and q. If*

$$|w| \geq \max\{n_1 q, m_{-1} p\},$$

then w is unary.

Proof. The word w has prefixes

$$u_1 \ldots u_{m_1} a = v_1 \ldots v_{n_1} \qquad \text{and} \qquad u_1 \ldots u_{m_{-1}} = v_1 \ldots v_{n_{-1}} a', \tag{2}$$

where the u_i are 2-abelian equivalent words of length p, the v_i are 2-abelian equivalent words of length q and a, a' are letters. Both a and a' are first letters of every u_i and v_i, so they are equal.

For any letter $b \neq a$, it follows from (2) that

$$m_1 |u_1|_b = n_1 |v_1|_b \qquad \text{and} \qquad m_{-1} |u_1|_b = n_{-1} |v_1|_b$$

and thus

$$m_1 n_{-1} |u_1|_b |v_1|_b = n_1 m_{-1} |u_1|_b |v_1|_b. \tag{3}$$

By (1), $m_1 p < n_1 q$ and $m_{-1} p > n_{-1} q$ and thus

$$m_1 n_{-1} < n_1 m_{-1}. \tag{4}$$

Both (3) and (4) can hold only if $|u_1|_b |v_1|_b = 0$. It follows that $w \in a^*$. □

Lower Bounds. The following lemma gives a lower bound in the 2-abelian case.

Lemma 4.3. *Let $q > p \geq 2$, $\gcd(p, q) = 1$, x, y be the smallest positive integers such that $xq - yp = \pm 1$. Then there exists a non-unary word w of length $(p - x)q$ in the case $xq - yp = +1$ (or $(q - y)p$ in the case $xq - yp = -1$) which has initial 2-abelian periods p and q.*

Proof. The pair (x, y) is either (n_1, m_1) or (n_{-1}, m_{-1}) (the one with smaller numbers), and the pair $(p - x, q - y)$ is the other one.

First we describe the construction (actually, the algorithm producing the word w), then give an example, and finally we prove that the algorithm works correctly, i.e. it indeed produces a word with initial 2-abelian periods p and q.

We need the following notion. Let $m \geq l \geq 0$, $c, d \in \{a, b\}$. Define $K_2(m, l, c, d)$ to be the set of binary words satisfying the following conditions:

- words of length m
- containing l letters b (and hence $m-l$ letters a)
- b's in them are isolated (i.e., with no occurrence of factor bb)
- the first letter being $c \in \{a,b\}$, the last letter being $d \in \{a,b\}$

The following properties are easy to conclude.

1. The set $K_2(m,l,a,a)$ is non-empty for $l < m/2$, and the set $K_2(m,l,a,b)$ is non-empty for $0 < l \leq m/2$.
2. $K_2(m,l,c,d)$ is a 2-abelian class of words. For $c = d = a$ it contains l occurrences of ab, l occurrences of ba, $m-2l-1$ occurrences of aa and no occurrences of bb. For $c = a$, $d = b$ it contains l occurrences of ab, $l-1$ occurrences of ba, $m-2l$ occurrences of aa and no occurrences of bb.
3. If $u \in K_2(m,l,c,d)$, $u' \in K_2(m',l',c',d')$, and at least one of the letters d and c' is a, then $uu' \in K_2(m+m', l+l', c, d')$.

Now, our construction is done as follows:

1. Find the smallest integers x, y satisfying $xq - yp = \pm 1$. In the case $xq - yp = -1$ we construct a word w with 2-abelian periods $K_2(p,x,a,a)$ and $K_2(q,y,a,b)$. Note that in this case $x < p/2$, $y \leq q/2$. If $xq - yp = 1$, we take the periods to be $K_2(p,x,a,b)$ and $K_2(q,y,a,a)$. In this case $x \leq p/2$, $y < q/2$. To be definite, assume that $xq - yp = -1$, in the other case the construction is symmetric.
2. Now we start building our word based on 2-abelian periods indicated in 1. We mark the positions ip and jq for $i = 0, \ldots, q-y$, $j = 0, \ldots, p-x$, and denote these positions by t_m in increasing order, $m = 0, \ldots, q-y+p-x$. Now we will fill in the factors $v_m = w[t_{m-1}, t_m - 1]$ one after another.
   - 2.1. Put v_1 equal to any word from the 2-abelian class $K_2(p,x,a,a)$ of p-period.
   - 2.2. If in v_m we have that $t_{m-1} = (i-1)p$ and $t_m = ip$ for some i, then simply put any word from the 2-abelian class $K_2(p,x,a,a)$ of p-period.
   - 2.3. If in v_m we have that $t_{m-1} = ip$ and $t_m = jq$ for some i and j, then fill it with any word from $K_2(jq-ip, jy-ix, a, b)$. Then the word $w_{t_m - q} \ldots w_{t_m - 1}$ is from the 2-abelian class $K_2(q,y,a,b)$ of the q-period.
   - 2.4. If in v_m we have that $t_{m-1} = jq$ and $t_m = ip$ for some i and j, then fill it with any word from $K_2(ip-jq, ix-jy, a, a)$. Then the word $w_{t_m - p} \ldots w_{t_m - 1}$ is from the 2-abelian class $K_2(p,x,a,a)$ of the p-period.

Example 4.4. $p = 7$, $q = 10$. We find $x = 2$, $y = 3$, so we take the 2-abelian class of the word $aaababa$ as p-period and the 2-abelian class of $aaaaababab$ as q-period, and the length of word is $p(q-y) = 49$. One of the words given by the construction is

$$aaababa.aab^{\cdot}aaba.aaabab^{\cdot}a.aaababa.ab^{\cdot}aaaba.aabab^{\cdot}aa.aaababa.$$

Here the lower dots are placed at positions $7i$, and the upper dots at positions $10j$. This word has initial 2-abelian periods 7 and 10. In the example each time we chose the lexicographically biggest word v_i, though we actually have some flexibility. E.g., one might take $v_1 = abaaaba$, so the word is not unique.

To prove the correctness of the algorithm, we will prove that on each step 2.3 and 2.4 the corresponding 2-abelian classes are non-empty, so that one can indeed choose such a word. This would mean that on each step m we obtain a word $v_1 \dots v_m$ such that all its prefixes of lengths divisible by p and q are 2-abelian p- and q-periodic, respectively (in other words, we have periodicity in full periods up to length t_m), and the last incomplete period (either p- or q-period) starts with a.

Correctness of step 2.3. At step 2.3, we should add a word $v_m \in K_2(jq - ip, jy - ix, a, b) = K_2(t_m - t_{m-1}, l, a, b)$, where the length $t_m - t_{m-1} = jq - ip < p$ and the number l of b's is as large as required so that $w_{t_m - q} \dots w_{t_m - 1}$ be 2-abelian equivalent to the 2-abelian q-period $K_2(q, y, a, b)$. In view of properties 1–3, these conditions are sufficient to guarantee that $w_{t_m - q} \dots w_{t_m - 1}$ is 2-abelian equivalent to the 2-abelian q-period $K_2(q, y, a, b)$. The only thing we should care about is that we can indeed choose such a word, i.e., that the set $K_2(jq - ip, jy - ix, a, b)$ is non-empty. So, we should check the required number $l = jy - ix$ of b's: it should not be larger than $|v_m/2|$ and it should not be less than 1.

Suppose $l = jy - ix \leq 0$ (negative values mean that we already have too many b's). The density of the letter b is $\rho_b^q = y/q$ in the q-period, and $\rho_b^p = x/p$, in the p-period. Since $xq - yp = -1$, we have $\rho_b^q = y/q > \rho_b^p = x/p$. On the other hand, since $jy < ix$ by assumption and $ip < jq$, we have a contradiction:

$$\rho_b^q = \frac{jy}{jq} \leq \frac{ix}{jq} < \frac{ix}{ip} = \rho_b^p.$$

Suppose $|v_m| > l > |v_m|/2$. By induction hypothesis, we have that the word $v_1 \dots v_{m-1}$ of length ip contains xi letters b, hence $v_1 \dots v_m$ of length jq should contain more than $xi + |v_m|/2$ letters b.

Consider a word u of length $(p - x)q$ with density $\rho_b^q = y/q$ having $y(p - x)$ letters b in it. Removing one letter b from it, we obtain a word of length $(p - x)q - 1 = (q - y)p$ with density

$$\frac{y(p - x) - 1}{(p - x)q - 1} = \frac{(q - y)x}{(q - y)p} = \frac{x}{p} = \rho_b^p$$

and with $x(q - y)$ letters b in it.

Now consider a word $v = v_1 \dots v_{m-1} v'$, where v' is of length $jq - ip$ and contains $jy - ix$ letters b (i.e., it is of the same length and with the same number of b's as v_m is supposed to be for 2-abelian q-periodicity). So, $|v| = qj$, $|v|_b = yj$. Now remove one letter b from the suffix v' of v. Compare v with the word u. Since v is shorter than u, after removing one letter b from v and from u the density of b's in the remaining part of v is smaller than in the remaining part of u, which is $\rho_b^p = x/p$. The remaining part of v consists of $v_1 \dots v_{m-1}$ with density $\rho_b^p = x/p$ and v' without one b and with density at least $1/2$. Since $\rho_b^p = x/p < 1/2$, we have that the density of b's in the remaining part of v is bigger than $\rho_b^p = x/p$. A contradiction.

The case $l \geq |v_m|$ leads to a contradiction in a similar way.

Correctness of step 2.4 is proved similarly to correctness of step 2.3.

So, we built a word w of length $(q-y)p$ having initial 2-abelian p-period and initial 2-abelian q-period till length $(q-y)p-q+1$ (within full periods). It remains to check that $\mathrm{suff}_{q-1}(w)$ can be extended till a word of the 2-abelian class $K_2(q,y,a,b)$ of the q-period. It is easy to see that it can be extended in this way by adding letter b. □

By a similar construction we find optimal words for the abelian case. Actually, in the abelian case the proof is simpler, since one has less restrictions than in the 2-abelian case; the only thing one should take care of is frequencies of letters. We construct such words satisfying additional condition, which we use later for k-abelian case. The construction is similar to the construction from Lemma 4.3, we omit the details due to space limit.

Lemma 4.5. *Let $q > p \geq 2$, $\gcd(p,q) = 1$. Then there exists a non-unary word w of length $pq-1$ which has initial abelian periods p and q, and moreover $w_{ip} = w_{ip+p}$ and $w_{jq} = w_{jq+q}$ for all i, j for which the indices are defined.*

Lemma 4.6. *Let $q > p > \gcd(p,q) = d = k$. Then there exists a non-unary word w of length $pq/d-1$ which has initial k-abelian periods p and q and no k-abelian period d.*

Proof. The word w is constructed from the word w' given by construction from Lemma 4.5 for p/d and q/d: $w = \varphi(w')a^{k-1}$, where the morphism φ is given by $\varphi(a) = a^k, \varphi(b) = a^{k-1}b$. □

Optimal Values. Combining the previous results gives two exact theorems.

Theorem 4.7. *Let $p, q > \gcd(p,q) = k$. Then*

$$L_k(p,q) = \frac{pq}{k} - 1.$$

Proof. Follows from Theorem 3.4 and Lemmas 4.5 and 4.6. □

Now we get a version of Fine and Wilf's theorem for initial 2-abelian periods.

Theorem 4.8. *Let $p, q > \gcd(p,q) = d$. Then*

$$L_2(p,q) = \begin{cases} \max\{m_1 p, n_{-1} q\} & \text{if } d = 1, \\ pq/2 - 1 & \text{if } d = 2, \\ \infty & \text{if } d \geq 3. \end{cases}$$

Proof. After some calculations the case $d = 1$ follows from Lemmas 4.2 and 4.3, the case $d = 2$ from Theorem 4.7, and the case $d \geq 3$ from Theorem 3.1. □

The size of $\max\{m_1 p, n_{-1} q\}$ depends a lot on the particular values of p and q. The extreme cases are $p = 2$, which gives $n_{-1} q = q$, and $q = p+1$, which gives $n_{-1} q = pq - q$. In general we get the following corollary.

Corollary 4.9. *Let $q > p > \gcd(p,q) = 1$. Then*

$$\frac{pq}{2} + \frac{p}{2} - 1 \leq L_2(p,q) \leq pq - q.$$

5 General Upper Bounds

In this section $L_k(p,q)$ is studied for $k \geq 3$. We are not able to give the exact value in all cases, but we will prove an upper bound that is optimal for an infinite family of pairs (p,q). We start with an example.

Example 5.1. Let $k \geq 2$, $p \geq 2k-1$ and $q = p+1$. The word

$$(a^{p-k+1}ba^{k-2})^{q-2k+2}a^{p-k+1}$$

of length $(q-2k+2)p + p - k + 1 = pq - 2kq + 3q + k - 2$ has initial k-abelian periods p and q but does not have period $\gcd(p,q) = 1$.

Recall the notation of Lemma 4.1: $m_i \in \{1,\dots,q'-1\}$ and $n_i \in \{1,\dots,p'-1\}$ are numbers such that $n_i q - m_i p = id$. This is used in the following lemmas and theorems. The proofs of Lemmas 5.2 and 5.3 are in some sense more complicated and technical versions of the proof of Lemma 4.2 and they are omitted because of space constraints.

Lemma 5.2. *Let $p = dp'$, $q = dq'$, $\gcd(p',q') = 1$ and $2 \leq p' < q'$. Let $k - 1 = dk'$ and $1 \leq k' < p'/2$. Let w have initial k-abelian periods p and q. Let $u = \mathrm{pref}_p(w)$, $v = \mathrm{pref}_q(w)$ and $s = \mathrm{pref}_d(w)$. If there are indices*

$$i \in \{-1,1\}, \qquad j \in \{-k',k'\}, \qquad l \in \{-2k'+1,\dots,-1\} \cup \{1,\dots,2k'-1\}$$

such that i,j,l do not all have the same sign and

$$m_i p, n_i q \leq |w| - k + 1 + d, \tag{5}$$
$$m_j p, n_j q \leq |w|, \tag{6}$$
$$m_l p, n_l q \leq |w| + k - 1 - d, \tag{7}$$

then

$$\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v) = s^{k'} \qquad \text{and} \qquad \mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v) = s^{k'}.$$

Lemma 5.3. *Let $p = dp'$, $q = dq'$, $\gcd(p',q') = 1$ and $2 \leq p' < q'$. Let $k - 1 = dk'$ and $1 \leq k' < p'/2$. Let w have initial k-abelian periods p and q. Let $u = \mathrm{pref}_p(w)$, $v = \mathrm{pref}_q(w)$ and $s = \mathrm{pref}_d(w)$. Let*

$$\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v) = s^{k'} \qquad \text{and} \qquad \mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v) = s^{k'}.$$

If there are indices

$$i,j \in \{-2k',\dots,-1\} \cup \{1,\dots,2k'\}$$

such that $m_i n_j \neq m_j n_i$ and

$$m_i p, n_i q, m_j p, n_j q \leq |w| + k - 1, \tag{8}$$

then w has period d.

Theorem 5.4. *Let $p = dp'$, $q = dq'$, $\gcd(p', q') = 1$ and $2 \le p' < q'$. Let $k - 1 = dk'$ and $1 \le k' \le p'/4$. Let w have initial k-abelian periods p and q. If*

$$|w| \ge \frac{pq}{d} - \frac{2(k-1)q}{d} + q + k - 1$$

then w has period d.

Proof. If $n_i q - m_i p = id$, then $(p' - n_i)q - (q' - m_i)p = -id$. It follows that for every i, either $m_i p, n_i q \le pq/(2d)$ or $m_{-i}p, n_{-i}q \le pq/(2d)$. Because $p \ge 4(k-1)$, $|w| \ge pq/(2d)$. Thus the indices i, j in Lemma 5.2 exist.

If the above indices i and j have a different sign, then l exists (for example, $l = i$ will do). If i and j have the same sign, then there are $2k' - 1$ candidates for l. All of these have the same sign, so for these l, the numbers n_l are different. If we select l so that n_l is as small as possible, then $n_l \le p' - 2k' + 1$. Now

$$m_l p, n_l q \le n_l q + |l|d \le (p' - 2k' + 1)q + (2k' - 1)d,$$

so in order for (7) to be satisfied, it is sufficient that

$$|w| \ge (p' - 2k' + 1)q + (2k' - 1)d - k + 1 + d.$$

This is the bound of the theorem, so l exists and Lemma 5.2 can be used.

We need to prove the existence of the indices i and j in Lemma 5.3; the other assumptions are satisfied by Lemma 5.2. By the argument that was used for the existence of the index l above, there exists $i \in \{1, \dots, 2k'\}$ such that $m_i p, n_i q \le |w| + k - 1$ and $j \in \{-1, \dots, -2k'\}$ such that $m_j p, n_j q \le |w| + k - 1$. Because $m_i p < n_i q$ and $n_j q < m_j p$, it follows that $m_i n_j < n_i m_j$ and Lemma 5.3 can be used to complete the proof. □

If $d = 1$, then Theorem 5.4 tells that $L_k(p, q) \le pq - 2kq + 3q + k - 2$. By Example 5.1, there is an equality if $q = p + 1$. The next example shows that for some p and q the exact value is much smaller.

Example 5.5. Let $k \ge 2$, $r \ge 2$, $p = rk+1$ and $q = rk+k+1$. Then $\gcd(p, q) = 1$. The word $w = ((a^{k-1}b)^r a)^{r+2} a^{k-2}$ has initial k-abelian periods p and q. Because $n_{-1} = r$, $m_{-1} = r + 1$, $m_{k-1} = r + 2$ and $n_{k-1} = r + 1$, it follows from Lemmas 5.2 and 5.3 and the above word w that

$$L_k(p, q) = (r + 2)p + k - 2 = \frac{pq}{k} + q - \frac{q}{k} - 1.$$

6 Conclusion

We conclude with a summary of the results related to initial k-abelian periods. Let $d = \gcd(p, q) < p < q$ and $d \le k$.

- By Theorem 4.7, if $d = k$, then

$$L_k(p, q) = \frac{pq}{d} - 1.$$

– By Theorem 4.8, if $d = 1$, then

$$L_2(p, q) = \max\{m_1 p, n_{-1} q\}.$$

– By Theorem 5.4, if $2 \leq k \leq p/4 + 1$ and $k - 1$ is divisible by d, then

$$L_k(p, q) \leq \frac{pq}{d} - \frac{2(k-1)q}{d} + q + k - 2.$$

This is optimal if $q = p + 1$.

– By Theorem 3.2, if $k \geq (p+1)/2$, then

$$L_k(p, q) \leq \max\{2p + 2q - 2k - 2, 2q - 2\}.$$

– By Lemma 2.4 and Theorem 2.1, if $k \geq (q+1)/2$, then

$$L_k(p, q) = p + q - d - 1.$$

References

1. Fine, N.J., Wilf, H.S.: Uniqueness theorems for periodic functions. Proc. Amer. Math. Soc. **16** (1965) 109–114
2. 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**(1) (1999) 83–94
3. Tijdeman, R., Zamboni, L.Q.: Fine and Wilf words for any periods II. Theoret. Comput. Sci. **410**(30–32) (2009) 3027–3034
4. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. Theoret. Comput. Sci. **218**(1) (1999) 135–141
5. Blanchet-Sadri, F.: Periodicity on partial words. Comput. Math. Appl. **47**(1) (2004) 71–82
6. Shur, A.M., Gamzova, Y.V.: Partial words and the period interaction property. Izv. Ross. Akad. Nauk Ser. Mat. **68**(2) (2004) 191–214
7. Kari, L., Seki, S.: An improved bound for an extension of Fine and Wilf's theorem and its optimality. Fund. Inform. **101**(3) (2010) 215–236
8. Constantinescu, S., Ilie, L.: Fine and Wilf's theorem for abelian periods. Bull. Eur. Assoc. Theor. Comput. Sci. EATCS **89** (2006) 167–170
9. Blanchet-Sadri, F., Tebbe, A., Veprauskas, A.: Fine and Wilf's theorem for abelian periods in partial words. In: Proceedings of the 13th Mons Theoretical Computer Science Days. (2010)
10. Choffrut, C., Karhumäki, J.: Combinatorics of words. In Rozenberg, G., Salomaa, A., eds.: Handbook of Formal Languages. Volume 1. Springer-Verlag (1997) 329–438
11. Huova, M., Karhumäki, J., Saarela, A.: Problems in between words and abelian words: k-abelian avoidability. Theoret. Comput. Sci. (To appear)
12. Karhumäki, J., Saarela, A., Zamboni, L.: On a generalization of Abelian equivalence (In preparation)