On k-Abelian Palindromic Rich and Poor Words

Juhani Karhumäki[1,⋆] and Svetlana Puzynina[1,2,⋆⋆]

[1] Department of Mathematics, University of Turku, FI-20014 Turku, Finland
{karhumak,svepuz}@utu.fi
[2] Sobolev Institute of Mathematics, Russia

Abstract. A word is called a *palindrome* if it is equal to its reversal. In the paper we consider a k-abelian modification of this notion. Two words are called *k-abelian equivalent* if they contain the same number of occurrences of each factor of length at most k. We say that a word is a *k-abelian palindrome* if it is k-abelian equivalent to its reversal. A question we deal with is the following: how many distinct palindromes can a word contain? It is well known that a word of length n can contain at most $n+1$ distinct palindromes as its factors; such words are called *rich*. On the other hand, there exist infinite words containing only finitely many distinct palindromes as their factors; such words are called *poor*. It is easy to see that there are no abelian poor words, and there exist words containing $\Theta(n^2)$ distinct abelian palindromes. We analyze these notions with respect to k-abelian equivalence. We show that in the k-abelian case there exist poor words containing finitely many distinct k-abelian palindromic factors, and there exist rich words containing $\Theta(n^2)$ distinct k-abelian palindromes as their factors. Therefore, for poor words the situation resembles normal words, while for rich words it is similar to the abelian case.

1 Introduction

The palindromicity of words is a widely studied area in formal languages. When a model of a computation is introduced, among the first questions is to ask whether the set of palindromes (or its infinite subset) can be recognized by the model. In other words, can the model identify whether it is irrelevant if words are read from left to right or from right to left? It is folklore that deterministic finite automata cannot do that. On the other hand it is among the simplest tasks for push-down automata, or on-line log-space Turing machines. A slightly different approach is to look at palindromic factors of words. They can be viewed as measuring how much the word is locally independent of the reading direction of a factor. The notion of palindromic complexity was formalized for infinite words in [6], and has been studied extensively ever since.

A problem related to our question of counting palindromes in a word is the problem of counting maximal repetitions in a word of length n, that is, runs in a

⋆ Supported in part by the Academy of Finland under grant 257857.
⋆⋆ Supported in part by the Academy of Finland under grant 251371 and by Russian Foundation of Basic Research (grants 12-01-00448, 12-01-00089).

A.M. Shur and M.V. Volkov (Eds.): DLT 2014, LNCS 8633, pp. 191–202, 2014.

word. It was shown in [18] that the maximal number of runs in a word is linear in n. Subsequently, there was much research performed to find the bound [8], which led to a conjecture that the bound should be n. Very close lower and upper bounds have been proved; however, the conjecture still remains open. Not only runs, but also various other questions concerning counting squares in a word have been considered, see, e.g., [13,14,19].

We recall that a word is a *palindrome*, if it is equal to its reversal. It is well known that the maximal number of palindromes a word of length n can contain is equal to $n + 1$, and such words are called *rich* in palindromes [10]. In some papers the same class of words was called *full* words (see, e.g., [2,6]). Lately, there is an extensive number of papers devoted to the study of rich words and their generalizations (see, e.g., [7,12]). This notion can be extended to infinite words: an infinite word is rich if each of its factors is rich. For example, Sturmian words are known to be rich. Note also that Sturmian words can be characterized via palindromic closures [9].

Recently the notion of palindromic poorness has been considered in [5,11]. Namely, an infinite word is called *poor* in palindromes if it contains only finitely many distinct palindromes. In particular, it has been shown that there exist poor words with the set of factors closed under reversal. Besides that, in [11] the authors found the minimal number of palindromes an infinite word satisfying different conditions (uniform recurrence, closed under reversal, etc.) can contain. In a related paper [23] words avoiding reversed subwords were studied.

In this paper the k-abelian version of the notion of a palindrome is studied. Two words are called *abelian equivalent* if they contain the same number of occurrences of each letter, or, equivalently, if they are permutations of each other. In the recent years there is a growing interest in abelian properties of words, as well as modifications of the notion of abelian equivalence [1,4,17,21,23]. One such modification is the notion of k-abelian equivalence: two words are called *k-abelian equivalent* if they contain the same number of occurrences of each factor of length at most k. For $k = 1$, the notion of k-abelian equivalence coincides with the notion of abelian equivalence, and when k is greater than half of the length of the words, k-abelian equivalence means equality. Therefore, the notion of k-abelian equivalence is an intermediate notion between abelian equivalence and equality of words. For more on k-abelian equivalence we refer to [15,16].

In analogy with normal palindromes, we say that a word v is a *k-abelian palindrome* if its reversal is k-abelian equivalent to v. For example, the word *aabaaabbaa* is a 3-abelian palindrome. We are interested in the maximal and minimal numbers of k-abelian palindromes a word can contain.

For $k = 1$, clearly, each word is an abelian palindrome, since it is abelian equivalent to its reversal. Therefore, there are no infinite 1-abelian poor words. But for $k > 1$ this no longer holds. We build infinite k-abelian poor words for $k > 1$ and sufficiently large alphabets. In fact, we provide a complete characterization of pairs (k, Σ) for which k-abelian poor words over the alphabet Σ exist.

Since a word of length n contains $\Theta(n^2)$ factors in total, a k-abelian rich word cannot contain more than $\Theta(n^2)$ abelian palindromes. However, it can indeed contain $\Theta(n^2)$ inequivalent abelian palindromes. We show that this extends to k-abelian palindromes when k is small compared to n.

The maximal and minimal numbers of inequivalent palindromes in the case of the equality, the k-abelian equality and the abelian equality are summarized in Table 1 (here C is a constant). We remark that in the minimal case, that is for poor words, infinite words are considered, while in the maximal case, that is in rich words, only finite words are considered. The message of the table is that in the big picture k-abelian equivalence behaves like equality for poor words, while it behaves like abelian equivalence for rich words.

Table 1. Maximal and minimal numbers of palindromes in the case of equality, abelian and k-abelian equivalence

	equality	k-abelian	abelian
poor	C	C	∞
rich	$n+1$	$\Theta(n^2)$	$\Theta(n^2)$

2 Definitions and Notation

Given a finite non-empty set Σ (called the alphabet), we let Σ^* and Σ^ω, respectively, denote the set of finite words and the set of (right) infinite words over the alphabet Σ. We will always assume $|\Sigma| \geq 2$. A word v is a *factor* (resp., a *prefix*, resp., a *suffix*) of a word w, if there exist words x, y such that $w = xvy$ (resp., $w = vy$, resp., $w = xv$). The set of factors of a finite or infinite word w is denoted by $F(w)$. The prefix and suffix of length k of w are denoted by $\mathrm{pref}_k(w)$ and $\mathrm{suff}_k(w)$, respectively. Given a finite word $u = u_1u_2 \dots u_n$ with $n \geq 1$ and $u_i \in \Sigma$, we let $|u| = n$ denote the length of u. The empty word is denoted by ε and we set $|\varepsilon| = 0$. An infinite word is called *recurrent* if each of its factors occurs infinitely often in it. An infinite word w is called *uniformly recurrent* if for each $v \in F(w)$ there exists N such that $v \in F(w_i \cdots w_{i+N})$ for every i. In other words, in a uniformly recurrent word each factor occurs with bounded gaps.

For each $v \in \Sigma^*$, we let $|u|_v$ denote the number of occurrences of the factor v in u. Two words u and v in Σ^* are said to be *abelian equivalent*, denoted $u \sim_{ab} v$, if and only if $|u|_a = |v|_a$ for all $a \in \Sigma$. For example, the words aba and aab are abelian equivalent. Clearly, abelian equivalence is an equivalence relation on Σ^*.

Let k be a positive integer. Two words u and v are *k-abelian equivalent*, denoted by $u \sim_k v$, if $|u|_t = |v|_t$ for every word t of length at most k. This is equivalent to the following conditions:

- $|u|_t = |v|_t$ 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 $|v| < k-1$).

For instance, $aabab \sim_2 abaab$, but $aabab \nsim_2 aaabb$. It is easy to see that k-abelian equivalence implies k'-abelian equivalence for every $k' < k$. In particular, it implies abelian equivalence, that is, 1-abelian equivalence.

For a finite word $v = v_1 \cdots v_n$ we let $v^R = v_n \cdots v_1$ denote its reversal. A word v is a *palindrome* if $v = v^R$. A word is a *k-abelian palindrome* (or briefly k-palindrome) if $v \sim_k v^R$. The empty word ε is considered as a palindrome and k-palindrome.

An infinite word is *k-abelian palindromic poor* (briefly k-poor) if there exists a constant C such that the word contains at most C inequivalent (in the sense of k-abelian equivalence) k-abelian palindromes. Clearly, it makes sense to consider only words having the set of factors closed under reversal, otherwise the example can be built in the obvious way, e.g., one can take $(abc)^\omega$ containing only 4 k-palindromes.

A word of length n is called *k-abelian palindromic rich* (briefly k-rich), if it contains at least $n^2/4k$ inequivalent k-abelian palindromes. Notice that the total number of factors contained in a word of length n is equal to $1 + \frac{n(n+1)}{2}$. Therefore, for a fixed k a k-abelian rich word contains the number of k-palindromes of the same order as the total number of factors when k is small relatively to n.

We emphasize that for poor words we consider infinite words, and for rich words we consider finite ones, and this is caused by the nature of the problem. Indeed, for poor words, since there exist infinite words containing only finitely many palindromes, all their factors have a uniformly bounded number of palindromes. On the other hand, the closed under reversal condition is not applicable to finite poor words, since it would imply a growing number of palindromes. Concerning rich words, an infinite word could easily contain infinitely many palindromes, so we are interested in maximal number of palindromes in finite ones. However, we propose an open problem concerning a modification of k-palindromic richness for infinite words (see Problem 3 in Section 5). In the next two sections we consider k-abelian poor and rich words, respectively.

3 *k*-Abelian Poor Words

In this section we show that there exist k-abelian palindromic poor words. This holds for almost all values of k and $|\Sigma|$, and we characterize those.

Theorem 3.1. *Let* $S = \{(1, l) | l \in \mathbb{N}\} \cup \{(2, 2), (2, 3), (4, 2), (3, 2)\}$.
I. *For* $(k, |\Sigma|) \notin S$ *there exist k-abelian palindromic poor words over* Σ *having a set of factors that is closed under reversal.*
II. *For* $(k, |\Sigma|) \in S$ *there are no k-abelian palindromic poor words over* Σ *having a set of factors that is closed under reversal.*

The results can be summarized in Table 2. Here + means that there exist k-abelian poor words having a set of factors that is closed under reversal over an alphabet Σ, and − indicates that there are no such words. In what follows, we will write (k, l)-poor words for k-abelian poor words over an alphabet of cardinality l for brevity.

Table 2. The classification of $(k, |\Sigma|)$ for the existence of k-poor words

$k\backslash\lvert\Sigma\rvert$	2	3	4	...
1	−	−	−	−
2	−	−	+	
3	−	+	+	
4	−	+	+	...
5	+	+	+	
...		...		+

Proof. First we prove Part I of the theorem by providing constructions of poor words, and then prove the non-existence for Part II of the theorem.

I. We remark that the existence of a (k, l)-poor word implies the existence of a (k', l')-poor word for each $k' \geq k$ and $l' \geq l$. Indeed, for $l' > l$ to build a (k, l')-poor word from a (k, l)-poor word one could split any letter into several letters in any way (i.e., for a chosen letter a, some of occurrences of a are substituted by one of the $l' - l$ new letters). The word remains k-poor, and closed under reversal condition can be preserved. For $k' > k$, the statement follows from the fact that every k'-abelian palindrome is also a k-abelian palindrome for any $k \leq k'$. Therefore, it is enough to build $(5, 2)$-, $(3, 3)$- and $(2, 4)$-poor words. We will provide the construction and a proof for the $(2, 4)$ case. The other cases are similar, so we just outline the constructions.

We construct an infinite recurrent $(2, 4)$-poor word as follows:

$$\begin{aligned} U_0 &= abca\ abda\ acda, \\ U_n &= U_{n-1}(abca)^{2^{2^n}}(abda)^{2^{2^n}}(acda)^{2^{2^n}}U_{n-1}^R. \end{aligned} \tag{1}$$

The required word is obtained as the limit $u = \lim_{n\to\infty} U_n$:

$$u = abca\ abda\ acda(abca)^4(abda)^4(acda)^4adca\ adba\ acba(abca)^{16}(abda)^{16}\dots$$

The set of factors of this word is closed under reversal since for each prefix the word contains its reversal as a factor by the construction.

To prove that it contains only finitely many 2-abelian palindromes, we will show that each factor of length greater than 12 contains either a unequal number of occurrences of factors bc and cb, or a unequal number of occurrences of factors bd and db, or a unequal number of occurrences of factors cd and dc.

A factor of the form $(abca)^t(abda)^t(acda)^t$ or $(adca)^t(adba)^t(acba)^t$ for $t = 2^{2^i}$ or $t = 1$ is called a *block*. In fact, the word u is a concatenation of blocks. A *subblock* is a factor of the form $(axya)^t$ for $t = 2^{2^i}$ or $t = 1$ and for $(x, y) \in \{(b, c), (c, b), (b, d), (d, b), (c, d), (d, c)\}$. Notice that each factor xy appears only in the corresponding subblock, and its reversal yx appears only only in the reversal of the corresponding subblock. Therefore, we have three pairs of subblocks, where a pair consists of a block and its reversal. We will say that subblocks $(axya)^{2^{2^i}}$ and $(ayxa)^{2^{2^i}}$ are of type (x, y).

Basically, the idea of the construction is based on the fact that we have three types of subblocks, each of them containing a specific factor, such that this factor and its reversal are contained only in subblocks of this type. Each long enough factor will contain a unequal number of occurrences of one of these specific factors and its reversals. Notice that it is important to have three pairs of such factors; two is not enough to guarantee absence of long k-abelian palindromes.

Take a factor v of length $|v| \geq 13$. The following cases are possible. The first case is the principal case when v intersects subblocks of all the three types. The second case is a special case when v intersects at most two different blocks.

Case 1: v intersects subblocks of all the three types. In this case the ends of v can cut at most two subblocks, and hence v contains only full subblocks of at least one type. We let (x, y) denote this type (or one of the types, if there are several of those). The sequence of numbers of occurrences of xy and its reversal yx in full subblocks of v is given by

$$z = 1, 4, -1, 16, 1, -4, -1, 256, 1, 4, -1, -16, 1, -4, -1 \ldots$$

Here positive numbers indicate occurrences of xy, and negative numbers mean the occurrences of yx. In fact, this sequence corresponds to the sequence of exponents of the corresponding block. For a subsequence $v = z(i), \ldots, z(j)$ of z we let $-v$ denote the sequence obtained by changing the sign of all numbers in it: $-v = -z(i), \ldots, -z(j)$. More formally, z is defined recursively as the infinite sequence starting with Z_n for all n:

$$\begin{aligned} Z_0 &= 1, \\ Z_n &= Z_{n-1}, 2^{2^n}, (-Z_{n-1})^R. \end{aligned} \tag{2}$$

If a factor v of u containing only full subblocks of type (x, y) is a palindrome, then the sum of consecutive elements corresponding to the subsequence v in the sequence z is equal to 0. We will prove the following auxiliary claim:

Claim. The sum of consecutive elements in the sequence z is never equal to 0.

Consider any subsequence of consecutive elements $z(i), \ldots, z(i + k)$ of z, and take the element $z(j)$ in it with the largest absolute value, $i \leq j \leq i+k$. We will prove that

$$|z(j)| > \sum_{l=i}^{l=i+k} |z(l)| - |z(j)|.$$

In other words, $|z(j)|$ is greater than the sum of absolute values of all other elements of the subsequence, and hence the sum of all elements in $z(i), \ldots, z(i+k)$ cannot be 0.

Let $|z(j)| = 2^{2^m}$ for some integer m. Clearly, either $z(i), \ldots, z(i + k)$ or $(-z(i), \ldots, -z(i + k))^R$ is a factor of the prefix Z_m of z, including the middle element 2^{2^m} of Z_m. It is enough to prove that

$$2^{2^m} > 2 \sum_{z(l)\in Z_{m-1}} |z(l)|. \tag{3}$$

By the construction,

$$\sum_{z(l)\in Z_i} |z(l)| = 2^{2^i} + 2 \sum_{z(l)\in Z_{i-1}} |z(l)|. \tag{4}$$

We prove (3) by induction. Straightforward computation shows that it holds for $m = 1$ and 2. Assume it holds for $m = i$. Then, using the induction hypothesis, we obtain

$$2^{2^{i+1}} = 2^{2^i} \cdot 2^{2^i} > 2^{2^i} \cdot 2 \sum_{z(l)\in Z_{i-1}} |z(l)| = 2^{2^i} \cdot \sum_{z(l)\in Z_{i-1}} |z(l)| + 2^{2^i} \cdot \sum_{z(l)\in Z_{i-1}} |z(l)|.$$

For $i \geq 2$ one has $\sum_{z(l)\in Z_{i-1}} |z(l)| \geq 2$, and $2^{2^i} \geq 4$. Applying these two inequalities to the first and the second summands, correspondingly, and then applying (4), we get

$$2^{2^i} \cdot \sum_{z(l)\in Z_{i-1}} |z(l)| + 2^{2^i} \cdot \sum_{z(l)\in Z_{i-1}} |z(l)| \geq 2^{2^i} \cdot 2 + 4 \cdot \sum_{z(l)\in Z_{i-1}} |z(l)| = 2 \sum_{z(l)\in Z_i} |z(l)|.$$

Combining the above inequalities, we obtain

$$2^{2^{i+1}} > 2 \sum_{z(l)\in Z_i} |z(l)|,$$

i.e., we get that (3) holds for $m = i + 1$, and hence we have the induction step. The claim is proved.

Therefore, the factor v contains different numbers of occurrences of xy and yx, and hence is not a 2-palindrome.

Case 2: The factor v intersects subblocks of at most two types.

Case 2.1: The factor v is contained entirely in a block, i.e., we have $v \in F((abca)^{2^{2^i}}(abda)^{2^{2^i}}(acda)^{2^{2^i}})$ or $v \in F((adca)^{2^{2^i}}(adba)^{2^{2^i}}(acba)^{2^{2^i}})$. In this case since $|v| \geq 13$, the word contains at least one of the factors xy for $(x, y) \in \{(b, c), (c, b), (b, d), (d, b), (c, d), (d, c)\}$ and does not contain its reversal, so v is not a 2-palindrome.

Case 2.2: The factor v intersects two blocks and subblocks of at most two types. By construction, every other block in u is of exponent 1; therefore, one of the two blocks is of exponent 1. So, in this case v is a factor of $(axya)^{2^{2^m}}(ax'y'a)$ (or that of its reversal) or a factor of $(ax'y'a)^{2^{2^m}}(axya)^{2^{2^m}}(ayxa)(ay'x'a)$, $m \geq 1$, $x \neq x'$, $y \neq y'$ (or that of its reversal). In the first case since $|v| \geq 13$, the word contains at least one occurrence of the factor xy and does not contains its reversal, so v is not a 2-palindrome. In the second case the same argument works if v contains at least two copies of xy. Otherwise v is a factor of $yaaxyaayxaay'x'a$. Straightforward checking shows that this word does not contain 2-abelian palindromes of length greater or equal to 13.

So, the case $(k, |\Sigma|) = (2, 4)$ is proved.

The proofs in the cases $(k, |\Sigma|) = (3, 3)$ and $(5, 2)$ are similar, so we only provide the constructions.

An infinite recurrent $(3, 3)$-poor word can be constructed as follows:

$$V_0 = bbacc\ aabcc\ bbcaa,$$
$$V_n = V_{n-1}(bbacc)^{2^{2^n}}(aabcc)^{2^{2^n}}(bbcaa)^{2^{2^n}}V_{n-1}^R.$$

The word is given by the limit $v = \lim_{n\to\infty} V_n$:

$$v = bbacc\ aabcc\ bbcaa(bbacc)^4(aabcc)^4(bbcaa)^4aacbb\ ccbaa\ ccabb(bbacc)^{16}\ldots$$

The proof is based on the fact that each sufficiently long factor contains either unequal numbers of occurrences of factors bac and cab, or unequal numbers of occurrences of factors abc and cba, or unequal numbers of occurrences of factors bca and acb, and hence is not a 3-palindrome. In other words, two letter factors of the case $(2, 4)$ are now replaced by suitable three-letter factors over ternary alphabet.

An infinite recurrent $(5, 2)$-poor word can be constructed as follows:

$$W_0 = bbbabaaabbb\ bbbabbaabbb\ bbbabaabbbb,$$
$$W_n = W_{n-1}(bbbabaaabbb)^{2^{2^n}}(bbbabbaabbb)^{2^{2^n}}(bbbabaabbbb)^{2^{2^n}}W_{n-1}^R.$$

The word is given by the limit $w = \lim_{n\to\infty} W_n$. The proof is based on the fact that each sufficiently long factor contains either unequal numbers of occurrences of factors $abaaa$ and $aaaba$, or unequal numbers of occurrences of factors $abbaa$ and $aabba$, or unequal numbers of occurrences of factors $abaab$ and $baaba$, and hence is not a 5-palindrome. Now the specific factors of two previous cases are five-letter binary words.

II. For the proof we split the set S of pairs into two parts with different types of proofs.

Case 1: $(k, |\Sigma|) \in \{(1, l) | l \in \mathbb{N}\} \cup \{(2, 2), (3, 2)\}$. For $k = 1$ (i.e., the abelian equivalence) each word is an abelian palindrome, since every word is abelian equivalent to its reversal. Therefore, all factors of any infinite word are abelian palindromes, and hence there are no abelian palindromic poor words.

In the 2-abelian case, each word starting and ending in the same letter is a 2-abelian palindrome. Indeed, without loss of generality let a word v start and end with a, and let it contains m blocks of b's. Then v contains m occurrences of the factor ab and m occurrences of the factor ba. Factors aa and bb do not affect 2-abelian palindromicity; hence v is a 2-palindrome. Since any infinite binary word contains infinitely many factors starting and ending with the same letter, there are no 2-abelian poor binary infinite words.

In the 3-abelian binary case the proof is similar, just a bit more technical. We omit the details of the proof.

Case 2: $(k, |\Sigma|) \in \{(2,3), (4,2)\}$. We provide a detailed sketch of the proof for the $(2,3)$ case. The idea of the proof in the case $(4,2)$ is similar, although it requires more thorough analysis.

First we introduce rewriting rules which do not affect the 2-palindromicity:

(1) for $x \in \Sigma$, substitute $xx \to x$
(2) for $x, y \in \Sigma$, substitute $xyx \to x$

Claim (i). Let v be ternary word, and let v' be obtained from v by applying a rewriting rule (1) or (2). Then v is a 2-palindrome if and only if v' is a 2-palindrome.

Indeed, after applying the rewriting rule (1), the multiset (the set with multiplicities) of factors of length 2 of v' is obtained from the multiset of factors of length 2 of v by removing one factor xx. Clearly, the resulting set coincides with its reversal if and only if the original set does. After applying the rewriting rule (2), the multiset of factors of length 2 of v' is obtained from the multiset of factors of length 2 of v by removing two factors xy and yx. Again, the resulting set coincides with its reversal if and only if the original set does. The claim follows.

Now take a ternary word v and apply rewriting rules until the word does not contain factors of the form xx and xyx. We call the resulting word the *reduced form* of v. We note that the reduced form of v is unique.

The following claim is straightforward:

Claim (ii). 1. The reduced form of any ternary word v is a factor of $(abc)^\infty$ or $(cba)^\infty$. 2. A ternary word v is a 2-palindrome if and only if its reduced form is a letter.

Now assume that an infinite ternary word w with its set of factors closed under reversal does not contain 2-palindromes of length greater than N for some integer N. Take a factor $v = w_i \cdots w_{i+N}$ of length $N+1$. Since the set of factors of w is closed under reversal, there exists an occurrence of $v^R = w_j \cdots w_{j+N}$. Without loss of generality we can assume that $j > i + N$ and that the reduced form of v is a word u of the form $(abc)^m(\text{pref}(abc))$ for some $m \geq 0$. Then the reduced form of v^R equals u^R. Now consider the factor $w_{i+N+1} \cdots w_{j-1}$; again without loss of generality its reduced form is of the form $(\text{suff}(cba))(cba)^r(\text{pref}(cba))$ for some $r \geq 0$. Consider the factor $w_{i+N+1} \cdots w_{j+N+1}$, which was rewritten to $(\text{suff}(cba))(cba)^r(\text{pref}(cba))(\text{suff}(cba))(cba)^m$. So, there exists s, $i + N + 1 \leq s \leq j + N + 1$, such that the reduced form of $w_{i+N+1} \cdots w_s$ is equal to $u^R = (\text{suff}(cba))(cba)^m$. It is straightforward that $w_i \cdots w_s$ is reduced to a, and hence is a 2-palindrome. The length of this 2-palindrome is greater than N, a contradiction. □

Remark 1. Notice that the examples of k-abelian poor words we build are recurrent, but not uniformly recurrent.

Remark 2. In the construction (1) in fact the powers can be made smaller (although growing), it is convenient for us to use 2^{2^n} in the proofs.

Remark 3. Our constructions are modifications of the so-called sesquipowers, see, e.g., [20, Chapter 4].

4 *k*-Abelian Rich Words

In this section we show that there exist words of length n which have the number of inequivalent k-abelian palindromic factors of the same order as the total number of their factors $\Theta(n^2)$. In this sense these words contain "many" k-palindromes and hence are considered as rich.

Theorem 4.1. *Let k be a natural number. There exists a positive constant C such that for each $n \geq k$ there exists a word of length n containing at least Cn^2 k-abelian palindromes. Actually, we can choose $C = 1/4k$.*

Proof. The word is defined by

$$v = a^l(ba^{k-1})^m,$$

where l and m are chosen to give maximal number of k-palindromes among words of this type. We let $\lfloor r \rceil$ denote the closest integer to r, we can take $m = \lfloor \frac{n-k+1}{2k} \rceil$. Let us count the number of inequivalent k-palindromes in the word $v = v_1 \cdots v_n$, $n = km + l$. The k-palindromes are the following:

- Starting from position 1, we get the following k-palindromes
  - ε (empty word)
  - v_1, v_1v_2, ..., $v_1 \cdots v_l$ (l k-palindromes consisting of only a's)
  - $v_1 \dots v_{l+k}$, $v_1 \dots v_{l+2k}$, ..., $v_1 \dots v_{l+mk}$ (m k-palindromes starting with a^{k-1}, of length $l + ik$ and containing i letters b, $i = 1, \dots, m$)
- Starting from each position j, $j = 2, \dots, l - k + 1$, we get the following new k-palindromes: $v_j \dots v_{l+k}$, $v_j \dots v_{l+2k}$, $\dots v_j \dots v_{l+mk}$ (m k-palindromes starting with a^{k-1}, of length $l-j+ik$ and containing i letters b, $i = 1, \dots, m$)
- Starting from each position j, $j = l - k + 2, \dots, l + 1$, we get the following new k-palindromes: $v_j \dots v_{2l-j+2}$, $v_j \dots v_{2l-j+2+k}$, ..., $v_j \dots v_{2l-j+2+(m-1)k}$ (m k-palindromes starting with a^{l+1-j}, of length $2l - 2j + 3 + (i - 1)k$ and containing i letters b, $i = 1, \dots, m$)

It is not hard to see that all the above k-palindromes are distinct up to k-abelian equivalence; in fact, they are abelian inequivalent. So, in total we have $(l + 1)(m + 1) = (n - mk + 1)(m + 1)$ distinct k-palindromes. Considering this as a function of m, we get that this function takes a maximal value when $m = \frac{n-k+1}{2k}$. Since all numbers are integer there, the actual maximal number of k-palindromes given by this construction is given by taking the closest integer value, i.e., $m = \lfloor \frac{n-k+1}{2k} \rceil$ (since the function is quadratic in m). Taking these values and taking into account the condition $n \geq k$, we derive that the number of k-palindromes is $(l + 1)(m + 1) \geq n^2/4k$. □

We remark that in the $\Theta(n^2)$ number of k-palindromic factors the constant actually depends on k, so it makes sense when k is small relatively to n.

5 Conclusions and Open Problems

We have considered the numbers of k-abelian palindromes in finite and infinite words. These numbers are always between a constant and a quadratic bound, corresponding to so-called poor and rich words. In the case of poor words to avoid trivialities we always assumed that the words are closed under reversal. Our main result was a construction of infinite words containing only finitely many k-abelian palindromes. This construction could be modified for different pairs (k, l), where k was a constant in k-abelian equivalence and l was the size of the alphabet. For the remaining pairs to show that such an infinite poor word does not exist, we used a different approach, based on rewriting rules preserving k-abelian palindromicity. We also gave an example showing the existence of rich finite words, that is words containing the maximal number of k-abelian palindromes up to a constant multiplicative factor. The bound we found is $n^2/4k$, that is of order Cn^2, where C is a constant independent of n.

A few natural open problems remain. We built recurrent, but not uniformly recurrent k-abelian poor words. The problem is the following:

Problem 1. Does there exist an infinite uniformly recurrent word having the set of factors that is closed under reversal and containing only finitely many k-abelian palindromes?

The second problem asks for optimal constants for rich and poor words.

Problem 2. What is the exact minimal number of k-abelian palindromes an infinite word having a set of factors closed under reversal can contain? What is the exact maximal number of distinct k-abelian palindromes a word of length n can contain?

Some bounds for the constants can be found in this paper, although we did not try to find the best constants.

In the case of equality and classical palindromes the notion of a rich word can be extended to infinite word. The question is whether this is possible for k-abelian palindromes:

Problem 3. Does there exist an infinite k-abelian rich word? More precisely, does there exist an infinite word w, such that for some constant C each of its factors of length n contains at least Cn^2 distinct k-abelian palindromes?

References

1. Adamczewski, B.: Balances for fixed points of primitive substitutions. Theor. Comput. Sci. 307(1), 47–75 (2003)
2. Ambrož, P., Frougny, C., Masáková, Z., Pelantová, E.: Palindromic complexity of infinite words associated with simple Parry numbers. Ann. Inst. Fourier (Grenoble) 56, 2131–2160 (2006)
3. Avgustinovich, S., Karhumäki, J., Puzynina, S.: On abelian versions of Critical Factorization Theorem. RAIRO - Theoret. Inf. Appl. 46, 3–15 (2012)

4. Avgustinovich, S., Puzynina, S.: Weak Abelian Periodicity of Infinite Words. In: Bulatov, A.A., Shur, A.M. (eds.) CSR 2013. LNCS, vol. 7913, pp. 258–270. Springer, Heidelberg (2013)
5. Berstel, J., Boasson, L., Carton, O., Fagnot, I.: Infinite words without palindrome, Arxiv: 0903.2382 (2009)
6. Brlek, S., Hamel, S., Nivat, M., Reutenauer, C.: On The Palindromic Complexity of Infinite Words. Internat. J. Found. Comput. Sci. 15, 293–306 (2004)
7. Bucci, M., De Luca, A., Glen, A., Zamboni, L.Q.: A new characteristic property of rich words. Theoret. Comput. Sci. 410, 2860–2863 (2009)
8. Crochemore, M., Ilie, L., Tinta, L.: The "runs" conjecture. Theor. Comput. Sci. 412(27), 2931–2941 (2011)
9. de Luca, A.: Sturmian words: structures, combinatorics and their arithmetics. Theoret. Comput. Sci. 183, 45–82 (1997)
10. Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy. Theoret. Comput. Sci. 255, 539–553 (2001)
11. Fici, G., Zamboni, L.Q.: On the least number of palindromes contained in an infinite word. Theor. Comput. Sci. 481, 1–8 (2013)
12. Glen, A., Justin, J., Widmer, S., Zamboni, L.Q.: Palindromic richness. European J. Combin. 30, 510–531 (2009)
13. Harju, T., Kärki, T., Nowotka, D.: The number of positions starting a square in binary words. Electron. J. Combin. 18(1) (2011)
14. Ilie, L.: A note on the number of squares in a word. Theor. Comput. Sci. 380(3), 373–376 (2007)
15. Karhumäki, J., Puzynina, S., Saarela, A.: Fine and Wilf's Theorem for k-Abelian Periods. Int. J. Found. Comput. Sci. 24(7), 1135–1152 (2013)
16. Karhumäki, J., Saarela, A., Zamboni, L.Q.: On a generalization of Abelian equivalence and complexity of infinite words. J. Comb. Theory, Ser. A 120(8), 2189–2206 (2013)
17. Keränen, V.: Abelian squares are avoidable on 4 letters. In: Kuich, W. (ed.) ICALP 1992. LNCS, vol. 623, pp. 41–52. Springer, Heidelberg (1992)
18. Kolpakov, R., Kucherov, G.: Finding Maximal Repetitions in a Word in Linear Time. In: FOCS 1999, pp. 596–604 (1999)
19. Kucherov, G., Ochem, P., Rao, M.: How Many Square Occurrences Must a Binary Sequence Contain? Electr. J. Comb. 10 (2003)
20. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press, Cambridge (2002)
21. Puzynina, S., Zamboni, L.Q.: Abelian returns in Sturmian words. J. Comb. Theory, Ser. A 120(2), 390–408 (2013)
22. Richomme, G., Saari, K., Zamboni, L.Q.: Abelian Complexity of Minimal Subshifts. J. London Math. Soc. 83, 79–95 (2011)
23. Rigo, M., Salimov, P.: Another Generalization of Abelian Equivalence: Binomial Complexity of Infinite Words. In: Karhumäki, J., Lepistö, A., Zamboni, L. (eds.) WORDS 2013. LNCS, vol. 8079, pp. 217–228. Springer, Heidelberg (2013)