On minimal factorizations of words as products of palindromes

A. Frid, S. Puzynina, L. Zamboni

June 17, 2012

Abstract

Given a finite word $u$, we define its palindromic length $|u|_{\text{pal}}$ to be the least number $n$ such that $u = v_1v_2\ldots v_n$ with each $v_i$ a palindrome. As each letter is a palindrome we have $|u|_{\text{pal}} \leq |u|$. Let $k$ be a positive integer. In this note we prove that if an infinite word $w$ is $k$-power free, then for each positive integer $n$ there exists a factor $u$ of $w$ whose palindromic length $|u|_{\text{pal}} > n$. We also obtain an analogous result in the context of privileged words recently introduced by Kellendonk, Lenz and Savinien.

Let $A$ be a finite non-empty set, and let $A^+$ denote the set of all finite non-empty words in $A$. A word $v = v_0v_1\ldots v_n \in A^+$ is called a palindrome if $v_i = v_{n-i}$ for each $i = 0, \ldots, n$. In particular each $a \in A$ is a palindrome. We also regard the empty word as a palindrome. For each finite word $u \in A^+$ we define its palindromic length, denoted $|u|_{\text{pal}}$ to be the least number $n$ such that $u = v_1v_2\ldots v_n$ with each $v_i$ a palindrome. Clearly $|u|_{\text{pal}} \leq |u|$ where $|u|$ denotes the length of $u$. For example, $|010010011010100|_{\text{pal}} = 1$ while $|0100111|_{\text{pal}} = 3$. Note that $0100111$ may be expressed as a product of 3 palindromes in two different ways: $(0)(1001)(1)$ and $(010)(0)(11)$. In [2], O. Ravsky obtains an intriguing formula for the supremum of the palindromic lengths of all binary words of length $n$.

Recently, J. Kellendonk, D. Lenz and J. Savinien [1] introduced a new class of words they call privileged words. Privileged words are defined recursively: First, the empty word and each element $a \in A$ is privileged. Next, a word $u \in A^+$ we define its palindromic length, denoted $|u|_{\text{pal}}$ to be the least number $n$ such that $u = v_1v_2\ldots v_n$ with each $v_i$ a palindrome. Clearly $|u|_{\text{pal}} \leq |u|$ where $|u|$ denotes the length of $u$. For example, $|0010110011010100|_{\text{pal}} = 1$ while $|0010111|_{\text{pal}} = 3$. Note that $0010111$ may be expressed as a product of 3 palindromes in two different ways: $(0)(010110011010)(0)$ or $(00)(1011001101)(00)$. We note that $00101100110100$ may be written as a product of 3 privileged words in more than one way: $(0)(010110011010)(0)$ or $(00)(1011001101)(00)$.

$^*$Corresponding author, Sobolev Institute of Mathematics SB RAS; supported in part by Presidential grant MK-4075.2012.1 and by RFBR grant 12-01-00089.

$^\dagger$Sobolev Institute of Mathematics SB RAS and University of Turku; supported in part by Academy of Finland (grant 251371) and by RFBR (grants 10-01-00424 and 12-01-00448)

$^\ddagger$Université Lyon I and University of Turku, supported in part by a FiDiPro grant from the Academy of Finland.
Let \( k \) be a positive integer. A word \( v \in A^+ \) is called a \( k \)-power if \( v = u^k \) for some word \( u \in A^+ \). An infinite word \( w = w_0 w_1 w_2 \ldots \in \mathbb{A}^\mathbb{N} \) is said to be \( k \)-power-free if no factor \( u \) of \( w \) is a \( k \)-power. For instance, the Thue-Morse word 011010011010110100110110010110... fixed by the morphism \( 0 \rightarrow 01, 1 \rightarrow 10 \) is 3-power free.

**Theorem 1** Let \( k \) be a positive integer and \( w = w_0 w_1 w_2 \ldots \in \mathbb{A}^\mathbb{N} \). If \( w \) is \( k \)-power free, then for each positive integer \( n \) there exist prefixes \( u \) and \( v \) of \( w \) with \( |u|_{\text{pal}} > n \) and \( |v|_{\text{priv}} > n \).

Given an infinite word \( w = w_0 w_1 w_2 \ldots \in \mathbb{A}^\mathbb{N} \) we set \( w[i..j] = w_i w_{i+1} \ldots w_j \). The proof of Theorem 1 will make use of the following lemmas:

**Lemma 2** Let \( u \) be a palindrome (respectively privileged). Then for every palindromic (respectively privileged) proper prefix \( v \) of \( u \), we have that \( u \) is \((|u| - |v|)\)-periodic.

**Proof.** First suppose that \( u \) and \( v \) are palindromes with \( v \) a proper prefix of \( u \). Then \( v \) is also a suffix of \( u \) and hence \( u \) is \((|u| - |v|)\)-periodic. Next suppose \( u \) and \( v \) are privileged words with \( v \) a proper prefix of \( u \). We will prove that \( v \) is also a suffix of \( u \). We proceed by induction on \(|u|\). The result is vacuously true for \(|u| = 1 \). Next suppose \(|u| > 1 \). Then \( u \) is a complete first return to a privileged word \( u' \) with \(|u'| < |u| \). We claim that \(|v| \leq |u'| \). In fact, suppose to the contrary that \(|v| > |u'| \). Then \( u' \) would be a proper prefix of \( v \) and hence by induction hypothesis \( u' \) is also a suffix of \( v \). This means that \( u' \) occurs at least three times within \( u \) (as a prefix of \( v \), as a suffix of \( v \), and as a suffix of \( u \)). This contradicts that \( u \) is a complete first return to \( u' \). Having established that \(|v| \leq |u'| \), it follows that \( v \) is a suffix of \( u' \). In fact, if \(|v| = |u'| \), then \( v = u' \) while if \(|v| < |u'| \), then by induction hypothesis \( v \) is a suffix of \( u' \). As \( u' \) is a suffix of \( u \) we obtain that \( v \) is a suffix of \( u \) as required. Whence, \( u \) is \((|u| - |v|)\)-periodic. \(\Box\)

**Lemma 3** Suppose the infinite word \( w \) is \( k \)-power-free. If \( w[i_1..i_2] \) and \( w[i_1..i_3] \) are palindromes (respectively privileged) with \( i_3 > i_2 \), then

\[
\frac{|w[i_1..i_3]|}{|w[i_1..i_2]|} \geq 1 + \frac{1}{k-1}.
\]

**Proof.** By Lemma 2, the word \( w[i_1..i_3] \) is \((i_3 - i_2)\)-periodic; at the same time, it cannot contain a \( k \)-power, so, \(|w[i_1..i_3]| < k(i_3 - i_2)\). Thus,

\[
\frac{|w[i_1..i_3]|}{|w[i_1..i_2]|} = \frac{|w[i_1..i_3]|}{|w[i_1..i_3]| - (i_3 - i_2)} > \frac{|w[i_1..i_3]|}{(1 - \frac{1}{k}) |w[i_1..i_3]|} = 1 + \frac{1}{k-1}.
\]

\(\Box\)

**Lemma 4** Let \( N \) be a positive integer. Then for each \( i \geq 0 \), the number of palindromes (respectively privileged words) of the form \( w[i..j] \) of length less or equal to \( N \) is at most \( 2 + \log_{k/(k-1)} N \).

**Proof.** For each \( i \geq 0 \), the length of the shortest non-empty palindrome (respectively privileged word) beginning in position \( i \) is equal to 1. By the previous lemma, the next palindrome (respectively privileged word) beginning in position \( i \) is of length at least \( \frac{k}{k-1} \), and the one after that is of length at least \( \left(\frac{k}{k-1}\right)^2 \), and so on. The longest one is of length at least \( \left(\frac{k}{k-1}\right)^n \leq N \), so that \( n \leq \log_{k/(k-1)} N \), and the total number \( n + 1 \) of such words is at most \( 1 + \log_{k/(k-1)} N \). Adding the empty word which is a palindrome gives the desired result. \(\Box\)

**Proof of Theorem 1.** Fix a positive integer \( n \) and let \( N \) be a positive integer satisfying

\[
(2 + \log_{k/(k-1)} N)^n < N.
\]
By the previous lemma, the number of prefixes of $w$ of the form $v_1v_2\ldots v_n$ where each $v_i$ is a palindrome (respectively a privileged word) of length less or equal to $N$ is at most $(2 + \log_2(k/(k-1))N)^n$, and hence at most $N$. But $w$ has $N$-many non-empty prefixes of length less or equal to $N$. This means that there exist prefixes $u$ and $v$ of length less or equal to $N$ such that $|u|_{\text{pal}} > n$ and $|v|_{\text{priv}} > n$.

As an immediate corollary we have:

**Corollary 1** Fix a positive integer $n$. For any Sturmian word $w$ whose slope has bounded partial quotients, there exist prefixes $u$ and $v$ of $w$ such that $|u|_{\text{pal}} > n$ and $|v|_{\text{priv}} > n$.

In fact, we are able to extend the result of Theorem 1 to a wider class of words, including the Sierpinski word. However the method of proof is considerably more involved.

Thus we propose the following open problem:

**Question 5** Does there exist an aperiodic word $w$ and a positive integer $n$ such that $|u|_{\text{pal}} \leq n$ (respectively $|u|_{\text{priv}} \leq n$) for each factor $u$ of $w$?

Even if the answer to the above question turns out to be false, this does not give a characterization of aperiodic words. In fact, some ultimately periodic words contain factors having arbitrarily large palindromic lengths, for example, $w = (110100)^\omega$.

We prove the following property of ultimately periodic words with a uniform bound on the palindromic length of its factors:

**Proposition 6** Let $n$ be an integer, $w$ an ultimately periodic word such that $|u|_{\text{pal}} \leq n$ for each factor $u$ of $w$. Then $w$ has a tail $w'$ of the form $w' = v(p_1p_2)^\omega$, where $p_1$ and $p_2$ are palindromes.

**Proof.** Let $w'$ be a tail of $w$ having period $t$. Consider a factor $u$ of $w'$ with $|u| > tn$. Then $u$ can be factored as $u = v_1v_2\ldots v_m$ with $m \leq n$ and each $v_i$ a non-empty palindrome. At least one of the palindromes $v_i$ in this factorization has length greater than $t$. So, $t$ is a period of this long palindrome. Now the proposition follows from the well-known fact that a period of a palindrome has the form $p_1p_2$, where $p_1$ and $p_2$ are palindromes.

Proposition 6 implies that if the answer to Question 5 is “no”, then an infinite word $w$ having with a uniform bound on the palindromic length of its factors is ultimately periodic, and moreover its period has the form $p_1p_2$, where $p_1$ and $p_2$ are palindromes.

The authors are grateful to Michelangelo Bucci and Alessandro de Luca for fruitful discussions.

**References**
