Palindromic factorization of rich words

Josef Rukavicka*

October 25, 2021
Mathematics Subject Classification: 68R15

Abstract

A finite word $w$ is called rich if it contains $|w| + 1$ distinct palindromic factors including the empty word. For every finite rich word $w$ there are distinct nonempty palindromes $w_1, w_2, \ldots, w_p$ such that $w = w_p w_{p-1} \cdots w_1$ and $w_i$ is the longest palindromic suffix of $w_p w_{p-1} \cdots w_i$, where $1 \leq i \leq p$. This palindromic factorization is called UPS-factorization. Let $\text{luf}(w) = p$ be the length of UPS-factorization of $w$.

In 2017, it was proved that there is a constant $c$ such that if $w$ is a finite rich word and $n = |w|$ then $\text{luf}(w) \leq c \frac{n}{\ln n}$. We improve this result as follows: There are constants $\mu, \pi$ such that if $w$ is a finite rich word and $n = |w|$ then

$$\text{luf}(w) \leq \mu \frac{n}{e^{\pi \sqrt{\ln n}}}.$$ 

The constants $c, \mu, \pi$ depend on the size of the alphabet.

Keywords: Rich words; Palindromic factorization; Palindromic length;

1 Introduction

A palindrome is a word that is equal to its reversal; for example “radar” or “noon”. A word $w$ is called rich if it contains $|w| + 1$ distinct palindromic
factors including the empty word. It is known that a word $w$ cannot have more than $|w| + 1$ distinct palindromic factors \cite{5}, thus less formally said, rich words are those words that contain the maximal number of distinct palindromic factors. The rich words have attracted quite a lot of attention over the last couple of decades; see, for instance, \cite{1, 4, 5, 7, 10, 12}. A well known example of rich words are Sturmian words \cite{2}.

In \cite{5, Definition 4 and Proposition 3} it was proved that the longest palindromic suffix of a rich word $w$ has exactly one occurrence in $w$. Using this property it was proved that:

**Lemma 1.1.** (see \cite{9, Lemma 1}) Let $w$ be a finite nonempty rich word. There exist distinct nonempty palindromes $w_1, w_2, \ldots, w_p$ such that

$$w = w_pw_{p-1} \cdots w_1$$

and $w_i$ is the longest palindromic suffix of $w_pw_{p-1} \cdots w_i$ for $i = 1, 2, \ldots, p$.

Based on the Lemma 1.1 the so-called *UPS-factorization* was introduced. In addition, for the current article, we define also the function $\text{luf}(w)$ (the length of UPS-factorization).

**Definition 1.2.** (see \cite{9, Definition 2}) We define UPS-factorization (Unioccurrent Palindromic Suffix factorization) to be the factorization of a finite nonempty rich word $w$ into the form $w = w_pw_{p-1} \cdots w_1$ from Lemma 1.1. Let $\text{luf}(w) = p$ be the length of UPS-factorization of the rich word $w$.

Let $R(n)$ denote the number of rich words of length $n$. In \cite{9}, the UPS-factorization has been used to prove that the function $R(n)$ grows subexponentially; formally

$$\lim_{n \to \infty} \left( \frac{R(n)}{n} \right)^{\frac{1}{n}} = 1. \quad (1)$$

It was proved in \cite{9, Theorem 3}) that there is a positive real constant $c$ such that for every finite rich word $w$ with $n = |w|$ we have that

$$\text{luf}(w) \leq c \frac{n}{\ln n}. \quad (2)$$

This can be interpreted as follows: A rich word $w$ is a concatenation of “small” number of palindromes, where “small” means that the length of UPS-factorization with regard to the length of the word $w$ tends to zero; formally we have that

$$\lim_{n \to \infty} \frac{c \frac{n}{\ln n}}{n} = \lim_{n \to \infty} \frac{c}{\ln n} = 0. \quad (3)$$
Since every palindrome is uniquely determined by its first half, the property (3) allowed in consequence to prove the subexponential growth of the number of rich words.

The formula (1) presents the currently best known upper bound on the number of rich words. In [8] it was conjectured for rich words on the binary alphabet that for some infinitely growing function $g(n)$ the following holds true

$$R(n) = O\left(\frac{n}{g(n)}\right)^{\sqrt{n}}.$$

Next to UPS-factorization, another palindromic factorizations have been studied; see, for instance, [3, 6]. A palindromic length $PL(v)$ of the finite word $v$ is the minimal number of palindromes whose concatenation is equal to $v$. It was conjectured in 2013 that for every aperiodic infinite word $x$, the palindromic length of its factors is not bounded [6]. Despite a considerable effort the conjecture remains open.

For every rich word $w$ we have obviously that $PL(w) \leq \text{luf}(w)$. Thus the result of the current article presents also an upper bound on the palindromic length of rich words. Note that, in general, for non-rich words, UPS-factorization does not need to exist.

The main result of the current article is the following theorem.

**Theorem 1.3.** For a given finite alphabet $A$, there are real positive constants $\mu, \pi$ such that if $w$ is a finite nonempty rich word over the alphabet $A$ and $n = |w|$ then

$$\text{luf}(w) \leq \mu \frac{n}{e^{\pi \sqrt{\ln n}}}.$$

**Remark 1.4.** Note that the constants $\mu, \pi$ in Theorem 1.3 and the constant $c$ in (2) depend on the size of the alphabet $A$.

To prove Theorem 1.3 we use principally the same ideas like in [9, Theorem 3]). The distinction is that in [9, Theorem 3]) we considered all palindromes of given length factors of the rich word in question. In the current article we apply the upper bound for the factor complexity of rich words from [11]. This allows the improvement of the upper bound (2).

Based on our research that we performed we present the following conjecture.

**Conjecture 1.5.** For a given finite alphabet $A$, there is a positive real constant $\lambda$ such that if $w$ is a finite nonempty rich word over the alphabet $A$ and $n = |w|$ then $\text{luf}(w) \leq \lambda \sqrt{n}$. 
2 Preliminaries

Let \( \mathbb{N}_1 \) denote the set of all positive integers, let \( \mathbb{R} \) denote the set of all real numbers, and let \( \mathbb{R}^+ \) denote the set of all positive real numbers.

Let \( \mathcal{A} \) denote a finite alphabet and let \( |\mathcal{A}| \) denote the size of the alphabet. Let \( \mathcal{A}^n \) denote the set of all words of length \( n \in \mathbb{N}_1 \) over the alphabet \( \mathcal{A} \). Let \( \mathcal{A}^* = \bigcup_{i \geq 1} \mathcal{A}^i \), let \( \epsilon \) denote the empty word, and let \( \mathcal{A}^\ast = \{ \epsilon \} \cup \mathcal{A}^* \). Let \( \mathcal{A}^\infty \) denote the set of all infinite words over the alphabet \( \mathcal{A} \); formally \( \mathcal{A}^\infty = \{ a_1 a_2 \cdots \mid a_i \in \mathcal{A} \text{ and } i \in \mathbb{N}_1 \} \).

Let \( \text{Pal} \subseteq \mathcal{A}^* \) denote the set of all palindromes and let \( \text{RW} \subseteq \mathcal{A}^* \) denote the set of all rich words. We have that \( \epsilon \in \text{Pal} \cap \text{RW} \).

Let \( \text{Fac}_w \subseteq \mathcal{A}^* \) be the set of all finite factors of the word \( w \in \mathcal{A}^* \cup \mathcal{A}^\infty \). Let \( \text{Fac}_w(n) = |\text{Fac}_w \cap \mathcal{A}^n| \). The function \( \text{Fac}_w(n) \) is the factor complexity of the word \( w \).

An infinite word \( w \in \mathcal{A}^\infty \) is rich if \( \text{Fac}_w \subseteq \text{RW} \); it means that all finite factors of \( w \) are rich. Let \( \text{RW}^\infty \subseteq \mathcal{A}^\infty \) denote the set of all infinite rich words.

3 Factor complexity

Let \( \delta = \frac{3}{2(\ln 3 - \ln 2)} \). Recall that \( q \) denote the size of the alphabet \( \mathcal{A} \). There is a subexponential upper bound on the factor complexity of rich words:

**Theorem 3.1.** (see [11, Theorem 1.1]) If \( w \in \text{RW} \cup \text{RW}^\infty \) and \( n \in \mathbb{N}_1 \) then

\[
|\text{Fac}_w(n)| \leq (4q^2n)^{\delta \ln 2 + 2}.
\]

**Remark 3.2.** In [11, Theorem 1.1], there is a condition that \( q > 1 \). But the theorem obviously holds also for \( q = 1 \), because in such a case we have that \( \text{Fac}_w(n) = 1 \) for every \( n \in \mathbb{N}_1 \). That is why we omit this condition in Theorem 3.1.

Let \( c_2 = \delta (\ln 4 + 2 \ln q) + \delta + \delta \ln 2 + 2 \). Let \( g \in \mathbb{R}^+ \) and let \( g < 1 \). Let \( \alpha = \sqrt{\frac{1}{c_2}} \) and let \( \beta = \frac{1 - g}{2c_2} \). Let \( c_1 \) be the minimal real constant such that \( c_1 \geq e^{(\delta \ln 2 + 2)(\ln 4 + 2 \ln q)} \) and \( c_1 e^{\beta + g \beta + c_2 \beta^2} \geq 1 \). It means that

\[
c_1 = \min \{ e^{(\delta \ln 2 + 2)(\ln 4 + 2 \ln q)}, e^{-(\beta + g \beta + c_2 \beta^2)} \}.
\]

**Remark 3.3.** The constant \( 0 < g < 1 \) can be choosen arbitrarily.

The constants \( g, \alpha, \) and \( \beta \) will be used in next sections. For Lemma 4.4 it is necessary that \( c_1 e^{\beta + g \beta + c_2 \beta^2} \geq 1 \). That is why we introduced these constants already in this section, where \( c_1 \) is defined.
The merit of the next corollary is to have the upper bound for factor complexity of rich words in a simple form.

**Corollary 3.4.** If \( w \in RW \cup RW^\infty \) and \( n \in \mathbb{N} \) we have that

\[
|\text{Fac}_w(n)| \leq c_1 n^{c_2 \ln n}.
\]

**Proof.** We have that

\[
(4q^2n)^{\delta \ln 2n+2} = e^{(\delta \ln n + \delta \ln 2+2)} = e^{\delta \ln n(\ln 4+2 \ln q) + \delta \ln n^2 + (\delta \ln 2+2)(\ln 4+2 \ln q) + (\delta \ln 2+2) \ln n} \leq e^{(\delta \ln 2+2)(\ln 4+2 \ln q)} e^{(\delta \ln 4+2 \ln q) + \delta \ln 2+2) \ln n}.
\]

Then the lemma follows from Theorem 3.1. This ends the proof. \( \square \)

## 4 UPS-length

The next technical lemma shows a lower bound on the sum \( \sum_{i=1}^{k} i \lfloor c_1 i^{c_2 \ln i} \rfloor \) using the constant \( g \). The sum \( \sum_{i=1}^{k} i \lfloor c_1 i^{c_2 \ln i} \rfloor \) has the following interpretation: It is the length of the word \( w \) which is concatenation of \( \lfloor c_1 i^{c_2 \ln i} \rfloor \) words of length \( i \) for \( i \in \{1, 2, \ldots, k\} \).

**Lemma 4.1.** If \( k \in \mathbb{N}_1 \) then

\[
\sum_{i=1}^{k} i \lfloor c_1 i^{c_2 \ln i} \rfloor \geq (k^g - 1)(k - k^g)c_1(k - k^g)^{c_2 \ln (k-k^g)} - \frac{k(k+1)}{2}.
\]

**Proof.** We have that

\[
\sum_{i=1}^{k} i \lfloor c_1 i^{c_2 \ln i} \rfloor \geq \sum_{i=1}^{k} i (c_1 i^{c_2 \ln i} - 1) = \sum_{i=1}^{k} ic_1 i^{c_2 \ln i} - \sum_{i=1}^{k} i =
\]

\[
\sum_{i=1}^{[k-k^g]} ic_1 i^{c_2 \ln i} + \sum_{i=[k-k^g]+1}^{k} ic_1 i^{c_2 \ln i} - \sum_{i=1}^{k} i \geq
\]

\[
\sum_{i=[k-k^g]+1}^{k} ic_1 i^{c_2 \ln i} - \sum_{i=1}^{k} i \geq [k^g](k-k^g)c_1(k-k^g)^{c_2 \ln (k-k^g)} - \sum_{i=1}^{k} i \geq
\]

\[
(k^g - 1)(k - k^g)c_1(k - k^g)^{c_2 \ln (k-k^g)} - \frac{k(k+1)}{2}.
\]
This completes the proof.

For the sum $\sum_{i=1}^{k} i [c_1 i^2 \ln i]$ we consider $k$ to be a function of $n \in \mathbb{N}_1$. Then for the lower bound from Lemma 4.1 we show a limit behavior as $n$ tends to infinity.

**Proposition 4.2.** If $n \in \mathbb{N}_1$, $\sigma : \mathbb{N}_1 \to \mathbb{N}_1$, $\lim_{n \to \infty} \sigma(n) = \infty$, and $k_n = e^{\sigma(n)}$ then

$$\lim_{n \to \infty} \frac{(k_n^g - 1)(k_n - k_n^g)c_1(k_n - k_n^g)c_2 \ln (k_n - k_n^g) - \frac{k_n(k_n+1)}{2}}{e^{(1+g)\sigma(n)}c_1e^{c_2(\sigma(n))^2}} = 1.$$  

**Proof.** Let $\tilde{\sigma}(n) = 1 - e^{(g-1)\sigma(n)}$. We have that

$$k_n - k_n^g = e^{\sigma(n)} - e^{g\sigma(n)} = e^{\sigma(n)}(1 - e^{(g-1)\sigma(n)}) = e^{\sigma(n)}\tilde{\sigma}(n).$$

It follows that

$$\begin{align*}
(k_n^g - 1)(k_n - k_n^g)c_1(k_n - k_n^g)c_2 \ln (k_n - k_n^g) & = \\
(e^{g\sigma(n)} - 1)e^{\sigma(n)}\tilde{\sigma}(n)c_1(e^{\sigma(n)}\tilde{\sigma}(n))c_2 \ln (e^{\sigma(n)}\tilde{\sigma}(n)) & = \\
(e^{g\sigma(n)} - 1)e^{\sigma(n)}\tilde{\sigma}(n)c_1(e^{\sigma(n)}\tilde{\sigma}(n))c_2(\sigma(n)+\ln \tilde{\sigma}(n)) & = \\
(e^{g\sigma(n)} - 1)e^{\sigma(n)}\tilde{\sigma}(n)c_1 e^{c_2(\sigma(n))^2}(\tilde{\sigma}(n))c_2\sigma(n) e^{c_2\sigma(n) \ln \tilde{\sigma}(n)}(\tilde{\sigma}(n))c_2\ln \tilde{\sigma}(n).
\end{align*}$$

We have that:

- $\lim_{n \to \infty} \tilde{\sigma}(n) = \lim_{n \to \infty} (1 - e^{(g-1)\sigma(n)}) = 1 - 0 = 1.$  

- $\lim_{n \to \infty} (\tilde{\sigma}(n))^{\sigma(n)} = \lim_{n \to \infty} (1 - e^{(g-1)\sigma(n)})^{\sigma(n)} = \\
\lim_{n \to \infty} \left(1 - \frac{1}{e^{(1-g)\sigma(n)}}\right)^{\sigma(n)} = 1.$  

- From (6) it follows that

$$\lim_{n \to \infty} \sigma(n) \ln \tilde{\sigma}(n) = \lim_{n \to \infty} \ln (\tilde{\sigma}(n))^{\sigma(n)} = \ln 1 = 0.$$  

6
\[
\lim_{n \to \infty} \left( \tilde{\sigma}(n)^{\ln \tilde{\sigma}(n)} \right) = \lim_{n \to \infty} \left( 1 - e^{(\ln 2) \sigma(n)} \right)^{\ln \sigma(n)} = \lim_{n \to \infty} \left( 1 - \frac{1}{(e^{1-g})^{\sigma(n)}} \right)^{\ln \sigma(n)} = 1. 
\]

(8)

\[
\lim_{n \to \infty} \frac{e^{\sigma(n)} - 1}{e^{\sigma(n)}} = 1. 
\]

(9)

From (5), (6), (7), (8), and (9) it follows that

\[
\lim_{n \to \infty} \frac{(e^{\sigma(n)} - 1)e^{\sigma(n)} \tilde{\sigma}(n)c_1e^{c_2(\sigma(n))^2}(\tilde{\sigma}(n))c_2\sigma(n)\ln \tilde{\sigma}(n)(\tilde{\sigma}(n))c_2\ln \tilde{\sigma}(n)}{e^{(1+g)\sigma(n)}c_1e^{c_2(\sigma(n))^2}} = 1. 
\]

(10)

Obviously

\[
\lim_{n \to \infty} \frac{k_n(k_n+1)}{2} = \lim_{n \to \infty} \frac{e^{2\sigma(n)} + e^{\sigma(n)}}{2e^{(1+g)\sigma(n)}c_1e^{c_2(\sigma(n))^2}} = 0. 
\]

(11)

The proposition follows from (4), (10), and (11). This ends the proof.

For \( \sigma(n) = \alpha \sqrt{\ln n} + \beta \) we show that the function \( e^{(1+g)\sigma(n)}e^{c_2(\sigma(n))^2} \) used in Proposition 4.2 is a multiple of \( n \) by a constant.

**Lemma 4.3.** If \( \sigma(n) = \alpha \sqrt{\ln n} + \beta \) and \( n \in \mathbb{N}_1 \) then

\[
e^{(1+g)\sigma(n)}e^{c_2(\sigma(n))^2} = e^{\beta + g\beta + c_2\beta^2}. 
\]

**Proof.** We have that

\[
e^{(1+g)\sigma(n)}e^{c_2(\sigma(n))^2} = e^{(1+g)(\alpha \sqrt{\ln n} + \beta) + c_2(\alpha \sqrt{\ln n} + \beta)^2 - \ln n} = e^{(c_2 \alpha^2 - 1) \ln n + \alpha(1 + g + 2c_2\beta)\sqrt{\ln n} + \beta + g\beta + c_2\beta^2} = (c_2 \left( \sqrt{\frac{1}{c_2}} \right)^2 - 1) \ln n + \alpha(1 + g + 2c_2\beta)\sqrt{\ln n} + \beta + g\beta + c_2\beta^2 = \beta + g\beta + c_2\beta^2. 
\]

(12)
The proposition follows from (12) and (13). This ends the proof.

Using Lemma 4.1, Lemma 4.3, and Proposition 4.2 we show that the sum $\sum_{i=1}^{k_n} i \lfloor c_1 i^{c_2 \ln i} \rfloor$ is bigger than $n$ as $n$ tends to infinity.

**Corollary 4.4.** If $\sigma(n) = \alpha \sqrt{\ln n} + \beta$ and $k_n = e^{\sigma(n)}$ then

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{k_n} i \lfloor c_1 i^{c_2 \ln i} \rfloor}{n} \geq c_1 e^{\beta + g \beta + c_2 \beta^2} \geq 1.$$  

**Proof.** Let $\tilde{k}_n = \lfloor k_n \rfloor$. Obviously we have that

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{\tilde{k}_n} i \lfloor c_1 i^{c_2 \ln i} \rfloor}{n} \geq c_1 e^{\beta + g \beta + c_2 \beta^2} \geq 1. \quad (14)$$

Then Lemma 4.1, Proposition 4.2 and (14) imply that

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{\tilde{k}_n} i \lfloor c_1 i^{c_2 \ln i} \rfloor}{n} \geq 1. \quad (15)$$

Recall that from the definition of $c_1$ we have that $c_1 e^{\beta + g \beta + c_2 \beta^2} \geq 1$. Hence Lemma 4.3 and (15) imply the corollary. This completes the proof. \qed

Let $\gamma = \beta + \ln c_1 + c_2 \beta^2 = \frac{1 - g}{2c_2} + \ln c_1 + c_2 \left(\frac{1 - g}{2c_2}\right)^2 \in \mathbb{R}$. The next lemma shows an upper bound for the sum $\sum_{i=1}^{k_n} c_1 i^{c_2 \ln i}$, where $k_n = \lfloor e^{\alpha \sqrt{\ln n} + \beta} \rfloor$.

**Lemma 4.5.** If $n \in \mathbb{N}_1$ and $\sigma(n) = \alpha \sqrt{\ln n} + \beta$ then

$$\sum_{i=1}^{e^{\sigma(n)}} c_1 i^{c_2 \ln i} \leq e^{\gamma} \frac{n}{e^{\alpha \sqrt{\ln n}}}.$$  

**Proof.** We have that

$$\sum_{i=1}^{e^{\sigma(n)}} c_1 i^{c_2 \ln i} \leq e^{\sigma(n)} c_1 \left( e^{\sigma(n)} \right)^{c_2 \ln e^{\sigma(n)}} = e^{\sigma(n)} c_1 e^{c_2 (\sigma(n))^2} = e^{\alpha \sqrt{\ln n} + \beta} e^{\ln c_1} e^{c_2 (\alpha \sqrt{\ln n} + \beta)^2} \quad (16)$$
and
\[
\alpha \sqrt{\ln n} + \beta + \ln c_1 + c_2(\alpha \sqrt{\ln n} + \beta)^2 = \\
\alpha \sqrt{\ln n} + \beta + \ln c_1 + c_2 \alpha^2 \ln n + c_2 2\alpha \beta \sqrt{\ln n} + c_2 \beta^2 = \\
c_2 \alpha^2 \ln n + \sqrt{\ln n}(\alpha + 2c_2 \alpha \beta) + \gamma.
\] (17)

Note that \(c_2 \alpha = c_2(\sqrt{\frac{1}{c_2}})^2 = 1\) and \(\alpha + 2c_2 \alpha \beta = \sqrt{\frac{1}{c_2}(1 + 2c_2 \frac{1-g}{2c_2})} = -g\alpha\).

Then it follows from (16) and (17) that
\[
\left\lfloor e^{\sigma(n)} \right\rfloor \sum_{i=1}^{\left\lfloor e^{\sigma(n)} \right\rfloor} c_1 i \omega(i) \leq e^{\ln n + \sqrt{\ln n}(-g\alpha) + \gamma} = e^\gamma \frac{n}{eg\alpha \sqrt{\ln n}}.
\]

This completes the proof. \(\square\)

In the proof of the main theorem we will need the following technical lemma.

**Lemma 4.6.** If \(\mathbf{t}, t \in \mathbb{N}_1, \ t > t, \ \omega, \omega : \mathbb{N}_1 \to \mathbb{N}_1, \ \overline{\omega}(i) \leq \omega(i) \) for every \(i \in \mathbb{N}_1\), and
\[
\sum_{i=1}^{\mathbf{t}} \overline{\omega}(i) > \sum_{i=1}^{t} \omega(i)
\]
then
\[
\sum_{i=1}^{\mathbf{t}} i \overline{\omega}(i) > \sum_{i=1}^{t} i \omega(i).
\]

**Proof.** We have that
\[
\sum_{i=1}^{\mathbf{t}} \overline{\omega}(i) - \sum_{i=1}^{t} \omega(i) = \\
\sum_{i=1}^{t} \overline{\omega}(i) + \sum_{i=t+1}^{\mathbf{t}} \overline{\omega}(i) - \left( \sum_{i=1}^{t} (\omega(i) - \overline{\omega}(i)) + \sum_{i=1}^{t} \overline{\omega}(i) \right) = \\
\sum_{i=t+1}^{\mathbf{t}} \overline{\omega}(i) - \sum_{i=1}^{t} (\omega(i) - \overline{\omega}(i)).
\]
Since \( \sum_{i=1}^{t} \overline{\omega}(i) > \sum_{i=1}^{t} \omega(i) \) it follows from (18) that
\[
\sum_{i=t+1}^{t} \overline{\omega}(i) > \sum_{i=1}^{t} (\omega(i) - \overline{\omega}(i)).
\] (19)

From (19) it is obvious that
\[
\sum_{i=t+1}^{t} i \overline{\omega}(i) > \sum_{i=1}^{t} i (\omega(i) - \overline{\omega}(i)).
\] (20)

We have that
\[
\sum_{i=1}^{t} i \overline{\omega}(i) - \sum_{i=1}^{t} i \omega(i) =
\sum_{i=1}^{t} i \overline{\omega}(i) + \sum_{i=t+1}^{t} i \overline{\omega}(i) - \left( \sum_{i=1}^{t} i (\omega(i) - \overline{\omega}(i)) + \sum_{i=1}^{t} i \overline{\omega}(i) \right) =
\sum_{i=t+1}^{t} i \overline{\omega}(i) - \sum_{i=1}^{t} i (\omega(i) - \overline{\omega}(i)).
\] (21)

It follows from (20) and (21) that \( \sum_{i=1}^{t} i \overline{\omega}(i) > \sum_{i=1}^{t} i \omega(i) \). This ends the proof.

**Remark 4.7.** Lemma 4.6 can be interpreted as follows. Suppose two finite sets \( \Omega_1, \Omega_2 \subset \mathbb{A}^+ \). Let \( \overline{\omega}(i) = |\Omega_1 \cap \mathbb{A}^i| \) and \( \omega(i) = |\Omega_2 \cap \mathbb{A}^i| \). Let \( \overline{w}, w \in \mathbb{A}^+ \) be words that are a concatenation of all words from \( \Omega_1, \Omega_2 \) respectively. Lemma 4.6 implies that if \( |\Omega_1| > |\Omega_2| \) and \( \overline{\omega}(i) \leq \omega(i) \) then \( |\overline{w}| > |w| \). This interpretation will simplify the comprehension of the proof of Proposition 4.8.

The next proposition shows a technique how to derive an upper bound for the length of UPS-factorization.

**Proposition 4.8.** If \( w \in \mathbb{R}W \cap \mathbb{A}^+ \), \( n = |w| \), \( t \in \mathbb{N}_1 \), and
\[
\sum_{i=1}^{t} i c_1 i^{c_2 i} \ln i \geq n
\]
then
\[
\text{luf}(w) \leq \sum_{i=1}^{t} c_1 i^{c_2 i} \ln i.
\]
Proof. Let $t \in \mathbb{N}_1$ be the minimal value such that
\[
\sum_{i=1}^{t} i \lfloor c_1 i^{c_2 \ln i} \rfloor \geq n. \tag{22}
\]
We claim that \(\text{luf}(w) \leq \sum_{i=1}^{t} c_1 i^{c_2 \ln i}\). To get a contradiction suppose that
\[
h = \text{luf}(w) > \sum_{i=1}^{t} c_1 i^{c_2 \ln i}. \tag{23}
\]
Let $w_1, w_2, \ldots, w_h \in \text{Fac}_w \cap \text{Pal} \cap A^+$ be distinct nonempty palindromes of UPS-factorization of $w$; it means that $w = w_1 w_2 \cdots w_h$. Let
\[
\phi(k) = |\{w_i \mid w_i \in A^k \text{ and } i \in \{1, 2, \ldots, h\}\}|
\]
be the number of palindromes $w_i$ of length $k$. Let $\bar{t} \in \mathbb{N}_1$ be the maximal value such that $\phi(\bar{t}) > 0$. Clearly we have that $h = \sum_{k=1}^{\bar{t}} \phi(k)$. Hence it follows from (23) that
\[
\sum_{k=1}^{\bar{t}} \phi(k) > \sum_{i=1}^{t} c_1 i^{c_2 \ln i}. \tag{24}
\]
Since $w$ is rich, Corollary 3.3 implies that
\[
\phi(k) \leq c_1 k^{c_2 \ln k} \quad \text{for every } k \in \{1, 2, \ldots, \bar{t}\}. \tag{25}
\]
Suppose that $\bar{t} > t$. Then from Lemma 4.6, (24), and (25) it follows that
\[
\sum_{k=1}^{\bar{t}} k \phi(k) > \sum_{i=1}^{t} i c_1 i^{c_2 \ln i},
\]
which is a contradiction to (22), since obviously $n = \sum_{k=1}^{\bar{t}} k \phi(k)$. We conclude that $\bar{t} \leq t$. Then (24) implies that there is at least one $k$ such that $\phi(k) > c_1 k^{c_2 \ln k}$, which contradicts to (25). We conclude that $\text{luf}(w) \leq \sum_{i=1}^{t} c_1 i^{c_2 \ln i}$. This completes the proof. \(\square\)

Now, we can step to the proof of the main theorem of the current article.
Proof of Theorem 1.3. We have that \( w \in RW \cap A^+ \) and \( n = |w| \). Let
\[
t = \lfloor e^{\sqrt{\ln n + \beta}} \rfloor.
\]
Then Corollary 4.4 implies that
\[
\sum_{i=1}^{t} i \lfloor c_1 i c_2 \ln i \rfloor \geq n
\]
and Lemma 4.5 implies that
\[
\sum_{i=1}^{t} c_1 i c_2 \ln i \leq e^\gamma \frac{n}{e^{\alpha \sqrt{\ln n}}} \text{ as } n \text{ tends to infinity.}
\]
Proposition 4.8 and (26) imply that
\[
\text{luf}(w) \leq \sum_{i=1}^{t} c_1 i c_2 \ln i.
\]
From (27) and (28) it follows that there are constants \( \mu, \pi \in \mathbb{R}^+ \) such that
\[
\text{luf}(w) \leq \mu \frac{n}{e^{\pi \sqrt{\ln n}}}
\]
This completes the proof.

Acknowledgments

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS20/183/0HK4/3T/14.

References

[1] K. Ago, B. Bašić, S. Hačko, and D. Mitrović, On generalized highly potential words, Theoretical Computer Science, 849 (2021), pp. 184–196.

[2] L. Balková, E. Pelantová, and Š. Starosta, Sturmian jungle (or garden?) on multilateral alphabets, RAIRO-Theor. Inf. Appl., 44 (2010), pp. 443–470.
[3] H. Bannai, T. Gagie, S. Inenaga, J. Kärkkäinen, D. Kempa, M. Piątkowski, S. J. Puglisi, and S. Sugimoto, *Diverse palindromic factorization is NP-complete*, in *Developments in Language Theory: 19th International Conference, DLT 2015, Liverpool, UK, July 27-30, 2015, Proceedings*, I. Potapov, ed., Springer International Publishing, 2015, pp. 85–96.

[4] M. Bucci, A. De Luca, A. Glen, and L. Q. Zamboni, *A new characteristic property of rich words*, Theor. Comput. Sci., 410 (2009), pp. 2860–2863.

[5] X. Droubay, J. Justin, and G. Pirillo, *Episturmian words and some constructions of de Luca and Rauzy*, Theor. Comput. Sci., 255 (2001), pp. 539–553.

[6] A. Frid, S. Puzynina, and L. Zamboni, *On palindromic factorization of words*, Adv. Appl. Math., 50 (2013), pp. 737–748.

[7] A. Glen, J. Justin, S. Widmer, and L. Q. Zamboni, *Palindromic richness*, Eur. J. Combin., 30 (2009), pp. 510–531.

[8] C. Guo, J. Shallit, and A. M. Shur, *Palindromic rich words and run-length encodings*, Inform. Process. Lett., 116 (2016), pp. 735–738.

[9] J. Rukavicka, *On the number of rich words*, Developments in Language Theory: 21st International Conference, DLT 2017, Liège, Belgium, August 7-11, 2017, Proceedings, Springer International Publishing, ISBN:978-3-319-62809-7, available at https://doi.org/10.1007/978-3-319-62809-7_26, (2017), pp. 345–352.

[10] J. Rukavicka, *A unique extension of rich words*, Theoretical Computer Science, (2021).

[11] Rukavicka, Josef, *Upper bound for palindromic and factor complexity of rich words*, RAIRO-Theor. Inf. Appl., 55 (2021), p. 1.

[12] J. Vestí, *Extensions of rich words*, Theor. Comput. Sci., 548 (2014), pp. 14–24.